Department of Urology, University Hospital of Ulm
Director: Prof Dr. med. Mark Schrader

Common germline variants for prostate cancer risk: Implication in DNA repair and *TMPRSS2-ERG* fusion formation

DISSERTATION

FOR THE DOCTORAL DEGREE OF HUMAN BIOLOGY (DR. BIOL. HUM.)
FACULTY OF MEDICINE, ULM UNIVERSITY

submitted by
Antje Eva Rinckleb
born in Chemnitz, Germany
2014

Acting Dean: Prof. Dr. Thomas Wirth

1. Correspondent: PD Dr. Christiane Maier

2. Correspondent: PD Dr. Dominic Varga

Day of Defense: 04.07.2014

Table of Contents

List of Abbreviations iii

1. Introduction 1

1.1 Prostate cancer 1

1.1.1 Epidemiology 1
1.1.2 Clinical background 2

1.2 Etiology - Genetic susceptibility is an important risk factor for prostate cancer . 4

1.2.1 High-risk genes identified by family based analyses 4
1.2.2 Risk loci found by genome-wide association studies (GWAS) 5

1.3 Somatic phenotype - Gene fusions denote separate entities 8

1.3.1 Discovery of recurrent oncogene rearrangements in prostate cancer 8
1.3.2 The *TMPRSS2-ERG* (T2E) fusion 8
1.3.3 Clinical relevance of the *TMPRSS2-ERG* fusion 9

1.4 Implication of DNA repair in prostate cancer 10

1.5 Rationale 11

2. Probands, Materials and Methods 13

2.1 Probands 13

2.1.1 The ULM sample cohort for case-control studies 13
2.1.2 Blood donors for an association study on common variants and DNA repair 13
2.1.3 Association study - common variants and the *TMPRSS2-ERG* fusion 14
2.1.4 Surgical specimens for gene expression analyses 17

2.2 Materials 21

2.2.1 Chemicals and reagents 21
2.2.2 Buffers, solutions and media 22
2.2.3 Antibodies, enzymes, reagent systems and kits 25
2.2.4 Cell lines 26
2.2.5 Bacterial artificial chromosomes (BACs) 26
2.2.6 Primer oligonucleotides and TaqMan® assays 26
2.2.7 Small interfering RNAs (siRNAs) 28
2.2.8 Laboratory equipment and consumables 29
2.2.9 Software and internet resources 30

2.3 Methods 31

2.3.1 Cultivation of human cells 31
2.3.2 Isolation of nucleic acids from different sources 32
2.3.3 Quantification of nucleic acids 35
2.3.4 Purification of nucleic acids 36
2.3.5 Agarose gel-electrophoresis 37
2.3.6 Polymerase chain reaction (PCR) 37
2.3.7 cDNA synthesis 39
2.3.8 Quantitative reverse transcriptase PCR (qRT-PCR) for gene expression analysis 39
2.3.9 Genotyping of single nucleotide polymorphisms (SNPs) 46
2.3.10 Determination of the somatic *TMPRSS2-ERG* fusion status in prostate cancer specimens 49
2.3.11 *TMPRSS2-ERG* fluorescence in situ hybridization (FISH) 50
2.3.12 *TMPRSS2-ERG* induction assay with integrated gene knock-down 57

2.3.13 Micronucleus assay 61
2.3.14 Mitotic delay assay 63
2.3.15 Statistical methods 65

3. Results 70

3.1 Contribution of common variants to prostate cancer risk 70

3.1.1 Verification of risk SNPs in a German case-control sample 70
3.1.2 Cumulative effects of common variants 72

3.2 Serial examinations of SNPs for mechanistic involvements 74

3.2.1 Association between common variants and cellular DNA repair capacity 74
3.2.2 Association between the *TMPRSS2-ERG* fusion and common variants in multiple populations 78

3.3 Utilization of mechanistic links to elucidate candidate risk regions 89

3.3.1 Analyses of prostate cancer candidate genes on 10q11 89
3.3.2 Expressional analysis of the 17q24 candidate gene *SOX9* 103
3.3.3 Expressional analysis of the 8q24 candidate gene *MYC* 107

4. Discussion 112

4.1 Epidemiology: Replication of single common variants and future directions ..113

4.2 Mechanistical annotation of prostate cancer risk loci as a strategy to overcome heterogeneity 115

4.2.1 Testing for DNA damage response intriguingly links risk region 10q11 to non-homologous end-joining repair 116
4.2.2 Subtyping for somatic *TMPRSS2-ERG* fusion status identifies 8q24 and 17q24 SNPs as the first two differentially associated germline variants 118

4.3 Consideration of candidate genes based on the implication of DNA repair and fusion gene phenotypes 121

4.3.1 The examination of multiple plausible candidate genes on 10q11 highlighted *PARG* as the most suspicious player 121
4.3.2 *SOX9* on 17q24 as a putative *TMPRSS2-ERG* fusion promoting gene 132
4.3.3 The implication of 8q24 variants in DNA repair and *TMPRSS2-ERG* negative prostate cancer - The role of *MYC* 134

4.4 Conclusions and prospects 137

5. Summary 139

6. Literature 141

Appendix 157

Acknowledgments 164

List of Abbreviations

A	adenine
aa	amino acids
ac	autoclaved
aq	aqua (water)
AR	androgen receptor
ARE	androgen responsive element
AUC	area under curve
BAC	bacterial artificial chromosome
BMI	body mass index
BNC	binucleated cell
bp	base pair
BPH	benign prostatic hyperplasia
C	cytosine
CAM	Chloramphenicol
cDNA	complementary deoxyribonucleic acid
CI	confidence interval
COGS	Collaborative Oncological Gene-Environment Study
Cs	cesium
Da	Dalton
DAPI	4',6-Diamidino-2-phenylindol-dihydrochloride
dd.	double distilled
DHT	dihydrotestosterone
DI	deionized
DIG	digoxigenin
DNA	deoxyribonucleic acid
dNTP	deoxynucleoside triphosphate
DSB	(DNA) double-strand break
E. coli	Escherichia coli
EDTA	Ethylendiamide tetraacetate
ERG	ETS related gene
EtBr	Ethidiumbromide
EtOH	Ethanol
ETS	E-twenty-six
FISH	fluorescence in situ hybridization
FCS	fetal calf serum
FFPE	formaline-fixed paraffin embedded
G	guanine
GWAS	genome-wide association study
Gy	Gray
h	hour(s)
H_2O	water
HR	homologous recombination
HUGO	Human Genome Organization
HWE	Hardy-Weinberg Equilibrium
IR	irradiation
ICPCG	International Consortium for Prostate Cancer Genetics
kb	kilo base pairs
LD	linkage disequilibrium
log	logarithm
M	molar
MAF	minor allele frequency
Mb	mega base pairs
MDI	mitotic delay index
MN	micronucleus
min	minute(s)
µ	micro-
MNT	micronucleus test
mRNA	messenger ribonucleic acid
n	nano-/number(s)

NAD+	Nicotinamide adenine dinucleotide
NCBI	National Center for Biotechnology Information
NFQ	non-fluorescent quencher
NHEJ	non-homologous end joining
NTC	non template control
ON	over night
OR	odds ratio
PAR	poly (ADP-ribose)
PCR	polymerase chain reaction
PBL	peripheral blood lymphocytes
PBS	phosphate buffered saline
pH	potentia hydrogenii
PIN	prostate intraepithelial neoplasia
PRACTICAL	Prostate Cancer Association Group to Investigate Cancer Associated Alterations in the Genome
PrCa	prostate cancer
PSA	prostate-specific antigen
qRT-PCR	quantitative reverse transcriptase PCR
RAF	risk allele frequency
RISC	RNA-induced silencing complex
RNA	ribonucleic acid
ROC	receiver operating characteristic
rpm	rounds per minute
RT	room temperature
SD	standard deviation
SEM	standard error of the mean
shRNA	short hairpin RNA
siRNA	small interfering RNA
sol.	solution
sf	sterile filtered
SNP	single nucleotide polymorphism
T	thymine
T1G4	a TMPRSS2-ERG transcript variant
T2E	TMPRSS2-ERG fusion
Tab.	table
TBE	Tris/Borat/EDTA (buffer)
TMA	tissue microarray
Tris	Trishydroxymethylamino-methane
UV	ultra violet
v/v	volume fraction
WT	wild type
w/v	mass concentration
yr.	year(s)

The official gene symbols and full names provided by the Human Genome Organization Gene Nomenclature Consortium are listed in Appendix A.

1. Introduction

1.1 Prostate cancer

The prostate is an exocrine gland that belongs to the male internal genital tract and is located underneath the bladder enclosing the urethra. The organ with the size of a chestnut secretes a slightly alkaline liquid which provides the major component of the seminal fluid and is important for the viability of the sperm cells. Prostate tumors are predominantly (~ 95 %) adenocarcinomas and derive to 70 % from the peripheral zone.

1.1.1 Epidemiology

Prostate cancer (PrCa) is the most commonly diagnosed malignant cancer in men in Western industrialized nations and is responsible for ~ 10 % of cancer-related deaths (third leading cause of death after cancers of lung (25 %) - and colorectal cancer (12 %)). In Germany, approx. 65,000 new cases and 12,000 PrCa-related deaths were recorded in 2010 [129]. As shown in Figure 1, the incidence has been increasing gradually, while the mortality rate has even slightly decreased. These trends are due to more sensitive diagnostics on the one hand, revealing an increasing number of smaller tumors, and on the other hand, improving treatment standards over time.

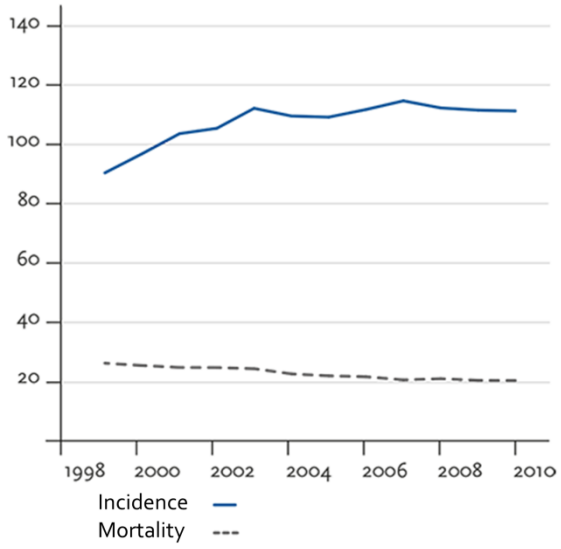


Figure 1: Age-standardized incidence and mortality of prostate cancer in Germany between 1999 and 2010 per 100,000 inhabitants. (modified from [129])

1.1.2 Clinical background

In Germany, urologists recommend to begin an annual PrCa screening for men at the age of 45. This routine examination usually comprises a rectal palpation of the prostate (digital rectal examination, DRE) and optionally, the determination of the blood PSA (prostate-specific antigen). The PSA protein is exclusively produced by the glandular cells of the prostate, and its levels are often increased when a tumor is present. As the PSA concentration could also be increased for other reasons, such as inflammation, benign hyperplasia or even physical activity, the test is not sufficiently specific to diagnose PrCa. According to the current guidelines of the German Society of Urology [33], a biopsy is indicated in case of a positive DRE, a PSA value greater than 4 ng/ml or a prominent annual PSA increase.

To find the optimal therapy, the tumor needs to be classified accurately with respect to the degree of severity and spreading. In addition to the PSA value, the TNM staging according to the UICC is used [143]. The letters indicate the status of the primary tumor (T), the lymph nodes (N) and, if applicable, distant metastases (M). The stages T1-2 N0 M0 denote locally confined carcinomas, T3-4 N0 M0 locally advanced tumors and the stages N1-3 and/or M1 are referred to as advanced or metastatic PrCa. To specify the histological status of a tumor, the Gleason-Grading according to the Society of Urological Pathology (ISUP) [40] is applied. The degree of cell differentiation is consecutively numbered from 1 (well- differentiated, resembles normal prostate tissue) to 5 (poorly-differentiated, no recognizable glands left). The pathologist determines the most common and the second most common tumor pattern and assigns the overall grade as the sum of these two scores resulting in values between 2 and 10.

Depending on the degree of disease severity, as well as the patient's age and preference, there are several therapeutic options. Locally confined and locally advanced prostate carcinomas are generally treated with the aim to remove the entire tumor, by means of radical prostatectomy, or alternatively, by various radiation therapies. 'Active surveillance' is a further strategy for handling local low-risk tumors. Here, the objective is to postpone a curative therapy as much as possible or until the patient decides otherwise. Men with a life expectancy of less than 10 years, accounted for age and co-morbidities, may not be burdened with an invasive therapy. For these cases, "watchful waiting" is an

option based on periodically monitoring, as long as the tumor does not impair live quality or survival. More advanced forms of PrCa, like recurrent or metastasized tumors, are in general treated systemically with different hormone therapies, e.g. androgen deprivation, or chemotherapies.

1.2 Etiology - Genetic susceptibility is an important risk factor for prostate cancer

Besides age and ethnicity, family history is one of the strongest risk factors for PrCa. The heritability, estimated by twin studies, is with 42 % the highest among the common carcinomas [86]. According to meta-analyses of family studies, the risk of developing PrCa rises to at least two-fold, if a family member has already been diagnosed with the tumor [74,76]. Men under 65 years are more prominently affected by this phenomenon than men older than 65. A dramatic increase in risk is also observed for men with several diseased relatives, especially if these were young at diagnosis. To date, the observed familial susceptibility cannot be sufficiently attributed to concrete genetic factors. Few high-risk genes are known for prostate cancer, as well as several functionally unexplained low-risk variants, which will be summarized briefly below.

1.2.1 High-risk genes identified by family based analyses

The classical linkage analysis has served for disease gene discovery, where inheritance patterns are retraced by genetic markers for cosegregation with the phenotype in families. The resulting linkage regions are then examined for candidate genes by fine-mapping and sequencing methods. The first candidate genes that derived from linkage analyses are *ELAC2* on 17p [153], *MSR1* on 8p [183] and *RNASEL* on 1q25 [24]. However, their relevance for PrCa risk seems limited after a series of replication studies [97,130,175].

Just recently, fine-mapping and subsequent next generation sequencing of an entire linkage region at 17q21-22 resulted in the identification of the first PrCa specific high-risk mutation [51,180]: A recurrent missense variant in the gene *HOXB13* (homeobox B13) [43], which is responsible for ~ 5 % of the PrCa families [181]. The gene is coding for a transcription factor which plays an essential role in prostate development and is capable to interact with the androgen receptor (AR) [115]. While *HOXB13* is currently investigated for its functional role and its utilization for predictive testing, the research community is enthusiastically continuing next generation sequencing for the identification of further

high risk genes. Prior to the identification of *HOXB13*, the early-onset breast cancer gene *BRCA2*, which is involved in homologous recombination repair, was considered for a long time to convey the strongest known genetic risk for PrCa. The first evidence for a substantial PrCa risk in *BRCA2* mutation carriers (OR = 4.65 [95% CI, 3.48–6.22]) was published by the Breast Cancer Linkage Consortium in 1999 [155].

1.2.2 Risk loci found by genome-wide association studies (GWAS)

GWAS are epidemiological approaches to discover genetic loci which are associated with specific phenotypes. For this purpose, case and control groups are analyzed for a very large number of genetic markers (single nucleotide polymorphisms, SNPs) spread all over the genome. A significant enrichment of a certain marker allele in the phenotype group is called an association and indicates a phenotype-related genetic cause within a range of neighboring alleles which have never undergone recombination in a population's history. Variants in genomic proximity, that are passed on conjointly over generations, are defined as blocks of linkage disequilibrium (LD) and these represent the smallest units of genetic resolution, e.g. where strategies of fine-mapping usually hit their limits.

Tagging virtually all LD blocks in the genome requires a hundreds of thousands to millions of examined SNPs, and in turn, strong adjustment of the significance threshold to account for a large number of performed tests. Under these conditions, GWAS are primarily capable for the identification of common variants, which provide only a small contribution to the phenotype as an inevitable consequence of their high allele frequencies. Rare high-risk variants remain concealed, unless favorably tagged by a marker SNP. Finally, it has been proven difficult to find a causal link between the phenotype and the associated marker SNPs, as many GWAS hits are located intergenic, within large gene deserts or in exceedingly gene-rich regions.

For PrCa, the first GWAS series were published in 2008 [37,80,156] and by these efforts, 77 prostate cancer-associated variants have been found to date. All these variants are relatively common in the population and contribute only a little effect size (Table 1), but nevertheless explain together ~ 30 % of the familiar prostate cancer risk [38].

Table 1: Overview of the 77 common risk variants identified for prostate cancer to date. The risk associated alleles and the risk allele frequencies (RAF) of the single nucleotide polymorphisms (SNP) and their effect sizes (odds ratio with 95% confidence interval (CI)) are taken from the original studies (references). The candidate genes are provided as suggested by the authors (gene symbols are given in the abbreviations). Please note that the list also includes SNPs which play a role in others than the population with European ancestry.

Genomic Region	*SNP*	*Allele(non-risk/risk)*	*RAF*	*per-allele odds ratios (95% CI)*	*Refe-rences*	*Candidate genes*
1q21	rs1218582	A/G	0.45	1.06 (1.03-1.09)	[38]	*KCNN3*
1q32	rs4245739	C/A	0.75	1.10 (1.05-1.14)	[38]	*MDM4*
2p11	rs10187424	G/A	0.59	1.09 (1.06-1.12)	[81]	*GGCX, VAMP8, VAMP5, RNF181*
2p15	rs721048	G/A	0.18	1.15 (1.10-1.21)	[56]	*EHBP1*
2p21	rs1465618	G/A	0.23	1.08 (1.03-1.12)	[36]	*THADA*
2p24	rs13385191	A/G	0.56	1.15 (1.10-1.21)	[152]	*C2orf43*
2p25	rs11902236	G/A	0.27	1.07 (1.03-1.10)	[38]	*TAF1B*, *GRHL1*
2q31	rs12621278	G/A	0.94	1.33 (1.25-1.43)	[7]	*ITGA6*
2q37	rs3771570	G/A	0.15	1.12 (1.08-1.17)	[38]	*FARP2*
2q37	rs2292884	A/G	0.25	1.14 (1.09-1.19)	[139]	*MLPH, PRLH, RAB17*
3p11	rs2055109	T/C	0.10	1.20 (1.13-1.29)	[3]	-
3p12	rs2660753	C/T	0.11	1.18 (1.06-1.31)	[37]	-
3q13	rs7611694	C/A	0.79	1.10 (1.08-1.14)	[38]	*SIDT1*
3q21	rs10934853	C/A	0.28	1.12 (1.08-1.16)	[54]	*EEFSEC*
3q23	rs6763931	C/T	0.44	1.04 (1.01-1.07)	[81]	*ZBTB38*
3q26	rs10936632	C/A	0.52	1.11 (1.08-1.14)	[81]	*CLDN11, SKIL*
4q13	rs1894292	A/G	0.52	1.10 (1.06-1.12)	[38]	*AFM*, *RASSF6*
4q22	rs12500426	C/A	0.46	1.08 (1.05-1.12)	[36]	*PDLIM5*
4q22	rs17021918	T/C	0.66	1.11 (1.08-1.15)	[36]	*PDLIM5*
4q24	rs7679673	A/C	0.55	1.10 (1.06-1.14)	[36]	*TET2*
5p12	rs2121875	T/G	0.34	1.05 (1.02-1.08)	[81]	*FGF10*
5p15	rs2242652	A/G	0.81	1.15 (1.11-1.19)	[81]	*TERT*
5p15	rs12653946	C/T	0.44	1.26 (1.20-1.33)	[152]	*IRX4*
5q35	rs6869841	G/A	0.21	1.07 (1.04-1.11)	[38]	*BOD1 (FAM44B)*
6p21	rs3096702	G/A	0.40	1.07 (1.04-1.10)	[38]	*NOTCH4*
6q21	rs2273669	A/G	0.15	1.07 (1.03-1.11)	[38]	*ARMC2*, *SESN1*
6p21	rs130067	T/G	0.21	1.05 (1.02-1.09)	[81]	*CCHCR1*
6p21	rs1983891	C/T	0.41	1.15 (1.09-1.21)	[152]	*FOXP4*
6q22	rs339331	C/T	0.64	1.22 (1.15-1.28)	[152]	*GPRC6A, RFX6*
6q25	rs9364554	C/T	0.29	1.17 (1.08-1.26)	[37]	*SLC22A3*
6q25	rs1933488	G/A	0.59	1.12 (1.09-1.15)	[38]	*RSG17*
7p15	rs12155172	G/A	0.23	1.11 (1.07-1.15)	[38]	*SP8*
7p15	rs10486567	A/G	0.77	1.35 (1.20-1.52)[a]	[156]	*JAZF1*
7q21	rs6465657	T/C	0.47	1.20 (1.07-1.34)	[37]	*LMTK2*
8p21	rs2928679	C/T	0.42	1.05 (1.01-1.09)	[36]	*SLC25A37*
8p21	rs1512268	G/A	0.45	1.18 (1.14-1.22)	[36]	*NKX3.1*
8p21	rs11135910	G/A	0.16	1.11 (1.07-1.16)	[38]	*EBF2*

Continued on following page

Table 1 continued:

Genomic Region	*SNP*	*Allele(non-risk/risk)*	*RAF*	*per-allele odds ratios (95% CI)*	*Refe-rences*	*Candidate genes*
8q24	rs1447295	C/A	0.07	1.51[b]	[10]	*MYC*
8q24	rs6983267	T/G	0.51	1.26 (1.13-1.41)	[187]	*MYC*
8q24	rs16901979	C/A	0.03	1.79 (1.36-2.34)	[55]	-
8q24	rs16902094	A/G	0.15	1.21 (1.15-1.26)	[54]	-
8q24	rs445114	C/T	0.64	1.14 (1.10-1.19)	[54]	-
8q24	rs10086908	C/T	0.70	1.15 (1.06-1.23)	[4]	-
8q24	rs12543663	A/C	0.31	1.08 (1.00-1.16)	[4]	-
8q24	rs620861	T/C	0.63	1.11 (1.04-1.19)	[4]	-
9q31	rs817826	T/C	0.08	1.41 (1.29-1.54)	[182]	*RAD23B-KLF4*
9q33	rs1571801	G/T	0.29	1.27 (1.10-1.48)[c]	[35]	*DAB21P*
10q11	rs10993994	C/T	0.37	1.25 (1.17-1.34)	[37]	*MSMB*
10q24	rs3850699	G/A	0.71	1.10 (1.06-1.12)	[38]	*TRIM8*
10q26	rs4962416	T/C	0.25	1.20 (1.07-1.34)[a]	[156]	*CTBP2*, *ZRANB1*
10q26	rs2252004	T/G	0.77	1.16 (1.10-1.22)	[3]	-
11p15	rs7127900	G/A	0.20	1.22 (1.17-1.27)	[36]	-
11q12	rs1938781	T/C	0.30	1.16 (1.11-1.21)	[3]	*FAM111A*
11q13	rs7931342	T/G	0.49	1.19 (1.11-1.27)	[37]	-
11q22	rs11568818	G/A	0.56	1.10 (1.06-1.14)	[38]	*MMP7*
12q13	rs10875943	T/C	0.31	1.07 (1.04-1.10)	[81]	*TUBA1C*, *PRPH*
12q13	rs902774	G/A	0.15	1.17 (1.11-1.24)	[37,139]	*KRT8*
12q24	rs1270884	A/G	0.49	1.07 (1.04-1.10)	[38]	*TBX5*
13q22	rs9600079	G/T	0.38	1.18 (1.12-1.24)	[152]	-
14q22	rs8008270	A/G	0.82	1.12 (1.08-1.16)	[38]	*FERMT2*
14q24	rs7141529	A/G	0.50	1.09 (1.06-1.12)	[38]	*RAD51B*
17p13	rs684232	A/G	0.36	1.10 (1.07-1.14)	[38]	*VPS53*, *FAM57A*
17q12	rs11649743	A/G	0.82	1.28 (1.07-1.52)[a]	[150]	*HNF1B*
17q12	rs4430796	G/A	0.49	1.22 (1.15-1.30)	[57]	*HNF1B*
17q21	rs11650494	G/A	0.08	1.15 (1.09-1.22)	[38]	*HOXB13*, *PRAC*, *SPOP*, *ZNF652*
17q21	rs7210100	G/A	0.05	1.51 (1.35-1.69	[60]	*ZNF652*
17q24	rs1859962	T/G	0.47	1.20 (1.14-1.27)	[57]	*SOX9*
18q23	rs7241993	A/G	0.70	1.09 (1.05-1.12)	[38]	*SALL3*
19q13	rs2735839	A/G	0.85	1.20 (1.10-1.33)	[37]	*KLK2*, *KLK3*
19q13	rs8102476	T/C	0.54	1.12 (1.08-1.15)	[54]	-
19q13	rs103294	T/C	0.24	1.28 (1.21-1.36)	[182]	*LILRA3*
20q13	rs2427345	A/G	0.63	1.06 (1.03-1.10)	[38]	*GATAS*, *CABLES2*
20q13	rs6062509	C/A	0.70	1.12 (1.09-1.16)	[38]	*ZGPAT*
22q13	rs5759167	T/G	0.53	1.16 (1.14-1.20)	[36]	*BIL, TTLL1*
Xp11	rs5945619	T/C	0.38	1.19 (1.071.31)	[37]	*GSPT2, MAGED1, NUDT10, NUDT11*
Xp22	rs2405942	G/A	0.79	1.14 (1.09-1.20)	[38]	*SHROOM2*
Xq12	rs5919432	G/A	0.81	1.06 (1.02-1.12)	[81]	*AR*

[a] Heterozygote odds ratio is given.

[b] No confidence interval available.

[c] Under assumption of a dominant model.

1.3 Somatic phenotype - Gene fusions denote separate entities

1.3.1 Discovery of recurrent oncogene rearrangements in prostate cancer

The first evidence for fusion genes in prostate tumors was revealed 2005 by the so-called COPA (cancer outlier profile analysis) method, which analyzes large amounts of microarray data for outlier expression profiles. More than a half of the examined prostate tumors (50–70 %) showed an overexpression of either *ERG* or *ETV1* [161], which were already known oncogenes in Ewing sarcoma and acute myeloid leukemia [69,116]. Further investigations of the transcript pattern in PrCa revealed a fusion between the 5'part of the androgen-dependent *TMPRSS2* gene and *ERG*/*ETV1* being the cause of the overexpression. To date, several fusion combinations have been identified. In addition to *TMPRSS2*, the 5'partners *HERPUD1*, *NDRG1*, *SLC45A3*, *ACSL3*, *HERV-K_22*, *HERVK17*, *CANT1*, *DDX5*, *KLK2*, *FOXP1*, *EST14*, *HNRPA2B1*, *HMGN2P46*, *OR51E2*, *UBTF* and the chromosomal region 14q13.3-14q21.1 are known [15,28,96,121]. Besides *ERG* and *ETV1*, further 3'genes are *ETV4* [160], *ETV5* [65] and *FLI1* [117], which belong altogether to the 'E Twenty-Six' (ETS) transcription factor family. In the further course of this work, I will focus on the most prevalent fusion, *TMPRSS2-ERG*, as it accounts for approx. 90 % of the ETS positive tumors.

1.3.2 The *TMPRSS2-ERG* (T2E) fusion

TMPRSS2 (21q22.3) is a member of the type II transmembrane serine protease (TTSP) family, which also includes HPN, TMPRSS2-5, Enteropeptidase, Corin, HAT (human airway trypsin) and MT-SP1 (membrane-type serine protease 1) [68]. The 70 kDa protein (492 aa) contains a transmembrane-, a serine protease-, a SRCR (scavenger receptor cysteine rich) - and a LDLRA (LDL receptor class A) domain. The *TMPRSS2* expression is androgen-dependent and thus predominantly restricted to the prostate gland.

ERG (ETS-related gene) on chromosome 21q22.3 belongs to the ETS (E-twenty-six) transcription factor family, which comprises approximately 30 members. The name ETS is derived from the homology to the viral oncogene v-ets of the avian erythroblastose virus E26. The functional spectrum of the ETS members reaches from regulation of

development, differentiation and proliferation to cell cycle control and apoptosis. All ETS proteins contain a highly-conserved DNA binding domain (Ets-domain) as well as a transcription activating or repressing domain.

As already mentioned, *TMPRSS2-ERG* is the most prevalent somatic fusion in prostatic carcinoma, possibly due to a chromosomal proximity of the fusion partners. Both genes are located approx. 3 megabases in distance and in the same orientation on the long arm of chromosome 21.

1.3.3 Clinical relevance of the *TMPRSS2-ERG* fusion

Since its discovery in 2005, many studies have addressed the role of the T2E fusion in PrCa tumorigenesis and progression. Most studies concluded that the presence of the fusion is neither associated with any of the clinical parameters (Gleason, stage, grade, PSA) nor with recurrence, progression and aggressiveness [5,30,42,49,107,108,131,164]. On the other hand, some authors have described an association between T2E and Gleason score [67], biochemical recurrence [110,111] and tumor grade [100]. Some approaches even lead to the conclusion that fusion-positive patients have a favorable prognosis [133]. In contrast, there is an interesting study that found the fusion significantly correlated with prostate specific death in a watchful waiting cohort of patients with localized prostate carcinoma [31]. In summary, the results are inconsistent and controversial, probably due to the mostly small sample sizes as well as the different approaches and endpoints. Thus, the prognostic value of ETS fusions is not clear to date.

Certainly, the detection of ETS fusions in urine or in biopsies has a diagnostic value due to the extremely high specificity. A negative biopsy finding in combination with a positive T2E urine test would suggest that the cancer has been missed by the biopsies. The fusion itself cannot be used as diagnostic marker because of its poor sensitivity, as roughly half of all prostate cancers are fusion negative. However, in multiplex approaches combined with *GOLPH2*, *PCA3*, and *SPINK1* the fusion gene adds promising predictive power to existing markers, such as serum *PSA* or *PCA3* alone [83].

1.4 Implication of DNA repair in prostate cancer

A cell is continually confronted with DNA damage of various kinds, whether endogenously by reactive oxygen species that arise in the course of metabolic processes, or exogenously by mutagenic substances such as cigarette smoke or energetic radiation. The cell's ability to respond to DNA damages and repair them or, if necessary, undergo apoptosis, is essential, as an enrichment of mutations promotes malignant transformation. Germline mutations in DNA repair-associated genes lead to an increased susceptibility to certain cancers. Though the etiology of PrCa is not well understood to date, an involvement of DNA damage response and repair genes is indisputable. As already mentioned in section 1.2.1, germline mutations in *BRCA2,* which plays a key role in homologous recombination repair, were found associated with PrCa risk [155]. Additionally, there is evidence for a risk modulation by variations in *PALB2*, that codes for a BRCA2 binding protein [41], in *OGG1* [184] and *XRCC1* [167], both genes of base excision repair and in *CHEK2*, which is important for damage-induced cell cycle arrest [34,141].

Considering that roughly half of all prostate carcinomas harbor an ETS rearrangement, the question of an implication of DNA repair in fusion susceptibility imposes itself, as the critical step to form a rearrangement is the occurrence of double strand breaks. The fusion positive prostate carcinoma may therefore represent a separate tumor entity driven by variation in DNA damage response and repair. And indeed, there is already evidence for an involvement of two repair associated genes, *POLI*, a component of translesion synthesis, and *ESCO1*, important for chromosomal maintenance, in predisposition for ETS positive PrCa [94,95].

In the process of fusion formation, hormones seem to play a major role. The dimeric androgen receptor was shown to co-localize genes with androgen response elements which can lead under genotoxic stress to double strand breaks and a subsequent formation of rearrangements [90]. Additionally, the topoisomerase 2B enzyme (TOP2B) was shown to be recruited androgen-dependent to breaking sites and generates double strand breaks irrespective of endogenous or exogenous damage [59].

Apart from the fact that fusion susceptibility might have its origin in a, whatever kind, enhanced formation of double strand breaks, a defective or predisposed double strand break repair could also be the reason. There are two major ways for this pathway available in the cellular machinery: The homologous recombination (HR) repair and non-homologous end joining (NHEJ). The error-prone NHEJ is active throughout the cell cycle, preferentially in the G0, G1 and early S phase [87]. The largely error-free HR repair requires a template sister chromatid and is therefore available after the cell replication in the late S- as well as in the G2 phase [132]. A possible mechanism for the formation of rearrangements could be an increased usage of the error-prone NHEJ pathway.

1.5 Rationale

Prostate cancer has undoubtedly a strong genetic component, which is, based on its heterogeneous character, difficult to clarify. To date, many independent common risk-modifying SNPs have been found for prostate cancer by genome-wide association studies, which are regarded as surrogate markers for defective genes or regulatory elements, but however little is known about the risk-mediating mechanisms behind them. Apart from fine-mapping of risk loci, which is routinely carried out with moderate success in genetic epidemiology, additional functional knowledge would be appreciated to dissect the causality of risk loci, especially the annotation of mechanistic features related to prostate cancer specific pathogenesis.

The aim of the present study is to gather functional evidence for an implication of common variants in a classical hallmark of carcinogenesis, the impairment of genome integrity. Two different strategies are used for this purpose: Common variants are investigated for a potential association with (1) the cellular DNA repair capacity assayed with two different tests, and (2) with the presence or absence of the somatic T2E fusion which is considered as a phenotypic reflection of an error-prone double strand break repair. The assignment of common variants to the T2E positive or negative tumor might also be of a great value since these subtypes are increasingly gaining attention as two different cancer entities which could also be distinct on the germline level.

The first focus is directed to a replication analysis of initially identified 25 common variants in a case-control association study to elucidate their general relevance in German case-control series. Within this setting, a subset of variants is defined for a cumulative approach, in order to fathom the collective risk contribution and to define high-risk groups.

For the second specific aim, the subset of variants with verified disease relevance is then subjected to mechanistic investigations. One functional approach is the genotyping of best-candidate variants in a collective of healthy probands, which have been assayed with two different tests on DNA repair capacity on the level of peripheral blood lymphocytes: The radiation-induced micronucleus test which measures unrepaired double strand breaks and the mitotic delay assay that depicts the extend of the delay in the G2 cell cycle phase after DNA damage. For a further mechanistic examination, a two-stage meta-analysis serves as a set-up to establish a relation between common variants and the T2E fusion. Six study centers from Europe and the USA could be achieved for participation, totaling up to accessible tumor specimen for T2E detection from 1,200 patients.

The final aim is the use of the established mechanistic knowledge for candidate gene analyses at regions of highlighted risk variants. Genes in the context of DNA damage response and repair are selected from functionally labeled loci and then analyzed with qRT-PCR methods in tumor and histologically normal prostate tissue specimens for expressional changes dependent on the tissue type, genotype and fusion status. Where applicable, the *TMPRSS2-ERG* induction assay can be applied, which measures the changes in fusion formation in a cell culture model after expressional manipulation of the gene of interest.

2. Probands, Materials and Methods

2.1 Probands

All prostate cancer patients and unaffected probands, which were included in the present studies, gave their written informed consent for participation. The studies were approved by the local institutional review boards. The sample cohorts for the particular substudies are described in the following paragraphs.

2.1.1 The ULM sample cohort for case-control studies

The PrCa cases (familial and sporadic) and controls are participants of the Prostate Cancer Genetics Project of the University of Ulm. The study was approved by the local ethical review committee (vote number 87/97). A general sample description can be found in section 2.1.3 and an overview of the clinical data is given in Table 2. The DNA from peripheral blood lymphocytes was previously extracted using the standard method (2.3.2.1).

For the association study concerning germline variants and PrCa risk, 708 cases (390 familial index and 317 sporadic patients) and 509 controls (214 age-matched and 295 population-based) were included.

2.1.2 Blood donors for an association study on common variants and DNA repair

A total of 165 healthy probands (128 female, 37 male) were recruited in the years 2008 and 2009 for examination of DNA repair properties. The study was approved by the local ethical review committee (vote number 08/2004) Due to the gender imbalance, the males were excluded. The mean age of the females was 24.9 years (± 7.7 SD). Blood was collected for DNA isolation (2.3.2.1) and for MNT and MD tests (2.3.13 and 2.3.14).

2.1.3 Association study - common variants and the *TMPRSS2-ERG* fusion

This study was carried out as a multicenter study consisting of six groups from Europe and the United States: FHCRC (USA), IPO-Porto (Portugal), TAMPERE (Finland), UKGPCS (UK), ULM (Germany) and Berlin (Germany). The participating sites were responsible to obtain their own local IRB votes, while the final collection of specimens and pseudonymized data by our group was approved by the ethical review committee of the University of Ulm (vote number 44/11 - genetics of T2E positive prostate cancer). Exclusively PrCa cases and controls with European ancestry were included. An overview of the sample sizes and the clinical features is given in Table 3. Each sample description was contributed by the respective research group and is cited verbatim as described in [127] in the following paragraphs in quotation marks:

"***FHCRC: Fred Hutchinson Cancer Research Center, Seattle, USA***

The study population consists of participants from two population-based case-control studies in Caucasian and African American residents of King County, Washington (Study I and Study II), which have been previously described [1,146]. Incident cases with histologically confirmed prostate cancer were ascertained from the Seattle-Puget Sound Surveillance, Epidemiology and End Results cancer registry. In Study I, cases were diagnosed between January 1, 1993, and December 31, 1996 and were 40-64 years of age at diagnosis. In Study II, cases were diagnosed between January 1, 2002, and December 31, 2005 and were 35-74 years of age at diagnosis. A comparison group of controls without a history of prostate cancer, residing in King County, Washington, was identified for each study using random digit telephone dialing. Controls were frequency-matched to cases by five-year age groups and recruited evenly throughout each ascertainment period for cases.

For the current study, 1,276 cases and 1,256 controls with blood DNA/genotype information were included. A subset of 392 cases was assayed for T2E, of which 174 were used for stage 1 and 218 for stage 2.

IPO-Porto: Instituto Português de Oncologia do Porto, Portugal

The IPO-Porto prostate cancer study includes patients with clinically localized prostate adenocarcinoma consecutively diagnosed and treated with open radical prostatectomy at the Portuguese Oncology Institute - Porto, Portugal, since 1999. The project involves sample collection of peripheral blood, urine and fresh-frozen tumor tissue. Relevant clinical data, namely Gleason grading, clinico-pathological staging and PSA level at diagnosis, were obtained from medical records.

For the current study, 329 cases and 201 controls with blood DNA/genotype information were included. A subset of 164 cases was assayed for T2E, of which 18 were included in stage 1 and 146 were used for stage 2.

TAMPERE: Finland

The first set of unselected cases and controls (PSA < 4 µg/ml) were collected in Tampere, Finland and all are of Finnish origin. The mean age of diagnosis was 68.7 years (range 36-94). The patients were diagnosed with prostate cancer from 1993-2008 at the Tampere University Hospital, Department of Urology. Tampere University Hospital is a regional referral center in the area for all patients with prostate cancer, which results in an unselected, population-based collection of patients. The other unselected set of cases and controls was collected in the Finnish arm of The European Randomized Study of Screening for Prostate Cancer, which was initiated in the early 1990s to evaluate the effect of screening with PSA testing on death rates from prostate cancer. These men were born in years 1933, 1937 and 1941 and were randomly assigned to a group that was offered PSA screening at an average of once every 4 years or to a control group that did not receive such screening. In addition to these two sporadic sample sets, familial cancer cases (mean age at diagnosis 70 years) from Finnish prostate cancer families were genotyped.

For the current study, 2,754 cases and 2,413 controls with blood DNA/genotype information were included. A subset of 174 cases was assayed for T2E and used in stage 1.

UKGPCS: UK Genetic Prostate Cancer Study

Blood DNA from prostate cancer cases was collected from patients throughout the UK aged ≤ 60 years at diagnosis and a systematic series from the prostate cancer clinic at The Royal Marsden NHS Foundation Trust. Diagnosis was confirmed from medical record or death certificate and 60 % of tumors were detected clinically.

For the current study, 4,549 cases and 2,413 controls with blood DNA/genotype information were included. A subset of 129 cases was assayed for T2E. These were used for stage 1.

ULM: Germany

[Author's note: This is a subset of the previously described ULM sample (2.1.1).]

Cases were recruited in two different ways. Familial prostate cancer probands (index cases) were ascertained from all over Germany from 1997-2007. They were advised by their attending physicians to contact the Clinic of Urology of Ulm. The positive family history was then verified by reviewing medical records or death certificates of family members. Only one member of each family (e.g. the index proband) was enrolled in the present study. Sporadic cases, who reported no relatives affected with prostate cancer, were collected in Ulm during their course of treatment (e.g. radical prostatectomy) at the Department for Urology of the University Hospital Ulm. The control group consists of age-matched healthy men and male population controls of unknown disease status.

For the current study, 698 cases and 354 male controls (209 age-matched, 145 unselected) with blood DNA/genotype information were included. A subset of 164 cases was assayed for T2E, of which 57 were used for stage 1 and 107 for stage 2.

Berlin: Germany

For this study total RNA of 210 prostate cancer tissues as well as lymphocyte DNA from patients who underwent radical prostatectomy between 2000 and 2006 at Charité – Universitätsmedizin, Klinikum Benjamin Franklin was obtained by standard isolation procedures. All patients gave their informed consent in accordance with guidelines of the local Ethics Review Board.

Finally, 198 cases with blood DNA/genotype information and known T2E status were included, all in stage 2."

2.1.4 Surgical specimens for gene expression analyses

Candidate gene mRNA expression analyses dependent on tissue type, T2E status and genotype were performed using an independent set of prostate tumor and histologically normal tissue pairs of 38 (35 for *MYC* and *SOX9*) PrCa patients from Erlangen and 49 (35 for *MYC* and *SOX9*) from Ulm. An overview of the clinical data is given in Table 4. All Erlangen specimens were kindly provided as RNA (treated with DNAse) and DNA samples by Prof. Helge Taubert. The Ulm set consisted of 35 fresh frozen tissue blocks of tumor and surrounding normal tissue, respectively, kindly adopted from Dr. Natascha Bachmann, which were used for DNA and RNA isolation (2.3.2.3). Fourteen additional sample pairs were kindly provided by Prof. Ralf Marienfeld as RNAs, which had been extracted using TRIzol® reagent. As no DNA was available, this RNA served also as a template for genotyping.

Table 2: Sample sizes and clinical data of the ULM prostate cancer cases. For the tumor stage, a distinction was made between localized (T1/T2, N0, M0) and advanced (T3/T4 and/or N1/M1). Abbreviations: adv = advanced, loc = localized, n = number, N/A = not available, SD = standard deviation, n-ukn = unknown (given as the number n of individuals)

Group	Number of individuals	Mean age at diagnosis / blood draw		PSA at diagnosis (ng/ml) in [%]			Stage [%]			Gleason sum [%]			
		age [± SD]	n-ukn	≤ 20	> 20	n-ukn	loc	adv	n-ukn	< 7	7	> 7	n-ukn
familiar cases	390	62.3 [± 6.7]	-	82.1	17.9	66	61.7	38.3	30	48.7	34.7	16.6	119
sporadic cases	317	63.8 [± 6.6]	-	84.0	16.0	24	55.2	44.8	9	47.3	36.4	16.3	53
all cases	707	63.0 [± 6.7]	-	83.0	17.0	90	58.7	41.3	39	48.0	35.5	16.4	172

Table 3: Sample sizes and clinical data of the prostate cancer and control samples included in the association study between common variants and *TMPRSS2-ERG* fusion status. The Stage 1 and 2 groups represent the samples with known *TMPRSS2-ERG* status or with available tumor material which were included in the identification and the replication stage of the meta-analysis, respectively. For the tumor stage, a distinction was made between localized (T1/T2, N0, M0) and advanced (T3/T4 and/or N1/M1). Abbreviations: adv = advanced, loc = localized, n = number, N/A = not available, SD = standard deviation, n-ukn = unknown (given as the number n of individuals)

Study Center	Group	Number of individuals	Mean age at diagnosis / blood draw		PSA at diagnosis (ng/ml) in [%]			Family history [%]			Stage [%]			Gleason sum [%]			
			age [± SD]	n-ukn	≤ 20	> 20	n-ukn	yes	no	n-ukn	loc	adv	n-ukn	< 7	7	> 7	n-ukn
Berlin (Germany)	Stage 2 (=all cases)	198	62.4 [± 5.7]	-	88.6	11.4	14	-	-	198	60.6	39.4	-	57.3	26.4	16.7	6
FHCRC (Fred Hutchinson Cancer Research Center, USA)	Stage 1	174	58.0 [± 7.2]	-	92.5	7.5	14	20.7	79.3	-	70.7	29.3	-	53.4	36.8	9.8	-
	Stage 2	218	58.4 [± 6.7]	-	91.5	8.5	7	25.7	74.3	-	69.3	30.7	-	48.2	47.2	4.6	-
	Stage 1+2	392	58.2 [± 7.1]	-	91.9	8.1	21	23.5	76.5	-	69.9	30.1	-	50.5	42.6	6.9	-
	all cases	1,276	59.9 [± 7.0]	-	90.3	9.7	98	21.9	78.1	-	78.1	21.9	-	57.1	33.2	9.7	4
	controls	1,256	59.6 [± 6.9]	619													
IPO-Porto (Instituto Português de Oncologia do Porto, Portugal)	Stage 1	18	61.5 [± 3.9]	-	100	-	-	-	-	18	50.0	50.0	-	22.2	66.7	5.6	-
	Stage 2	146	63.6 [± 6.1]	-	98.6	1.4	-	27.3	72.7	135	53.4	46.6	-	34.2	58.2	7.5	-
	Stage 1+2	164	63.3 [± 5.9]	-	98.8	1.2	-	27.3	72.7	153	53.0	47.0	-	32.9	59.1	7.9	-
	all cases	329	61.2 [± 6.0]	-	98.0	2.0	22	27.3	72.7	318	43.5	56.5	-	24.3	63.8	11.9	-
	controls	201	N/A														
TAMPERE (Finland)	Stage 1	174	71.2 [± 8.1]	-	73.7	26.3	3	-	-	174	66.9	33.1	2	45.3	29.2	25.5	13
	cases	2,754	68.2 [± 8.0]	-	80.4	19.6	150	-	-	2,754	77.1	22.9	136	55.6	29.0	15.4	389
	controls	2,413	N/A														
UKGPCS (UK Genetic Prostate Cancer Study)	Stage 1	129	66.5 [± 6.0]	-	97.2	2.8	93	100	-	123	97.6	2.4	4	81.6	17.6	0.8	4
	cases	4,549	63.8 [± 8.0]	36	75.2	24.8	1,721	94.9	5.1	4,020	62.4	37.6	772	50.7	32.6	16.7	1,020
	controls	4,180	58.2 [± 5.3]	24													
ULM (Germany)	Stage 1	57	61.2 [± 5.4]	-	83.6	16.4	2	96.5	3.5	-	64.3	35.7	1	47.4	38.6	14.0	-
	Stage 2	107	63.7 [± 6.7]	-	83.3	16.7	53	70.9	29.1	4	50.0	50.0	1	40.0	39.0	21.0	7
	Stage 1+2	164	62.9 [± 6.4]	-	83.5	16.5	55	80.0	20.0	4	54.9	45.1	2	42.7	38.9	18.5	7
	cases	698	63.8 [± 6.7]	11	83.7	16.3	116	47.8	52.7	4	57.1	42.9	40	46.5	36.8	16.7	135
	controls	354	58.4 [± 11.8]	146													

Table 4: Sample sizes and clinical data of the prostate cancer patients from Ulm and Erlangen with available RNA/tumor material for gene expression analyses. For the tumor stage, a distinction was made between localized (T1/T2, N0, M0) and advanced (T3/T4 and/or N1/M1). Abbreviations: adv = advanced, loc = localized, n = number, SD = standard deviation, n-ukn = unknown (given as the number n of individuals)

Study Center	Group	Number of individuals	Mean age at diagnosis		PSA at diagnosis (ng/ml) in [%]			Family history [%]			Stage [%]			Gleason sum [%]			
			age [± SD]	n-ukn	≤ 20	> 20	n-ukn	yes	no	n-ukn	loc	adv	n-ukn	< 7	7	> 7	n-ukn
Erlangen (Germany)	all cases	38	66.1 [± 5.8]	-	77.1	22.8	3	-	-	38	43.2	56.8	1	13.5	68.6	21.6	1
Ulm (Germany)	Sample 1	35	64.2 [± 5.5]	-	-	-	35	3.2	96.8	4	48.6	51.4	-	35.3	55.9	8.8	1
	Sample 2	14	64.5 [± 4.3]	6	60.0	40.0	9	-	-	14	16.7	83.3	8	57.1	14.3	28.6	7
	all cases	49	64.2 [± 5.2]	6	60.0	40.0	44	3.2	96.8	18	43.9	56.1	8	39.0	48.8	12.2	8
	Total	87	65.1 [± 5.5]	6	75.0	25.0	47	3.2	96.8	56	43.6	56.4	9	26.9	56.4	16.7	9

2.2 Materials

2.2.1 Chemicals and reagents

Acetic acid	AppliChem, Darmstadt, Germany
Agarose SeaKem® LE	Lonza, Rockland, USA
Boric acid	J.T. Baker, Deventer, The Netherlands
Bromophenol blue	Merck, Darmstadt, Germany
Cleaning Solution	Partec, Münster, Germany
Chloramphenicol	Sigma-Aldrich, St. Louis, USA
Chloroform	Merck, Darmstadt, Germany
Cot 1 DNA, human (1mg/ml)	Life Technologies, Carlsbad, USA
Cytochalasin-B (Helminthosporium dermatioideum)	Sigma-Aldrich, St. Louis, USA
DAPI Staining Solution	Partec, Münster, Germany
Deoxyribonucleoside triphosphates (dNTPs)	Fermentas, St. Leon-Rot, Germany
Dextran Sulfate	Sigma-Aldrich, St. Louis, USA
4', 6-Diamidino-2-phenylindoldihydrochloride (DAPI)	Sigma-Aldrich, St. Louis, USA
Dihydrotestosterone (DHT)	Sigma-Aldrich, St. Louis, USA
Ethylenediaminetetraacetic acid (EDTA)	Fluka, Neu-Ulm, Germany
Ethanol, absolute	VWR Prolabo, Darmstadt, Germany
Ethidium bromide (EtBr) solution, 10mg/ml in H_2O	Sigma-Aldrich, St. Louis, USA
fetal calf serum (FCS)	Biowest, Nuaillé, France
fetal calf serum (FCS), charcoal stripped	Biowest, Nuaillé, France
Formamide, deionized (DI)	Amresco, Solon, USA
Gentamicin	Biochrom, Berlin, Germany
Glycerine, 87 %	J.T. Baker, Deventer, The Netherlands
H_2O Ultrapure (Chromanorm)	VWR Prolabo, Darmstadt, Germany
Hydrochloric acid (HCl), 1M	Merck, Darmstadt, Germany
Isoamyl alcohol	Roth, Karlsruhe, Germany
Isopropanol	Roth, Karlsruhe, Germany
Lipofectamine® RNAiMAX Reagent	Life Technologies, Carlsbad, USA
Magnesium Chloride ($MgCl_2$)	Sigma-Aldrich, St. Louis, USA
Magnesium Chloride ($MgCl_2$), 25 mM for PCR	GE Healthcare, Munich, Germany
2-Mercaptoethanol	Roth, Karlsruhe, Germany
Methanol	Sigma-Aldrich, St. Louis, USA
Milk powder	Seeberger, Ulm, Germany
Opti-MEM® I Reduced-Serum Medium (Gibco)	Life Technologies, Carlsbad, USA
PCR Buffer, 10 x	GE Healthcare, Munich, Germany

Penicillin-Streptomycin Solution, 100X	PAA Laboratories GmbH, Pasching, Austria
Phosphat buffered saline (PBS) powdered buffer	PAA Laboratories GmbH, Pasching, Austria
Phosphat buffered saline (PBS) without Ca^{2+} and Mg^{2+}	PAA Laboratories GmbH, Pasching, Austria
Phenol:Chlorophorm:Isoamyl alcohol 25:24:1	Fluka, Neu-Ulm, Germany
Phytohaemagglutinine (M-form) (PHA)	Life Technologies, Carlsbad, USA
Potassium chloride (KCl)	Merck, Darmstadt, Germany
Potassium dihydrogen phosphate (KH_2PO_4)	Merck, Darmstadt, Germany
RPMI-1640 (Gibco)	Life Technologies, Carlsbad, USA
RPMI-1640 with L-Glutamine and 25mM HEPES (Gibco)	Life Technologies, Carlsbad, USA
Salmon DNA, ultrapure (10 mg/ml)	Life Technologies, Carlsbad, USA
SOC medium	Life Technologies, Carlsbad, USA
Sodium chloride (NaCl)	AppliChem, Darmstadt, Germany
Sodium chloride (NaCl) 0.9 %	B.Braun, Melsungen, Germany
Sodium dodecyl sulfate (SDS)	Serva, Heidelberg, Germany
Sodium hydrogen carbonate ($NaHCO_3$)	Merck, Darmstadt, Germany
Trisodium citrate dihydrate (Na_3-Citrat x $2H_2O$)	Merck, Darmstadt, Germany
Tris(hydroxylmethyl)aminomethane (TRIS)	Sigma-Aldrich, St. Louis, USA
Trypsin-EDTA 1x	PAA Laboratories GmbH, Pasching, Austria
Tween® 20	Sigma-Aldrich, St. Louis, USA
VectaShield® Mounting Medium	Vector Labs, Burlingame, USA
VectaShield® Mounting Medium with DAPI	Vector Labs, Burlingame, USA
Xylene cyanol	Sigma-Aldrich, St. Louis, USA
Xylene	Fisher Scientific, Schwerte, Germany
Yeast extract	Sigma-Aldrich, St. Louis, USA

2.2.2 Buffers, solutions and media

Agarose gel electrophoresis loading buffer	Glycerol Bromophenol blue Xylene cyanol EDTA in aq. dd. *ad pH 8.0*	50 % 0.25 % (w/v) 0.25 % (w/v) 10 mM
Chloramphenicol stock solution	Chloramphenicol in aq. dd. *storage at -20°C*	12.5 mg/ml
Cot1/Salmon DNA Mix (10x)	Cot1 DNA Salmon DNA 4 M NaCl EtOH, absolut *mix, centrifuge full speed, 20 min, 4°C* *wash pellet with 80 % EtOH, air dry,* *resuspend in 50 µl aq. dd., 15 min at 55°C,* *storage at -20°C*	500 µl 50 µl 27.5 µl 2.5 volumes

Cytochalasin-B Solution	Cytochalasin-B in DMSO *storage at -20°C*	2 mg/ml
DAPI solution	20 x SSC DAPI in aq. dd. *storage at 4°C, protected from light*	20 % (v/v) 1 µg/ml
Dihydrotestosterone (DHT) stock solution, 10 mM	DHT in absolute EtOH *storage at -20°C*	10 mM
DIG 10x dNTP Mix	dATP dCTP dGTP dTTP in aq. dd. *storage at -20°C*	1 mM 1 mM 1 mM 0.65 mM
EDTA, 0.5 M	EDTA in aq dd. *ad pH 8.0 (NaOH), ac.*	0.5 M
Fixation Solution (FISH) *(prepare fresh)*	formamide, DI 20x SSC in aq. dd.	50 % (v/v) 1 % (v/v)
Fixative solution 1 *(prepare fresh)*	acetic acid methanol NaCl (0.9 %)	1 part 5 parts 6 parts
Fixative solution 2 *(prepare fresh)*	acetic acid methanol	1 part 5 parts
Hybridization Mix (FISH)	formamide, DI dextran sulfate 20x SSC EDTA salmon DNA in aq. dd. *Incubation ON 4°C, 20 min 42°C, storage at -20°C*	60 % (v/v) 12 % (w/v) 12 % (v/v) 0.14 mM 400 µg/ml
KCl solution, 0.56 %	KCl in aq. dd. (ac.)	0.56 % (w/v)
LB CAM (chloramphenicol) medium	tryptone NaCl yeast extract in aq. dd. *ad pH 7.5, ac.* chloramphenicol *storage at 4°C*	1 % (w/v) 1 % (w/v) 0.5 % (w/v) 12.5 mg/l
LB CAM freezing medium	glycerol in LB CAM medium *storage at -20°C*	50 % (v/v)
LB CAM medium for agar plates	agar in LB CAM medium *(after pH adjustment, before ac.)*	1.5 % (w/v)

Lymphocyte culture medium (short-time)	FCS PHA Gentamicin in RPMI-1640 with L-Glutamine and 25 mM Hepes *storage at -20°C*	20 % (v/v) 1 % (v/v) 1 % (v/v)
Lysis Buffer (DNA Isolation)	NH_4Cl $KHCO_3$ EDTA in aq. dd. *ad pH 7.4*	155 mM 10 mM 0.1 mM
NaCl Solution, saturated (6 M)	NaCl in aq. dd.	6 M
Proteinase K Solution	Proteinase K in aq. dd. *Storage at -20°C*	10 mg/ml
RNase A Solution	RNase A in aq. dd. *Storage at -20°C*	10 mg/ml
RPMI medium (standard or hormone-reduced)	FCS, normal or charcoal stripped Penicillin-Strepto-mycin solution in RPMI 1640 *storage at 4°C*	15 % (v/v) 1 % (v/v)
SDS Solution, 20 % (DNA Isolation)	SDS in aq. dd.	20 % (w/v)
SE - Buffer (DNA Isolation)	$MgCl_2$ EDTA in aq. dd. *ad pH 8.0 (NaOH), ac.*	75 mM 25 mM
SSC (20x)	NaCl $Na_3Citrat \times H_2O$ in aq. dd. *ad pH 7.0 (HCl), ac.*	3 M 0.3 M
SSCT	20x SSC Tween 20 in aq. dd.	20 % (v/v) 0.15 % (v/v)
SSCTM *(prepare fresh)*	milk powder in SSCT *15 min 42°C in a shaking water bath,* *centrifugation 5 min, full speed,* *filtrate with 0.4 µm pore size*	5 % (w/v)
TBE (1x)	TRIS Boric acid EDTA in aq. dd.	89 mM 89 mM 2 mM
TE (10x)	TRIS-HCl pH8 EDTA in aq. dd. *ad pH 8.0, ac.*	10 mM 1 mM

TE - Buffer for TaqMan® probe dilution	TRIS EDTA in RNase-free H_2O (do not use DEPC) *pH ad 8.0 (HCl)*	1 mM 10 µM
Tissue Disruption Buffer sufficient for one DNA isolation from tissue *(prepare fresh)*	SE - Buffer SDS Solution, 20 % Proteinase K sol. RNase A solution	2 ml 200 µl 9.8 µl 40 µl

2.2.3 Antibodies, enzymes, reagent systems and kits

100 bp DNA Ladder	New England Biolabs, Ipswich, USA
Anti-digoxigenin-fluorescein, Fab fragments	Roche Diagnostics, Mannheim, Germany
Anti-fluorescein, goat IgG fraction, Alexa Fluor 488	Life Technologies, Carlsbad, USA
BACMAX® DNA Purification Kit	Epicentre Biotechnologies, Madison, USA
BioPrime® DNA Labeling System	Life Technologies, Carlsbad, USA
Digest-All 3 (Pepsin)	Life Technologies, Carlsbad, USA
GeneRuler™ Ultra Low Range DNA Ladder	Fermentas, St. Leon-Rot, Germany
Illustra GenomiPhi V2 DNA amplification Kit	GE Healthcare, Munich, Germany
Illustra ProbeQuant G-50 Micro Columns	GE Healthcare, Munich, Germany
PAXgene Blood RNA Kit	PreAnalytiX GmbH, Hombrechtikon, Switzerland
PAXgene Blood RNA Tube	PreAnalytiX GmbH, Hombrechtikon, Switzerland
Proteinase K	Sigma-Aldrich, St. Louis, USA
RNase A, DNase free	Fermentas, St. Leon-Rot, Germany
RNase free DNase Set	Qiagen, Hilden, Germany
RNeasy Mini Kit	Qiagen, Hilden, Germany
RNeasy Mini Kit Plus	Qiagen, Hilden, Germany
SPoT-Light Tissue Pretreatment Kit	Life Technologies, Carlsbad, USA
Superscript™ III Reverse Transkriptase	Life Technologies, Carlsbad, USA
Taq DNA Polymerase (cloned)	GE Healthcare, Munich, Germany
TaqMan® Genotyping Master Mix	Life Technologies, Carlsbad, USA
QIAamp DNA FFPE Tissue Kit	Qiagen, Hilden, Germany
QIAprep Spin Miniprep Kit	Qiagen, Hilden, Germany
QIAshredder spin columns	Qiagen, Hilden, Germany
QuantiFast® Multiplex RT-PCR Kit + R	Qiagen, Hilden, Germany

2.2.4 Cell lines

For cell culture applications, the human PrCa cell line **LNCaP** was used (purchased from the American Type Culture Collection (ATCC), Manassas, USA), which originated from a lymph node metastasis and are androgen sensitive.

2.2.5 Bacterial artificial chromosomes (BACs)

BACs, needed for probe construction for the *TMPRSS2-ERG* fluorescence in situ hybridization (FISH) application, were kindly provided by Ph.D. Jeremy Clark (Institute of Cancer Research (IRC), Sutton, UK) and were originally obtained from the Children's Hospital Oakland Research Institute (CHORI). The BACs are listed in Table 5.
The BACs were cultured in the *Escherichia coli* strain DH10B.

Table 5: Information on the Bacterial artificial chromosomes (BACs) used in this work. A schematic map for the human genomic inserts is presented in Figure 6.

BAC name	Size [kbp]	Position	Application
CTD-2511E13 RP11-95G19 RP11-720N21	220 185 210	centromeric from *ERG*	FISH probe preparation , labeling with DIG
RP11-115E14 RP11-372O17 RP11-729O4	175 190 185	between *TMPRSS2* and *ERG*	FISH probe preparation , labeling with Biotin

2.2.6 Primer oligonucleotides and TaqMan® assays

Primer oligonucleotides for standard PCR applications and for custom TaqMan® gene expression assays were purchased from biomers.net (Ulm, Germany). The sequences are listed in the Tables 6 and 7. 100 µM stock solutions and 10 µM working solutions were prepared using H_2O Ultrapure.

TaqMan® probes for custom TaqMan® gene expression assays were purchased from Life Technologies (Carlsbad, USA) and were diluted in working solutions (10 µM) using a TE buffer with low EDTA concentration (2.2.2) as recommended by the supplier. The sequences are listed in Table 7.

The ***predesigned and custom TaqMan® SNP Genotyping Assays*** (in 40 x concentration, listed in Table 8) were also obtained from Life Technologies (Carlsbad, USA).

Table 6: Primers for qualitative transcript analyses. The names and sequences of the forward (fwd) and reverse (rev) primers are given. The next columns contain the optimal annealing temperatures (Temp.), the optimal $MgCl_2$ concentrations and the expected product length (in base pairs). The PCR of the putative *PARG* revealed no product and in consequence, the optimal reaction conditions are not known.

Target	Type	Identifier	Sequences (5' - 3')	Temp.	MgCl	Length
MSMB	fwd	h_n4_MSMB	GTGATCTTTGCCACCTTCGT	55°C	2.5 mM	408/ 302
	rev	r_n3_MSMB	GATAGGCATGGCTACACAATCA			
putative *PARG*	fwd	h_putPARG	CTGCAATCTCGGCACTTTG	?	?	286
	rev	r_putPARG	ACGGTGGGACCCTGAACT			

Table 7: Primers and TaqMan® probes for the quantitative real-time PCR. The names and sequences of the forward (fwd) and reverse (rev) primers are given. The 5' - fluorescence dyes (FAM or VIC) and the 3' - modifications (minor groove binder (MGB) and non-fluorescent quencher (NFQ)) of the TaqMan® probes are indicated in dark blue. The last two columns contain the temperatures (Temp.) for the annealing/extension step and the amplicon lengths (in base pairs).

Target gene	Type	Identifier	Sequences (5' - 3') and conjugates	Temp.	Length
ALAS1	fwd	f_n3-ALAS1	TGATGAACTACTTCCTTGAGAATC	60°C	70
	rev	r_n3_ALAS1	GAATGAGGCTTCAGTTCCA		
	probe	p_n3_ALAS1	FAM-CTAGTCACATGGAAGCAA-MGB-NFQ		
ESCO1	fwd	f_n4c_ESCO1_QPCR	ATTTGGCCAATGAGATAAAACCT	60°C	98
	rev	r_n4b_ESCO1_QPCR	CAAACATTGGCTGAAATTCTTAT		
	probe	p_ESCO1	FAM-ATGCTGAATCAAAAGAATGT-MGB-NFQ		
G6PD	fwd	f_n1c_G6PD	CCGGGCATGTTCTTCAA	60°C/ 62°C	78
	rev	r_n1_G6PD	AGGGAGCTTCACGTTCTTGTAT		
	probe	p_n1_G6PD	VIC-TCCAGCTCCGACTCC-MGB-NFQ		
MSMB	fwd	h_n2_MSMB	CTGTCTATAAGGAGTCCTGCTTATCA	60°C	88
	rev	r_n8_MSMB	TGCATTGCATAAAGTCACGA		
	probe	p_MSMB	FAM-ATCTTTGCCACCTTCGT-MGB-NFQ		
MYC	fwd	f_n2b_MYC	GTCCTCGGATTCTCTGCTCTC	60°C	123
	rev	r_n2_MYC	TCATCTTCTTGTTCCTCCTCAGA		
	probe	p_MYC	FAM-ATGAGGAGACACCGCC-MGB-NFQ		
NCOA4	fwd	h_n2_NCOA4	CAGCAGCTCTACTCGTTATTGG	60°C	103
	rev	r_n2_NCOA4	TCTCCAGGCACACAGAGACT		
	probe	p_NCOA4	FAM-TCAATTGTCTTACTCATCAAC-MGB-NFQ		
PARG	fwd	h_n3_PARG	GCACTGAGATCGTTGCCAT	60°C	106
	rev	r_n2_PARG	AGAAATCCACAGTAAGCCTTGTT		
	probe	p_PARG	FAM-CTTCACTTCAGACGCTAC-MGB-NFQ		
SOX9	fwd	f_n3k_SOX9	CACTTGCACAACGCCGA	60°C	65
	rev	r_n3e_SOX9	TCGCTCTCGTTCAGAAGTCTC		
	probe	p_SOX9	FAM-CAAGACGCTGGGCAA-MGB-NFQ		
TIMM23	fwd	h_n3_TIMM23	ATTTGGTGTCATCATTGAGAAAAC	60°C	115
	rev	r_n3_TIMM23	TGCTATCCCTCGAAGACCA		
	probe	p_TIMM23	FAM-ATGACCTTAACACAGTAGCA-MGB-NFQ		
TIMM23B	fwd	h_n1_TIMM23	TGCATTTGGTGTCATCATTG	60°C	121
	rev	r_n1_TIMM23B	GGAATCCAAAGCCATCTCTG		
	probe	p_TIMM23B	FAM-ATGACCTTAACACAGTAGCA-MGB-NFQ		
TMPRSS2	fwd	f_n5h_TMPRSS2	CAGGGACATGGGCTATAAGAA	60°C/ 62°C	81
	rev	r_n5c_TMPRSS2	AGTTTCATAAAGCTGGTGGATC		
	probe	p_n5b_TMPRSS2	FAM-CTATTCCTTGGCTAGAGTAA-MGB-NFQ		
TMPRSS2-ERG(T1G4)	fwd	f_n6_T1G4_QPCR	TGGAGCGCGGCAGGAAG	62°C	67
	rev	r_n14_T1G4_QPCR	TCCGTAGGCACACTCAAACA		
	probe	p_T1G4-FAM	FAM-TCCTCACTCACAACTGATAA-MGB-NFQ		

Table 8: TaqMan® SNP Genotyping Assays. Given are the SNP IDs, their corresponding genomic region, as well the allele classification and the assay ID from the supplier (Life Technologies, Carlsbad, USA). All SNPs were kindly provided by the PRACTICAL consortium. The green highlighted SNPs were also used for further functional analyses.

SNP ID	Genomic region	Non-risk/ risk alleles	Assay ID (Applied Biosystems)
rs721048	2p15	G/A	C____579489_20
rs1465618	2p21	G/A	C___7523218_10
rs12621278	2q31	G/A	C__27373730_10
rs2660753	3p12	C/T	C__16059981_10
rs17021918	4q22	T/C	C___1569779_10
rs12500426	4q22	C/A	C__11346761_20
rs7679673	4q24	A/C	C___2551995_10
rs9364554	6q25	C/T	C___2737073_10
rs10486567	7p15	A/G	C___2616676_20
rs6465657	7q21	T/C	C___2691796_20
rs1447295	8q24 (region 1)	C/A	C___2160574_30
rs6983267	8q24 (region 3)	T/G	C__29086771_20
rs16901979	8q24 (region 2)	C/A	C__33280526_10
rs2928679	8p21	C/T	C___1382529_10
rs1512268	8p21	G/A	C___1593767_10
rs10993994	10q11	C/T	C____178879_10
rs4962416	10q26	T/C	C__29032910_20
rs7931342	11q13	T/G	C____146313_10
rs7127900	11p15	G/A	C__30685144_20
rs4430796	17q12	G/A	C__30685144_10
rs11649743	17q12	A/G	C___2559889_10
rs1859962	17q24	T/G	C__11942243_10
rs2735839	19q13	A/G	C__26638655_10
rs5759167	22q13	T/G	Custom: AHRRS8U
rs5945619	Xp11	T/C	C__26215761_10

2.2.7 Small interfering RNAs (siRNAs)

All siRNAs were obtained from Qiagen (Hilden, Germany) and are listed in Table 9. So-called FlexiTubes were preferred, which contain four siRNAs against each target, complementary to different motives within the RNA sequences of the target gene. The AllStars Negative Control siRNA has no homology to any known mammalian gene.

Table 9: List of small interfering RNAs (siRNAs).

Target gene	Product name
-	Allstars Negative Control
ESCO1	Hs_ESCO1_5/6/7/8
MSMB	Hs_MSMB_5/7/8/9
NCOA4	Hs_NCOA4_1/3/4/5
PARG	Hs_PARG_2/5/6/7
TIMM23	Hs_LOC100287932_1/2/3/4

2.2.8 Laboratory equipment and consumables

2.2.8.1 Equipment

Axioplan 2 Imaging System	Carl Zeiss, Oberkochen, Germany
Epoch microplate spectrophotometer incl. Take 3 micro-volume plate	BioTek, Winoosky, USA
Fast Real-Time PCR System 7900 HT incl. 96-Well Fast Block and 384-Well Block modules	Life Technologies, Carlsbad, USA
Flowcytometer CCA	Partec, Münster, Germany
Matrix 8-channel 384 Equalizer Pipette, 1-30 µl	Thermo Fisher Scientific, Hudson, USA
Microtom Leica CM1900 cryostat	Leica Microsystems, Nussloch, Germany
NanoDrop Photometer	Implen, Munich, Germany
Packard Multiprobe II Robotic Liquid Handling System	PerkinElmer, Waltham, USA
Gammacell 2000 Cs-137 source, 3.3 Gy/min	Molsgaard Medical, Heorsholm, Denmark
SpotLight® CISH™ Hybridizer	Life Technologies, Carlsbad, USA
Thermocycler Biometra T-Gradient	Biometra, Göttingen, Germany
ViiA™ 7 Real-Time PCR System incl. 96-Well Fast Block and 384-Well Block modules	Life Technologies, Carlsbad, USA

2.2.8.2 Consumables

Standard plastic ware such as pipet tips and reaction tubes were purchased from Sarstedt (Nümbrecht-Rommelsdorf, Germany), Eppendorf (Hamburg, Germany) and BD Falcon (Franklin Lakes, USA). The following are the more special consumables.

Cell culture vessels (t25 and t75 flasks, flat bottom plates)	Sarstedt, Nümbrecht-Rommelsdorf, Germany
Cell strainer, 40 µm, nylon	BD Falcon, Franklin Lakes, USA
Cover slips, 22 x 50 mm, #0 and #1.5	Menzel, Braunschweig, Germany
Flat bottom cell culture tubes	Nunc, Thermo Fisher Scientific, Waltham, USA
Impact 384 Tips, sterile	Thermo Fisher Scientific, Hudson, USA
MicroAmp® Fast Optical 96-Well Reaction Plates	Life Technologies, Carlsbad, USA
MicroAmp® Optical 384-Well Reaction Plates	Life Technologies, Carlsbad, USA
MicroAmp® Optical Adhesive Films	Life Technologies, Carlsbad, USA
PAP pen for immunostaining	Sigma-Aldrich, St. Louis, USA
RoboRacks, 20 µl Clear Non-Conductive Tips	PerkinElmer, Waltham, USA
S-Monovette® Potassium-EDTA (7.5 or 9 ml)	Sarstedt, Nümbrecht-Rommelsdorf, Germany
S-Monovette® Lithium-Heparin (7.5 or 9 ml)	Sarstedt, Nümbrecht-Rommelsdorf, Germany
Superfrost glass slides	Menzel, Braunschweig, Germany

2.2.9 Software and internet resources

2.2.9.1 Software

Flow cytometer histogram display	
WinMDI v. 2.9	Flow Cytometry Core Facility, The Scripps Research Institute, La Jolla, USA
Microscopy	
Metafer 4 – MetaCyte v. 3.6.7 with TMA-Tool	Metasystems, Altlussheim, Germany
Metafer 4 – Msearch v. 3.6.4	Metasystems, Altlussheim, Germany
Office applications	
Microsoft® Office 2010	Microsoft, Redmont, USA
Reference Manager v. 10	Thomson Reuters, Carlsbad, USA
QRT-PCR/genotyping	
7900 HT Sequence Detection Systems (SDS)	Life Technologies, Carlsbad, USA
Primer Express v. 3.0	Life Technologies, Carlsbad, USA
RQ Manager v. 1.2	Life Technologies, Carlsbad, USA
ViiA™7 Software v. 1.2	Life Technologies, Carlsbad, USA
Statistics	
Minitab v. 16.2.3	Minitab Inc., State College, Pennsylvania, USA
PLINK v. 1.0711	Shaun Purcell at CHGR, MGH and the Broad Institute of Harvard & MIT http://pngu.mgh.harvard.edu/purcell/plink/
Review Manager v. 5.1.7	Copenhagen: The Nordic Cochrane Centre, The Cochrane Collaboration, 2012 http://ims.cochrane.org/revman
SAS v. 9.3	SAS Institute Inc., Cary, USA
Statview v. 5.0.1	SAS Institute Inc., Cary, USA

2.2.9.2 Bioinformatics online tools, genome browsers and databases

Primer 3 Plus	http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi
NCBI Blast	http://blast.ncbi.nlm.nih.gov/Blast.cgi
Reverse complement tool	http://www.bioinformatics.org/sms/rev_comp.html
Two state melting tool	http://mfold.rna.albany.edu/?q=DINAMelt/Two-state-melting
Homodimer simulation tool	http://mfold.rna.albany.edu/?q=DINAMelt/Homodimer-simulations
HUGO Gene Nomenclature Committee (HGNC)	http://www.genenames.org/
Online Mendelian Inheritance in Man (OMIM) database	http://www.ncbi.nlm.nih.gov/omim
Ensembl genome browser	http://www.ensembl.org/Homo_sapiens/Info/Index

Z-Score to Percentile Calculator	http://www.measuringusability.com/pcalcz.php
Pubmed literature database	http://www.ncbi.nlm.nih.gov/pubmed
SNP database	http://www.ncbi.nlm.nih.gov/snp/
UCSC genome browser	http://genome.ucsc.edu/

2.3 Methods

2.3.1 Cultivation of human cells

2.3.1.1 LNCaP prostate cancer cell line

The adherent growing LNCaP cell line was maintained in RPMI medium containing 15 % FCS and 1 % Penicillin/Streptomycin. The cultures were incubated at 37°C in a humidified atmosphere with 5 % CO_2. Cells were cultivated in t75 cell culture flasks and were passaged when they reached 60–80 % confluence to keep up exponential growth by the following standard procedures: The culture medium was aspirated; the cells were then gently washed with PBS and trypsinized with 3 ml of a trypsin EDTA solution at 37°C until the cells appeared detached from the surface (~ 5 min). The trypsinization reaction was stopped with 7 ml of RPMI medium; the cell suspension was transferred into a 15 ml tube and centrifuged for 5 min at 1000 rpm at RT. After removing the supernatant, the cell pellet was resuspended thoroughly in 2 ml RPMI medium and seeded in the ratio 1:3 to 1:5 depending on the current growth rate. For the *TMPRSS2-ERG* induction assay, charcoal-stripped FCS was used for media from the time point of transfection.

2.3.1.2 Short-time lymphocyte cultures

Short-time lymphocyte cultures obtained from whole blood were needed for the micronucleus assay and the mitotic delay assay. For this purpose, 500 µl of heparin-stabilized whole blood was added to 5 ml lymphocyte culture medium in flat bottom culture tubes. The incubation was carried out in a horizontal position at 37°C. For mitogenic stimulation, the growth medium was provided with phytohaemagglutinine (PHA).

2.3.2 Isolation of nucleic acids from different sources

2.3.2.1 DNA extraction from blood lymphocytes

Due to its high yield, quality, integrity and availability, DNA from peripheral blood lymphocytes is commonly used for genetic studies. The isolation was performed according to a modified protocol from Miller and colleagues [106]. To lyse the erythrocytes, the whole blood was transferred from the EDTA-monovette (7.5 or 10 ml) into a 50 ml centrifuge tube, mixed with 30 ml lysis buffer and incubated on ice in a horizontal position for 20 min followed by a centrifugation for 15 min at 1200 rpm at room temperature. After the supernatant was discarded, the pellet was suspended in 5 ml lysis buffer and centrifuged again using the same conditions. This washing step was repeated until the pellet appeared clean (milky-white). For protein lysis, 5 ml SE buffer was added to the pellet, suspended thoroughly and mixed with 50 µl proteinase K (10 mg/ml) and 250 µl 20 % SDS solution. The tube was incubated in a shaking water bath overnight at 55°C. To purify the DNA, an alcohol precipitation (a modified version from 2.3.4.1) was performed: 1.8 ml saturated NaCl solution (6 M) was added to the protein-DNA mixture, vortexed rigorously for 15 sec and centrifuged for 15 min at 5000 rpm at room temperature. The supernatant was transferred into a new 50 ml tube and approximately two volumes of absolute ethanol were added. Subsequently, the DNA was precipitated by gentle inversion of the tube. The precipitate was then collected with a plastic spatula and washed in 70 % EtOH. Depending on the amount of the precipitate, the DNA was dissolved in 200 to 800 µl TE buffer overnight. The DNA solution was stored at 4°C in a screw-cap tube for long-term storage.

2.3.2.2 RNA extraction from short-time lymphocyte cultures

RNA from lymphocyte cultures (set-up described in section 2.3.1.2) was prepared as follows: The culture medium was transferred into a 50 ml centrifuge tube and was mixed with 35 ml PBS (to remove heparin). After a centrifugation step for 10 min at 1,200 rpm at room temperature, the supernatant was carefully removed. The pellet was then treated with lysis buffer analogously as described in section 2.3.2.1 until all erythrocytes were removed. The RNA isolation was performed using the RNeasy Mini Kit (QIAGEN). Therefore, the clean pellet was treated with 350 µl RLT buffer (Kit component), supplemented with 1 % 2-mercaptoethanol. The lysate was then transferred onto a

QIAShredder spin column and centrifuged at 14,000 rpm for 2 minutes. Afterwards, the supernatant was transferred into a new tube, supplied with 1 volume of 70 % EtOH and pipetted onto the RNeasy spin column. The further purification and washing steps, as well as an on-column DNA digestion with the RNase free DNase Set (QIAGEN), were performed according to manufacturer's instructions. For the final elution, 2 x 30 µl of RNase-free H_2O were used. The concentration was determined by means of a NanoDrop photometer (2.3.3) and the integrity and purity was checked using a 1 % agarose gel (2.3.5). The RNA samples were stored at -70°C until further use.

2.3.2.3 DNA and RNA from fresh frozen prostate tissue

Tissue slicing using a Leica microtome cryostat

The tissue bank of the molecular genetics project for PrCa consists of tumor- and normal tissue specimen from 35 PrCa cases. The material was cut into blocks of 0.5–1 cm lateral length after prostatectomy, subsequently embedded in Tissue-Tek® and stored at -70°C. A Leica CM1900 cryostat was used to prepare tissue slices of 10 µm thickness, which are suitable to homogenize for nucleic acid extraction. Prior to tissue preparation, the chamber of the cryostat was precooled to -20°C and the specimen head to -30°C. The tissue block was attached on the specimen disk by freezing up with 0.9 % NaCl solution and was subsequently placed on the specimen head. Both, DNA and RNA extraction required approx. 10 mg, maximum 20 mg fresh tissue. For this purpose, an appropriate number of 10 µm frozen sections were prepared and transferred into 200 µl Tissue Disruption buffer for DNA extraction or 350 µl RLT buffer (taken from RNeasy Mini Kit, supplemented with 1 % 2-mercaptoethanol) for RNA extraction. The specimen was kept at -70°C until further processing.

DNA extraction from frozen tissue sections

The tissue was homogenized using a plastic mortar and mixed with additional 2 ml Tissue Disruption buffer. In order to disrupt the cells and denature proteins, the samples were incubated overnight in a gently shaking water bath at 37°C. For salt precipitation of DNA, 600 µl of a saturated NaCl solution was added, followed by vigorously vortexing for 15 sec and a subsequent centrifugation step (15 min, RT, 4000 rpm). The supernatant was transferred into a fresh 50 ml centrifuge tube without traces of the pelleted cell debris. If

the solution appeared turbid, the centrifugation step was repeated. The precipitation and the subsequent steps were performed according to 2.3.2.1.

RNA extraction from frozen tissue sections

All steps were performed at room temperature using the RNeasy Mini Kit (QIAGEN). The tissue kept in RLT buffer was homogenized using a 0.9 mm needle and syringe (20-G) and subsequently transferred onto a QIAShredder spin column and centrifuged at 14,000 rpm for 2 min. The lysate was repeatedly centrifuged at 14,000 rpm for 3 min to pellet down any cell debris leftover. The supernatant was transferred into a new tube, supplied with 1 volume of 70 % EtOH and pipetted onto the RNeasy spin column. The further purification and washing steps, as well as an on-column DNA digestion with the RNase free DNase Set (QIAGEN), were performed according to manufacturer's instructions. For the final elution, 2 x 20 µl of RNase-free H_2O were used. The concentration was determined by means of a NanoDrop photometer (2.3.3) and the integrity and purity was checked using a 1 % agarose gel (2.3.5). The RNA samples were stored at -70°C until further use.

2.3.2.4 RNA from prostate cancer cell lines

RNA from PrCa cell lines was used for gene expression analyses (2.3.8). Untreated cells from a t25 or one third of a t75 cell culture flask were trypsinized as indicated in 2.3.1.1 and washed with PBS. The clean cell pellet was suspended in 350 µl RLT buffer, supplemented with 1 % 2-mercaptoethanol (taken from RNeasy Mini Kit). For the extraction from cells cultivated on 6-well plates, the RLT buffer was applied directly on the adherent cells, which have been previously exempted from the culture medium and washed with PBS. The cells were shredded by pipetting up and down and the lysate was then transferred into 1.5 ml reaction tubes and kept at -70°C until further processing. In order to homogenize the cells, the lysate was pipetted onto a QIAShredder spin column (QIAGEN) prior centrifugation at 14,000 rpm for 2 min. The flow-through was then mixed with 350 µl 70 % EtOH and transferred onto an RNeasy spin column. The following purification and washing steps, as well as the on-column digestion with the RNase free DNase Set (QIAGEN) were performed according to manufacturer's instructions. For the final elution, 2 x 30 µl (final volume 60 µl) of RNase-free H_2O were used. In the case of cells undergoing T2E quantification, 2 x 30 µl using the same flow-through two times

(final volume 30 µl) to concentrate the RNA were used. The RNA samples were stored at -70°C until further use.

2.3.2.5 DNA from FFPE prostate tissue

The extraction of DNA from FFPE tissue was required for individual cases when no blood sample was available for genotyping. This exceptional method was used on a series of patients with known T2E status, which should be genotyped for common variants to determine possible associations with fusion status.

For this purpose, the QIAamp DNA FFPE Tissue Kit from QIAGEN with minimal deviations from the manufacturer's protocol was used: Depending on the size of the tissue area, 5–8 10 µm sections were prepared prior placing on a mounting slide and exempting from the surrounding paraffin with a scalpel roughly. If possible, only normal tissue was dissected for genotyping, and this was transferred into 1.5 ml reaction tubes. As specified in the manufacturer's protocol, the tissue was then treated with xylene to dissolve the remaining paraffin. This step was, as a modification of the procedure recommended in the manual, carried out several times. After a washing step with absolute EtOH for xylene removal, the reaction tubes were incubated with open lid at 37°C until all EtOH was completely evaporated. The dried tissues were then resuspended in a special lysis buffer and treated with proteinase K (both kit components). The incubation time of 1 h at 56°, as specified in the manual, had often to be exceeded considerably in order to ensure a sufficient enzymatic digestion. Afterwards, a further incubation for 1 h at 90°C allowed specific buffer components to reverse at least some of DNA modifications caused by formaldehyde. The final purification was carried out according to manufacturer's instructions using filter columns. The DNA was eluted in 50 µl ATE buffer (kit component).

2.3.3 Quantification of nucleic acids

The concentrations of nucleic acid solutions were determined photometrical using a one-channel NanoDrop photometer (Implen) or a microplate spectrophotometer with 16-channel micro-volume plate (Biotec).

The principle of the method is based on the characteristic absorption of nucleic acids at a wavelength of 260 nm. Deriving from the Lambert-Beer law E = ε*c*d, the extinction (E) of a solution depends on the extinction coefficient (ε), the concentration (c) and the path length (d). Applying the extinction coefficient for DNA and RNA of ε = 0.02 and 0.025 $\mu l*ng^{-1}*cm^{-1}$, respectively, an extinction of 1 and a path length of 1 cm, results in a concentration of 50 ng/µl for double stranded DNA and 40 ng/µl for RNA at an OD of 260 nm. As a linear relationship is reliable only at extinctions between 0.1 and 1, standard photometers require sample dilution when higher ranges are anticipated. However, the micro-volume technique implements an extremely small path length, which allows measuring high-concentrated solutions. Thus, a pre-dilution is not necessary and pipet errors can be avoided.

2.3.4 Purification of nucleic acids

2.3.4.1 Alcohol precipitation

Nucleic acids precipitate in the presence of alcohol due to the withdrawal of the hydrate envelope. This effect is further enhanced by salt addition as the salt shields their negative phosphate groups.

The standard procedure is as follows: 1/10 volume NaAc and 2.5 volumes of ice-cold absolute ethanol were added to the nucleic acid solution and incubated at -20°C over night. Then, the solution was centrifuged at 14,000 rpm for 15 min at 4°C, and the pellet washed twice with 70 % ethanol. Afterwards, the pellet was dried at room temperature or at 37°C, and finally dissolved in an appropriate amount of nuclease-free water or TE buffer.

Within the preparing process of the T2E FISH probes (2.3.11.1), a modified precipitation method was used: The nucleic acid solution was added with 1/20 volume of 4M NaCl and 1 volume of isopropanol, incubated for 20 min at -20 ° C and subsequently centrifuged at 14,000 rpm for 10 min at room temperature. The resulting pellet was then washed with 80 % ethanol and dried before dissolving in an appropriate volume TE buffer.

2.3.4.2 Phenol-chloroform extraction

This method is used when a highest purity of nucleic acids is required. It uses the property of phenol and chloroform for extremely efficient denaturation of proteins, and in addition, the effect of nucleic acids accumulating selectively in the aqueous phase for isolation.

In the first step, one volume of a phenol:chloroform:isopropanol (25:24:1) mixture was added to the sample solution and incubated for 5 min at RT under constant inverting, then centrifuged for 10 min at 14,000 rpm and then the upper, aqueous phase was carefully transferred to a new reaction tube. The carry-over was then mixed with one volume of chloroform and treated again as described above, including incubation and centrifugation. Subsequently, the aqueous phase was transferred to a new reaction tube. The obtained nucleic acid solution was then subjected to an ethanol precipitation (2.3.4.1) to remove any solvent residues.

2.3.5 Agarose gel electrophoresis

The agarose gel electrophoresis is a simple method to separate nucleic acids by size to examine a PCR product or a RNA isolate or obtain DNA fragments for subsequent cloning experiments. The principle of this method is the retention of negatively charged phosphate groups of nucleic acids in an electric field through an agarose gel matrix. The steric hindrance is more effective on longer molecules, which therefore move through the gel slower than small fragments. Depending on the pore size of the matrix, i.e. according on the agarose content, different size ranges of nucleic acids can be separated. Mainly 1.5 % agarose gels were used, since they are well suited for DNA fragments in the size range of 300 bp to 3 kb. For smaller fragments, such as qRT-PCR amplicons, 2.5 % agarose gels were used and for larger fragments and for checking RNA isolates, 0.8 or 1.0 % gels.

2.3.6 Polymerase chain reaction (PCR)

The polymerase chain reaction is a common method for the in vitro amplification of specific DNA fragments and it was the basis for many applications used in the present work, e.g. genotyping and gene expression analysis. The amplified region is determined

by two flanking primer DNA oligonucleotides. They anneal during the PCR reaction to the single stranded DNA template and are extended at their free 3'-OH ends by a DNA-dependent DNA polymerase. For the primer extension (so called elongation), free deoxyribonucleoside triphosphates (dNTPs) are required in the reaction tube, as well as a suitable PCR buffer, which ensures optimum reaction conditions. A typical PCR cycle consists of three steps in an iterative procedure. First, the double-stranded DNA is heat-denatured at 95–98 °C prior to an annealing phase in which the primers bind to the complementary sequence of the template. The optimal annealing temperature (usually between 50°C and 70°C) depends on the melting temperature of the primer and must be established for each pair of primers. In the third step, the elongation phase, the DNA polymerase extends the primer using the dNTPs according to the complementary sequence. The most commonly used polymerase in PCR applications is the thermostable Taq polymerase. The enzyme was originally cloned from the thermophilic bacterium *Thermus aquaticus* and has a reaction optimum at 72°C. PCR usually is terminated after 30–40 cycles to avoid interfering product accumulation and increasing amplification of unspecific fragments.

The standard PCR reaction was used for transcript analysis with cDNA as a template (synthesis see section 2.3.7). Components and conditions are given in Tables 10 and 11. The primer oligonucleotides are listed in Table 6.

Table 10: Standard PCR reaction mix.

Component	Concentration of the stock solution	Final con-centration/amount	Pipetting volume [µl]
10 x PCR buffer	10 x	1x	2.5
dNTPs	20 mM	200 µM	0.25
forward primer	10 pmol/µl	0.25 pmol/µl	0.625
reverse primer	10 pmol/µl	0.25 pmol/µl	0.625
$MgCl_2$	25 mM	1.5–2.5 mM, as needed*	
Taq polymerase	2.5 U/ µl	0.1 U/µl	0.1
template DNA	50–500 ng/µl	< 250 ng	1
nuclease free H_2O			ad 25

* The 10 x buffer contains 15 mM $MgCl_2$ corresponding to 1.5 mM end concentration. For an increase of 0.5 mM, 0.5 µl $MgCl_2$ solution was added

Table 11: Standard PCR cycling program.

Step	Time	Temperature	Repeats
Initial denaturation	2 min	95°C	1
Denaturation	30 sec	95°C	
Annealing	40 sec	variable	35
Elongation	1 min/kb	72°C	
Final elongation	10 min	72°C	1

2.3.7 cDNA synthesis

The synthesis of cDNA (copy DNA) is necessary for all objectives in which the analysis of mRNA with PCR methods is desired. The central enzyme for this process is an RNA-dependent DNA polymerase, which synthesizes the first DNA strand starting from the 3' end of a DNA primer. The resulting single stranded DNA is suitable as template for downstream PCR applications.

In this work, two different cDNA synthesis methods were used. Both are based on universal priming by random DNA hexamers, which facilitate cDNA synthesis on a broad variety of RNA species due to their shortness and sequence ambiguity. The first method was carried out with use of the Superscript™ III Reverse Transkriptase Kit (Life Technologies), for the qualitative analysis of the *MSMB* and *PARG* transcripts.

The second cDNA synthesis method was part of the one-step QuantiFast Multiplex RT-PCR +R Kit (QIAGEN, Hilden), as described in 2.3.8, which means that the cDNA first strand synthesis was performed prior PCR amplification. According to this concept, all involved reagents and enzymes were assembled in one reaction tube and processed as recommended by the manufacturer.

2.3.8 Quantitative reverse transcriptase PCR (qRT-PCR) for gene expression analysis

Within this work the qRT-PCR was used for a multitude of analyses. First, it was used to characterize the expressional pattern of the genes *MSMB*, *NCOA4*, *PARG*, *TIMM23*, *TIMM23B* (all located on 10q11), as well as *MYC* (8q24) and *SOX9* (17q24) in prostate tumor- and histologically normal tissue specimen of a total of 87 patients, in order to identify possible relationships between gene expression, genotypes and the somatic T2E fusion status. Second, qRT-PCR was utilized to phenotype the T2E fusion status by a

simple qualitative evaluation of the *T1G4* fusion transcript expression in prostate tumor RNA when no material was available for FISH analyses (2.3.11). Moreover, qRT-PCR played an essential role for the *TMPRSS2-ERG* induction assay (2.3.12), in which the knockdown of the target genes *ESCO1*, *MSMB*, *NCOA4*, *PARG* and *TIMM23* had to be verified as well as the de novo formation of the T2E fusion had to be quantified by qRT-PCR in the primary fusion negative LNCaP cells.

2.3.8.1 General background

The quantification of mRNA via qRT-PCR is a useful tool to study gene expression pattern of specific genes. As a prerequisite cDNA has to be generated from the mRNA (see 2.3.7). Then the abundance of a target transcript is quantified in a relative manner in relation to the expression levels of reference genes ("housekeeping genes") or absolutely with use of standard curves resulting from known target concentrations. Most of experimental approaches use the formerly mentioned relative quantification, as the comparison of gene expression between samples usually requires no absolute numbers and different measurement series are more comparable to each other. Reference genes should be selected for their stable expression, being independent from cell cycle stage or environmental influences. Since the quantification of the target and/or reference gene expression by analyzing the intensity of the PCR product bands on a gel matrix (endpoint quantification) is quite inaccurate, qPCR is usually performed at "real-time" by implementing a reporter dye in the reaction that produces a signal equivalent to the arising PCR product. This in turn can be carried out using amplicon-specific probes that are labeled with reporter dyes, or a dye which becomes detectable when intercalating into double strands, e.g. Sybr Green.

In the present work, gene expression analyses were performed using specific TaqMan® MGB probes. These short DNA probes are labeled with a fluorescence dye on 5' end (here: VIC or FAM) and on 3' end with a non-fluorescent quencher (NFQ), which suppresses the fluorescence signal emitted from the dye via fluorescence resonance energy transfer (FRET). Additionally, they are modified with a minor groove binder (MGB) that binds with high affinity to the minor groove of the DNA, increasing the melting temperature of the probe and thus allowing the design of shorter probes with a better quenching property (closer proximity between dye and NFQ). Flanked by the PCR primers,

the probe gets degraded during the elongation step of the PCR by the 5' - 3' exonuclease activity of the polymerase resulting in the emission of the reporter signal (Figure 2A). The abundance of target cDNA is proportional to the amount of dismantled probe, and thus to the intensity of the generated fluorescence signal. In order to minimize variations due to multiple pipetting steps, the QuantiFast Multiplex RT-PCR Kit (QIAGEN) was used, which allows cDNA synthesis and quantification within the same reaction. For PCR and detection, two platforms were used: The "Fast Real-Time PCR System 7900HT" and the "VIIA7 Real-Time PCR System" (Life Technologies), which measure the fluorescence signal after each PCR cycle. It is generally not recommended to change the measuring system during the work, but in this case it was unavoidable due to a move of the working group.

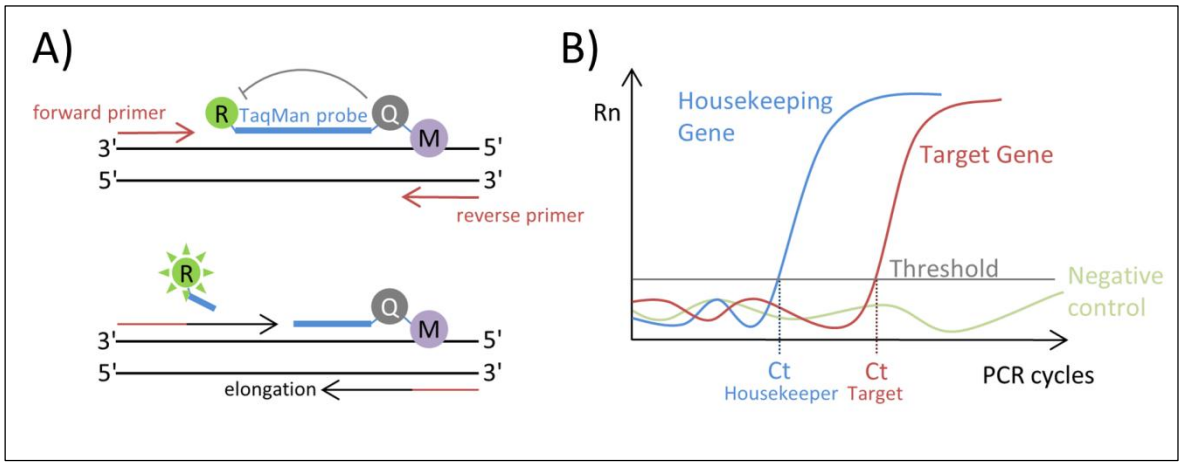


Figure 2: Overview of the mRNA quantification method using TaqMan® probes. A) Shown is a typical TaqMan® MGB probe and the principle of the quantification. The emission of the fluorescent reporter dye (R) on the 5' end of the probe is absorbed by the 3' non-fluorescent quencher (Q), as long as they are in close proximity to each other. Once the probe has been destroyed by the DNA polymerase in the course of the ongoing PCR, the reporter dye emits light which can be detected by appropriate devices. The special feature of this probe is the minor groove binder (M) which increases the affinity to the target sequence allowing for shorter probes and thus for an enhanced quenching efficiency (closer proximity of dye and quencher). **B) Schematic illustration of a typical curve profile of the fluorescence intensities.** The threshold marks that value on the y-axis where the normalized reporter intensity (Rn) reaches an intensity distinctly above the background corresponding to the start of the exponential phase of the PCR. The threshold cycle (Ct) which is used for the analysis is read at this point on the x-axis.

2.3.8.2 Reaction set-up

The expressional pattern of *MSMB*, *NCOA4*, *PARG*, *TIMM23*, *TIMM23B*, *MYC* and *SOX9* in tissue specimen were analyzed relative to the reference genes *ALAS1* and *G6PD*. In order to achieve highly accurate values and to avoid template inhibition, the RNA was diluted 1:50 prior adding 2 µl to the reaction. For all other experiments, undiluted RNA templates

were used, so for the assessment of the somatic T2E status in prostate tissue specimen by detection of the *T1G4* transcript, as well as the verification of the *ESCO1*, *MSMB*, *NCOA4*, *PARG* and *TIMM23* knockdown with solely *G6PD* as reference gene. All reactions were set as duplicates and as indicated in Table 12. The cycling programs were set as shown in Table 13.

Table 12: Standard reaction mix for the quantitative real-time PCR using custom TaqMan® assays and the QuantiFast Multiplex RT-PCR +R Kit. The concentration of the 2 x QuantiFast Master Mix is effectively 1.92 as ROX solution has to be added according to the manufacturer's recommendation. The primer/probe combinations are listed in Table 7.

Component	Concentration of the stock solution	Final concentration/amount	Pipetting volume [µl]
Forward primer	10 pmol/µl	0.1 (0.05*) µM	0.2 (0.1*)
Reverse primer	10 pmol/µl	0.1 (0.05*) µM	0.2 (0.1*)
TaqMan® probe	10 pmol/µl	0.1 (0.05*) µM	0.2 (0.1*)
2 x QuantiFast Multiplex RT-PCR Master Mix[a]	1.92 x	1 x	10.4
QuantiFast Multiplex RT Mix	2 U/µl	0.02 U/µl	0.2
RNA	50 –500 ng/µl	< 250 ng	1–3
RNase free H_2O			ad 20

[a] including 4 % ROX solution

Table 13: Cycling program for the quantitative real-time PCR using the QuantiFast Multiplex RT-PCR +R Kit. The optimal temperature for the annealing/elongation step for each assay is listed in Table 7.

Step	Time	Temperature	Repeats
1. cDNA synthesis	20 min	50°C	1
2. Heat-activation of DNA polymerase	5 min	95°C	1
3. Denaturation	15 sec	94°C	45
4. Annealing, elongation and data collection	1 min	60°C (or 62°C)	

A slightly different setup was chosen for the *T1G4* quantification within the *TMPRSS2-ERG* induction assay (2.3.12). Because of the very low abundance of the fusion transcript, the *T1G4* measurements were performed in quadruplicates using 3 µl undiluted template RNA per well. A possible template inhibition was tolerated in favor of a greater chance to detect a fusion transcript. To compensate for side-effects on the androgen pathway, the *T1G4* expression was quantified in relation to the wild-type transcript of the *TMPRSS2* gene, since both, the *T1G4* fusion and the wild-type *TMPRSS2* gene, share the same promoter. In order to detect possible irregularities in cell numbers, also *G6PD* was determined.

2.3.8.3 Data analysis using the $\Delta\Delta C_t$ method

The $\Delta\Delta C_t$ method [92] is a common approach for the analysis of gene expression data. It estimates the target gene expression relative to a reference gene (ΔC_t) and allows its comparison with a reference sample ($\Delta\Delta C_t$), e.g. an untreated control. Prerequisite for the ΔΔCt method is an almost equivalent PCR efficiency of reference and target gene assays (see 2.3.8.4), which has to be determined beforehand.

An essential term of the $\Delta\Delta C_t$ method is the so called threshold cycle (C_t), which is defined as the PCR cycle in which the fluorescence signal significantly advances from the background noise (Figure 2B). Typically, the threshold is determined individually for every single assay during assay establishment and is then kept in all related experiments.

For data analyses the mean C_t values of the reference gene were subtracted from the ones of the target gene, resulting in the ΔCt:

$$\Delta C_t = C_{t\ \text{target gene}} - C_{t\ \text{reference gene}} \qquad \textit{(Equation 1)}$$

To set the sample in relation to another sample (e.g. an untreated control), the $\Delta\Delta C_t$ is calculated:

$$\Delta\Delta C_t = \Delta C_{t\ \text{experiment}} - \Delta C_{t\ \text{control}} \qquad \textit{(Equation 2)}$$

Finally, the linear differential value is set in an exponential relationship:

$$\text{Relative expression} = 2^{-\Delta\Delta Ct} \qquad \textit{(Equation 3)}$$

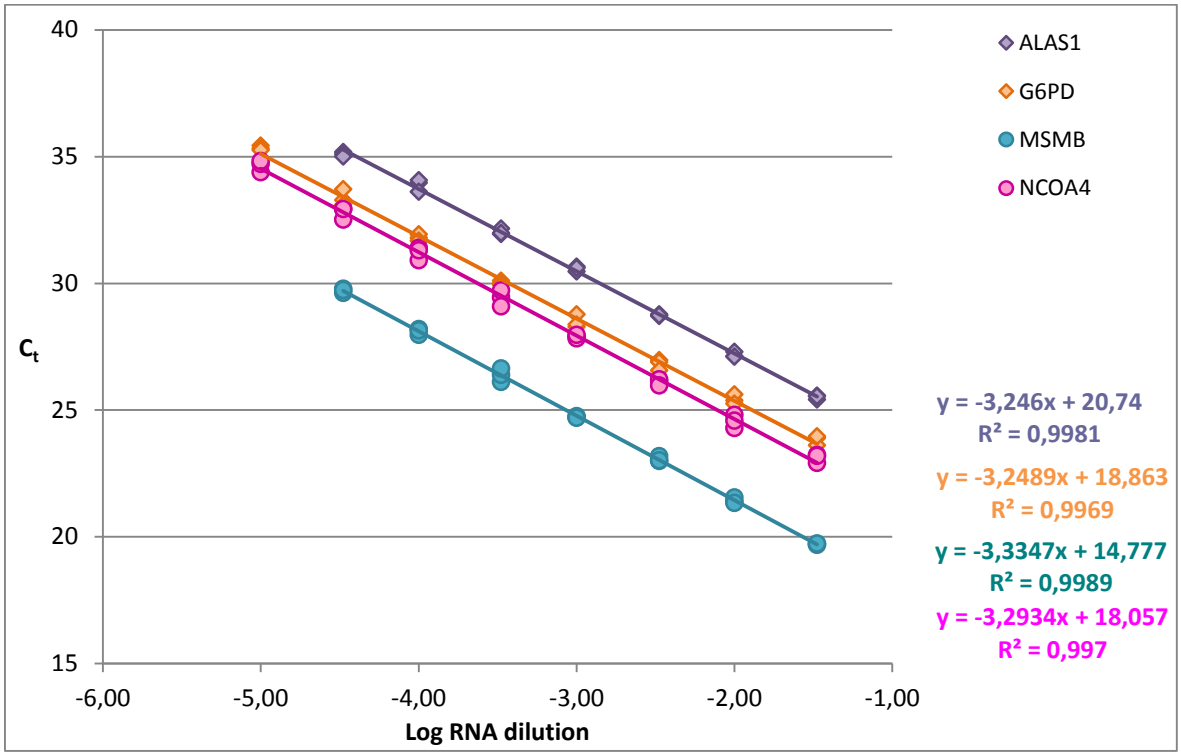
An experimental control sample is not always mandatory, e.g. if comparing expression values in different specimen. In such settings, only the $2^{-\Delta Ct}$ values are used.

For the T2E quantification in the course of the *TMPRSS2-ERG* induction assay, the *T1G4* values had to be corrected for failed measurements, which arose from the very low abundance of *T1G4* transcript: The $2^{-\Delta Ct}$ values of *T1G4* were divided by the total number of measurements and multiplied with the number of valid measurements.

2.3.8.4 Gene expression assay design and establishment

The primers and probes were designed according to the guidelines of the QuantiFast Multiplex RT-PCR +R Kit. When possible, exon-spanning primers were chosen to maximize cDNA specificity and to avoid unspecific amplification of residual genomic DNA, which is present in the RNA specimens. Prior the probe design, the PCR performance was evaluated by agarose gel electrophoreses (2.3.5). For that purpose 1:10, 1:100 and 1:1000 dilutions of suitable RNA templates were prepared, as well as a non-template control and a sample containing 50 ng of genomic DNA to uncover unspecific primer binding. The reaction setup of this pre-test and the PCR program were according to Tables 12 and 13, respectively. Probes were designed only for primer combinations, which exhibited a specific and robust amplification. The probes were designed using the Primer Express v3.0 software and purchased from Life Technologies.

As mentioned in 2.3.8.3, a prerequisite for the ΔC_t method is that the reference and target gene expression assays comprise similar PCR efficiencies. To verify this, standard curves were generated with different RNA dilutions in quadruplicates (1:10, 1:30, $1:10^2$, 1:300, $1:10^3$, etc.; optionally with further intermediate steps). As a general rule for the ΔC_t method, the PCR efficiencies should not differ more than 5 % and the slopes of the standard curves not more then 0.1 (logarithm of the RNA dilution factor versus C_t), respectively. A slope of 3.32 corresponds to an efficiency of 100 %. The standard curves for the 10q11 gene assays, as well as for *MYC* and *SOX9*, are displayed in the Figures 3 and 4. The assays for *TMPRSS2*, *TMPRSS2-ERG* and *ESCO1* were previously validated for comparability with the reference genes and were kindly provided by Dr. Manuel Luedeke [94]. All assays are listed in detail in Table 7.



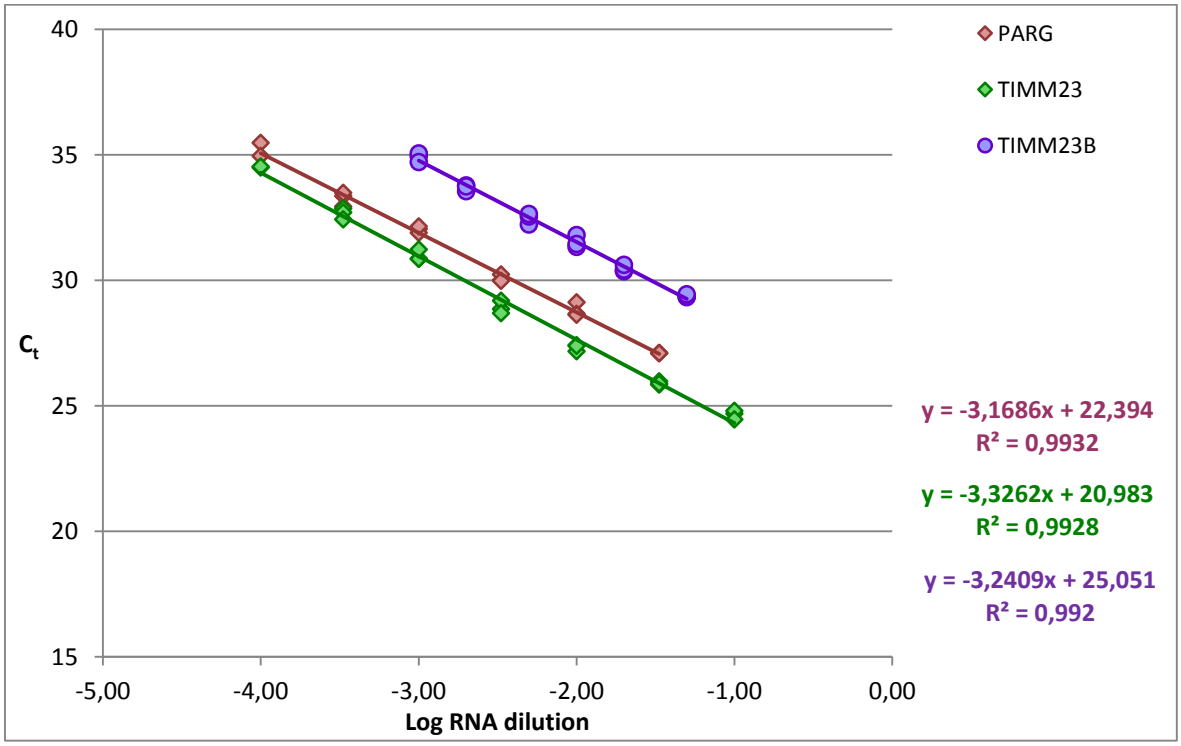


Figure 3: Standard curves for the reference genes *ALAS1* and *G6PD* and the target genes *MSMB*, *NCOA4*, *PARG*, *TIMM23* and *TIMM23B*. Displayed are the C_t values dependent on the logarithm (Log) of the RNA dilution factor (e.g. -2.0 correspond to 1:100 dilution). The equations are the linear functions of the standard curves. All target genes are comparable to *ALAS1* and *G6PD* as their slopes differ less than 0.1. The coefficients of determination (R^2) are close to one, indicating that the data points fit very well to the curves in a linear model.

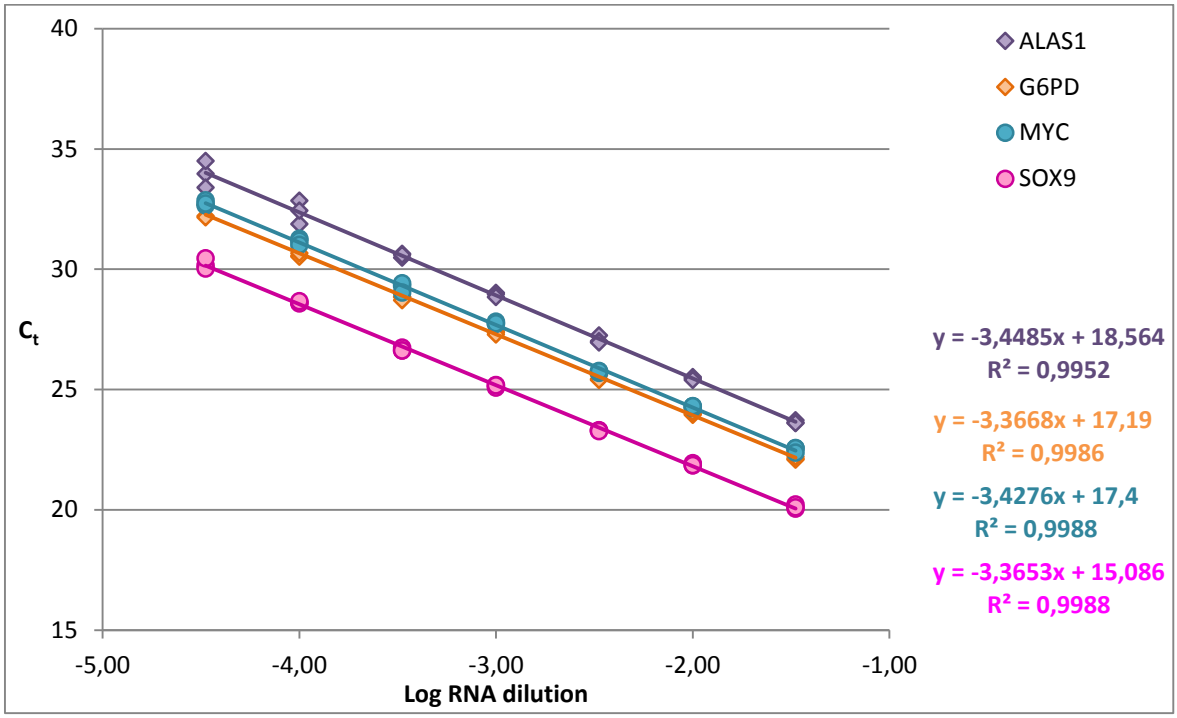


Figure 4: Standard curves for the reference genes *ALAS1* and *G6PD* and the target genes *MYC and SOX9.* Displayed are the C_t values dependent on the logarithm (Log) of the RNA dilution factor (e.g. -2.0 correspond to 1:100 dilution). The equations are the linear functions of the standard curves. All target genes are comparable to *ALAS1* and *G6PD* as their slopes differ less than 0.1. The coefficients of determination (R^2) are close to one, indicating that the data points fit very well to the curves in a linear model.

2.3.9 Genotyping of single nucleotide polymorphisms (SNPs)

The SNP genotyping plays a central role in the present work since a large number of PrCa risk associated common variants had to be determined in different sample sets to correlate them with DNA repair capacity and the presence of the somatic T2E fusion.

2.3.9.1 Chip-based genotyping using Illumina Infinium HD technology

The genotype data for the hypothesis generating sample set (Stage 1) for the association study of common variants and the T2E fusion were taken from a large chip-based genotyping initiative using the iCOGs chip [38]. Though these data were not generated in our lab, a brief overview about the methodology is given in the following section:

The collaborative approach of the COGS consortium used the Illumina Infinium HD technology. It is based on a microchip with small silica beads, on which oligonucleotide probes, complementary to the sequence right beside the SNP, are linked with (one SNP per bead). The amplified and fragmented genomic DNA is hybridized with the chip,

resulting in complementary binding of the SNP regions to the probes. The SNP alleles are then detected by a single-base extension with differently labeled nucleotides and subsequent immune-fluorescent visualization of the incorporated nucleotides. Homozygous genotypes result in red or green fluorescent signals, and heterozygous genotypes appear in yellow stain.

The iCOGS chip is a custom array conceived by research collaborations for PrCa, breast cancer and ovarian cancer genetics [112]. It is composed of more than 200,000 SNPs which were previously identified by GWAS approaches for any of the cancer types, regional SNPs for fine mapping of previous GWAS hits as well as a series of variants in generally known cancer genes.

2.3.9.2 Genotyping using TaqMan® probes

General background

TaqMan® genotyping was the method of choice for medium-throughput genotyping of the large number of SNPs and samples. It is a PCR-based method, in which specific primers that flank the polymorphism of interest are combined with two different fluorescence dye (VIC and FAM) labeled TaqMan® MGB probes (structure as depicted in Figure 2A). One of the probes binds complementary to the wild-type allele and the other one to the variant allele of the examined SNP. The principle of fluorescence detection is analogous to the qRT-PCR as described above (2.3.8.1). In contrast to expression analysis, however, the curve profile of the fluorescence intensity is not so much of interest as rather the end point (displayed as the allelic discrimination plot, see Figure 5). Depending on the genotype, either the VIC or the FAM probes (homozygous), or both (heterozygous) bind to the sequence and are fragmented due to the exonuclease activity of the DNA polymerase in the course of the PCR.

Procedure

The SNP genotyping was performed using TaqMan® SNP genotyping assays (Table 8) and TaqMan® Genotyping Master Mix from Life Technologies on the Fast Real-Time PCR System 7900 HT and on the ViiA™7 Real-Time PCR System. The mastermix contains all necessary components for PCR and additionally ROX dye for internal standardization. Depending on the number of samples, either a robot assisted 384-well platform with a

final reaction volume of 5 µl was used, or a 96-well format with 10 µl final volume. The reaction set-ups for both scaling schemes are listed in Table 14. Since the probes in the assays are photosensitive, exposure to light was kept to a minimum during pipetting. The master mix was prepared, mixed appropriately and dispensed into a MicroAmp® plate of the chosen scale. After adding DNA, the plate was gently spinned down and sealed with MicroAmp® Optical Adhesive Film. The PCR cycling program was performed as shown in Table 15. It does not differ substantially between the two devices which were used in this work.

For quality control, genotyping projects in 384-well format were conducted with at least 2 % duplicate samples. In addition, at least one non-target control per 96-well plate was included.

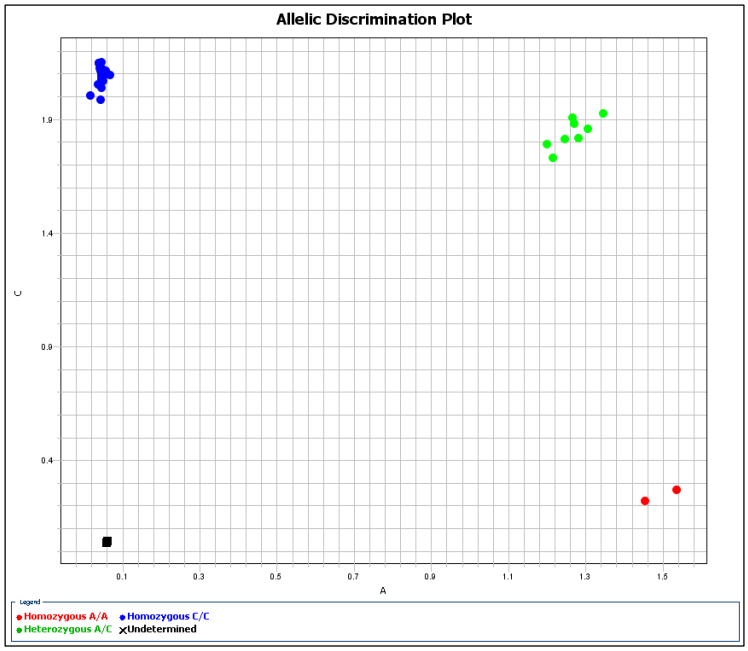


Figure 5: Allelic discrimination plot resulting from TaqMan® SNP genotyping using the ViiA7 device. The fluorescence intensity of the VIC dye is plotted on the x-axis, that of the FAM dye on the y-axis. Each dot represents one sample. The samples within the green cluster have both colors meaning that they are heterozygous for the SNP, in contrast to the homozygous blue and red samples. The black squares are the non-target controls.

Table 14: Standard reaction mix for SNP genotyping. The genotyping assays, which are available in different concentrations, are listed in Table 8. The assay performance is robust to varying DNA stock concentrations, but templates of less than 5 ng/µl should be avoided. If possible, it was adjusted to ~ 25 ng/µl.

Component	Pipetting volume [µl]	
	for 5 µl final volume	for 10 µl final volume
TaqMan® Genotyping Master Mix (2 x)	2.5	5.0
TaqMan® SNP Genotyping Assay (40 x / 80 x)	0.125 / 0.0625	0.25 / 0.125
DNA	1.0	1.0
Nuclease free H_2O	ad 5.0	ad 10.0

Table 15: Cycling program for TaqMan® genotyping.

Step	Time	Temperature	Repeats
Pre-Read	30 sec (2 min*)	60°C	1
Initial denaturation	10 min	95°C	1
Denaturation	15 sec	92°C	45†
Annealing and elongation	1 min	60°C	
Post-Read	30 sec (2 min*)	60°C	1

* For the Fast Real-Time PCR System 7900 HT
† could be increased, if necessary

The data analysis (allele calling) was automatically carried out by the respective software. Nevertheless, the data was always checked in the allelic discrimination plot whether the genotype clusters were correctly assigned (Figure 5), the non-target controls showed no amplification and the duplicates were congruent.

2.3.10 Determination of the somatic *TMPRSS2-ERG* fusion status in prostate cancer specimens

A large set of PrCa cases had to be phenotyped for the presence of the T2E fusion in order to test for associations between fusion status and common genetic variants. Cases from Berlin, FHCRC, Tampere, UKGPCS, Ulm and Porto were included in the meta-analysis (refer to section 2.1.3 and Table 3 for detailed sample description). Due to spatial and temporal differences, it was not possible to examine the fusion status of all of the samples in the same way and only a part of the phenotyping was performed in our lab. The different cohorts and the method by which the fusion status was determined are listed in Table 16.

Table 16: Detection methods for the assessment of the somatic *TMPRSS2-ERG* fusion status. Some of the samples were determined by quantitative real-time PCR (qRT-PCR) using RNA from fresh frozen tissue, while others were assessed with fluorescence in situ hybridization (FISH) on formalin fixed and paraffin embedded (FFPE) tissue.

Study center	n	Detection method	Template	Reference
Berlin	198	qRT-PCR	RNA from fresh frozen tissue	present work, par. 2.3.8
FHCRC	392	dual-color FISH	FFPE tissue microarray	present work, par. 2.3.11 and [147]
IPO-Porto	164	qRT-PCR (all)/ triple-color FISH (29 cases for validation)	FFPE tissue microarray/ RNA from fresh frozen tissue	[117]
Tampere	174	triple-color FISH	FFPE tissue microarray	[133]
UKGPCS	129	dual-color FISH	FFPE whole mount sections	[147]
ULM	129	dual-color FISH	FFPE tissue microarray	[67,120]
	35	qRT-PCR	RNA from fresh frozen tissue	present work, par. 2.3.2.3 / 2.3.8

2.3.11 *TMPRSS2-ERG* fluorescence in situ hybridization (FISH)

The FISH method is based on the complementary binding of fluorescent labeled DNA probes to the chromatin regions of interest. A particular FISH type was used for the determination of the T2E fusion status of FFPE prostate tumor tissue specimen: the *ERG* break-apart assay. The principle of this method is the binding of different labeled probes flanking the *ERG* locus on chromosome 21 (here: green and red, see Figure 6). In the wild type situation, the green and red signals are close together or overlap forming a yellow merger. If a gene fusion of *TMPRSS2* and *ERG* is present, the red signal is separated (if the intergenic region was inserted elsewhere into the genome) or lost (if the intergenic region was deleted) (Figure 8).

For this work, the break-apart assay was used to analyze FFPE prostate tumor samples on tissue microarrays (TMAs) provided by the FHCRC group. TMAs allow the simultaneous analysis of a large number of samples. Figure 7A shows one of the TMAs in the process of imaging.

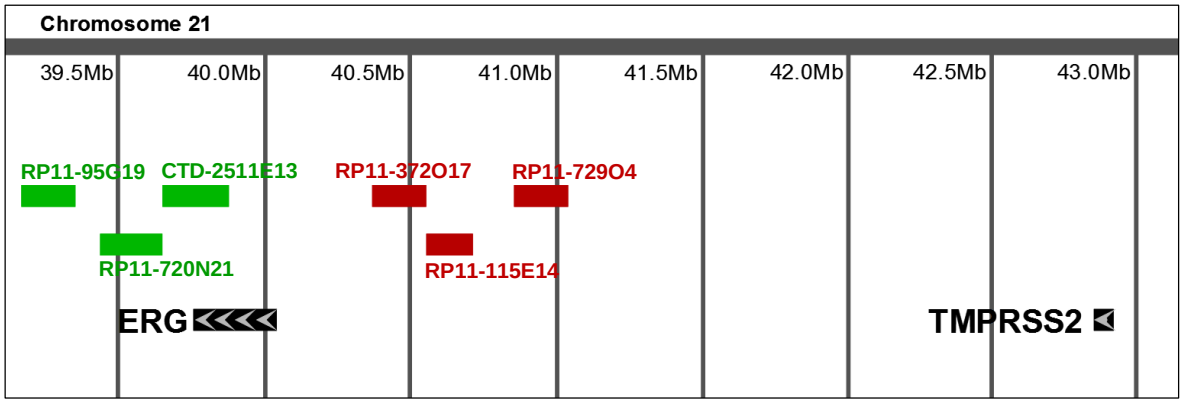


Figure 6: Overview of the BAC (bacterial artificial chromosomes) probe locations for the *TMPRSS2-ERG* FISH break-apart assay. The green bars represent the DIG labeled, the red bars the Biotin labeled probes, respectively. Please find more information on the BACs in section 2.2.5.

2.3.11.1 Probe preparation

BAC (bacterial artificial chromosome) preparation

Six genomic fragments from the human *ERG* and *TMPRSS2* locus were obtained as *Escherichia coli* DH10B BAC clones. For probe preparation, each bacterial clone was inoculated in 2 ml CAM-containing LB medium and incubated at 37°C (~ 16 h) on a shaking device. The BACS were then purified with the BACMAX™ DNA Purification Kit (Epicentre), which is based on alkaline lysis. The procedure was conducted according to the manufacturer's instructions. In brief, the bacteria were pelleted, resuspended, and then lysed by alkaline treatment. A neutralization step was performed immediately after lyses in order to prevent DNA damage. After a centrifugation step, the BAC DNA was present in the supernatant, while the sediment contained cell debris and the bacterial genomes anchored to proteins of the cell walls. The supernatants were collected, subjected to nucleic acid precipitation with use of isopropanol, and then treated with the supplemented RNase enzyme mix 'RiboShredder Blend' in order to remove bacterial RNA from the sample. Finally, the BAC DNA was precipitated a second time and dissolved in 20 µl TE buffer. The quality and purity was verified photometrically (2.3.3) and by agarose gel electrophoresis (2.3.5).

Amplification of BAC DNA (GenomiPhi)

The amplification of the BAC DNA was performed with the Illustra GenomiPhi V2 DNA amplification kit from GE Healthcare. It is based on the principle of continuous elongation of randomized binding oligonucleotide hexamers. In the course of the polymerizing reactions, the already existing strands are displaced (‚strand displacement').

First, the BAC DNA was incubated for 5 min at 55°C and centrifuged 10 min at 13,000 rpm at room temperature. One µl of the DNA was mixed with 9 µl sample buffer (kit), denatured for 3 min at 95°C and then placed on ice for 5 min. Afterwards, a mix of 9 µl reaction buffer and 1 µl polymerase (both kit components) were added and incubated in a thermocycler for 90 min at 30°C (polymerization) and 10 min at 65°C (inactivation of the enzyme). The amplified material was then purified by a phenol-chloroform extraction (2.3.4.2), followed by an ethanol precipitation with NaCl (2.3.4.1). The concentration was checked photometrically (2.3.3).

Probe labeling

The amplified material of the BACs served as the raw material for probe labeling. The three probes located centromeric of ERG were labeled with digoxigenin (DIG), the three telomeric ones with biotin. For this purpose, the BioPrime® DNA Labeling System from Life Technologies was used.

For each reaction, 300 ng amplified BAC DNA was diluted in 24 µl TE buffer and mixed with 20 µl 2.5 x random primer solution (kit). The further procedure is given in Table 17.

Table 17: Procedure of the BAC DNA labeling using the BioPrime®DNA Labeling System (Life Technologies)

	Biotin labeling	**Digoxigenin labeling**
2,5x Random Primer Solution (Kit)		20 µl
Incubation step	5 min at 95°C, then 10 min on ice	
10x dNTP Mixture (Kit)	5 µl	-
DIG 10x dNTP Mix	-	5 µl
DIG-11-dUTP	-	1.75 µl
Klenow Fragment, 40 U/µl (Kit)	1 µl	1 µl
Incubation step	3 h or overnight at 37°C	
Stop Buffer (Kit)		5 µl

After the reactions were finalized, they were purified using G50 micro spin columns (Illustra) according to the manufacturer's recommendations. The equilibration was performed with 1 x TE buffer. The flow-through was quantified photometrically, then supplied with 30 µl Cot1 DNA (1µg/µl) and 1 µl salmon DNA (10 µg/µl) and subjected to a NaCl precipitation (2.3.4.1). The pellets were solved in an appropriate volume of H_2O to obtain a final probe concentration of 200 ng/µl. The probe solutions were stored at -20°C.

2.3.11.2 Pretreatment of the TMA mounting slides

Pretreatment of the FFPE tissue was carried out in order to increase the accessibility of preserved chromatin for probe hybridization. The first step is the removal of paraffin. For this purpose, the slides were incubated under occasionally gentle shaking in xylene at 55°C for 10 min (three times) and then two times for 3 min in absolute EtOH at room temperature. Afterwards the tissue underwent a special treatment, the so-called "chemical aging", which enhanced the accessibility of the tissue: Immediately after the last EtOH step (without removing the residual EtOH), the slides were placed on a heating plate with 94–98°C, doused with 250 µl absolute EtOH and covered with a coverslip (22 x 50 mm #1.5). Once all ethanol was evaporated, the slides were collected and the coverslips were removed (should detach without any force). The slides were then incubated 2 x 3 min in H_2O at room temperature. For the next pretreatment step, the SPoTlight tissue pretreatment kit was used. The slides were placed in a plastic cuvette which was then filled with preheated (microwave) pretreatment buffer. The tightly closed plastic cuvette was then kept totally submerged in a boiling water bath for 20 min. After that, the slides were incubated for 2 x 3 min in H_2O at room temperature. If necessary, all previous steps were repeated. After the slides were dried at room temperature, the TMA areas were bordered with a liquid blocker pen, leaving an approx. 1 cm recess for draining. These borders facilitated an economical use of all reagents incubated on the tissue in the further course. In general, a protease treatment has been proven to be beneficial for FISH analyses. In this work, Digest-All 3 (Pepsin) from Life Technologies was used. An appropriate volume (~ 600 µl per slide) was preheated to 30°C on a heating block and incubated for 4 min on the slide on the CISH hybridizer at 30°C. Subsequently, the slides were incubated for 2 x 3 min in H_2O and dried at room temperature.

2.3.11.3 Probe hybridization

Depending on the size of the TMA or the tissue on the slide, an appropriate amount of probe/hybridization mix was prepared. For a whole slide, 2 µl of each probe were mixed with 2 µl 10x Cot1/Salmon DNA Mix and 38.4 µl hybridization mix. The components were mixed well and pipetted onto a coverslip, which was subsequently inverted onto the slide avoiding bubbles. The slides were placed on a CISH hybridizer and incubated for 2 min at 85°C (denaturation) and then over night at 37°C (hybridization). The next day, the coverslips were detached by incubating the slides in 2x SSC buffer for 10 minutes (or as long as necessary) at 42°C. The coverslips have to slip off without any force, as forced detachment could smudge the liquid blocker marking. The slides were then subjected to stringent washing. For this purpose, the slides were submerged twice in the formamide-containing fixation solution for 5 min at 42°C, then twice in 2x SSC buffer for 5 min at 42°C and subsequently in shaking SSCT buffer for 3 min at room temperature.

2.3.11.4 Probe detection

The probes are detected by antibodies, which are coupled with a fluorescent dye. Immediately after the SSCT washing step, the slides with the hybridized probes were placed in a chamber with a humidified atmosphere. Since the DIG probes were detected by fluorescent coupled antibodies, the slides were initially incubated with a milk powder containing SSCTM buffer for 15 min at 37°C, in order to minimize an unspecific binding of the antibodies (blocking). In the meantime, antibody and streptavidin solutions were centrifuged for 5 min at room temperature and diluted 1:200 in SSCTM buffer. Per slide, 3 µl were taken from the liquid surface and added to 600 µl SSCTM. After the blocking step, the slides were washed with SSCT and incubated at room temperature for 3 min in a horizontal position while still covered in SSCT. To detect the DIG labeled probes, SSCTM with FITC dye labeled Anti-DIG antibodies were applied on the slides followed by incubation for 10 min at 37°C. After rinsing and incubating twice with SSCT for 3 min at room temperature, the streptavidin-Cy3 conjugates in SSCTM were applied to bind the biotinylated probes and incubated for 10 min at 37°C. After a further SSCT washing round, as described, an anti-FITC antibody with Alexa 488 dye was applied and incubated in the same way as before, to enhance the weak FITC signal with the more stable Alexa dye. The slides were then rinsed again with SSCT and incubated one last time for 2 min at room temperature. Finally, the slides were taken from the moistening chamber and washed

twice in PBS for 5 min and twice in 70 % EtOH for 2 min at room temperature. The slides were air-dried in a light-protected environment, sealed with DAPI-containing Vectashield mounting medium and kept at 4°C for at least 1 h before analysis.

2.3.11.5 FISH Analysis

The fluorescence signals were analyzed using an Axioplan 2 Imaging System with an automated slide scanning device and Metafer/MetaCyte software. The TMA maps with the patient identifier were imported into the TMA tool which was implemented in the MetaCyte software. After a prescan of the whole TMA slide using the DAPI filter in 10 x magnification, the tool was used to define the center of each core to the TMA map (Figure 7A). If necessary, the center assignment was manually adjusted. A subsequent FITC and Cy3 scan in 40x magnification was automatically performed in 6 x 9 grids (54 single images) around the core centers (Figure 7B). Scanning parameters, such as exposure time and the focus planes (at least 3) have been adjusted according to the quality of the FITC and Cy3 signals.

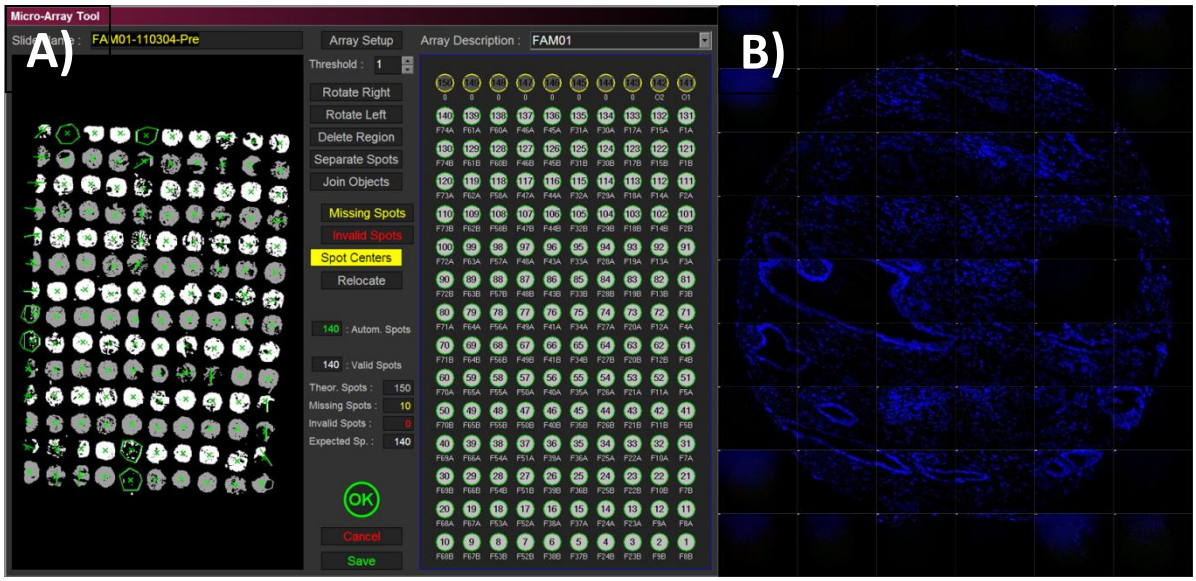


Figure 7: Overview of the TMA scanning for FISH analysis. A) The TMA tool was used to assign the DAPI scan in 10x magnification (left) to the TMA map with the tissue identifier (right). Many of the cores are incomplete either because of tissue detachment, or because of excessive stromal tissue content, which is incapable for bright DAPI staining. Nevertheless, each of the centers had to be defined in order to provide a reference point for the 40x scan. **B) Example of a TMA core composed of 54 single pictures** in 40x magnification. FITC (ALEXA) and CY3 signals can be visualized by inspection of single pictures and virtual zooming in (see Figure 8).

The assessment of the signals on the merged single pictures was carried out by two independent experimenters. The different constellations indicative for the presence or absence of a T2E fusion are shown in Figure 8. Tissue cores which have been interpreted differently by scorers were re-evaluated together to find a consensus score. In general, a core was excluded if less than 25 % of material was available. If it was obviously fusion positive, the score was nevertheless recorded, but was not included in frequency calculation as it would have biased towards over-estimation of fusion positive cases. Furthermore, a single fusion positive looking cell was not sufficient to call a lesion positive, based on the notion that discrete tumor foci arose clonal and thus adjacent cells should harbor the same abnormality.

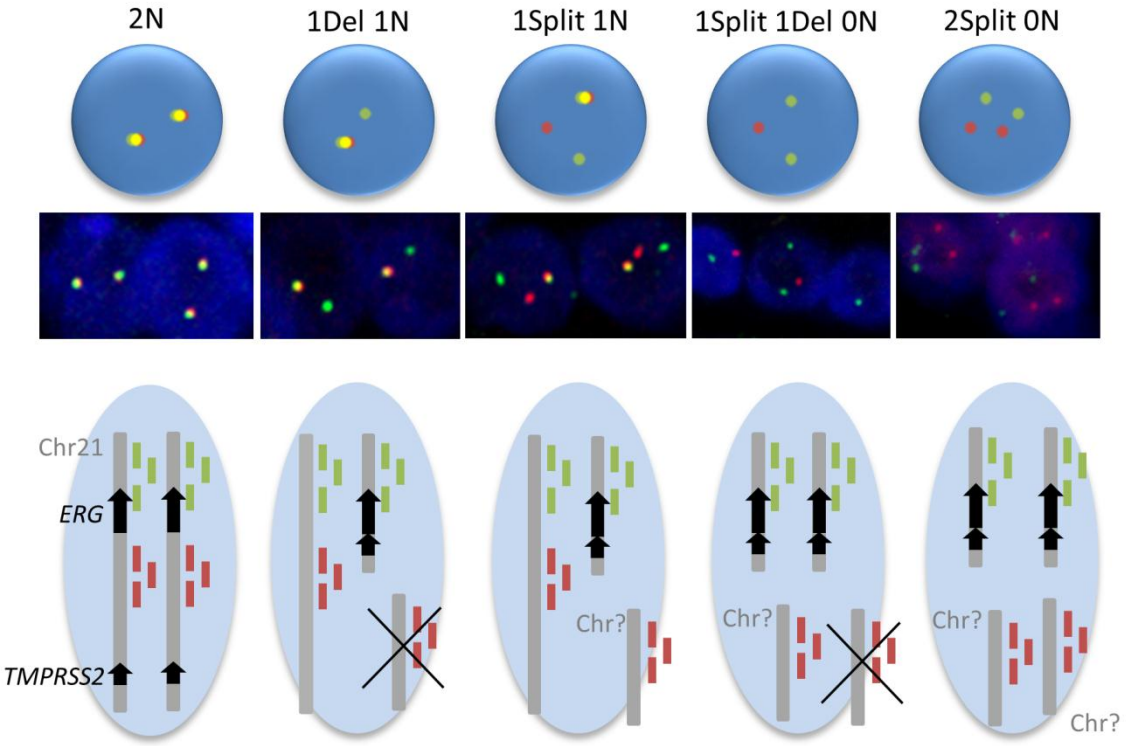


Figure 8: Overview of the most common signal constellations of the *ERG* break-apart fluorescence in situ hybridization (FISH) assay. Schematic FISH signal constellations are shown in the upper panel, corresponding example pictures in the middle panel, and the underlying genomic rearrangements are depicted in the lower panel. The green and red labeled probes, here schematically illustrated through small bars, flank the *ERG* gene in the wild-type situation and appear as yellow signals. 2N stand for two normal signals, Del denotes deletion of the intergenic material between *TMPRSS2* and *ERG* and Split its translocation elsewhere in the genome. All constellations are also found in aneuploid forms, indicated by 4N, 2Del 2N etc.

2.3.12 *TMPRSS2-ERG* induction assay with integrated gene knock-down

2.3.12.1 Background

An in-vitro assay for the induction of rearrangements in primarily fusion negative cells was initially introduced in 2009 independently by Mani et al. [98] and Lin et al. [90]. The authors observed that the ETS negative PrCa cell line LNCaP is able to develop a T2E fusion after DNA damaging exposure (γ-irradiation) in the presence of dihydrotestosterone (DHT). Furthermore, there is evidence that the fusion occurs more frequently after DNA damage, if the error-prone non-homologous end joining (NHEJ) repair pathway is used instead of the error-free homologous recombination (HR) repair [90] (Figure 9). Thus, the *TMPRSS2-ERG* induction assay represents an excellent method for investigating effects of genes on the T2E formation, and consequently, the genes role in DNA repair processes. In our lab, the assay has been already applied in a modified version to test the relevance of two DNA repair genes (*POLI* and *ESCO1*) for PrCa and their involvement in *TMPRSS2-ERG* formation [94]. In the present work, this assay was used to test genes of the 10q11 region (*MSMB*, *NCOA4*, *PARG* and *TIMM23*) on their impact on the *TMPRSS2-ERG* formation. In this context, *ESCO1* was chosen as a positive control. As LNCaP cells are notoriously difficult to be transfected with plasmids, the present work refrained from overexpression experiments and was confined to gene silencing using siRNAs.

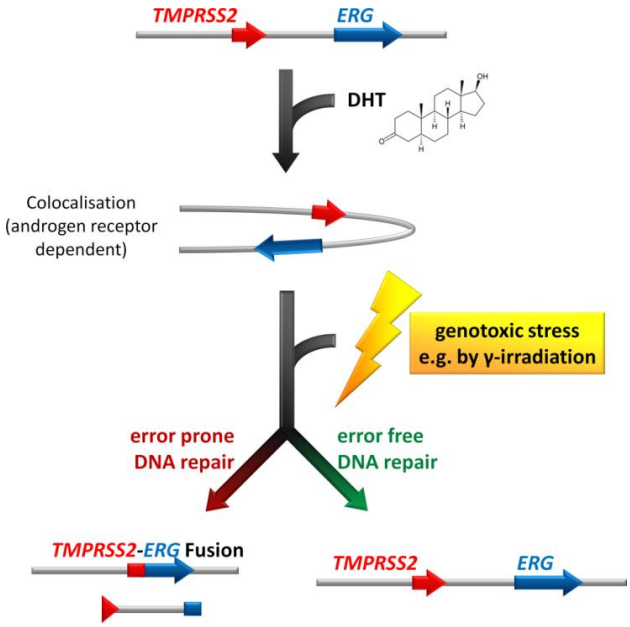


Figure 9: Principle behind the *TMPRSS2-ERG* induction assay. (from [94]) Dihydrotestosterone (DHT) leads to an androgen receptor dependent colocalisation of the *TMPRSS2* and the *ERG* gene loci. After genotoxic stress, inflicted by γ-irradiation, the cellular repair machinery removes the double strand breaks. Depending on cell cycle phase and various intracellular factors, more or less fusions are generated by error prone DNA repair pathways

2.3.12.2 Experimental overview

As mentioned above, the *TMPRSS2-ERG* induction assay was previously established in our lab [94]. In brief, LNCaP cells were seeded in t25 cell culture flasks and transfected with siRNA simultaneously to the culture start (2.3.12.3). At this stage, only hormone-reduced culture medium was used, allowing for defined application of a specific DHT dose before irradiation. At day 4, one batch of transfected cells was harvested for the RNA isolation (2.3.2.4) and quantitative expression analysis to check the knockdown efficiency (2.3.8) and the other batch underwent an irradiation of 300 Gy (2.3.12.4) to inflict DNA strand breaks. Ninety min before irradiation, the cells were treated with 1 µM DHT. At day 6, these cells were harvested for RNA isolation (2.3.2.4) and subsequent analysis of the induced amount of T2E fusion transcript (2.3.8). For both, the knockdown analysis and the *TMPRSS2-ERG* induction assay, one culture flask per gene was analyzed. The experimental series was repeated independently six times. Each series consisted of duplicate measurements for the reference experiments (negative control siRNA) and single measurements per candidate gene.

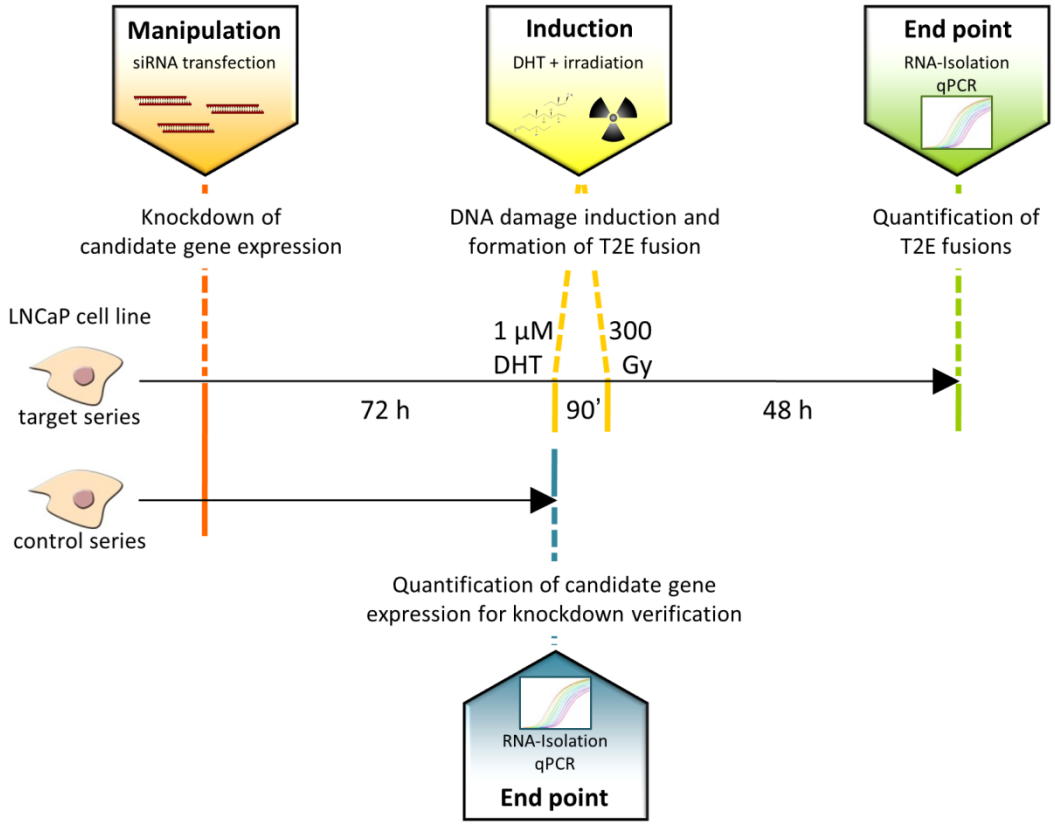


Figure 10: Experimental overview of the *TMPRSS2-ERG* induction assay (modified from [94]). Description see section 2.3.12.2. DHT = Dihydrotestosterone, Gy = Gray, h = hours, T2E = *TMPRSS2-ERG*

2.3.12.3 siRNA transfection

Background

RNA interference (RNAi) is one of the regulatory mechanisms for gene expression in the eukaryotic cell. The principle is that small, 20–21 nucleotides long stretches of double-stranded RNA, known as siRNAs (small interfering RNAs) are capable to specifically induce cleavage of the target mRNA. In general, these siRNAs arise from long double-stranded RNAs through the nuclease activity of Dicer, a component of the RISC (RNA-induced silencing complex) enzyme complex. The siRNA then binds to RISC and the sense strand becomes degraded. The remaining antisense strand can bind to the complementary target mRNA and leads to its cleavage. For molecular biological applications, this mechanism is utilized to silence specific target genes for diverse downstream investigations.

In this work, sets of four different siRNA against each of the genes *ESCO1*, *MSMB*, *NCOA4*, *PARG* and *TIMM23* were purchased from QIAGEN. In general, the use of several siRNAs for a given target gene can improve the knockdown efficiency and reduce unspecific off-target effects. As a negative control, the AllStars Negative Control siRNA (QIAGEN) was used, which is not complementary to any known human gene. The transfection of the RNAs into LNCaP cells was performed with the Lipofectamine® RNAiMAX Reagent (Life Technologies) using the “fast forward” reverse transfection protocol, which enhances the transfection efficiency of LNCaP cells. The difference to the standard protocol is that the transfection complexes are not added to the already adhered cells, but rather applied in the course of the cell seeding.

Procedure

Experiments were set-up in t25 cell culture flasks with a starting amount 1,500,000 LNCaP cells per flask. For the *TMPRSS2-ERG* induction assay, the negative siRNA control was performed in duplicate and the target gene knockdowns in single samples. To check the transfection efficiency at the time of irradiation, an additional transfection for each gene/control was set-up in parallel. Per transfection reaction, a total of 25 nM siRNA was mixed into a 1.5 ml reaction tube (for four siRNAs: 6.25 nM each) and kept on ice until application. The transfection complexes were prepared by adding 960 µl Opti-MEM® I Reduced-Serum Medium to the siRNAs followed by 10 µl Lipofectamine® RNAiMAX

Reagent. After each pipetting step, the mixture was mingled and spinned down. It was kept for 15–20 min at room temperature to allow transfection complex formation. During this process, LNCaP cells, maintained as indicated in 2.3.1.1, were harvested, suspended in 10 ml RPMI with hormone-reduced FCS and separated thoroughly with the aid of a 40 µm cell strainer. After the cell count was determined, the suspension was diluted to 300,000 cells/ml. Finally, the cell culture flask bottoms were covered with the transfection complex mixtures followed by adding 5 ml LNCaP cell suspension. The flasks were gently shaken for a uniform distribution of the cells. The transfected cells were kept in a CO_2 incubator at 37°C for 3 days until the time point of harvesting for the check of the knockdown efficiency or until the γ-irradiation, respectively.

2.3.12.4 Irradiation and harvesting

Ninety minutes before irradiation, the culture medium was gently removed and replaced with 5 ml RPMI with hormone-reduced FCS supplemented with 1 µM DHT. The cells were kept in the CO_2 incubator at 37°C until irradiation. Just before irradiation, the medium was carefully aspirated. Two t25 flasks were simultaneously placed vertically into a Gammacell 2000 Cs-137 source device (without medium to prohibit cell detachment at the liquid border) and were irradiated with 300 Gy in 90 min corresponding to a dose rate of 3.3 Gy/min. Subsequently, the cells were gently covered with 5 ml fresh hormone-reduced medium containing 1 µM DHT. After 48 h, in which the LNCaP had time to repair the inflicted DNA damage and to express the T2E transcript, the cells were harvested. RNA was isolated according to 2.3.2.4 and expression analysis was performed as described in 2.3.8.

2.3.13 Micronucleus assay

2.3.13.1 Background

Micronuclei arise from acentric chromosomal fragments, caused by unrepaired double strand breaks or dysfunction of the mitotic apparatus that could not be incorporated in the daughter cell during mitosis and were enveloped with a nucleic membrane [114]. The micronucleus assay utilizes the count of micronuclei as a surrogate marker for DNA repair capacity.

In the present work, a modified protocol from Fenech and Morley [45] was used, in order to assess the relative DNA repair capacity of peripheral blood lymphocytes (PBL) in individuals and to correlate the results with existing germline variants. A standardized clastogenic noxe, here 2 Gy, of ionizing radiation allowed the comparison of the DNA repair capacity of different proband groups. A defined timespan (e.g. two days) is conceded to the cells to make use of their innate DNA repair machineries. To restrict counting of micronuclei in cells which underwent exactly one round of mitosis, the cells were treated in the course of the procedure with cytochalasin-B, resulting in the formation of binucleated cells (BNCs). This substance blocks cell division by inhibition of the actin polymerization while allowing the division of nuclei.

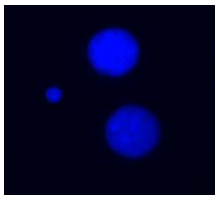


Figure 11: Example of a binucleated lymphocyte cell with micronucleus, which was stained with 4'-6-Diamindino-2-phenylindoldihydrochloride (DAPI).

2.3.13.2 Culture set-up and treatment

Two short-time lymphocyte cultures per proband were set up in flat bottom culture tubes as indicated in 2.3.1.2 and were irradiated immediately with 2 Gy of ionizing radiation from a Caesium-137 source (dose rate 1 Gy/min). After 44 h incubation at 37°C, Cytochalasin-B was added to a final concentration of 6 µg /ml and the cultures were incubated for a further 24 h.

2.3.13.3 Cell harvesting and slide preparation

To harvest the lymphocytes from the culture tubes, the blood cultures were suspended and centrifuged for 4 min at 2000 rpm at room temperature. Subsequently, the supernatant was removed while leaving approx. 1 ml residual buffer zone on the cell pellet. In order to remove erythrocytes, a hypotonic treatment using 5 ml 0.56 % KCl (4°C) was applied, suspended and immediately centrifuged for 4 min at 2000 rpm. The supernatant was removed again and was supplied with 5 ml fixative solution 1. The suspension was mixed again, incubated for 10 min at room temperature and centrifuged for 4 min at 2000 rpm. The same procedure was repeated at least three times (without incubation step) with fixative solution 2 until the supernatant was completely clear. In order to ensure a proper fixation, the lymphocytes were stored in 5 ml fixative solution 2 at 4°C at least overnight. For slide preparation, the cells were dropped on pretreated slides (kept in H_2O and wiped only to the extent that a thin water film still remained) and were spread by gently tilting the slide. After the surface had dried, the slides were examined using a light microscope, whether the lymphocyte swelling and cell density on the slide was sufficient. For visualization under the fluorescence microscope, cells were stained with DAPI by incubating in a DAPI solution for 5 to 10 min at room temperature and subsequently washed twice with H_2O. Finally, slides were dried and sealed with Vectashield®.

2.3.13.4 Counting of micronuclei

For the evaluation of the micronuclei counts, an Axioplan 2 fluorescence microscope with an automated imaging system (Metafer 3.1.2) was used. To standardize on cells that have passed exactly one division cycle, the classifier software was programmed to count only BNCs. The classifier parameters are provided in detail in Appendix B. At least two slides per proband were used for counting. Depending on the cell density on the slides, on average approx. 4000 BNCs were assessed per individual. The micronuclei counts are conventionally recorded per 1000 BNCs.

2.3.14 Mitotic delay assay

2.3.14.1 Background

The principle of this test is based on the ability of mitotic cells to undergo an arrest at the G2/M checkpoint. When exposed to genotoxic stress, cells accumulate in G2 either to repair DNA damage, or alternatively, to initiate apoptosis in order to avoid a division of cells with defect genetic material. Depending on the individual's efficiency of damage recognition and DNA repair, the extent of the delay is variable. In this work, peripheral blood lymphocytes from healthy subjects were treated with a defined damage amount (2 Gy of ionizing radiation) and were investigated for their cell cycle distribution in relation to the respective untreated samples. For this purpose, the cells were stained with DAPI and analyzed using a flow cytometer. This device is able to detect every single cell separately and determines its amount of DAPI signal, which corresponds to the DNA content. The resulting record is a histogram in which the DNA content (x-axis) is plotted against the cell number (y-axis) (Figure 12). The first peak represents cells in G1, possessing half as much chromatin as cells in G2 phase (second peak). The region between the G1 and G2 peak represents a continuous range of cells in S-phase, which are in different stages of replication, and thus have different amounts of chromatin.

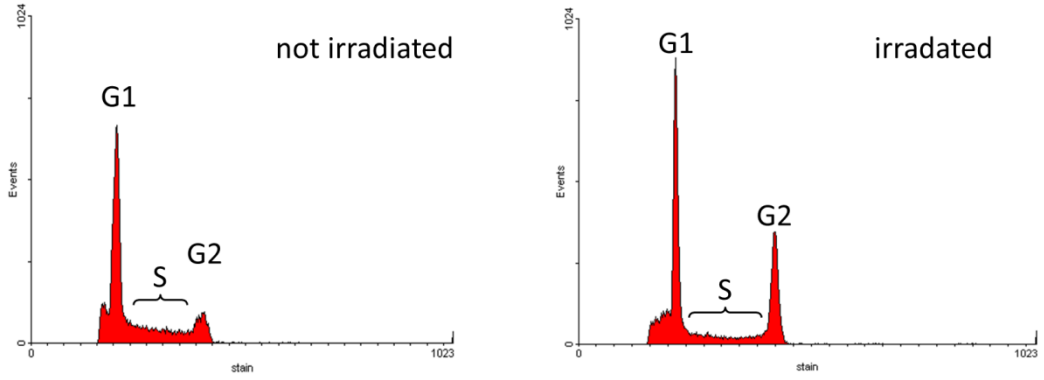


Figure 12: Flow cytometer histograms of peripheral blood lymphocytes. A flow cytometer counts single cells (y-axis) and detects their DAPI stained nucleic acid content which is assigned to 1024 intensity channels (x-axis). The first peak represents the cells in the G1 phase of the cell cycle and the second peak the cells in G2 phase with twice of the DNA content. Between them are the cells differently progressed in the DNA replication (S) phase with variable DNA content. The γ-irradiated cell populations are characterized by an enrichment of G2 phase cells due to a delayed mitosis entry.

2.3.14.2 Culture set-up and treatment

Two lymphocyte cultures per proband were set up according to 2.3.1.2. After 54 h at 37°C, one culture sample was irradiated with 2 Gy (induced sample) as already described in 2.3.13.2, while a parallel sample was left untreated (basal sample). The tubes were incubated for further 18 h at 37°C and were then prepared for flow cytometry. For this purpose, 500 µl of the cell suspension was added to 1.5 ml of DAPI Staining Solution (Partec) in flow cytometer tubes, sealed with Parafilm®, inverted gently and kept at 4°C overnight.

2.3.14.3 Flow cytometry

After the samples have been stained, they were gently inverted before examination with a CCA flow cytometer (Partec). Before and after every measurement, the capillary was flushed with Cleaning Solution (Partec). In order to obtain comparable histograms without artifacts, at least 20,000 cells were counted, and the flow rate was usually set to 100 cells per second. The upper and lower thresholds were defined to exclude shredded or clumped in order to increase the resolution for the real histogram. The standard parameters are listed in Table 18.

Table 18: Standard parameters for flow cytometric measurement of blood lymphocytes using a CCA flow cytometer (Partec).

Parameter	Value
Channel resolution	1024
Photomultiplier gain	380–420
Total cell count maximum	25,000
Total count volume maximum	5 ml
Lower Level (channel)	40
Upper Level (channel)	999

2.3.14.4 Data analysis

The raw data files, exported as cell counts per DAPI intensity channel, were transformed into tab delimited text files using the freeware program WinMDI. The mitotic delay between different probands was then compared with help of the mitotic delay index (MDI), which is defined by the quotient of the G2/S phase ratios of the radiation-induced (i) and the basal (b) sample:

$$MDI = [G2(i)/S(i)]/[G2(b)/S(b)] \quad \textit{(Equation 4)}$$

For the definition of the boundaries, e.g. the channels corresponding to each cell phase in a given sample, our group has established an Excel-based tool for standardized MD analysis.

2.3.15 Statistical methods

2.3.15.1 Hardy-Weinberg Equilibrium (HWE) - quality control for case-control analyses

The HWE states that the genotype frequencies of a genomic variation remains constant in an ideal population and that the expected genotype distributions can be calculated on the basis of observed allele frequencies. For case-control studies, the HWE should be fulfilled in order to obtain valid and unbiased results. The chi-square test was used to compare the genotype frequencies expected under HWE, deduced from allele frequencies, with the observed genotype frequencies. If significant deviations are present, the results of the association study should be taken with caution.

2.3.15.2 Association analyses

Association analyses were carried out when a particular risk factor, such as a genetic variant, should be tested for enrichment in a given population. In case-control studies, the strength of an association is represented by the odds ratio (OR), which can be interpreted as the effect size of a factor. An OR of 5.0 means, that the odds of a factor being present are 5-fold within a particular group compared to the reference group. The estimation of the ORs can be calculated from 2 x 2 contingency tables as shown below.

The numbers of individuals, which are positive and negative for the parameter of interest, are counted in the case- and in the control population, respectively:

	Number of Cases	**Number of Controls**
Risk factor present	a	b
Risk factor not present	c	d

The OR is calculated as the cross product of the contingency table:

$$OR = \frac{a \times d}{b \times c}$$ *(Equation 5)*

In the present work, ORs were calculated by logistic regression with use of the software StatView 5.1. In addition to the OR, the 95% confidence interval (95% CI) was estimated, corresponding to a range of the OR, where the 'true' value would be found with a probability of 95 %. In addition, p-values were calculated using the chi-square test for independence.

2.3.15.3 Meta-analyses for association in multiple cohorts

For the association analysis of common germline variants with the somatic T2E status, patients from different international study centers were recruited. Due to unequal sample sizes as well as different allele and phenotype distributions among the groups, simple pooling of all patients for a combined analysis may be invalid. Therefore, the Mantel-Haenszel meta-analysis was used, which analyses the study groups separately and afterwards combines the results into meta-data. The fixed-effects model assuming similar effects among the particular groups was selected. In this model, the true effect size is a value X and the variance is primarily explained by random effects resulting from small sample sizes. In contrast, the random-effects model assumes that unknown covariates lead to different effects among the groups and thus, there is a range of true effect sizes. In this setting, all sub-groups are approximately equally weighted, regardless of their particular sample sizes. Crucial parameters are the particular weights of each input sample, along with their individual ORs, as well as an estimate for the extent of heterogeneity present in the total sample. Typically, all individual effect sizes and the combined result of a risk factor are displayed by a so-called "forest-plot".

The detailed procedure can be found in the handbook of the Review Manager 5.1 software, which was used for this part of the study. Figure 13 shows the Stage 1 meta-analysis for the SNP rs1512268.

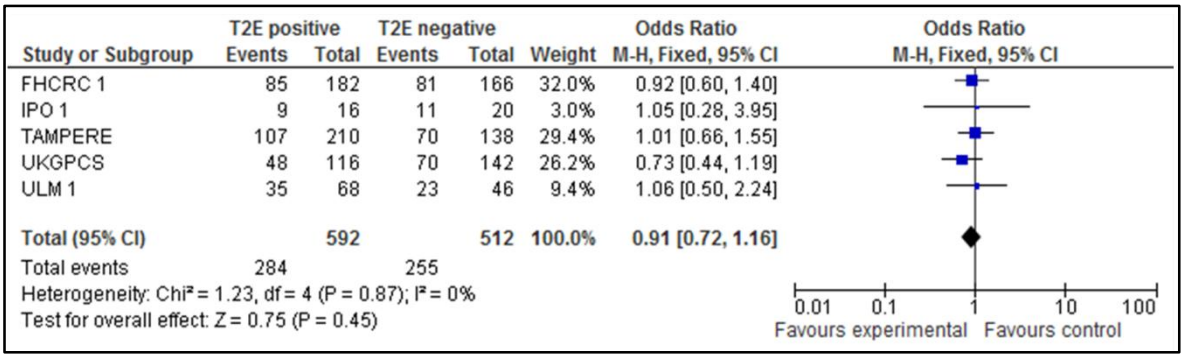


Figure 13: Exemplary meta-analysis of the SNP rs1512268 using the Review Manager v. 5.1. The program calculates the odds ratios, the corresponding 95% confidence intervals (CI) and the respective p-value with Mantel-Haenszel (M-H) methods based on the observed genotype and phenotype data. In this case, the input was the number of the prostate cancer risk alleles ("Events") and total number of alleles ("Total") of the respective SNP in *TMPRSS2-ERG* (T2E) fusion positive and negative prostate cancer cases. On the right side the graphical illustration of the data, the so-called forest-plot, is shown. The blue squares, proportional sized to the weight of the respective samples, represent the odds ratios of the sub-studies and the flanking lines the 95% CIs. The diamond displays the meta-odds ratio including the 95% CI.

2.3.15.4 Gene expression data analysis

Certain data, e.g. some gene expression data, is not necessarily normally distributed, and therefore, inappropriate for t-test analysis.

The Wilcoxon Signed Rank Test was applied for paired analyses, e.g. expression of a gene in tumor and normal tissue. Cochran-Mantel-Haenszel statistics were used to determine the association of the number of risk alleles (n = 0, 1, 2 corresponding to the three possible genotypes) with the quantitative values of the expression analysis. This method corresponds to a stratified Jonckheere-Terpstra-Test. For comparisons that include small subgroups (e.g. a subgroup with a rare genotype), an exact Jonckheere-Terpstra-Test was performed in addition. In the present work, Cochran-Mantel-Haenszel statistics were primarily used to stratify by the two centers, Erlangen and Ulm, for which the comparability of the expression data could not be safely assumed.

In order to display the data graphically as box plots, z scores were calculated for each center. For correlation analyses between the expression levels of two genes, the non-parametric Spearman rank correlation coefficient (Spearmans Rho) was calculated. For

fold-change analyses, to compare the paired expressional changes from normal to tumor tissue between different subgroups, a “percent-change” value x was estimated as follows:

If $\Delta Ct_{tumor} > \Delta Ct_{normal}$: $x = ((\Delta Ct_{tumor}/\Delta Ct_{normal}) -1) *100$ *(Equation 6)*

If $\Delta Ct_{tumor} < \Delta Ct_{normal}$: $x = (-(\Delta Ct_{tumor}/\Delta Ct_{normal}) +1) *100$ *(Equation 7)*

The resulting percentage value indicates the expressional change on the basis of the normal tissue. Values > 0 imply an increased expression in tumor versus normal tissue, while values < 0 indicate a decreased expression in tumor versus normal tissue. As an example, x = 10 means, that the examined gene expression is 10 % higher in tumor versus normal tissue, and x = -10 that it is 10 % lower expressed in tumor tissue.

2.3.15.5 Analysis of DNA repair test results

In order to uncover relationships between results from the micronucleus test and the mitotic delay assay with genotype data, in a primary analysis the measurements were modelled by means of a linear regression approach assuming an additive effect of n = 0, 1 and 2 alleles. To account for recessive and dominant effects, also unpaired two-tailed t-tests were performed comparing the contrasting genotype groups nn + nr vs. rr and nn vs. nr + rr (n = non-risk allele, r = risk allele), respectively.

2.3.15.6 Welch’s t-test to compare *TMPRSS2-ERG* induction

The heteroscedastic Welch’s t-test was used to compare two groups of measurements with differing variances.

For the T2E induction assay, the effects of silencing particular genes on T2E fusion formation in LNCaP cells were compared to a negative control treated with AllStars negative siRNA. This control has in contrast to the experiments no variance as it was set = 1. In this setting, Welch’s t-test is equivalent to a corresponding one-sample-t-test.

2.3.15.7 Significance and correction for multiple testing

In general, p-values of < 0.05 were termed statistically significant.

However, statistic stringency and issues of multiple testing were modulated according to particular requirements of the specific questions under study, e.g. the prior evidence for variants to be associated with disease.

Verification of previously associated SNPs

Candidate SNPs, which were significantly associated with PrCa in previous studies (GWAS), were considered "significantly verified", if, in the independent study sample of Ulm, a nominal p-value of 0.05 was observed in the case-control comparison. Thus, although series of SNPs were analyzed simultaneously for the purpose of verification, no adjustment for multiple testing was required.

Testing PrCa variants in novel hypotheses

Phenotypes of DNA repair and subtypes of PrCa were considered as new hypothesis outcomes, without prior evidence if (or which) risk variants are involved in. To identify SNPs that are specifically correlated with a phenotype under study, achievement of global significance levels was mandatory. For these study series, correction for multiple testing according Bonferroni was applied. Therefor, 0.05 was divided by the number of tests resulting in the new significance threshold.

Selection of variants for exploring cumulative risk models

The rationale of modelling cumulative risk effects was to elucidate whether multiple variants could have a significant contribution to PrCa when analyzed together; in other words, if the anticipated multifactorial mode of inheritance actually exists. However, no claim was ascribed for a specific set of SNPs to play this definite role, neither that every particular variant must itself represent a true risk factor for PrCa. For the selection of candidate risk SNPs to be included in the cumulative analysis (section 3.1.2), a generous significance threshold of p = 0.1 was arbitrarily chosen.

3. Results

The present study focused on the role of common PrCa associated variants, especially on the identification of functional evidence which could refine the identification of predisposing mechanisms. The first part, as presented in chapter 3.1, assessed the relevance of the 25 initial identified variants in a German case-control series with classical approaches of genetic epidemiology, as well as by exploring a multifactorial mode of susceptibility. The second part, chapter 3.2, describes the examination for mechanistic involvements, where subsets of common variants were tested for association with DNA repair capacity and somatic T2E fusion. Finally, chapter 3.3 outlines analyses of candidate genes from risk loci, which have been investigated because of - and with the use of - the newly established functional implication in DNA repair and/or fusion gene formation.

3.1 Contribution of common variants to prostate cancer risk

3.1.1 Verification of risk SNPs in a German case-control sample

A series of three independent pioneer GWAS [37,80,156] have identified the first 25 candidate SNPS for PrCa risk. These initially known common variants were genotyped in the Ulm sample set composed of 708 (390 familial indices, 317 sporadic) cases and 509 controls using TaqMan assays, to estimate their contribution to PrCa risk in Germany. Table 19 shows the results for familial and sporadic cases alone and for all cases merged. Nine of the candidate variants could be verified to be associated with PrCa risk in this study population by a significance of $p < 0.05$. The highest effect size exerts rs1447295 in 8q24 region 1 with a per-allele odds ratio of 1.6 and a corresponding p-value of 0.0003 in all cases versus controls. This dataset was contributed to collaborative approaches in the course of validation of GWAS results in multicenter case-control studies [9], as well as for assessing the role of common variants in familial clustering of PrCa [73,154].

Table 19: Case-control analysis results of the 25 SNPs investigated for the ULM prostate cancer sample set. Shown are the SNP IDs, the genomic region, the alleles and the risk allele frequencies (RAF) in controls. Every SNP was compared in familial, sporadic and all cases, respectively, with controls. The respective per-allele odds ratios (OR) and the corresponding p-values are shown. Nominally significant results ($p < 0.05$) are printed in bold, SNPs which were included in the subsequent cumulative analysis (comparison with all cases: $p < 0.1$) are highlighted in green.

SNP	*Genomic region*	*Allele (non-risk/risk)*	*RAF*	*familial cases*		*sporadic cases*		*all cases*	
				OR	*p-value*	*OR*	*p-value*	*OR*	*p-value*
rs721048	2p15	G/A	0.137	0.98	0.8742	0.90	0.4379	0.95	0.5938
rs1465618	2p21	G/A	0.385	1.10	0.4295	1.03	0.7976	1.06	0.5431
rs12621278	2q31	G/A	0.960	1.19	0.4644	1.30	0.3112	1.29	0.2217
rs2660753	3p12	C/T	0.102	**1.34**	**0.0425**	1.33	0.0583	**1.33**	**0.0216**
rs17021918	4q22	T/C	0.646	1.07	0.4799	0.99	0.9543	1.05	0.5598
rs12500426	4q22	C/A	0.460	1.15	0.1583	1.11	0.3020	1.11	0.2491
rs7679673	4q24	A/C	0.624	1.11	0.2710	1.22	0.5480	1.16	0.0875
rs9364554	6q25	C/T	0.274	1.16	0.1608	1.15	0.2193	1.15	0.1178
rs10486567	7p15	A/G	0.752	1.22	0.0912	1.10	0.4545	1.16	0.1349
rs6465657	7q21	T/C	0.509	0.92	0.3646	1.03	0.7742	0.97	0.6811
rs1447295	8q24 (reg1)	C/A	0.071	**1.65**	**0.0005**	**1.52**	**0.0063**	**1.59**	**0.0003**
rs6983267	8q24 (reg3)	T/G	0.487	**1.40**	**0.0005**	1.22	0.0502	**1.36**	**0.0009**
rs16901979	8q24 (reg2)	C/A	0.031	1.38	0.1582	1.55	0.0635	1.46	0.0607
rs2928679	8p21	C/T	0.456	0.99	0.9407	0.97	0.7723	1.01	0.9011
rs1512268	8p21	G/A	0.420	**1.21**	**0.0453**	**1.27**	**0.0195**	**1.26**	**0.0112**
rs10993994	10q11	C/T	0.341	**1.31**	**0.0052**	1.20	0.0706	**1.26**	**0.0057**
rs4962416	10q26	T/C	0.257	0.98	0.8414	1.11	0.3191	1.04	0.6651
rs7931342	11q13	T/G	0.531	1.17	0.1129	1.02	0.8249	1.10	0.2608
rs7127900	11p15	G/A	0.235	**1.34**	**0.0147**	**1.45**	**0.0025**	**1.39**	**0.0014**
rs4430796	17q12	G/A	0.491	1.15	0.1396	1.06	0.5989	1.11	0.2189
rs11649743	17q12	A/G	0.757	1.22	0.0965	1.23	0.1088	**1.23**	**0.0491**
rs1859962	17q24	T/G	0.473	**1.43**	**0.0002**	1.17	0.1286	**1.30**	**0.0015**
rs2735839	19q13	A/G	0.863	1.22	0.1572	**1.49**	**0.0088**	**1.33**	**0.0165**
rs5759167	22q13	T/G	0.549	**1.29**	**0.0097**	1.04	0.6864	1.17	0.0663
rs5945619	Xp11	T/C	0.385	0.93	0.6019	1.17	0.2731	1.03	0.7689

3.1.2 Cumulative effects of common variants

As each particular common variant has only a small influence on a carrier's disease risk, it is important to consider, if multiple risk alleles may accumulate to a relevant risk. Such scenario was explored on a set of common variants, which are likely relevant in our population. Applying non-stringent selection criteria, all SNPs which reached a significance level less than p = 0.1 were investigated for cumulative effects without applying any assumption or model of inheritance. Twelve SNPs reached this threshold and were thus included in the study (highlighted in green in Table 19). For the cumulative analysis, exclusively controls/patients with genotype information for all 12 SNPs could be evaluated (497 controls, 679 cases). The risk alleles were counted in every case and control individual. The distributions of cases and controls according to their individual number of risk alleles are shown in Figure 14. The corresponding numbers are given in Appendix C. Apparently, both groups resemble a Gaussian distribution, but the numbers in cases have shifted in the direction of more risk alleles as compared to controls. As the median risk allele number in controls was 10, this value was assumed as 'normal' in the population and served as the reference point. The probands were categorized according to their risk allele numbers as indicated in the y-axis of Figure 15. All groups except that with 8 or 9 risk alleles differed significantly from the reference group with 10 risk alleles. Men with less than seven risk alleles (3.6 % of cases) have less than half of the risk of developing PrCa (OR = 0.44). For those with 11 or 12 risk alleles the risk is almost doubled (OR = 1.86), and three-fold for men with 13 or 14 risk alleles (OR = 3.04). The risk of disease over 14 alleles is even increased to 5.5-fold (OR = 5.44). 4.7 % of cases are attributable to this high risk group. This shows that cumulative effects of common variants play a relevant role in PrCa susceptibility, especially for a fraction of men who carry an extraordinary high load of risk alleles.

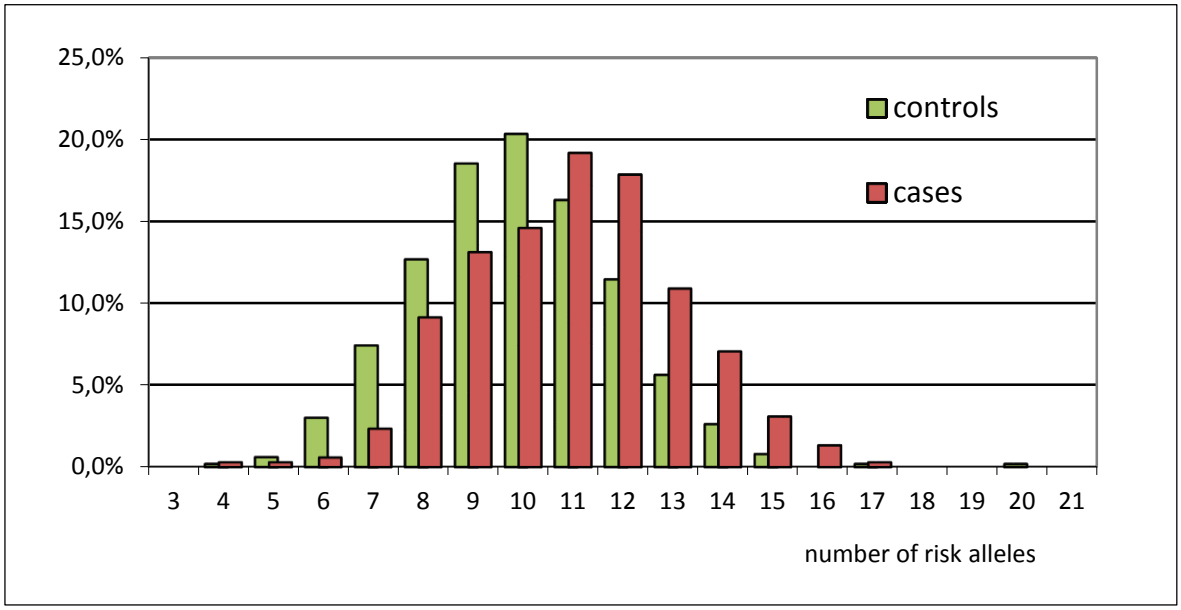


Figure 14: Cumulative numbers of risk alleles of the 12 selected SNPs in prostate cancer cases and controls. Theoretically, for the 12 autosomal loci, numbers between 0 and 24 are possible per individual. The median number of risk alleles in controls is 10, in cases 11.

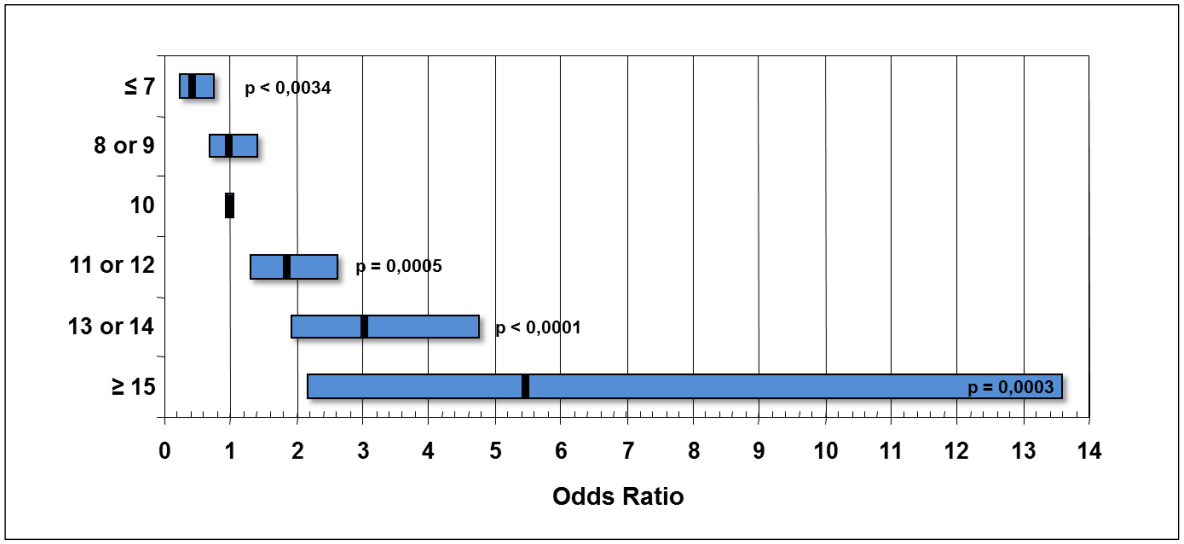


Figure 15: Cumulative effects of 12 common risk variants for prostate cancer. Shown are the odds ratios (OR) as black bars with confidence intervals (blue) for groups of individuals with different risk allele numbers (categories on the y-axis). The resulting p-values are indicated beside the bars. As the median risk allele number of controls was 10, this group served as reference resulting in an OR = 1.

A secondary question was, if particular risk allele groups (according to Figure 15) are correlated with clinical parameters. Logistic regression analyses revealed no evidence for a correlation of the risk allele groups with any of the clinical/pathological parameters: age at diagnosis (p = 0.3692), PSA at diagnosis (p = 0.2252), Gleason Score (p = 0.9040), tumor stage (dichotomized in local and advanced) and disease aggressiveness (p = 0.8449).

3.2 Serial examinations of SNPs for mechanistic involvements

As explained in detail in section 1.2.2, the functional modes of action of common variants of PrCa are still poorly understood. With the aim to facilitate the investigation of the variants' underlying genes and pathogenic mechanisms, this section deals with the question of whether common variants are connected to pathomechanistic features of PrCa. For this purpose two strategies were pursued. The first, focused on a possible association of common variants with the cellular DNA repair capacity (3.2.1), based on the hypothesis that defects within the DNA repair machinery may lead to an increased susceptibility for PrCa. The second approach investigated a correlation of the common variants with the somatic T2E fusion (3.2.2), which is the most prevalent chromosomal rearrangement present in PrCa and which is thought to be an early event in the carcinogenesis of PrCa.

3.2.1 Association between common variants and cellular DNA repair capacity

Maintenance of genome integrity is a crucial property of cells to prevent transformation, and DNA repair mechanisms are functionally connected to several malignancies, including PrCa. To seek for correlations between PrCa risk SNPs and DNA repair capacity, two endpoints of DNA damage response (measured with the micronucleus test (MNT) and mitotic delay (MD) assay, respectively) were applied on peripheral blood lymphocyte cultures of 128 healthy probands (data already published in [128]). The results of the MNT are given in micronucleus (MN) frequency per 1000 binucleated lymphocytes and those from the MD assay with the mitotic delay index (MDI), which represents the mitotic delay caused by γ-irradiation relative to a non-irradiated control culture of the same patient. The mean MN frequency (± SD) was 236.9 (± 43.0) and the mean MDI (± SD) 3.91 (± 1.00). For both assay results, the Kolmogorov-Smirnov test showed no evidence for deviation from normal distribution in the study population. The correlation analysis revealed a weak influence of the probands' age on MDI ($r = -0.185$, $p = 0.0470$) resulting in higher test values in younger individuals, and a weak correlation between the MDI and MN frequencies ($r = 0.240$, $p = 0.0095$) indicating that the two different tests are not completely independent.

A subset of 14 common variants was genotyped and analyzed for associations with the test results according to additive (primary analysis), dominant and recessive models (secondary analyses). The genotype frequencies of the 14 investigated SNPs observed in the sample set as well as all respective results for the two repair tests are listed in Table 20. Boxplots for all comparisons are given in Appendix D. The test results were considered as globally significant, when the respective p-values were less than 0.0018 (according to Bonferroni adjustment for multiple testing accounting for 28 independent tests). Due to the small sample size, also nominally significant results were taken into consideration. Using the linear regression analysis assuming an additive model (the number of risk alleles is essential) the rs10993994 variant at 10q11 yielded the most striking results. The PrCa risk allele T was globally significantly associated with a decreased MN frequency (p = 0.0003) as well as nominally significant with a lower MDI (p = 0.0353). Considering the respective box plots (Figure 16), it is remarkable that the MD results may fit not well into an additive model, as the CC and CT groups are very similar to each other, while the homozygous risk allele carriers differ. Indeed, the secondary analyses revealed the recessive as the better fitting model (p = 0.0035). For the MNT, the additive model seems to fit best (dominant: p = 0.0016, recessive: p = 0.0240), wherein the dominant p-value also reaches global significance level.

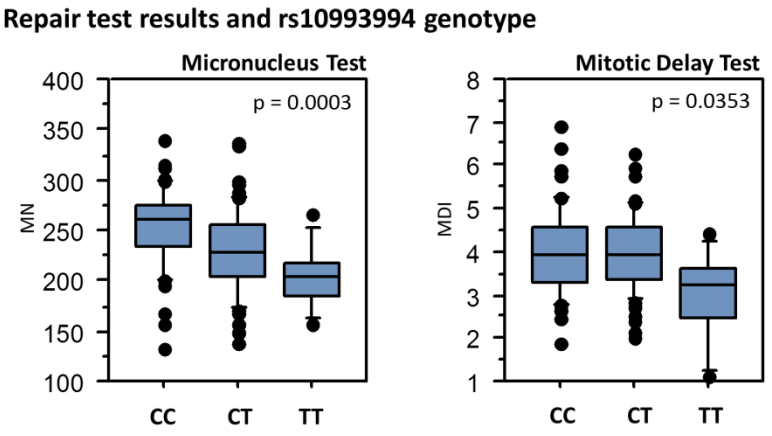


Figure 16: The micronuclei (MN) frequencies and the mitotic delay indices (MDI) dependent on the rs10993994 genotypes (modified from [128]). The T allele is the prostate cancer risk associated allele. The p-values were calculated with linear regression (additive model). See Table 20 for additional calculations under a dominant and recessive model.

Table 20: Results of the association study between prostate cancer risk associated germline variants and DNA repair capacity. [modified from [128]] Listed are the 14 prostate cancer risk tagging single nucleotide polymorphisms (SNPs) with genotype frequencies (freq), the mean mitotic delay indices and micronucleus (MN) frequencies observed in the test group (n = 128). The p-values result from calculations under an additive (linear), dominant (dom) and recessive (rec) model, respectively. According to Bonferroni adjustment for multiple testing (28 comparisons), $p < 0.0018$ was considered as globally significant (printed in bold). All SNPs with nominally or globally significant results are highlighted in green.

SNP genotypes		*freq [%]*	*mitotic delay index*				*induced MN frequencies*			
			mean	*p-value additive*	*p-value dom*	*p-value rec*	*mean*	*p-value additive*	*p-value dom*	*p-value rec*
rs2660753	CC	83.6	3.85				237			
(3p12)	CT	14.1	4.06	0.06	0.15	0.05	235	0.94	0.88	0.89
	TT	02.3	5.02				240			
rs7679673	AA	12.5	3.98				250			
(4q24)	AC	52.3	3.92	0.66	0.75	0.69	234	0.53	0.21	0.98
	CC	35.2	3.86				237			
rs10486567	AA	04.7	3.85				214			
(7p15)	AG	36.7	3.89	0.83	0.88	0.85	238	0.42	0.19	0.69
	GG	58.6	3.92				238			
rs1512268	GG	39.1	3.87				225			
(8p21)	GA	45.3	3.89	0.56	0.72	0.51	243	0.0164	0.0123	0.21
	AA	15.6	4.04				248			
rs1447295	CC	81.3	3.89				237			
(8q24)	CA	18.8	3.96	0.76	0.76	-	235	0.83	0.83	-
	AA	00.0	-				-			
rs6983267	TT	26.6	4.09				243			
(8q24)	TG	43.8	4.03	0.0276	0.23	0.0143	236	0.31	0.35	0.44
	GG	29.7	3.57				232			
rs16901979	CC	92.9	3.86				236			
(8q24)	CA	07.1	4.52	0.07	0.07	-	249	0.44	0.44	-
	AA	00.0	-				-			
rs10993994	CC	36.7	4.03				253			
(10q11)	CT	56.3	3.95	0.0353	0.30	0.0035	230	**0.0003**	**0.0016**	0.0240
	TT	07.0	2,98				204			
rs7127900	GG	64.8	3.96				235			
(11p15)	GA	32.0	3.85	0.39	0.48	0.47	242	0.50	0.42	0.94
	AA	03.1	3.55				235			
rs7931342	TT	25.0	4.13				227			
(11q13)	TG	42.2	3.77	0.43	0.16	0.99	240	0.19	0.13	0.46
	GG	32.8	3.91				241			
rs11649743	AA	04.7	3.96				220			
(17q12)	AG	28.1	3.92	0.88	0.92	0.89	241	0.88	0.34	0.80
	GG	67.2	3.90				236			
rs1859962	TT	20.3	4.21				252			
(17q24)	TG	54.7	3.76	0.50	0.10	0.61	233	0.16	0.05	0.69
	GG	25.0	3.99				234			
rs2735839	AA	03.1	4.06				196			
(19q13)	AG	28.1	4.07	0.26	0.79	0.23	234	0.15	0.10	0.29
	GG	68.8	3.83				240			
rs5759167	TT	26.6	3.81				239			
(22q13)	TG	53.9	3.95	0.65	0.53	0.95	238	0.50	0.70	0.47
	GG	19.5	3.92				231			

The second finding was a nominally significant genotype dependence of the MD test results with rs6983267 on 8q24 (Figure 17) under the additive model (p = 0.0276), which, similar to rs10993994, fits better with the recessive model (p = 0.0143).

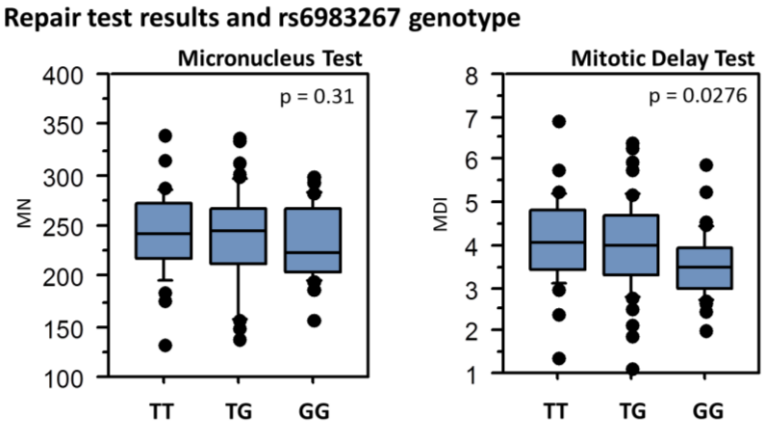


Figure 17: The micronuclei (MN) frequencies and the mitotic delay indices (MDI) dependent on the rs6983267 genotypes. (modified from [128]) The G allele is the prostate cancer risk associated allele. The p-values were calculated with linear regression (additive model). See Table 20 for additional calculations under a dominant and recessive model.

The third-most conspicuous SNP was rs1512268 at 8p21 that exerted weak influence on the MNT results (Figure 18). The PrCa risk allele A showed nominally significant association with an increased MN frequency under the additive model (p = 0.0164) and similarly under the dominant model (p = 0.0123).

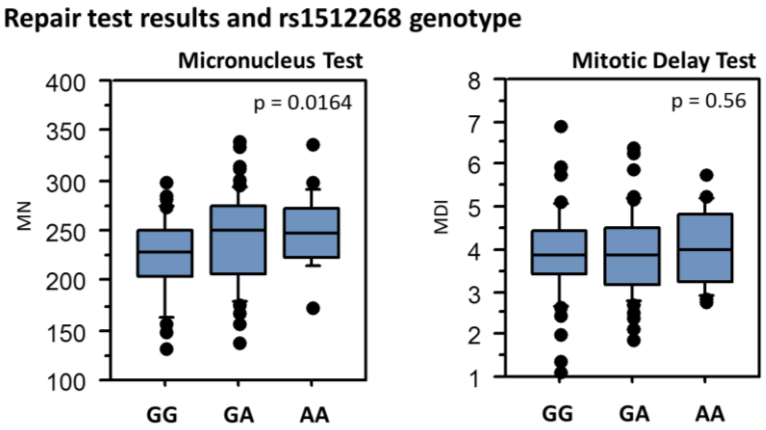


Figure 18: The micronuclei (MN) frequencies and the mitotic delay indices (MDI) dependent on the rs1512268 genotypes. (modified from [128]) The A allele is the prostate cancer risk associated allele. The p-values were calculated with linear regression (additive model). See Table 20 for additional calculations under a dominant and recessive model.

In summary, assessment of repair capacity in peripheral blood lymphocytes revealed compelling evidence for one PrCa risk region, rs10993994 at 10q11, to be mechanistically involved in DNA damage response. Two further loci, rs6983267 on 8q24 and rs1512268 at 8p21, could serve as candidates after additional functional validation.

3.2.2 Association between the *TMPRSS2-ERG* fusion and common variants in multiple populations

In this part the question was raised, if there are common PrCa variants which influence the formation of the somatic T2E fusion. This hypothesis is interesting in several respects, because (1), assuming fusion events as a consequence of DNA repair deficits, genetic variants associated with this phenotype could probably be assigned to the DNA damage response and repair pathway, and (2), the distinctness of T2E positive and negative prostate tumors could be supported by identification of a specific germline footprint.
This study, which is currently submitted for publication elsewhere [127], was performed in a two stage meta-analysis design. In brief, a total of 1,221 PrCa cases with somatic T2E fusion status were analyzed for associations with 27 independent common variants.

3.2.2.1 *TMPRSS2-ERG* phenotype assessment

In the course of this work, the two largest tumor samples of the collaborating groups had to be phenotyped for the T2E fusion status, the study cohorts FHCRC (FISH analysis on FFPE tissue, see 2.3.11) and Berlin (real-time based using RNA of fresh tissue, see 2.3.8). The fusion data of all other cohorts (IPO-Porto, Tampere, Ulm and UKGPCS) were already available from earlier investigations. The fusion frequencies in these samples are given in Table 21.

TMPRSS2-ERG assessment by FISH on tissue-microarray specimen (FHCRC sample)

In the total number of 392 cases, 168 (42.9 %) stained for the wild-type situation in the T2E region (fusion negative) and 224 (57.1 %) showed signals indicative for the presence of a gene fusion. Among the materials with rearrangements, 110 had a loss of the intergenic region between *TMPRSS2* and *ERG*, accounting for roughly half (49.1 %) of fusion positive cases (Figure 19). Nearly the same fraction (n = 104, 45.5 %) exhibited a deletion of the intergenic region at chromosome 21, but the fragment had translocated somewhere else into the genome. A fraction of 10 cases (5.4 %) had fusion signals of both types, deletion and translocation. Aside from specific chromosomal rearrangements between the syntenic genes *TMPRSS2* and *ERG*, the FISH technique also allows assessment of copy number alterations for the targeted region. The exact mechanism, such as fragmental amplifications or aneuploidy, cannot be distinguished. Overall, 148 specimens on the investigated TMA displayed more than two signals of 21q. As shown in the inlay chart of Figure 19, aneuploidy/amplification occurred proportionally similar among fusion negative and positive PrCa, and also among different subtypes of rearranged cases.

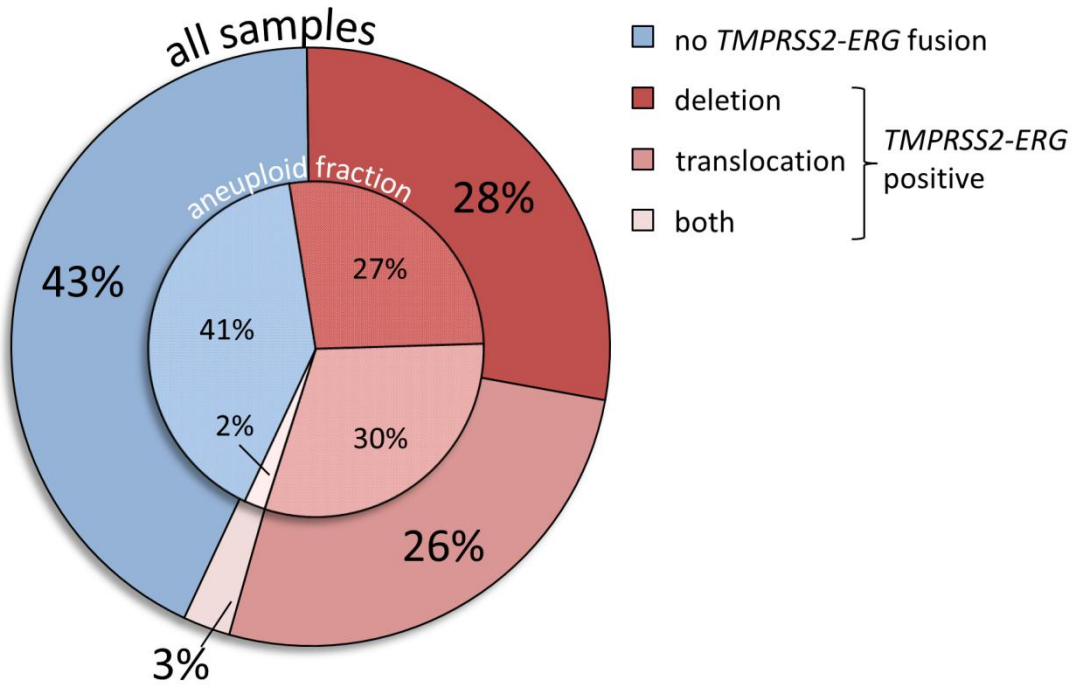


Figure 19: Results of the *TMPRSS2-ERG* (T2E) fluorescence in situ hybridization (FISH) of the FHCRC sample set. Shown is the percentage distribution of the fusion status of the *TMPRSS2/ERG* gene region of the 392 prostate cancer cases from FHCRC that were analyzed using the T2E break apart assay on a tissue microarray (2.3.11). The outer circle represents the total sample set (n = 392), the inner circle the fraction with copy number alterations (n = 148).

TMPRSS2-ERG assessment by qRT-PCR in fresh frozen materials (Berlin sample)

The most abundant mRNA isoform of T2E, designated T1G4, was targeted by qRT-PCR as a surrogate for fusion events in native (fresh-frozen) tissues. This method directly proofs the presence of a functional gene fusion, but is blind to specific rearrangement patterns. Out of 198 cases with RNA sample from tumor tissue available, 111 (56.1 %) were considered positive for *T1G4*. A smaller sample of cases (n = 35) from Ulm were also analyzed by this technique. The fusion positive fraction was 60.0 % (21 cases).

Table 21: Sample numbers and *TMPRSS2-ERG* (T2E) fusion frequencies in the included prostate cancer patient subgroups of the respective study stages. The patients are provided from Fred Hutchinson Cancer Research Center (FHCRC), Seattle, USA; from the Instituto Português de Oncologia (IPO-Porto) do Porto, Portugal; from the University of Tampere, Finland; The UK Genetic Prostate Cancer Study (UKGPCS) and from Berlin and Ulm, Germany

	sample cohort	cases with *T2E* status	*T2E* positive cases	*T2E* negative cases	*T2E* frequency
Stage 1 - Identification	FHCRC I	174	91	83	0.52
	IPO-Porto I	18	8	10	0.44
	TAMPERE	174	105	69	0.60
	UKGPCS	129	58	71	0.45
	ULM I	57	34	23	0.60
	Subtotal	**552**	**296**	**256**	**0.54**
Stage 2 - Replication	FHCRC II	218	133	85	0.61
	IPO-Porto II	146	79	67	0.54
	ULM II	107	65	42	0.61
	BERLIN	198	111	87	0.56
	Subtotal	**669**	**388**	**281**	**0.58**
Stage 1 + 2	**Total**	**1221**	**684**	**537**	**0.56**

In summary, somatic phenotyping for T2E was carried out with different techniques, due to practicability and availability of asservated materials. Despite of the non-standardized approach, the frequencies of fusion positive cases among contributing sites were comparable, and in a similar range as reported in literature. For the upcoming association studies concerning fusion status and genotype correlations, the phenotyping results were considered acceptable.

3.2.2.2 The analysis of 27 SNPs on a hypothesis generating sample set (Stage 1)

Genotyping for a first step on T2E phenotyped cases was conducted within the framework of a superordinate multi-centric chip-based genotyping project of the Collaborative Oncological Gene-Environment Study (COGS) [38]. The chip enclosed genotype data for 211,155 SNPs, including 27 confirmed PrCa GWAS hits, which were subject of the present work. A total of 552 PrCa cases with European ancestry from five study centers with T2E phenotype data (FHCRC, IPO-Porto, Tampere, UKGPCS and Ulm) with T2E phenotype data could be included (the fusion frequencies are listed in Table 21) None of the patient groups deviated significantly (threshold p = 0.0018) from Hardy-Weinberg Equilibrium for any of the 25 autosomal SNPs. A test for a potential sampling bias was performed by comparing the SNP frequencies between the subgroups of T2E phenotyped patients and the unselected collective (n = 8,141) using Mantel-Haenszel analyses. This quality control pointed out one SNP, rs7127900 on chromosome 11p15 (p = 0.0056) (Table 22). This means, in consequence, that the investigated sub-collective is not representative for the overall population for this SNP, and a reliable evaluation could not be guaranteed.

A requirement for the Mantel Haenszel meta-analysis using a fixed-effects model was an adequate homogeneity among the different study centers. Chi-square statistics confirmed that there was no considerable heterogeneity ($I^2 < 75$ %, Table 22).

In order to determine differences in the frequencies of the 27 SNPs between positive and negative prostate carcinomas, Mantel-Haenszel ORs were estimated comparing these two groups, leading to effect sizes above 1 (SNP risk allele is overrepresented in fusion positive patients) or below 1 (SNP risk allele is overrepresented in fusion negative patients). Due to the limited power of this comparatively small sample size, p-values < 0.05 were considered as a sufficient criterion for inclusion in Stage 2 analysis. All in all, four SNPs were significantly enriched in one subgroup: rs10993994 on 10q11, rs1859962 on 17q24 and rs2735839 on 19q13 were enriched in T2E positive, and rs16901979 on risk region 2 of 8q24 in T2E negative patients (Table 22, Figure 20). Conspicuously, a second independent SNP on 8q24 (region 1) was also enriched in T2E negative cancer with a barely not significant p-value of 0.0891. This variant was also included in Stage 2.

Table 22: Overview of the investigated PrCa risk loci and Stage 1 results of the *TMPRSS2-ERG* (*T2E*) positive versus *T2E* negative comparison.

SNPs that were selected for Stage 2 analysis are highlighted in green. The SNP highlighted in orange was not considered for the analysis due to the possible sampling bias.

SNP	*Region*	*Position*	*Allele(non-risk/risk)*[a]	*Risk allele frequency*[b]				*Heterogeneity*		*Sampling Bias*[c]	*Identification Stage I Results*[d]	
				Controls	*Cases*	*T2E +*	*T2E -*	*I^2[%]*	*p-value*		*OR [95% CI]*	*p-value*
rs721048	2p15	62985235	G/A	0.18	0.19	0.19	0.20	8	0.36	0.8619	1.02 [0.75 - 1.38]	0.8966
rs1465618	2p21	43407453	G/A	0.21	0.23	0.22	0.21	59	0.05	0.5148	1.08 [0.81 - 1.44]	0.6171
rs12621278	2q31.1	173019799	G/A	0.94	0.96	0.95	0.95	4	0.38	0.3489	1.07 [0.62 - 1.85]	0.8103
rs2660753	3p12	87193364	C/T	0.10	0.10	0.12	0.09	0	0.44	0.5603	1.30 [0.88 - 1.92]	0.1936
rs6763931	3q23	142585523	C/T	0.44	0.45	0.45	0.45	0	0.45	0.7667	1.04 [0.82 - 1.32]	0.7339
rs12500426	4q22.3	95733632	C/A	0.47	0.49	0.48	0.47	0	0.60	0.1278	0.97 [0.76 - 1.24]	0.8259
rs17021918	4q22.3	95781900	T/C	0.65	0.68	0.70	0.65	65	0.02	0.5629	1.19 [0.92 - 1.53]	0.1770
rs7679673	4q24	106280983	A/C	0.58	0.61	0.60	0.60	15	0.32	0.9382	0.98 [0.77 - 1.25]	0.8572
rs130067	6p21	31226490	T/G	0.20	0.21	0.20	0.19	0	0.77	0.3516	1.03 [0.77 - 1.40]	0.8259
rs9364554	6q25	160753654	C/T	0.29	0.31	0.32	0.30	26	0.25	0.8917	1.17 [0.90 - 1.52]	0.2301
rs10486567	7p15	27943088	A/G	0.75	0.78	0.79	0.82	34	0.19	0.3013	0.83 [0.61 - 1.11]	0.2113
rs6465657	7q21	97654263	T/C	0.47	0.51	0.52	0.48	52	0.08	0.7292	1.02 [0.80 - 1.29]	0.8966
rs2928679	8p21.2	23494920	C/T	0.42	0.44	0.45	0.42	49	0.09	0.8646	1.07 [0.84 - 1.37]	0.5687
rs1512268	8p21.2	23582408	G/A	0.42	0.48	0.48	0.49	0	0.87	0.8732	0.91 [0.72 - 1.16]	0.4533
rs1447295	8q24 (R1)	128554220	C/A	0.12	0.17	0.16	0.19	31	0.21	0.5627	**0.76 [0.56 - 1.04]**	**0.0891**
rs6983267	8q24 (R3)	128482487	T/G	0.51	0.57	0.56	0.56	0	0.79	0.5447	1.05 [0.83 - 1.33]	0.6965
rs16901979	8q24 (R2)	128194098	C/A	0.03	0.05	0.04	0.07	0	0.62	0.9614	**0.53 [0.31 - 0.91]**	**0.0214**
rs10993994	10q11	51219502	C/T	0.37	0.43	0.46	0.39	0	0.45	0.9444	**1.35 [1.06 - 1.72]**	**0.0151**
rs4962416	10q26	126686862	T/C	0.25	0.27	0.26	0.24	0	0.92	0.3549	1.13 [0.86 - 1.49]	0.3843
rs7127900	11p15.5	2190150	G/A	0.19	0.23	0.27	0.26	71	0.007	0.0056	1.06 [0.81 - 1.39]	0.6527
rs7931342	11q13	68751073	T/G	0.49	0.54	0.51	0.56	0	0.75	0.5500	0.86 [0.68 - 1.10]	0.2225
rs10875943	12q13	47962277	T/C	0.30	0.32	0.30	0.32	0	0.82	0.4858	0.89 [0.69 - 1.16]	0.3898
rs11649743	17q12	33149092	A/G	0.82	0.83	0.85	0.81	4	0.38	0.6375	1.31 [0.96 - 1.80]	0.0930
rs1859962	17q24	66620348	T/G	0.47	0.52	0.56	0.50	19	0.29	0.8569	**1.29 [1.01 - 1.64]**	**0.0375**
rs2735839	19q13	56056435	A/G	0.85	0.88	0.91	0.85	0	0.53	0.7079	**1.73 [1.20 - 2.51]**	**0.0035**
rs5945619	Xp11.22	51258412	T/C	0.38	0.43	0.42	0.40	0	0.59	0.5032	1.01 [0.72 - 1.43]	0.9442
rs5919432	Xq12	66938275	G/A	0.81	0.83	0.82	0.82	67	0.02	0.9930	1.07 [0.70 - 1.65]	0.7490

[a] risk alleles for prostate cancer

[b] estimated in controls (n = 8,404), unselected cases (n = 8,681) and *T2E* phenotyped cases from stage 1 from all cohorts merged

[c] fixed-effects Mantel-Haenszel meta-analysis comparing *T2E* typed (n = 552) versus somatically untyped cases (n = 8,141) to evaluate whether the subset is representative

[d] fixed-effects meta-analysis of five sample sets from the USA (Seattle, FHCRC), Finland (University of Tampere), Portugal (IPO-Porto), UK (UKGPCS) and Germany (University of Ulm) with a total number of 296 *TMPRSS2-ERG* fusion positive and 256 *TMPRSS2-ERG* fusion negative prostate cancer cases

3.2.2.3 Replication (Stage 2) and combined analysis

To validate the results from the Stage 1 meta-analysis, 669 further PrCa cases from four study centers with known somatic T2E fusion status (IPO-Porto II and ULM II) or with available tumor material (FHCRC II and BERLIN) were included (Table 21). 388 of them were T2E positive (58.2 %) and 279 T2E negative (41.8 %). These cases were not genotyped as part of the above mentioned iCOGS chip, so this was done with conventional TaqMan probes (2.3.8) for the five SNPs of interest (highlighted green in Table 22).

The results of the Stage 2 Mantel-Haenszel meta-analysis are displayed combined with Stage 1 as forest plots from the Review Manager 5.1 software (Figure 20). Three variants, rs1859962 on 17q24 ($p = 0.0178$), rs16901979 ($p = 0.0121$) and rs1447295 ($p = 0.0085$) on 8q24 could be replicated in stage 2. In the combined analysis of Stage 1 and 2, rs1859962 and rs16901979 discriminated with global significance ($p < 0.0019$ according Bonferroni adjustment for multiple testing) between T2E positive and negative prostate carcinoma.

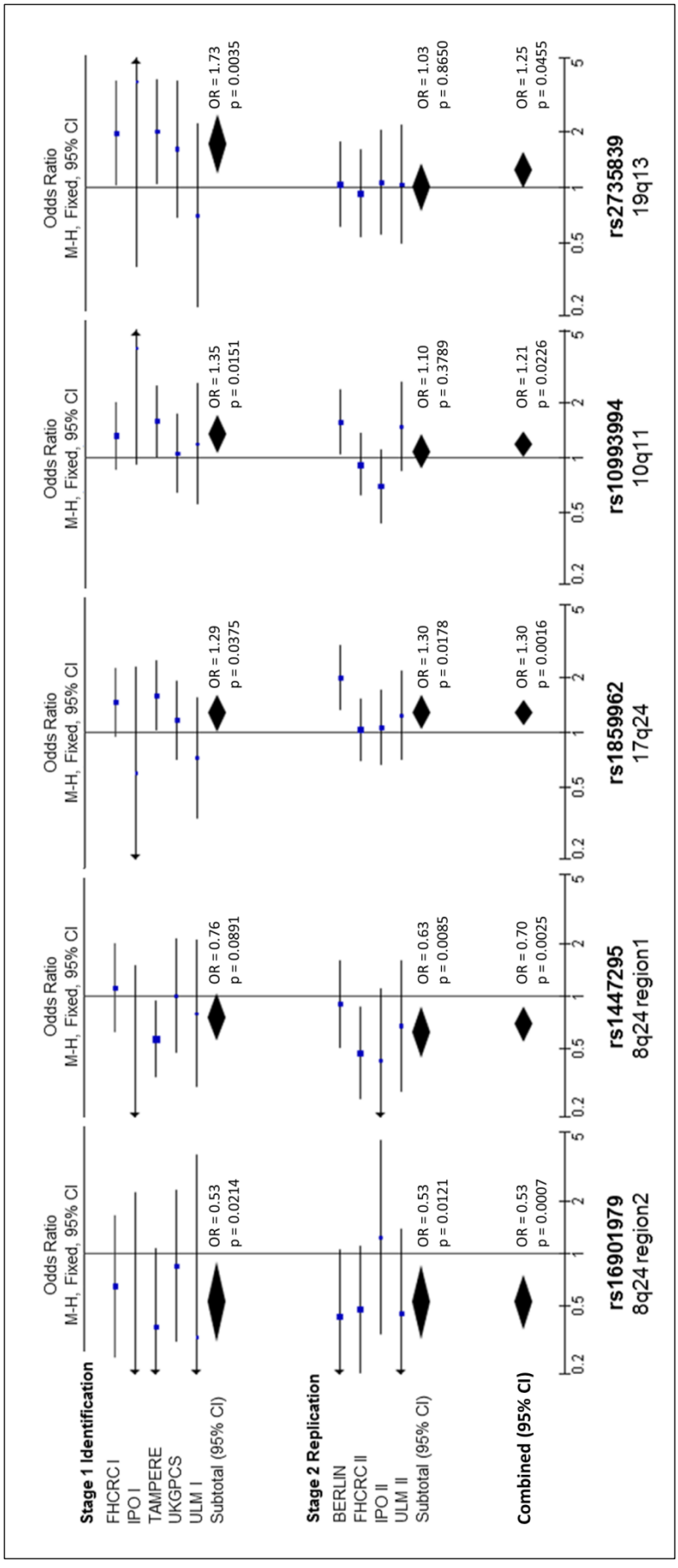


Figure 20: The results from the association study between prostate cancer associated germline variants and the *TMPRSS2-ERG* (T2E) fusion visualized as forest plots. The blue squares represent the odds ratios (OR) estimated for the prostate cancer risk allele in T2E positive versus negative cases using Mantel-Haenszel (M-H) methods. The flanking lines are the respective 95% confidence intervals (CI). For the very small IPO I sample, the ORs are in part out of range and the directions are indicated by arrows. The diamonds represent the OR and the 95% CI resulting from the meta-analyses. ORs above one correspond to an overrepresentation of the prostate cancer risk allele in T2E positive cases, those below one to an overrepresentation in T2E negative cases. The sample sizes and information are given in Tables 3 and 21.

3.2.2.4 *TMPRSS2-ERG* positive and negative cases, respectively, compared to controls

Comparing fusion positive versus fusion negative cases, as done in the primary analysis, is the most accurate, albeit underpowered practice to identify subtype specific risk factors. However, this approach does not answer the intuitive question for the true risk effect a particular germline factor would have, if the variant had been occasionally analyzed for the particularly relevant tumor type. So far, by convention, risk effects of PrCa-associated variants were reported from case versus control analyses, and these results may be underestimated because of the imprecise definition of PrCa. Since contributing centers to the present study were able to provide all genotype data from controls, all unselected PrCa cases and the fractions of cases with fusion status (samples listed in Table 3), the relevant comparison could be lined up for the five SNPs, which were selected for Stage 2. As shown in Table 23, both 8q24 SNPs that were found associated with T2E negative PrCa, yielded much stronger effect sizes within the case-control set selected for T2E negative cases, than in the unselected. Moreover, they forfeit the significance when considering T2E positive cases only. Tumor subgroup-specific effects were also observed for variants on 10q11 and 17q24 that were found associated with T2E positive PrCa, while association with the fusion negative tumor subtype was marginally. For rs2735839 on 19q13, which was borderline significant for T2E positive PrCa, the effect size was not increased considering T2E positive PrCa only compared to the unselected case-control set. The latter finding may be regarded indicative for a false positive result in the primary analysis.

Table 23: Case-control association meta-analysis of the five SNPs genotyped in Stage 2. The analysis was performed for both, unselected cases and cases separated by T2E fusion status. All available samples from the six study centers were included (see Table 3). Allelic odds ratios (OR) with the respective 95% confidence intervals (95% CI) and the p-values are shown. For the unselected comparisons, all estimated p-values were < 0.00001.

SNP ID	*Region*	*Cases vs. controls*	*T2E positive vs. controls*		*T2E negative vs. controls*	
		OR [95% CI][a]	*OR [95% CI]*	*p-value*	*OR [95% CI]*	*p-value*
rs16901979	8q24	1.68 [1.50 - 1.88]	1.47 [0.82 - 2.64]	0.1971	2.25 [1.69 - 3.01]	< 0.0001
rs1447295	8q24	1.44 [1.35 - 1.53]	1.17 [0.97 - 1.41]	0.0949	1.70 [1.41 - 2.04]	< 0.0001
rs1859962	17q24	1.22 [1.17 - 1.27]	1.47 [1.30 - 1.66]	< 0.0001	1.14 [1.00 - 1.30]	0.0536
rs10993994	10q11	1.27 [1.22 - 1.33]	1.39 [1.24 - 1.57]	< 0.0001	1.16 [1.01 - 1.32]	0.0300
rs2735839	19q13	1.31 [1.23 - 1.39]	1.32 [1.01 - 1.72]	0.0455	1.00 [0.84 - 1.20]	0.9681

3.2.2.5 Secondary findings - *TMPRSS2-ERG* fusion and clinical parameters

Due to recent publications which indicated a correlation between the T2E fusion and the patients' age at diagnosis [135,178], this aspect was also investigated in the present cohort. In addition, PSA at diagnosis, Gleason Score and tumor stage, dichotomized in local (T1/T2) and advanced (T3/T4 and/or N1/N2) carcinoma, as well as the tumor aggressiveness was tested for correlation with the T2E fusion using logistic regression. The criterion for aggressive disease was once defined by Schaid et al. [137] as having at least one of the following characteristics: advanced tumor stage, Gleason Score ≥ 7, PSA at diagnosis ≥ 20 ng/ml and prostate-specific death before age 65 years. The latter criterion was not considered, as this information was not available for most of the patients. Since clinical parameters, especially age at diagnosis (Table 24), were heterogeneous among the present study centers, all comparisons were adjusted for center effects.

As shown in Table 25, the chance to be T2E negative is slightly, but significantly increased in older cases (OR = 0.966 per year, $p < 0.0001$). According to Table 24, the mean age at diagnosis in T2E positive cases is decreased for all center subgroups, except for IPO-Porto and Ulm. Also the PSA value at diagnosis was found associated with fusion status, though only borderline significant ($p = 0.0422$). Accordingly, patients with PSA ≥ 4 are generally less likely to have aT2E fusion (ORs below one) relative to the reference group with values < 4. The proportion of T2E negative tested patients was approx. 1.77 fold increased in the group with PSA values > 10. In contrast, no evidence for an influence on T2E fusion status was shown for Gleason score, tumor stage and aggressiveness (Table 25).

Table 24: Age at diagnosis in prostate cancer cases dependent on *TMPRSS2-ERG* (T2E) fusion status. Sample numbers are provided in n and the mean at diagnosis in years with the standard deviation (SD). The patients are derived from Fred Hutchinson Cancer Research Center (FHCRC), Seattle, USA; from the Instituto Português de Oncologia (IPO-Porto) do Porto, Portugal; from the University of Tampere, Finland; The UK Genetic Prostate Cancer Study (UKGPCS) and from Berlin and Ulm, Germany

study center	cases with *T2E* status		*T2E* positive cases		*T2E* negative cases	
	n	mean age [± SD]	n	mean age [± SD]	n	mean age [± SD]
BERLIN	198	62.42 [± 5.69]	111	61.94 [± 5.96]	87	63.03 [± 7.03]
FHCRC	392	59.84 [± 6.99]	224	56.84 [± 6.85]	168	60.07 [± 7.03]
IPO-Porto	164	61.21 [± 6.01]	87	63.68 [± 5.67]	77	62.94 [± 6.22]
TAMPERE	174	68.18 [± 7.96]	105	70.53 [± 7.95]	69	72.32 [± 8.32]
UKGPCS	129	61.12 [± 7.37]	58	65.72 [± 6.23]	71	67.22 [± 5.67]
ULM	164	62.52 [± 8.46]	99	62.84 [± 6.09]	65	62.88 [± 6.01]
Total	**1,221**	**62.95 [± 7.94]**	**684**	**62.25 [± 7.03]**	**537**	**63.83 [± 7.72]**

Table 25: The *TMPRSS2-ERG* fusion and clinical/pathological characteristics in 1,221 prostate cancer patients. For more information on the parameters' distribution, please refer to Table 3. Fusion status was defined as dependent, and the clinical continuous or categorical parameters as independent variables in a logistic regression model, adjusted for study center. Given are the odds ratios and 95% confidence intervals for the chance to find a *TMPRSS2-ERG* fusion in a patient's carcinoma. The p-values were estimated using a logistic likelihood ratio test.

parameters	*classification*	*odds ratio [95% CI]*	*p-value*
age at diagnosis	per year	0.965 [0.948-0.982]	< 0.0001
PSA at diagnosis, ng/ml	< 4	ref.	0.0422
	4 - 10	0.687 [0.445 - 1.061]	
	> 10	0.563 [0.355 - 0.893]	
Gleason Score	< 7	ref.	0.1623
	7	0.781 [0.604 - 1.011]	
	> 7	0.847 [0.578 - 1.241]	
Stage	local (T1/T2)	ref.	0.1589
	advanced (T3/T4 and/or N1/M1)	1.198 [0.931 - 1.538]	
aggressive carcinoma	no	ref.	0.7368
	yes	0.959 [0.753 - 1.222]	

Since particular risk variants as well as clinical parameters appeared associated with the T2E fusion in the present work, interactions between the genetic and clinical parameters with respect to their influence on fusion status were investigated next. For this purpose, a logistic regression model was set up including T2E status as dependent variable, clinical variables and genetic variants as covariates, and with adjustment for center effects. The analysis was restricted to the two global significant SNPs. As shown in Table 26, the estimated per-allele ORs for rs1859962 and rs16901979 hardly differ from the results derived from the unconditional Mantel-Haenszel meta-analysis (previous significance: p = 0.0015 and p = 0.0006, respectively), but there is evidence for an influence of PSA on the association between rs1859962 and T2E fusion status as indicated by a weaker p-value (p = 0.0169) and to a lower extent, also of Gleason Score (p = 0.0040). However, these two parameters are correlated with each other (p < 0.0001).

Table 26: Influence of clinical/pathological parameters on the association results of rs1859962 and rs16901979 with the *TMPRSS2-ERG* fusion status in 1,221 prostate cancer patients. For more information on the parameters' deviation, please refer to the patients' description in Table 3. Fusion status was modeled as dependent variable with the SNP and the respective covariate as independent variables using a logistic regression. All analyses were adjusted for study center effects. The per-allele odds ratios and 95% confidence intervals are given for the chance to find a *TMPRSS2-ERG* fusion in a patient's carcinoma. The p-values were estimated using a logistic likelihood ratio test

parameters	*odds ratio [95% CI]*	*p-value*
rs1859962		
crude per-allele odds ratio	1.301 [1.106 - 1.531]	0.0015
adjusted for age	1.303 [1.106 - 1.535]	0.0015
adjusted for PSA	1.239 [1.039 - 1.478]	0.0169
adjusted for Gleason	1.274 [1.080 - 1.502]	0.0040
adjusted for stage	1.302 [1.105 - 1.533]	0.0015
adjusted for aggressiveness	1.298 [1.103 - 1.529]	0.0017
rs16901979		
crude per-allele odds ratio	0.533 [0.370 - 0.768]	0.0006
adjusted for age	0.519 [0.359 - 0.750]	0.0004
adjusted for PSA	0.503 [0.335 - 0.756]	0.0008
adjusted for Gleason	0.528 [0.364 - 0.766]	0.0007
adjusted for stage	0.534 [0.371 - 0.770]	0.0007
adjusted for aggressiveness	0.531 [0.368 - 0.765]	0.0006

In summary, the association study between the somatic T2E fusion status and germline variants pointed out two genomic loci, rs16901979 on 8q24 and rs1859962 on 17q24, which are globally significantly enriched in T2E negative and positive subgroups, respectively. In addition, the previously reported associations between age at diagnosis as well as PSA and fusion status could be verified also in the present sample set.

3.3 Utilization of mechanistic links to elucidate candidate risk regions

The mechanistic analyses of common variants (see 3.2) revealed four independent regions, which were therefore investigated for candidate genes: rs10993994 at 10q11 (3.3.1), rs1859962 at 17q24 (3.3.2) and the two 8q24 SNPs rs1601979 and rs1447295 (3.3.3).

3.3.1 Analyses of prostate cancer candidate genes on 10q11

The PrCa associated region 10q11, represented by the SNP rs10993994, appeared as a strong candidate locus for T2E positive tumors, although with a lack of global significance after the replication stage in the described association study (see section 3.2.2). Nevertheless, the link between genomic fusion rearrangements and DNA repair is intriguing. As evidence has been found for a DNA repair modifying factor within 10q11 (see section 3.2.1), the region was analyzed for potential candidate genes. The rs10993994 variant is located within the promoter of the seminal fluid protein coding *MSMB* gene and is known to influence its expression [25,124,163]. The gene *NCOA4*, which is located 16 kb telomeric, is also affected by rs10993994 [124] and it seemed likely that the variation could have a general effect on the transcriptional activity of the entire region. Thus, further genes in region were taken into consideration, with a particular focus on candidates that could be involved in DNA repair and/or the formation of oncogenic rearrangements.

3.3.1.1 Selection of candidate genes

Schematic overviews of the region around the SNP rs10993994 are given in Figure 21. There are several genes present in the region which are interesting in the context of DNA damage response and repair. *MSMB* (microseminoprotein β, β-MSP, PSP94) is coding for the third most prevalent protein in the seminal fluid, and though there is no obvious connection to DNA damage response, it is the most favorite candidate due to the localization of rs10993994. *NCOA4* (nuclear receptor coactivator 4), approx. 16 kb telomeric from rs10993994, encodes a coactivator of the androgen receptor, also known

as ARA70. Its expression was previously shown to depend on the genotype of rs10993994. Thus it was also included in the analyses. The *PARG* gene encodes the poly (ADP-ribose) glycohydrolase, which is located ~ 175 kb centromeric of rs10993994. Though the relatively long distance, the gene was taken into consideration due to (1) an alternative transcription start site ~ 180 kb telomeric of the SNP (Figure 21) and (2) the interesting function as counterpart of the PARP protein (poly (ADP-ribose) glycohydrolase). A further conspicuous gene is *TIMM23* and its genomic duplicate *TIMM23B* ((translocase of inner mitochondrial membrane 23 homolog (yeast)), respectively, though previous investigations for expressional changes depending on the rs10993994 genotype observed no differences [124]. On the one hand, the authors have not distinguished between the two genomic copies which probably blurred the results, and on the other hand, the protein represents a mitochondrial channel protein, a plausible player in the apoptosis pathway [52]. The screenshot from the UCSC genome browser (Figure 21) illustrates the situation on the genomic region of 10q11, which is schematically simplified in Figure 21. The *TIMM23* copy, which shares a transcription start site with *PARG* [103], is referred as *TIMM23B*.

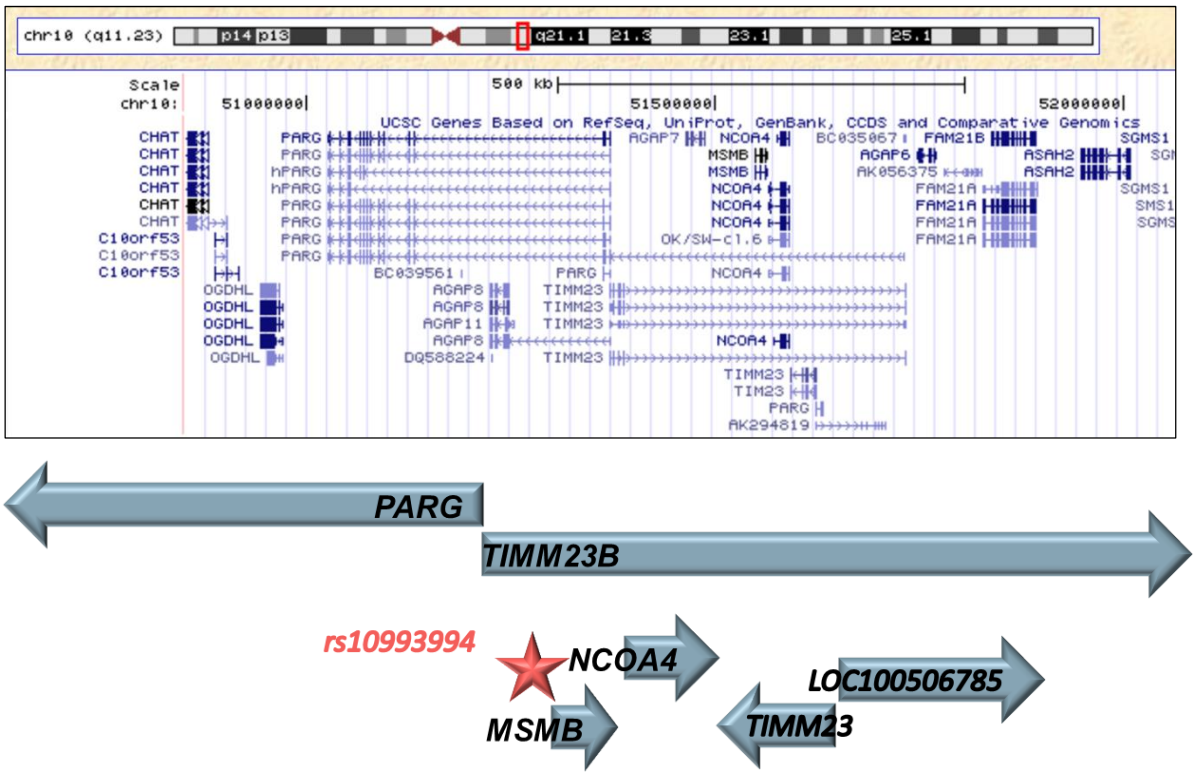


Figure 21: Schematic overview of the prostate cancer susceptibility region on chromosome 10q11. The prostate cancer risk SNP is located within the promoter region of *MSMB*.

3.3.1.2 Transcript analysis of candidate genes in 10q11

Prior qRT-PCR analyses of prostate tissue specimen, the transcripts of the nominated genes of interest were analyzed with standard PCR methods in order to clarify some questions. As the rs10993994 variant is located within the proximal promoter region of *MSMB*, it is the most obvious candidate gene for the genotype-dependent changes in DNA damage response and repair (3.2.1). The first objective is to check if it is transcriptionally active in blood lymphocytes, since it is actually coding a mucous protein. The *MSMB* gene consists of four exons and, according to the ensemble genome browser, there are two splice isoforms, one full-length (ENST00000358559) and one product lacking exon 3 (ENST00000298239). The PCR on cDNA of different cell types revealed strong bands in prostate tumor and normal tissue and weak in LNCaP cells (Figure 22). CDNA from peripheral blood lymphocytes yielded no detectable product.

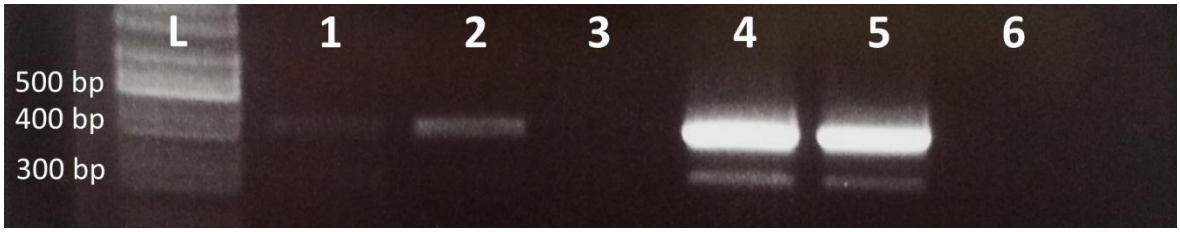


Figure 22: PCR results indicating the presence of the *MSMB* transcript in RNA of different cell types. The RNA was transcribed in cDNA prior PCR reaction. The primers used for this approach target the two splice variants of the *MSMB* transcript resulting in a 408 bp and a 302 bp product. **L =** 100 bp Ladder; **1** = LNCaP, γ-irradiated with 410 Gy; **2** = LNCaP, untreated; **3** = Peripheral blood lymphocytes from a healthy 30-year old proband with rs10993994 genotype CC; **4** and **5** = prostate tumor tissue and corresponding histologically normal prostate tissue from a patient with rs10993994 genotype TT; **6** = negative control.

The *PARG* gene, which is an interesting candidate gene with respect to its functions, is located rather distant from rs1099399 (approx. 175 kb centromeric). According to the UCSC genome browser, a transcription start site is annotated ~ 180 kb telomeric (no accession number available) resulting in a long transcript spanning the SNP region. The long transcript is also displayed in the middle of the screenshot in Figure 21. The attempt to detect this putative isoform using PCR primers failed in all tested cell types (data not shown). As the first exon, the only distinctive feature, was only 45 bp in size, there was no scope to design alternative primers in order to prove or rule out its existence.

For quantitative expression analyses, TaqMan assays for *MSMB*, *NCOA4*, *PARG*, *TIMM23* and *TIMM23B* (Table 7) were designed for qRT-PCR (2.3.8). As the initial evidence for a DNA damage response and repair contribution of the 10q11 gene locus was found in peripheral blood lymphocytes (3.2.1), this cell type was tested for genotype specific expressional pattern first. Thirteen healthy volunteers from the Institute of Human Genetics, Ulm donated a blood sample for rs10993994 genotyping and for a lymphocyte culture with subsequent expressional analyses. The results are displayed in Figure 23. Due to the expressional lack in lymphocytes (Figure 22), *MSMB* was not tested any more. Though the informative value is limited by the small number of probands, it becomes apparent, that the expression of *TIMM23* and *TIMM23B*, as well as of *NCOA4*, by trend, could possibly be increased in homozygous risk-allele (T) carriers of rs10993994. None of these differences were statistically significant, neither using the Pearson (problematically due the small sample size), nor the Spearman rank coefficient. For the *PARG* expression, no tendency for a genotype-dependent difference was observed. As indicated in Table 27, however, the expressions of the individual genes are related to each other.

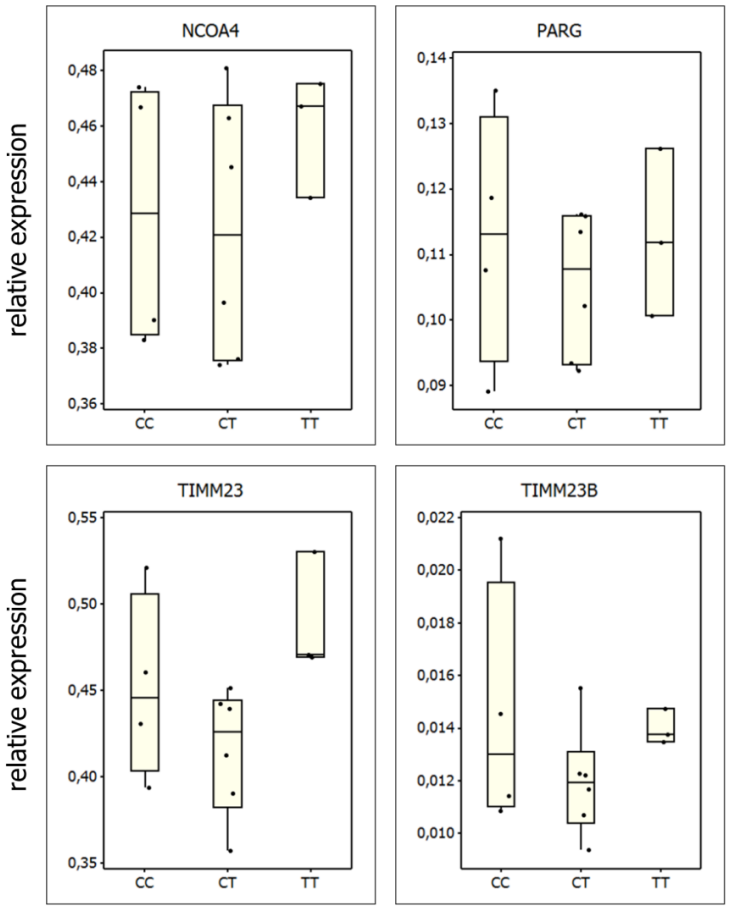


Figure 23: Relative expression of *NCOA4*, *PARG*, *TIMM23* and *TIMM23B* in peripheral blood lymphocytes dependent on the rs10993994 genotype. The lymphocyte cultures were allowed to grow 3 days at 37 (as the basal culture described in 2.3.14.2) Gene expressions were measured relative to the reference gene *G6PD* in n = 4 probands with CC, n = 6 with CT and n = 3 with TT genotype. The T allele is the prostate cancer risk associated allele.

Table 27: Correlation matrix for the relative expressions of *PARG*, *NCOA4*, *TIMM23* and *TIMM23B* in peripheral blood lymphocytes measured in 13 healthy volunteers. Given are the Spearman coefficients (rho) and the respective p - values. P-values < 0.05 are printed in bold.

	PARG	*NCOA4*	*TIMM23*
TIMM23B	rho = 0.549 p = 0.0570	rho = 0.824 **p = 0.0043**	rho = 0.769 **p = 0.0077**
TIMM23	rho = 0.566 **p = 0.0499**	rho = 0.698 **p = 0.0156**	
NCOA4	rho = 0.489 p = 0.0903		

3.3.1.3 Gene expression analysis in prostate tissue

An important step to evaluate the relevance of the 10q11 genes in PrCa tumorigenesis is the examination of expressional differences between tumor and normal tissue. *MSMB*, *NCOA4*, *PARG*, *TIMM23* and *TIMM23B* expression were measured in RNA derived from prostate tumor- and histologically normal tissue from 87 patients (49 from Ulm and 38 from Erlangen). The boxplots from the analysis on differences between normal- and tumor tissue is shown in the upper panel of Figure 24. The Wilcoxon Test for dependent samples showed *MSMB* highly significant down-regulated in tumor compared to normal tissue ($p < 0.0001$), *PARG* and *TIMM23*, on the other hand, were up-regulated (both $p < 0.0001$). *NCOA4* was borderline significant up-regulated in tumor tissue ($p = 0.0515$). No difference between normal- and tumor tissue was observed for *TIMM23B* ($p = 0.6190$).

Since 10q11 was found to play a role especially in T2E positive PrCa, the question was raised, if expressional changes of the local candidate genes were subtype-specific. To test this hypothesis, 38 specimens from Erlangen and six from Ulm were assessed for the presence or absence of the fusion by qRT-PCR (2.3.8). Within this collective, the T2E frequency was 38.3 % (n = 17). As expected, the T allele frequency among fusion positive patients was higher (44.1 %) than in negative (40.7 %). Thirty-five Ulm samples were already typed for the meta-association study between T2E fusion and common variants (3.2.2). For eight Ulm specimens, the assessment of the fusion status was not possible. In total, fusion status was available for 79 patients (positive: n = 38, negative: n = 41)

The subgrouping in T2E positive and negative carcinoma (boxplots displayed in the lower panel of Figure 24) and a consequently splitted comparison using the Wilcoxon Test for paired comparisons showed several differences between the two somatic subtypes. The

increased expression of *NCOA4* appears to be restricted to T2E positive specimen (positive: p = 0.0057, negative: p = 0.9250). More significantly different in T2E positive patients than in negative were also *PARG*, *TIMM23* and to a less extent *MSMB* (p-values are indicated in Figure 24).

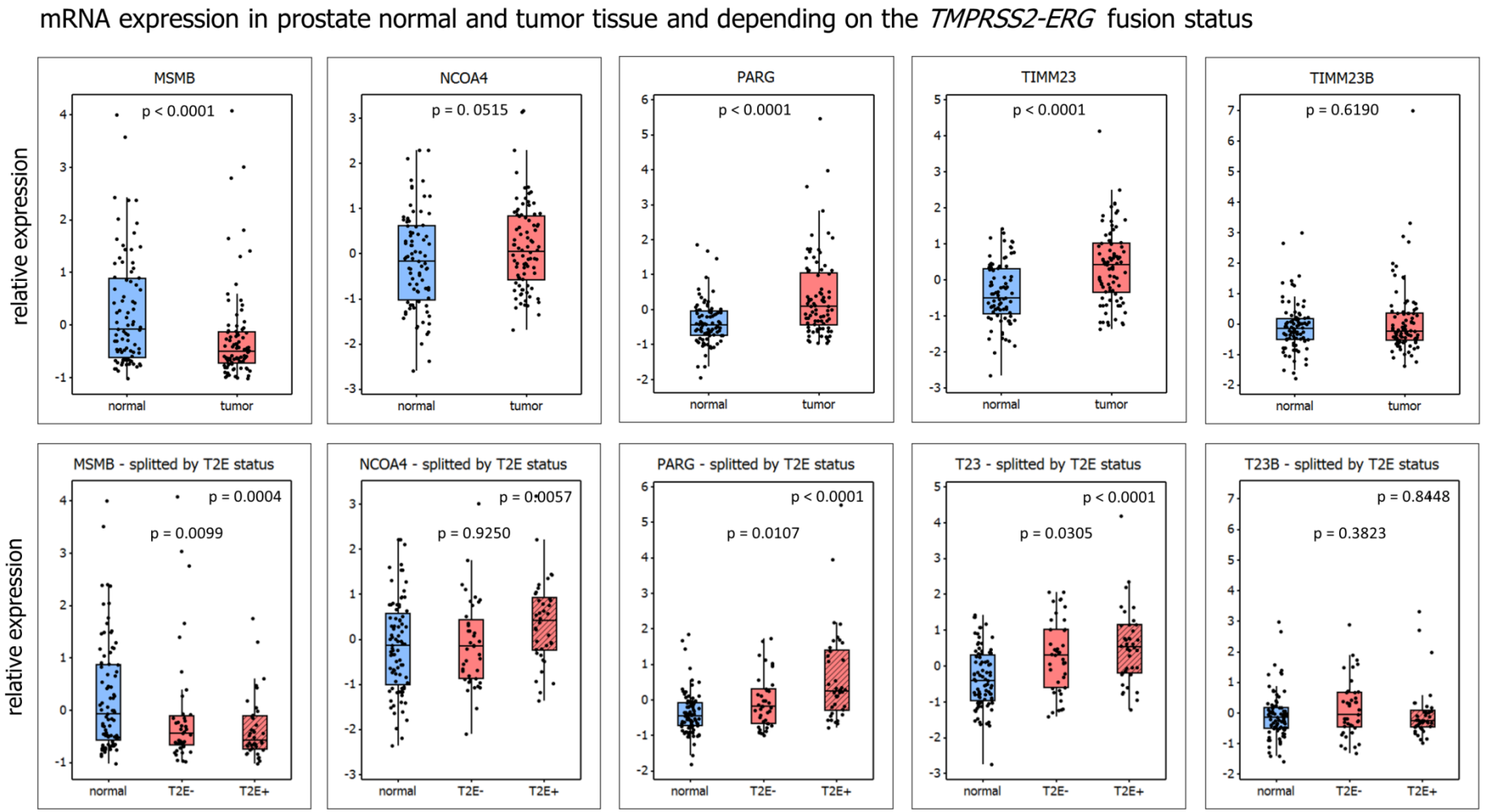


Figure 24: Relative Expression of 10q11 region genes in prostate normal- and tumor tissue and dependent on the *TMPRSS2-ERG* status of 87 patients. Presented are the box plots as well as the single values from the z transformed results of normal (blue) and tumor (red) tissue. In the bottom panel, the expression in tumor tissue is splitted in *TMPRSS2-ERG* negative (T2E-) and positive (T2E+). The indicated p-values result from testing on differences between the normal/tumor tissue pairs with the Wilcoxon signed-rank test, either among the total sample, or splitted by fusion status.

To identify possible interactions between the genes on 10q11, a correlation matrix was generated. As displayed in Table 28, there were multiple significantly correlated gene pairs. Notably, all of them were correlated positively (positive rho values). *TIMM23* is co-expressed with all investigated genes, except with *MSMB* in tumor tissue, *TIMM23B* is correlated with *PARG* and *TIMM23*.

Table 28: Correlation matrix for the relative expressions of the investigated genes on chromosome 10q11 in normal (green) and tumor (red) tissue. The Spearman coefficients (rho) and the respective p- values are given. P-values < 0.05 are printed in bold.

	MSMB	*PARG*	*NCOA4*	*TIMM23*	*TIMM23B*
MSMB	-	rho = 0.173 p = 0.1129	rho = 0.407 **p = 0.0002**	rho = 0.257 **p = 0.0185**	rho = 0.104 p = 0.3402
PARG	rho = -0.192 p = 0.0770	-	rho = 0.335 **p = 0.0021**	rho = 0.525 **p < 0.0001**	rho = 0.398 **p = 0.0003**
NCOA4	rho = 0.114 p = 0.2912	rho = 0.380 **p = 0.0005**	-	rho = 0.384 **p = 0.0004**	rho = 0.154 p = 0.1590
TIMM23	rho = -0.029 p = 0.7925	rho = 0.599 **p < 0.0001**	rho = 0.443 **p < 0.0001**	-	rho = 0.473 **p < 0.0001**
TIMM23B	rho = -0.020 p = 0.8510	rho = 0.316 **p = 0.0036**	rho = 0.001 p = 0.9894	rho = 0.523 **p < 0.0001**	-

The gene expression patterns dependent on the rs10993994 genotype are shown in Figure 25. The first impression of the blots is the decreased expression of *MSMB* and the increased expression of *NCOA4* with the risk genotypes. However, for the sake of objectiveness, the data were analyzed under a linear model (expression dependent on the number of risk alleles = 0, 1, 2). For this purpose, Cochran-Mantel-Haenszel statistics, which correspond to a stratified Jonckheere-Terpstra Test, were applied using 'study center' (Erlangen and Ulm) as a stratum. These tests revealed a highly significant linear relationship for *MSMB* in normal and tumor tissue (p = 0.0050 and p = 0.0024), for *TIMM23B* in normal tissue (p = 0.0062) and by trend in tumor tissue (p = 0.0581), and a less significant correlation for *NCOA4* in tumor tissue (p = 0.0354) and by trend in normal tissue (p = 0.0900). For the other two genes, there was no evidence for a linear relationship (*PARG* normal p = 0.2798, tumor p = 0.1669; *TIMM23* normal p = 0.1943, tumor p = 0.2748).

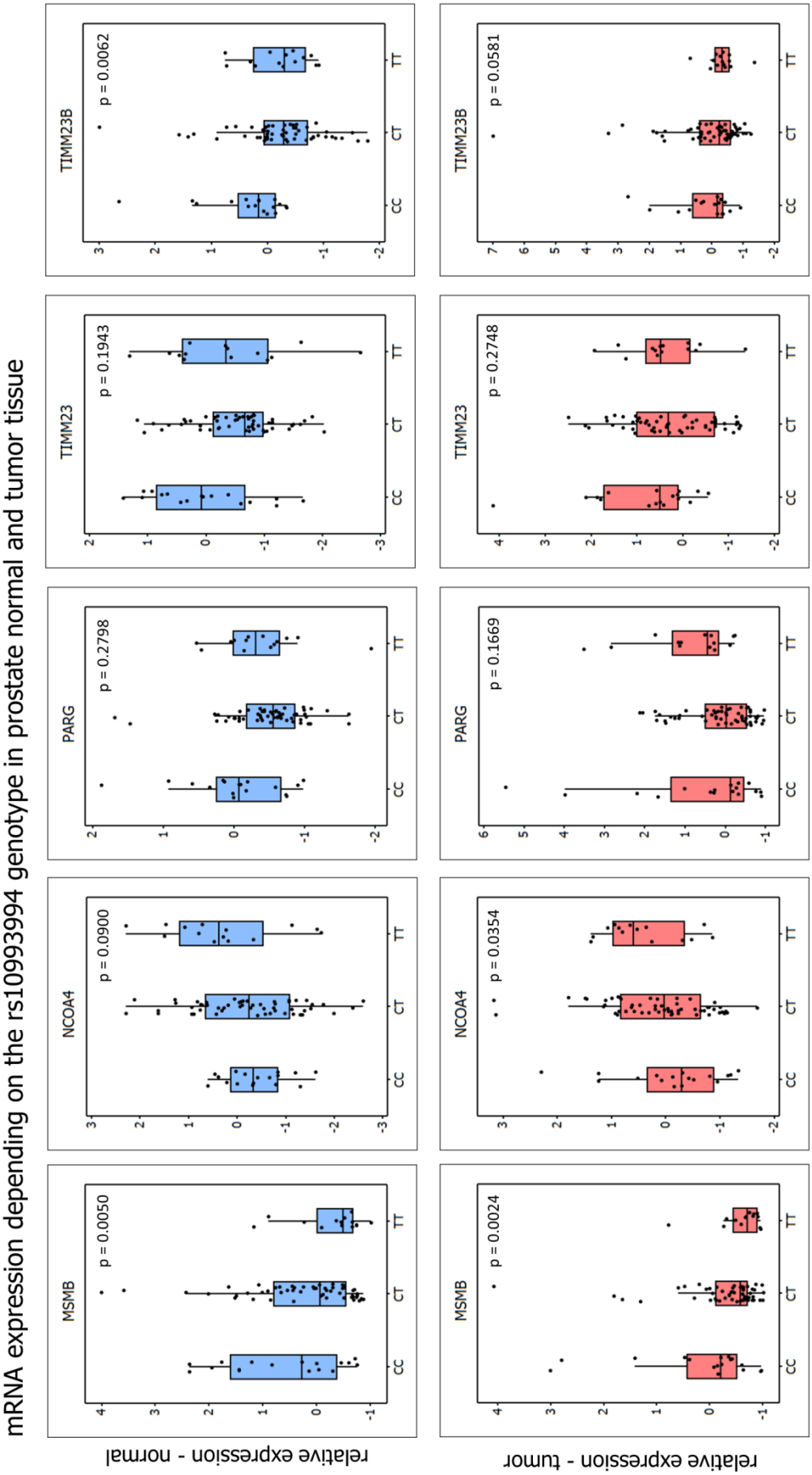


Figure 25: Relative expression of the 10q11 region genes dependent on the rs10993994 genotype in prostate normal- and tumor tissue of 87 patients. Shown are the box plots as well as the single values from the z transformed results of normal (blue) and tumor (red) tissue. The T allele represents the prostate cancer risk associated allele. The given p-values were calculated with the Cochran-Mantel-Haenszel statistics assuming a linear model (expression dependent on the number of risk alleles n = 0, 1, 2).

Genotype-dependent expression patterns were further analyzed in tissues in the subtypes of T2E positive and negative cases. Linear correlation was estimated analogous as described above. All box plots with the respective p-values are provided in Appendix E. As expected, due to the strong subgrouping, there were no highly significant results. The expressional decrease for *MSMB* linked to the rs10993994 T allele (see Figure 25), is again apparent when stratified for T2E fusion status, in fact for both fusion negative and positive tumors, however genotype differences are not significant in normal tissue.

The next question was focused on the dynamics of expressional changes in tumor versus normal tissue, in other words, if any of the genes' observed downregulation (*MSMB*) or upregulation (*NCOA4*, *PARG*, *TIMM23*) could be influenced by genotypes of the risk SNP.
In this alternative approach, the fold-change between normal- and tumor tissue pairs was assessed in dependence of the genotype and fusion status. The correlation between the percent-change values and number of risk alleles was not estimated with Cochran-Mantel-Haenszel statistics as there was no stratification on center necessary. Instead, Spearman rank correlation with correction for ties was used and revealed a borderline significant association between the number of rs10993994 T alleles and the percent-change values of *PARG* (rho = 0.208, p = 0.0585) (Figure 26A). Splitting for T2E fusion status, this trend remained solely in T2E negative (rho = 0.272, p = 0.0893), and not in T2E positive cases (rho = -0.028, p = 0.8669) (Figure 26B). No significant correlation was observed for *MSMB*, *NCOA4*, *TIMM23* and *TIMM23B*, neither in the total sample, nor in the T2E negative and positive subgroups. The box plots with the corresponding correlation results are attached in the Appendix F.

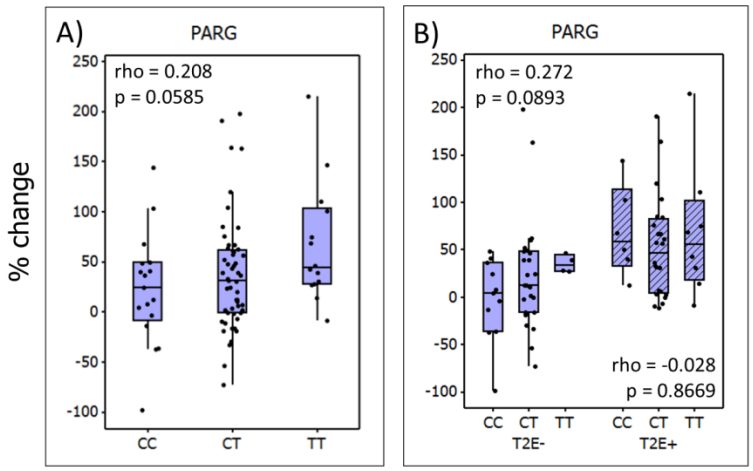


Figure 26: Percentage change of *PARG* expression in tumor versus normal tissue dependent on rs10993994 genotype. The expressional change is indicated in %; values < 0 indicate a decreased, values > an increased expression in tumor compared to normal tissue. The stated p- and rho-values were estimated with a Spearman rank correlation with correction for ties to correlate the number of risk alleles (T) with the expression data. **A)** Expressional changes observed in the total sample set with paired expression data (n = 84) dependent on genotype. **B)** Expressional changes observed in the *TMPRSS2-ERG* negative (T2E-, n = 40) and positive (T2E+, n = 38) subgroups dependent on the genotype (n = 78).

Taken together, a whole series of interesting candidate genes on the 10q11 gene locus was selected and assessed for expressional changes in dependence on the rs10993994 genotype, on tissue type and T2E fusion status. The analyses have led to an abundance of different findings, thus the identification of a particular gene/pathomechanism is still difficult. On the one hand, the expression levels of *MSMB*, *NCOA4* and *TIMM23B* in tumor and/or normal tissues show static dependence on genotypes. On the other hand, the *PARG* gene remains an interesting candidate with respect to the dynamics of expression, since the observed up-regulation in tumor (vs. normal) tissue could be driven by risk alleles.

3.3.1.4 Genes influencing *TMPRSS2-ERG* formation (*TMPRSS2-ERG* induction assay)

The influence of an mRNA knockdown on the T2E fusion formation was tested for a further evidence for a participation of candidate genes in DNA repair and/or in development of a fusion positive PrCa. Therefore, LNCaP cells were treated with siRNAs against the transcripts of *MSMB*, *NCOA4*, *PARG* and *TIMM23* and were subsequently subjected to the T2E induction assay. *TIMM23B* was not studied since no *TIMM23B*-specific siRNA was available at this time. An mRNA knockdown of *ESCO1* was previously shown to increase T2E fusion formation, and thus, this gene was assayed simultaneously as a positive control.

At first, the previously established irradiation protocol [94] was re-evaluated, as the radiation sensitivity of the newly thawed LNCaP cell line may be different. Figure 27 shows the relative T2E fusion transcript formation depending on the γ-irradiation intensity in Gray. The results showed that the previously established irradiation strength of 300 Gy led to a robust formation of the fusion transcript. Though at 400 Gy the fusion induction was even higher, the results were not that stable (increased standard deviation).

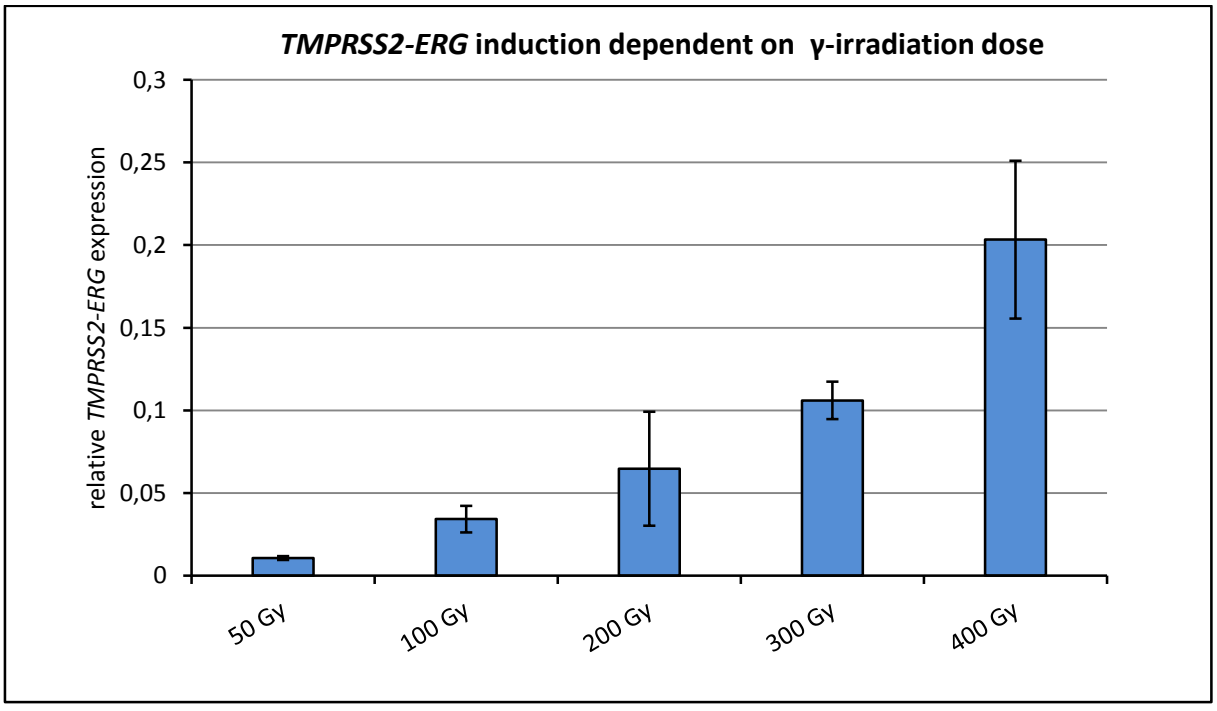


Figure 27: γ-irradiation dose-dependent formation of the *TMPRSS2-ERG* fusion transcript in LNCaP. The *TMPRSS2-ERG* expression was measured relative to the wild-type *TMPRSS2.* The bars represent the mean of n = 2 experiments (± standard deviation).

A time response experiment was performed in order to ensure, that the mRNA knockdown of candidates is present before, after and at the time point of irradiation (72 h after siRNA transfection). For *ESCO1*, a sufficient knockdown efficiency over the time has been shown previously [94]. As displayed in Figure 28, all genes except *MSMB* were silenced from the point of 24 h up to 120 h after transfection on a residual level of less than 13 % compared to the initial expression. The *MSMB* knockdown declined slower and finally reached a robust and stable value of ~ 85 % at 72 h. As *MSMB* is only weakly expressed in LNCaP (Figure 22), within a range where the measurements tend to fluctuate, a knockdown is naturally more difficult to detect. The high *MSMB* transcript abundance at 24 h is probably also due to such fluctuations. Nevertheless, this experiment shows that the siRNA approach is suitable for a sufficient knockdown already before the time of irradiation, scheduled at 72 h, which increases the chance that the respective protein levels were also decreased. At the time period of recovering, 48 h after irradiation until cell harvest, the knockdown was definitely maintained.

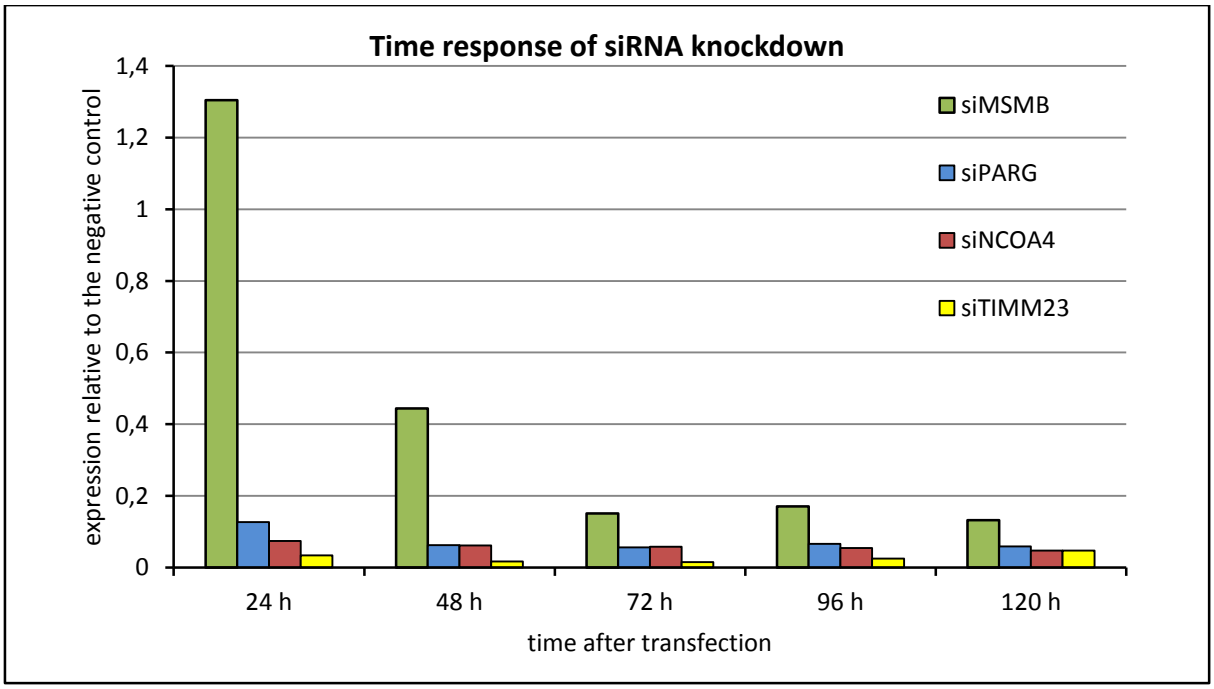


Figure 28: Control of the knockdown persistence of siRNA treatment over an extended time period in LNCaP. The mRNA expressions of the genes of interest were measured at different time points after siRNA transfection and normalized to the housekeeper gene *G6PD*. The bars represent the data relative to the expression of the respective genes in LNCaP which were treated with AllStars negative control siRNA (set = 1). The *ESCO1* knockdown was verified previously [94].

The main experiments of candidate gene silencing in the course of the T2E induction assay were repeated six times (Figure 29). They were always accompanied by knockdown controls (Figure 30) which were transfected simultaneously and harvested at the time of radiation induction. The average knockdown efficiency at the time of irradiation was 76 % for of *ESCO1*, 93 % for *PARG*, 88 % for *MSMB*, 98 % for *TIMM23* and 91 % for *NCOA4*. The downregulation of *ESCO1* (positive control) in LNCaP led to a weak but significantly increased T2E fusion induction compared to the LNCaP treated with AllStars negative control siRNA (1.8 fold, p = 0.0248). The silencing of *PARG* even resulted in a significant, 3.0 fold increased T2E fusion formation (p = 0.0015). The knockdown of all other investigated genes increased the fusion formation compared with the control slightly, but not significantly (*MSMB*: 1.3 fold, p = 0.0660; *TIMM23*: 1.4 fold, p = 0.1105; *NCOA4*: 1.7 fold, p = 0.0825).

Taken together, the *TMPRSS2-ERG* induction assay qualifies *PARG* as a strong candidate gene at the PrCa risk locus 10q11, as a knockdown significantly enhanced the relative T2E formation after γ-irradiation in LNCaP cells.

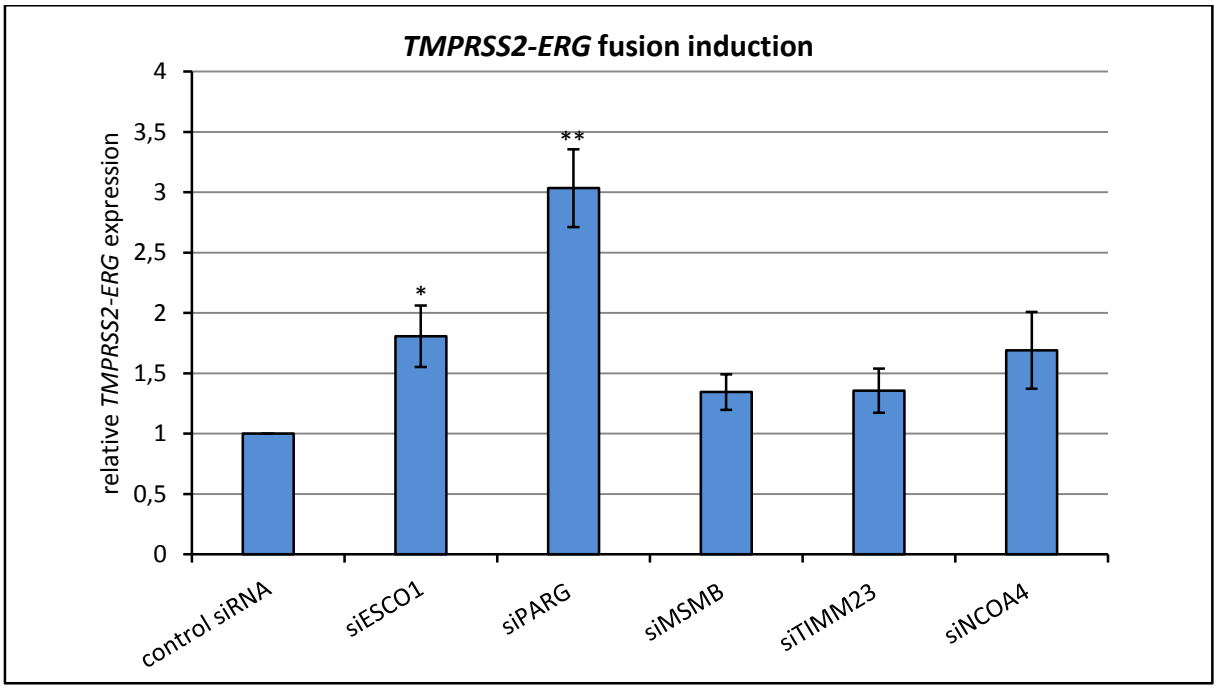


Figure 29: The Induction of *TMPRSS2-ERG* fusions in LNCaP cells treated with varying siRNAs. The most predominant *TMPRSS2-ERG* transcript variant *T1G4* was measured relative to the *TMPRSS2* expression. The results of the respective experiments are displayed in relation to the AllStars negative control siRNA. The bars represent the mean of n = 6 independent experiments (± SEM, * $p < 0.05$, ** $p < 0.01$).

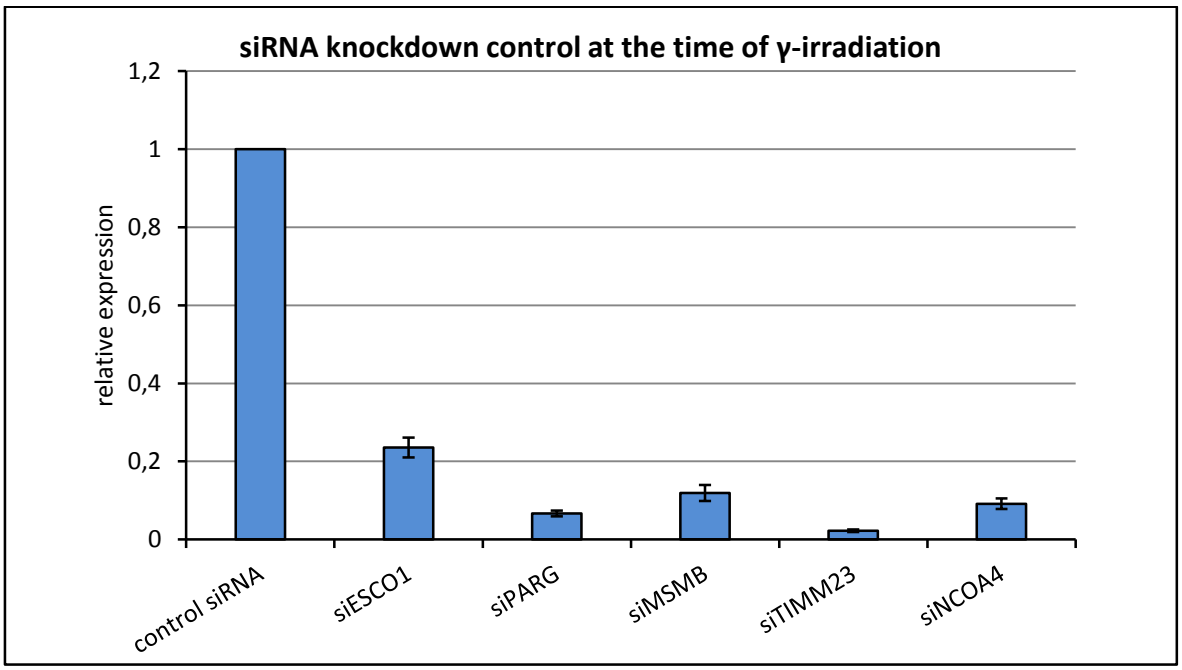


Figure 30: Control of the knockdown efficiency of the siRNA treatment in LNCaP. The expression of the respective genes was measured relative to the housekeeper *G6PD* and normalized to the expression in LNCaP cells treated with the AllStars negative control siRNA. The bars represent the mean of n = 6 independent experiments (± SEM).

3.3.2 Expressional analysis of the 17q24 candidate gene *SOX9*

The rs1859962 PrCa risk variant was shown to be overrepresented in fusion positive carcinoma (3.2.2). The respective region 17q24 is a prolonged gene desert with no gene within ~ 1 megabase up- and downstream of the SNP (Figure 31). The closest gene in the centromeric direction of chromosome 17 is *KCNJ2* (potassium inwardly-rectifying channel, subfamily J, member 2), approx. 900 kb in distance. Approx. 1 megabase in the telomeric direction, the *SOX9* (SRY (sex determining region Y)-box 9) gene is located, which has already been shown to be influenced by an enhancer element within the rs1859962 LD block by long-range interactions [189]. Thus, *SOX9* appeared as a tempting candidate for an examination of its expression levels concerning genotype and fusion status.

SOX9 gene expression was measured by qRT-PCR in histologically normal and tumor tissue of a total of 70 patients (35 from Erlangen and 35 from Ulm). For the tissue set from Erlangen, the rs1859962 genotypes were determined by TaqMan genotyping (2.3.9.2), while for the Ulm series the phenotype and genotype data were already available from the association study (3.2.2).

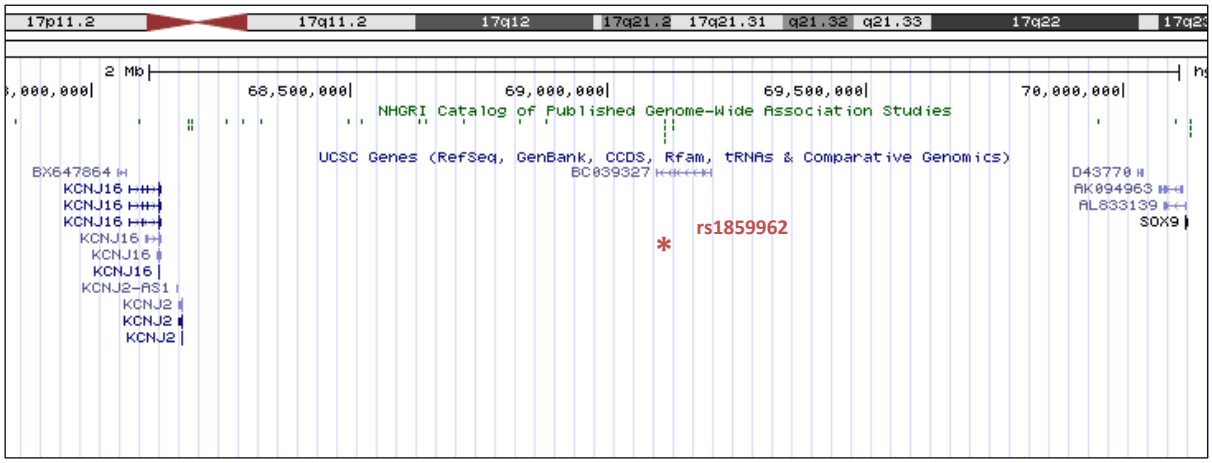


Figure 31: Modified screenshot of the 17q24 region displayed by the UCSC genome browser.

The rs1859962 genotype distribution of complete series was as follows: 22.1 % (n = 15) were homozygous for the PrCa non-risk allele T, 54.4 % (n = 37) were heterozygous and 23.5 % (n = 16) were homozygous for the risk allele G. The resulting risk allele frequency in this entire group was 50.7 %, and as expected, enriched in T2E positive (53.1 %) versus negative cases (48.6 %). A complete data set, including *SOX9* expression data, T2E and genotype information could be generated for 95.7 % (n = 67) of the patients. For the Erlangen dataset, the risk allele frequency was quiet low (42.9 %), and not enriched in positive (40.9 %) versus negative (68.2 %).

The *SOX9* expression was increased in tumor versus normal tissue (Figure 32A) by approx. 50 %. The paired comparison using the Wilcoxon Test yielded a borderline significant p-value of 0.047. When splitting the sample according to the T2E fusion status, the overexpression of *SOX9* was not present in fusion negative cases (p = 0.94), but was highly significant in fusion positive cases (p = 0.0049) (Figure 32B). The relative *SOX9* expression showed no significant evidence for a genotype-dependent pattern, neither regarding normal and tumor tissue (Figure 32C), nor further splitted by the T2E subgroups (Figure 32D/E). Solely in T2E negative tumor tissue, a slight trend towards a decreased expression correlated with the rs1859962 risk allele G could be suggested (Figure 32E).

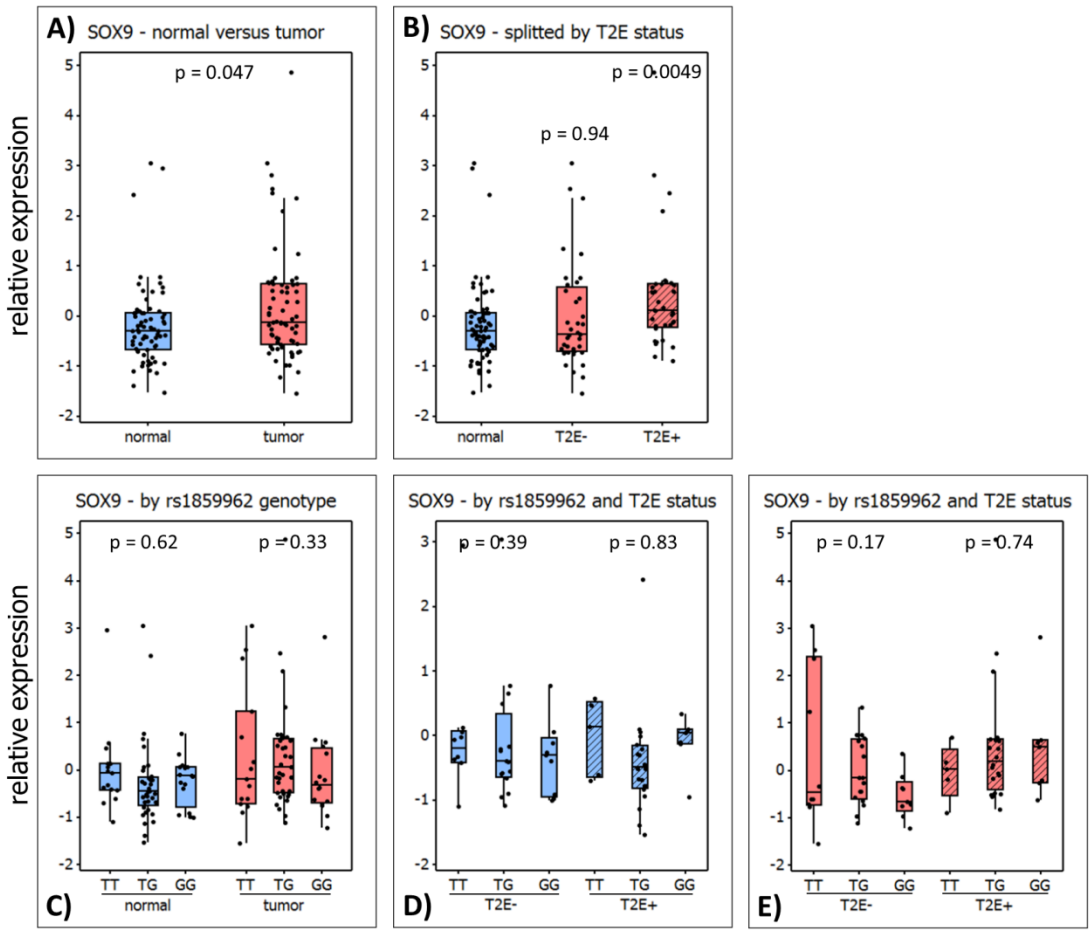


Figure 32: Relative mRNA expression of *SOX9* in prostate normal- and tumor tissue of 70 patients. Represented are the box plots as well as single values from the Z transformed results. The paired comparisons were calculated with the Wilcoxon Test, and the genotype-dependent expression under a linear model using Cochran-Mantel-Haenszel statistics. **A)** *SOX9* expression in normal and tumor tissue. **B)** *SOX9* expression in normal tissue compared to *TMPRSS2-ERG* negative (T2E-) and *TMPRSS2-ERG* positive (T2E+) tumor tissue. **C)** *SOX9* expression dependent on the rs1859962 genotype in normal- and tumor tissue. **D)** *SOX9* expression in normal tissue dependent on the rs1859962 genotype and on the T2E status. **E)** *SOX9* expression in tumor tissue dependent on the rs1859962 genotype and on the T2E status.

A further approach to elucidate a correlation between *SOX9* expression, the rs1859962 genotype and the T2E fusion status was the paired analysis of the expressional change between normal and tumor tissue as already explained in section 3.3.1.3. There was no significant correlation between the values of the percentage change values and the genotypes, neither in the unselected sample set (Figure 33A), nor if splitted by the T2E fusion status (Figure 33B).

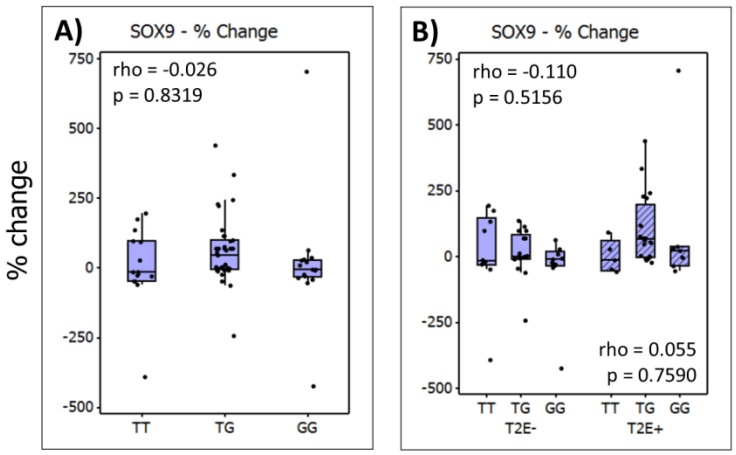


Figure 33: Percentage change of *SOX* expression in tumor versus normal tissue dependent on rs1859962 genotype. The value indicates the expressional change in %; values < 0 indicate a decreased, values > 0 an increased expression in tumor compared to normal tissue. The stated p- and rho-values were estimated with a Spearman rank correlation with correction for ties to correlate the number of risk alleles (G) with the expression data. **A)** Expressional changes observed in the total sample set with paired expression data (n = 68) dependent on genotype. **B)** Expressional changes observed in the *TMPRSS2-ERG* negative (T2E-, n = 36) and positive (T2E+, n = 32) subgroups dependent on the genotype.

In summary, the putative candidate gene of the rs1859962 locus on 17q24, *SOX9* was analyzed for expressional changes dependent on tissue type, genotype and fusion status. An expressional correlation with the number of risk alleles can only be suggested by trend in T2E negative, especially in tumor tissue. With respect to the up-regulation in tumor tissue, which was mainly observed in the T2E positive subgroup, it can nevertheless not be excluded as a candidate gene.

3.3.3 Expressional analysis of the 8q24 candidate gene *MYC*

Two independent PrCa risk regions on 8q24 (region 1 represented by rs1447295 and region 2 represented by rs16901979) were found associated with somatic T2E fusion status (3.2.2). The PrCa risk alleles of both SNPs are overrepresented in T2E negative carcinomas. 8q24 is known as gene desert that spans approx. 1.2 Mb (Figure 34). The most prominent gene at the telomeric border is *MYC*, coding for the transcription factor and proto-oncogene c-Myc. Literature research revealed inconsistent information whether 8q24 PrCa risk variants influence *MYC* expression in prostate tissue [123,144]. Thus, the hypothetical relationship between the 8q24 SNPs and the *MYC* expression was investigated along with the T2E fusion status as additional criterion.

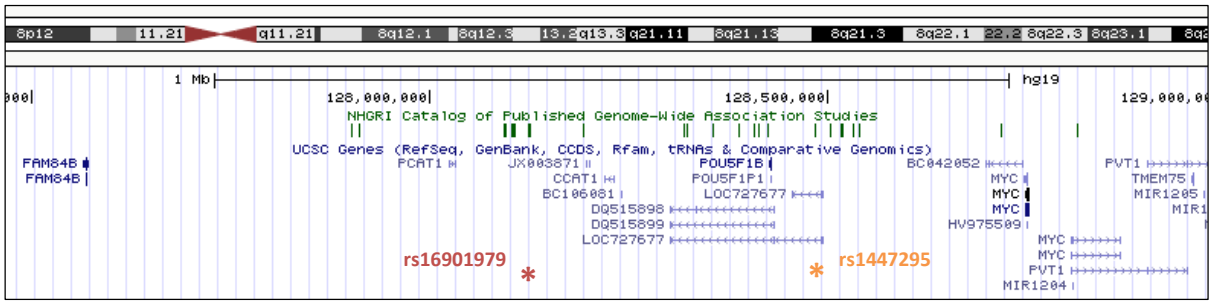


Figure 34: Modified screenshot of the 8q24 region displayed by the UCSC genome browser.

Analogously to 3.3.2, the *MYC* mRNA expression was measured in histologically normal and tumor tissue of 70 patients from Erlangen and Ulm with known T2E fusion status, and the rs1447295 and rs16901979 genotypes were determined.

The genotyping call rate was 97.1 % (n = 68) for rs16901979 and 95.7 % (n = 67) for rs1447295. For the cumulative evaluation of both SNPs, 94.3 % of cases (n = 66) could be included. The overall risk allele frequency of rs16901979 was 4.4 % (3.1 % in T2E positive, 5.6 % in T2E negative) and 15.7 % (14.5 % in T2E positive, 16.7 % in T2E negative) for the rs1447295 SNP, respectively. Taken both SNPs together, the risk allele number in fusion negative cases was increased 1.2 fold as compared to positive cases. In the Erlangen sample set, which was not included in the previous association analysis between SNPs and T2E fusion, the risk allele frequency was 1.5 % for rs16901979, and 20.3 % for rs1447259. Enrichment for rs16901979 risk allele A in fusion negative cases could not be validated here, as there was solely one variant carrier. The rs1447295 risk allele was not

overrepresented in negative (18.2 %) versus positive (25.0 %) cases. However, fluctuations due to small sample size are not uncommon.

The *MYC* expression in tumor versus normal tissue was strongly increased (Figure 35A) by approx. 140 %. The paired comparison using the Wilcoxon test derived a highly significant p-value of < 0.0001. When splitting the sample according to the T2E fusion status, an overexpression was present in both somatic subgroups, which was less significant in fusion negative (p = 0.0041) than in fusion positive cases (p = 0.0002) (Figure 35B).

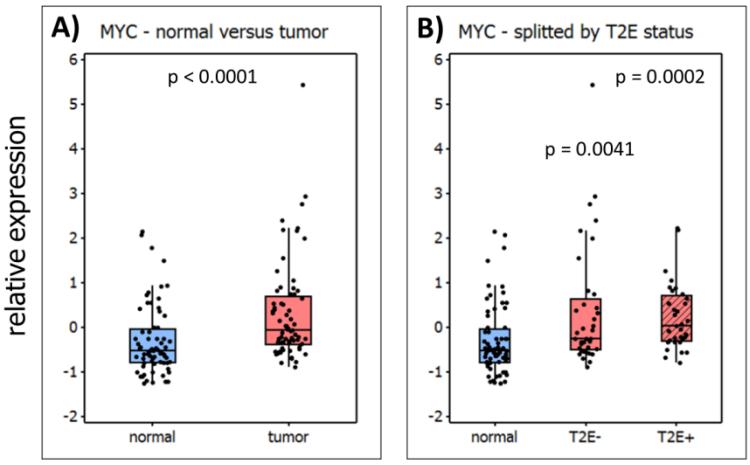


Figure 35: Relative mRNA expression of *MYC* in prostate normal- and tumor tissue of 70 patients. Box plots as well as single values from the Z transformed results are presented. The p-values were calculated with a Wilcoxon Test for paired comparisons between the normal and tumor tissues **A)** *MYC* expression in normal and tumor tissue. **B)** *MYC* expression in normal tissue compared to *TMPRSS2-ERG* negative (T2E-) and *TMPRSS2-ERG* positive (T2E+) tumor tissue.

The crucial question, if the *MYC* mRNA expression is influenced by the rs1447295 or rs16901979 genotype, was difficult to answer due to the low risk allele frequency. Three patients are homozygous for the rs1447295 PrCa risk allele A, and only one patient for the rs16901979 risk allele A. To account for the small subgroups, the exact Jonckheere-Terpstra-Test was used for computations. Figure 36A suggest an increase in *MYC* expression dependent on the number of rs1447295 risk alleles, which can not be statistically proven (normal: p = 0.13, tumor: p = 0.49). The SNP rs16901979 seemed to exert an influence in the same direction, but it is questionable whether this issue can be answered with this small sample set, or not. In order to increase the size of the subgroups, the risk alleles of the two SNPs were pooled hereinafter and used together for analysis. According to Figure 36A, the risk allele number seems to tend to increase the

MYC expression. However, there was still weak statistical support for this finding (normal: p = 0.14, tumor: p = 0.55). Considering the somatic tumor subgroups based on the T2E status, the rs1447259 risk allele A was significantly (p = 0.027) associated with an increased *MYC* expression in T2E negative cases in normal tissue (Figure 36B). Such a linear trend could not be observed in the tumor, however, the homozygous risk allele carriers seemed to be increased (Figure 36C). It could be suggested that the rs16901979 variant also causes an increased *MYC* expression, but the low risk allele frequency allows no statistical significant results. The combined analysis of both SNP alleles revealed a borderline significant increase of the *MYC* expression in T2E negative cases in normal tissue (p = 0.050), but not in tumor tissue (p = 0.50).

In addition to the standard evaluation procedures, the percentage change values were estimated, which indicate the percentage by which the tumor tissue is differentially expressed compared with normal tissue (Figure 37). However, no statistically significant change was observed.

In summary, the *MYC* expression was assessed in 70 prostate normal and tumor tissue pairs, in order to evaluate expressional effects in relation to tissue type, 8q24 genotypes and T2E fusion status. Though the results should be handled with care due to the small sample sizes and accordingly small subgroups of specimens with rare genotypes, there is evidence for an influence of risk alleles, particularly of the rs1447295 SNP, towards an increased *MYC* expression, which was detected predominantly in T2E negative cases.

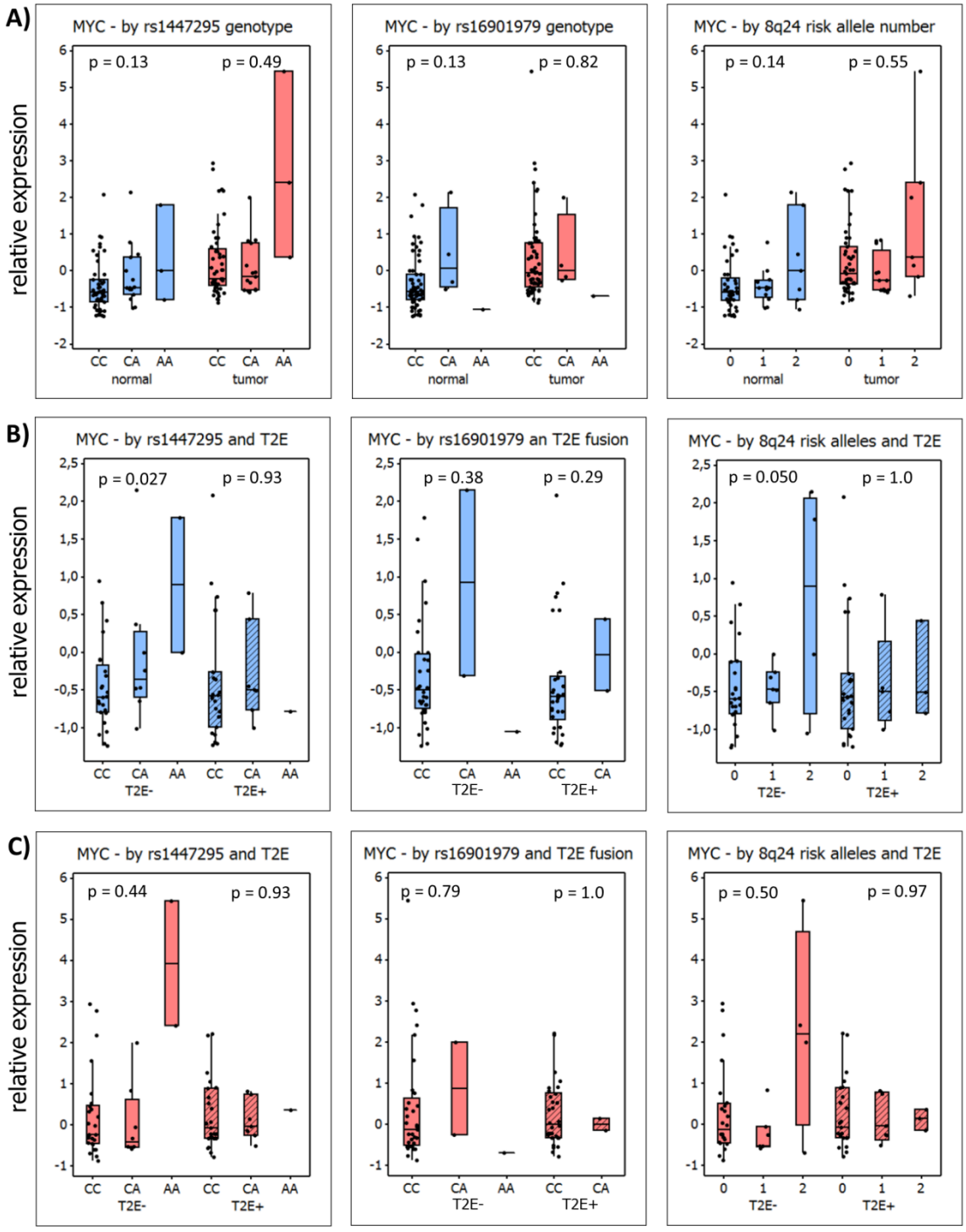


Figure 36: Relative mRNA expression of *MYC* in prostate normal- and tumor tissue of 70 patients depending on 8q24 variants and *TMPRSS2-ERG* fusion. Represented are the box plots as well as single values from the Z transformed results. The expressions were plotted dependent on rs1447295, rs16901979 and on the cumulative risk allele (A) number of both variants. Due to the partly small subgroups, the p-values were estimated using the exact Jonckheere-Terpstra test. **A)** *MYC* expression dependent on genotypes in normal and tumor tissue. **B)** *MYC* expression dependent on the genotypes in normal tissue in *TMPRSS2-ERG* negative (T2E-) and positive (T2E+) subgroups. **C)** *MYC* expression dependent on the genotypes in tumor tissue in *TMPRSS2-ERG* negative (T2E-) and positive (T2E+) subgroups.

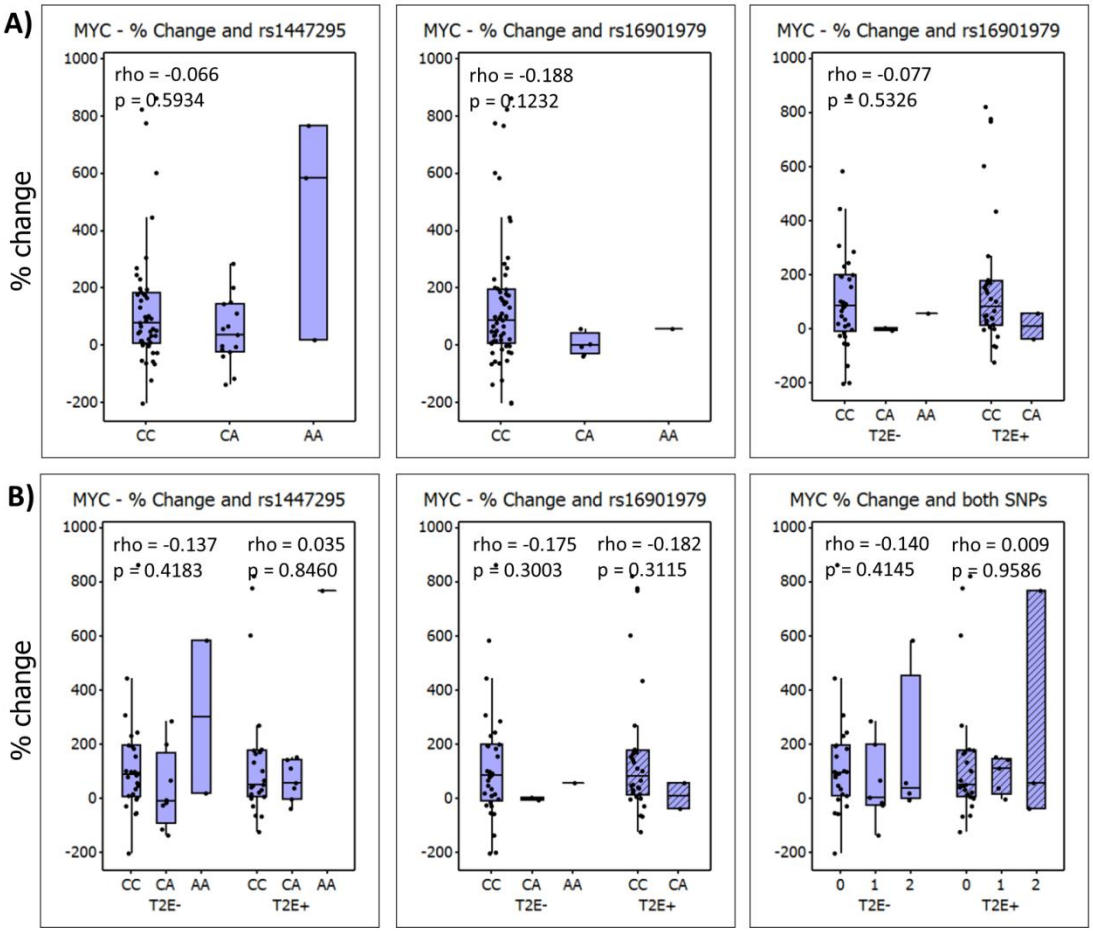


Figure 37: Percentage change of *MYC* expression in tumor versus normal tissue dependent on genotypes of the 8q24 SNPs rs1441950, rs16901979 and dependent of the risk alleles (A) of both SNPs combined. The value indicates the expressional change in %; values < 0 indicate a decreased, values > 0 an increased expression in tumor compared to normal tissue. The stated p- and rho-values were estimated with a Spearman rank correlation with correction for ties to correlate the number of risk alleles (A) with the expression data. **A)** Expressional changes observed in the total sample set with paired expression data (n = 68 for rs16901979, n = 67 for rs1447295, n = 66 for the combined analysis)) dependent on genotype and risk alleles. **B)** Expressional changes observed in the *TMPRSS2-ERG* negative (T2E-) and positive (T2E+) subgroups dependent on the genotype.

4. Discussion

Except for the recently identified first high-risk gene for PrCa, *HOXB13* [43], genetic epidemiology has so far failed to uncover the genetic causes of this extremely heterogeneous disease. However, association studies facilitated the discovery of a steadily increasing amount of common low-risk SNPs, to date 77 in number (see Table 1). The problem is that these variants often raise more questions than they answer. For example, the most prominent susceptibility locus 8q24 is a broad gene desert containing several independent common variants, which are differentially associated, aside from PrCa, with breast, ovarian and colorectal cancer [50]. In fact, most variants are not located within or even in the vicinity of candidate genes, complicating the recognition of causality.

The first aim of the present work was to evaluate the significance of 25 previously identified common variants for the German population and to assess their cumulative value for PrCa risk prediction. To elucidate possible mechanisms how these SNPs might exert disease risk, associations of the tumor predisposing alleles with a classical cancer-related pathway, the DNA damage response and repair function, were investigated with help of the following strategies. On the one hand, tests on DNA repair capacity with two different endpoints were performed on peripheral blood lymphocytes of healthy volunteers to uncover general genotype-specific correlations. On the other hand, common variants were analyzed in a multi-centric association study for their implication in T2E fusion positive prostate carcinoma, a somatic subtype that is postulated to be predisposed by deficits or variation in cellular DNA repair capacity.

4.1 Epidemiology: Replication of single common variants and future directions

All of the 25 previously identified common variants could be verified in large international replication studies under participation of the ULM center [9,73,154]. The data presented in the current work show that nine of these variants were replicated with a $p < 0.05$ for their significance in the German population. Although there are certainly population-specific differences between the risk contributions of the single SNPs, the limited number of significant SNPs within the ULM cohort is probably mainly due to the comparable small sample size. By now, GWAS analyses have extended the list of PrCa risk common variants to a number of 77. This up-dated list is estimated to explain about 30 % of the heritability of PrCa risk [38].

Twelve SNPs, which were associated with $p < 0.1$ in the German cohort, were selected for a cumulative analysis in order to explore multifactorial susceptibility, an inheritance model that has been repeatedly postulated for PrCa. Indeed, significant deviations from "normal" disease risk were already observed for proband groups with only a few less or a few more risk alleles than the median of 10 alleles. The most extreme group, defined by 15 and more showed a 5.5-fold risk increase as compared to the normal population. This magnitude is comparable to a high risk gene effect. Given the limited frequency of carriers with such a high load - only about 5 % of men are in the utmost category - excessive common variants can resemble the epidemiology of a single, highly penetrant gene. In principle, men and families could thus certainly benefit from genotyping common risk alleles in the context of predictive testing.

One of the limits of this study is that the effect sizes of this explorative approach might be overestimated, as the same sample set was used for the identification of the most significant SNPs, which were in turn picked for investigating cumulative effects. However, the strength of the study is the definition of an adequate reference category. While most of the previous cumulative approaches, like the pioneer study from Zheng and colleagues [190], estimated the odds ratio versus controls with a minimum of risk alleles [63,149,190], the present work's reference was defined as the median number of risk alleles in controls, which gives rise to an intuitive and much more robust estimate for the

true effect. Therefore, individuals harboring 10 risk alleles were considered as normal, with a baseline PrCa risk on population level (OR = 1). Nevertheless, though the risk effect sizes of the here presented and the currently published common variants may be inaccurate to some extent, the cumulative risk modifying nature of the common variants remains undeniable, which emphasizes the hypothesis of a multifactorial inherited predisposition to PrCa. In consequence, several studies dealt with the question whether the knowledge of the common variants can be used for risk prediction of PrCa. A very detailed review analyzed 14 of these previously published studies and concluded that so far no SNP panel (in part extended by other parameters) fulfilled the requirement of an AUC of 0.75 or higher for a diagnostic use [91]. Anyway, insufficient AUC just indicates that cumulative testing would be of limited information for the majority of individuals, as if applied for a general population. Further indications, such as family history could be considered to identify individuals with a high load of common variant risk alleles.

No association was observed between risk allele number and age of onset or clinicopathological parameters. In fact, this secondary analysis was not motivated by a high prior expectation, since most risk loci have been proven not to be associated with clinical parameters at all, or occasionally, with the more favorable fraction of disease [9]. Missing homogeneity in this respect could lead to dilution or even compensation of particular disease effect, rather than to a cumulative enhancement. As an example, while some SNPs are associated with a higher PSA at diagnosis, the rs1447295 risk allele was found preferred in cases with lower levels [9]. One further disturbing factor when testing common variants for aggressive disease parameters could be the lack of knowledge of less frequent variants or high-risk mutations within the case group. Intuitively, severe forms of PrCa are yet reserved for hidden high-risk genes.

Importantly, multifactorial approaches for PrCa are presently performed without any assumption of functional models, due to the absence of biological knowledge for most risk loci. In other words, variants have been analyzed for conjointly cumulating on risk, without any evidence that a physiologic interaction of involved factors truly exists. Accuracy and predictive value of such models would benefit from a more comprehensive understanding of the functional question, which particular sets of common variants were definitely involved in the same lines of pathogenicity. Adding mechanistic knowledge to common PrCa risk variants, as claimed also by the excellent review of Freedman and

colleagues [48], was one major aim of the present work. A first strong insight has been provided now by showing two genomic loci, 8q24 and 17q24, differentially associated with T2E fusion positive PrCa (see paragraph 4.2.2). When accounting for the appropriate tumor entity (e.g. fusion status), the effect sizes considerably increased in case-control comparisons to OR = 2.25 and 1.47 for 8q24 and 17q24, respectively, whereas weaker effects (OR = 1.68 and 1.22, respectively) were assessed from unselected association studies.

4.2 Mechanistical annotation of prostate cancer risk loci as a strategy to overcome heterogeneity

While PrCa is frequently classified by the phrase “heterogeneous disease” in various respects, it is though usually treated as one uniform tumor entity. In a genetic sense, so called “locus heterogeneity” indicates the co-existence of multiple genes that predispose to a (seemingly) common outcome. Heterogeneous is also used in the context of etiology and tumorigenesis, and is reflected by a discernible diversity of somatic mutation patterns, hormonal requirements and growth behaviors. Finally, malignancies can be heterogeneous in a clinical view, linked to the question why some patients manifest moderate progression, while others with the same diagnosis (and treatment) develop a fast growing, life threatening tumor. Beyond doubt, all meanings of heterogeneity are true for PrCa, and dissecting particular entities in one of the contexts could also provide clues for a refinement of sub-categories among all other research and clinical disciplines. The core aim of this thesis was the attribution of a yet unsorted bulk of germline risk variants to presumptively distinct pathomechanistic and somatic entities. The referring phenotypes of DNA repair capacity and oncogene fusion, as outlined in the following paragraphs, have both been proven to denote a fraction of PrCa. Apart from the self-evident notion, that oncogenic rearrangements, such as T2E, would likely arise from impaired DNA damage response, interconnection of these two endpoints in prostate cancer has by now been substantially supported by epidemiologic and experimental evidence. First, association studies in T2E fusion positive PrCa have identified germline variants in DNA-repair genes [67,95]. Second, the influence of particular factors which

represent essential DSB repair pathways could be demonstrated by the model of the *TMPRSS2-ERG* induction assay [90,94,98].

4.2.1 Testing for DNA damage response intriguingly links risk region 10q11 to non-homologous end-joining repair

Defects in DNA repair genes that lead to genomic instability are a hallmark of cancer. While certain malignancies, such as breast or colon cancer, are broadly attributed to specific pathways of DSB repair and mismatch repair, respectively, prostate cancer in general has a less strong and less uniform connection to DNA repair. Initial links to DSB repair became evident with an increased prostate cancer risk in male relatives of *BRCA2*-dependant high-risk breast cancer families [155]. Recently, Leongamornlert and colleagues highlighted the significance of germline mutations in genes of the DNA damage response pathway for PrCa, as ~ 7 % of patients harbored putative loss-of-function mutations in selected genes [84]. Assays for measuring DNA repair capacity in body cells have pointed at a subgroup of sporadic PrCa patients with increased radiosensitivity [171].

The strategy of the present work was comprehensive and impartial testing of all common variants with a proven risk for PrCa for an association with DNA damage response (see also [128]). The most remarkable result was observed for the SNP rs10993994 on 10q11. For this locus the PrCa risk associated allele T led to a decreased micronucleus frequency and the genotype TT to a shorter mitotic delay after radiation-induced DNA damage. This result seems difficult to interpret in view of the fact that a cancer risk allele, if there is any connection to the DNA damage response and repair pathway, is intuitively expected to increase, rather than to decrease the micronucleus frequency after damage infliction. Moreover, lymphocytes that carry the suspicious T allele seem to repair the inflicted double strand breaks in a shorter time. There are several conceivable scenarios which could explain the findings: The T allele could promote a premature damage-caused apoptosis induction, resulting in an overrepresentation of cells that survived due to less damage. For example, Ban et al. observed a lower radiation-induced micronucleus frequency for cervical cancer patients than for controls, and explained this by enhanced induction of apoptosis, possibly caused by infection with the human papillomaviruses

(HPV) [14]. However, total cell numbers were recorded throughout the present micronucleus assays and with regard to the counted cell numbers, there is no evidence for an enhanced apoptosis rate in risk allele carriers. Alternatively, the T allele could, directly or by other mechanisms, lead to a hyper-responsive repair activity, which is reflected by a faster, but not necessarily precise repair. It has been shown in early cancer precursor lesions that deregulated DNA replication and DNA damage can cause a steadily activated DNA damage response [16]. A further plausible explanation of the present findings might lie in an impaired or less activated homologous recombination repair (HR) in T-allele carriers, resulting in an increased usage of the more error-prone non-homologous end joining (NHEJ) pathway. In this regard, the shorter G2 delay could be explained by its faster processing. According to Mao et al., NHEJ can be completed in 30 min, while HR takes seven hours or longer [99]. Both pathways are active during S and G2 phase [18,142]. In general, little is known about how the cell decides using either NHEJ or HR, but several factors are suggested to modulate this choice, including chromatin status, [11,18], the complexity of the lesion [142] and various damage response proteins [44]. This hypothesis of rs10993994 as a modulator of the DSB repair pathway usage seems to be the most plausible one, particularly since the SNP was also found, by trend, associated with T2E positive PrCa. The candidate genes behind the SNP are discussed in paragraph 4.3.3.

The limitation of this study is a comparable small sample size, which may conceal effects of moderate magnitude. Given, that the investigated common variants only marginally increase the risk for PrCa, no severe differences for the risk SNPs were expected. Furthermore, some risk mediating mechanisms may require prostate tissue specific environment, e.g. certain hormone levels. Here, peripheral blood lymphocytes of healthy volunteers were chosen, as blood samples can easily be taken and the cellular DNA repair capacity should be represent a ubiquitously observable endpoint. Earlier studies dealing with different cancer entities have repeatedly demonstrated that the damage endpoints in lymphocytes can reflect cancer susceptibility [13,151,170]. Finally, this was a screening-like approach with no claim for completeness.

4.2.2 Subtyping for somatic *TMPRSS2-ERG* fusion status identifies 8q24 and 17q24 SNPs as the first two differentially associated germline variants

The somatic T2E fusion was used as a second phenotype for association with common PrCa variants (submitted for publication, see [127]). As present in about 50 % of all prostate tumors, it has been previously used successfully as a splitting criterion to reduce disease heterogeneity for the identification of PrCa candidate genes [67,95]. Once a tentative hypothesis, the assumption of a genetic predisposition for ETS fusion has been meanwhile supported by accumulating substantial evidence for ETS fusion positive and negative PrCa as distinct tumor entities on somatic level [12,178]. According to recent deep sequencing efforts by Baca and colleagues, the genomes of ETS fusion positive tumors are often characterized by a pattern of consecutive structural rearrangements, a phenomenon named “chromoplexy” [12]. The cascade occurs in a non-independent manner and apparently often begins with T2E fusion or with a loss of the transcription factor gene *NKX3-1* [12]. As the breakpoints are mainly found within transcriptionally active genes and preferentially involve androgen-dependent elements, the androgen receptor represents a plausible co-driver [12,178]. Sun and colleagues demonstrated that the androgen-sensitive VCaP (Vertebral Cancer of the Prostate) cell line [79], which expresses the *T1G4* transcript variant [161], grows in a highly *ERG*-dependent manner [148]. According to the authors, cells that underwent siRNA mediated *ERG* knockdown were less proliferative in vitro and showed strikingly less tumorigenicity in mice [148]. ETS negative PrCa cells, in contrast, do not harbor excessive genomic rearrangements and therefore seem to be driven by other mechanisms. According to Boerno and colleagues, the genomes of ETS negative tumors are characterized by a significantly increased chromatin methylation [20]. A progressive hypermethylation, which involves an increased expression of the Histone-lysine N-methyltransferase *EZH2*, is a typical feature of PrCa evolution [169]. In line with this notion, Boerno and colleagues found a stronger abundance of *EZH2* in ETS negative compared to positive tumors [20].

The results of the present study confirmed the hypothesis, that ETS positive and negative PrCa were also distinct by germline. Two loci were differentially associated with T2E fusion: The risk allele G of the rs1859962 variant on 17q24 was found enriched in T2E positive cases, while the rs16901979 (8q24) risk allele A was overrepresented in T2E negative cases. Interestingly, a second independent variant in 8q24, rs1447295, was also

found enriched in T2E negative cases, supporting the significance of 8q24 in this subtype. Both loci, 17q24 and 8q24, are large gene deserts, where the search for candidate genes has been challenging although comprehensive fine-mapping efforts (see paragraph 4.3).

Recent studies found additional evidence for distinctness of ETS positive and negative PrCa also on the level of clinicopathological characteristics. Accordingly, ETS positive tumors are more frequent in PrCa patients with an earlier age and lower PSA at diagnosis [135,178]. These findings could be verified in the present investigation, although with marginal significance for PSA. The higher incidence of ETS alterations in early onset patients is a stunning finding, especially with respect to the hypothesis of a germline predisposition. By common sense, hardly any other etiologic factor could trigger age of onset. Nevertheless, several alternative explanations have been proposed for the age phenomenon. Detection bias has been taken into consideration, as ETS positive tumors are supposed to develop faster, and thus could simply lead to an earlier clinical detection [32,135]. A rather odd theory, albeit presented in a top-class journal, makes use of the critical role of androgens in T2E fusion formation: Because androgen levels are undoubtedly at a maximum in earlier decades of lifetime, younger men would be at higher risk for developing fusion positive PrCa [178]. As an interesting side aspect, age and PSA levels at diagnosis were analyzed as covariates along with the major germline risk factor of fusion positive PrCa identified in the present work. The association of rs1859962 at 17q24 with T2E positive PrCa was less significant, when using PSA at diagnosis as a covariate. This might possible reflect such suspected interdependencies, as the SNP could indirectly modulate PSA levels. On the other hand, rs1859962 was not reported to be associated with PSA, but instead with an earlier age of onset [9]. For the present study sample, however, age at diagnosis and rs1859962 were found independently associated with T2E status. This means, in summary: 1) rs1859962 is an independent risk factor for T2E fusion PrCa without or minor confounding effects by age and PSA, respectively. 2) rs1859962 may be involved in a detection bias for T2E positive tumors, because an interaction with diagnostic PSA could not be ruled out. 3) rs1859962 does not predispose to early onset, fusion positive PrCa, since the factor of age at diagnosis is independent.

Apart from age and germline susceptibility, there are several more etiological risk factors currently under study for differences between fusion positive and negative PrCa. The series of almost 600 prostatectomy cases from FHCRC (Seattle), which were phenotyped

in the course of the present work, enables profound investigations of T2E status along with comprehensively recorded epidemiologic parameters in this cohort. With this opportunity, we were able to demonstrate for body-mass index (BMI) an inverse correlation with PrCa risk in T2E positive but not negative cases [39]. In fact, obese men are known for long to be at lower risk of developing PrCa [125], most likely due to hormonal conditions such as low testosterone levels [7]. The recent finding that the protective effect of obesity apparently operates particularly on fusion positive PrCa is thus well in line with the superior role of androgens especially for this tumor subtype. A second protective effect in this cohort, again on the fusion positive subtype, was observed for aspirin intake [179]. This result is exciting with respect to the hypothesis of T2E as a result of DNA damage. Aspirin has been shown to reduce cellular levels of radical oxygen species and could therefore be beneficial to prevent from DNA lesions [27]. Besides the etiological differences of ETS positive and negative PrCa, subtypes could also respond differentially to treatments. As a first step towards targeted therapy, recent studies have provided evidence for a sensitivity of ETS positive, but not negative tumors to *PARP1* inhibition [21,61].

The present study collective comprising ~ 1,200 cases might be comparable small for performing a meta-association study with six sub-populations. Nowadays, association studies for PrCa risk usually involve several thousands of cases and controls to accommodate the notoriously small effect sizes of common variants. Nevertheless, the sample used here represents one of the biggest case cohorts with known T2E fusion status available to date. As the importance of the molecular subtyping is still undervalued, the assessment of this phenotype has not yet found the way into the laboratory routines of translational research. The fact that - on top of the small sample sizes - the T2E phenotypes of the different subgroups were assessed with different methods, additionally strengthens the significance of the identified loci. Due to the comparable high effect sizes of rs1859962 and rs16901979 among the common variants, it is not surprising that these were the first identified ones associated with the T2E fusion status. Future studies including more common variants and more cases would probably help to find whole sets of SNPs that better characterize the somatic subtypes. In consequence, the cumulative value of SNPs could be assessed in a hypothesis driven way resulting in more robust and realistic prediction models. Vice versa, assigning SNPs to a subtype might also

facilitate the search for possible candidate genes on the nearly unmanageable bulk of common variants identified to date.

Taken together, this study identified two germline variants differentially associated with T2E fusion and represents the first proof of the principle that somatic subtypes are characterized by a unique fingerprint also on germline level.

4.3 Consideration of candidate genes based on the implication of DNA repair and fusion gene phenotypes

Functional assignment of common PrCa risk variants identified three loci of interest: (1) rs10993994 on 10q11 as a modulator of DNA repair capacity and also associated with T2E positive PrCa; (2) rs1859962 on 17q24 as significantly associated with T2E positive PrCa and (3) the complex genomic region 8q24 with two variants (rs16901979 and rs1447295) independently associated with T2E negative PrCa subtype. One further region in 8q24 (rs6983267) tended towards an association with a shorter mitotic delay. In order to identify the underlying genes, mRNA gene expression analysis dependent on tissue type, genotype and T2E fusion status in fresh frozen tissue specimen, was the method of choice.

4.3.1 The examination of multiple plausible candidate genes on 10q11 highlighted *PARG* as the most suspicious player

Based upon present data, as further discussed in 4.3.1.4, *PARG* excelled as the most plausible gene in 10q11. Nevertheless, novel findings for every alternative candidate gene are briefly outlined in the following sections, along with respective pieces of evidence suggested by previous studies.

4.3.1.1 The historical prime suspect *MSMB* - convicted because of presence at the scene

The PrCa risk variant rs10993994 in the region 10q11 has been known since 2008 [37,156]. In contrast to most other known variants, it is not located in an intergenic region, but within the promoter region of a gene, *MSMB*, which seemed just directly to explain the pathomechanism. And indeed, the risk allele T was shown to decrease the *MSMB* mRNA expression in prostate normal and tumor tissue [124] and in reporter gene experiments [26,93]. Expressional dependency on the genotype could be irrevocably replicated in the present study. The gene encodes the Beta-microseminoprotein (β-MSP), also known as PSP94 (prostate secretory protein of 94 amino acids), which is, besides PSA and PAP, one of the three most abundant proteins of the seminal fluid [89]. It is expressed mainly in prostate, but also found in other secretory tissues, and in contrast to PSA and PAP, it is androgen independent. The function is not fully understood. However, the significance of *MSMB* as tumor suppressor for PrCa was not questioned, as it was repeatedly found down-regulated in prostate tumor tissue compared to normal tissue on both, mRNA and protein level [25,124,163,168]. It is unclear, if down-regulation of *MSMB* is an early event in tumorigenesis and the mechanism how *MSMB* might exert its tumor suppressor function is still elusive. There are indications that MSMB can bind on different cell surface receptors and could participate in the regulation of apoptosis. A downregulation of MSMB, whether through silencing by EZH2 [17], or also caused by the rs10999394 T allele, could therefore lead to uncontrolled cell growth, which could in turn enhance tumorigenesis.

To return to the findings of the current work, the question is how *MSMB* could modify the cellular DNA damage response and repair of peripheral blood lymphocytes, although it was found weakly or not expressed at all in this cell type. β-MSP, from wherever it is secreted, is proven to be present in blood and the quantity of circulating β-MSP also depends on the rs10993994 genotype [185]. Possibly, it binds on cell surface receptors of lymphocytes and thus initiates intracellular processes. In both, micronucleus assay and mitotic delay test, whole blood with all its components was used, and extracellular effects have to be taken into consideration.

The question for the capability of *MSMB* as a driver for somatic rearrangements could not be definitely clarified by the T2E induction assay utilized in the present work. In this case,

the LNCaP cell line represented no perfect model organism, as it expressed only low levels of *MSMB* intrinsically. A further knockdown of the weak expression, however, did not cause a significant change in fusion formation frequency. With the novel pieces of evidence accumulated here, *MSMB* may not be considered as the most plausible candidate gene in 10q11.

4.3.1.2 *NCOA4* - co-accused for genotype-dependent expression pattern

More likely, another gene of the 10q11 locus could be involved. Since Pomerantz and colleagues showed that, besides *MSMB*, also the neighbored gene *NCOA4* (nuclear receptor coactivator 4, often referred to *ARA70*) was influenced by the rs10993994 genotype [124], it became clear that the investigations may require expansion to a larger list of candidates. In the present work it was hypothesized that also other genes in region could be influenced and should be taken into consideration as potential PrCa susceptibility genes. *NCOA4* was first identified in papillary thyroid carcinoma as a fusion partner of the RET oncogene [19]. An increasing number of functions have been assigned to *NCOA4*, including - with its molecular function as a positive coregulator of the androgen receptor NcoA-4 [188] - processes such as development, inflammation, erythrogenesis, proliferation and cell cycle progression (see [78] for review). Due to these important characteristics, *NCOA4* was previously analyzed for its properties as a PrCa candidate gene. Pomerantz et al. focused on *MSMB*, *NCOA4* and *TIMM23* with respect of mRNA splice variants and investigated 121 tumor/normal tissue pairs and 59 additional tumor tissues [114]. Consistent with the present results, the authors reported a decreased expression level of *MSMB* in tumor compared to histologically normal prostate tissue, and in rs10993994 T allele carriers in tumor and normal tissue. Contrary to the present results, they also found *NCOA4* decreased in tumor compared to normal tissue. I observed a borderline significant increase in *NCOA4* mRNA in tumor tissue. Indeed, different publications proclaim different results for *NCOA4*. While Hu and colleagues described, in compliance with the present work, an increased NcoA-4 protein staining in prostate tumor versus normal tissue [70], FitzGerald and her co-authors found no differences in protein quantity with similar methods [46]. Different antibodies could certainly play an additional role for the apparent discrepancies. No significant differences were also reported by Mestayer and colleagues [102]; however, they used a very small sample size (n = 7) and semi-quantitative PCR. A further publication from Li et al.

reported, consistent with Pomerantz and colleagues, a decreased *NCOA4* transcript quantity using quantitative *in situ* RNA hybridization [85]. Usage of primers for the variety of *NCOA4* isoforms could also be an issue. Pomerantz and colleagues, for example, showed the downregulation in tumor tissue for all five investigated transcriptional isoforms using different primer pairs. The primers used in the present work were designed to target most of the splice variants simultaneously (four of the five Pomerantz variants) and should thus come to the same results [124]. Supporting the present results implicating *NCOA4* as an oncogene, its expression was slightly increased in presence of the rs10993994 risk allele (in contrast to Pomerantz et al. [124]).

However, *NCOA4* overexpression was demonstrated to suppress the tumor growth in LNCaP xenografts [88], and enhance anchorage-independent growth of prostate epithelial cells [124], suggesting a tumor suppressor rather as an oncogenic mode of action. On the other hand, a possible explanation for these different contrasting results could be the presence of two different main isoforms, *NCOA4α* (full-length, 70 kDa) and *NCOA4*β (35 kDa) [6], which differ in cellular localization [118] and thus might exert different functions.

Regardless of the conflicting arguments concerning tumor suppressor versus oncogene potential, *NCOA*4 seems incapable to explain the suggested involvement of 10q11 in fusion positive PrCa, as the *TMPRSS2-ERG* induction assay provided no evidence for *NCOA4* to influence fusion formation.

4.3.1.3 *TIMM23* and *TIMM23B* - two innocent brothers?

A further candidate gene in the risk locus 10q11, more precisely a pair of genes, is *TIMM23* and its genomic duplication *TIMM23B*. The *TIMM23/TIMM23B* gene (translocase of inner mitochondrial membrane 23 homolog (yeast), also known as *TIM23*) codes for Tim23, the essential pore-forming component of the Tim23 translocase complex (additionally comprising Tim17, Tim21 and Tim44). It is well-known that mitochondrial permeability plays a role in caspase dependent and independent apoptosis (see [53] for review). According to Goemans et al., the autodigestion of the Tim23 protein could represent an initial event for mitochondrial destruction and caspase independent apoptosis [52].

TIMM23 was not considered as a candidate gene for PrCa so far, except in the gene expression study together with *MSMB* and *NCOA4* by Pomerantz and colleagues [124]. The authors found no expressional changes between prostate tumor and normal tissue as well as no dependency on the rs10993994 genotype. Importantly, however, they have not distinguished between the two genomic copies [124]. Though both genes are coding for the same protein, they were assessed separately in the present work to account for a possible selective influence of rs10993994. And indeed, the expressional pattern was found dissimilar between *TIMM23* and *TIMM23B* in prostate tissue. While *TIMM23B* showed no differences in tumor compared to normal tissue, *TIMM23* was strongly up-regulated in tumor tissue, more significantly in T2E positive than in negative tumors. Considering the involvement of *TIMM23* in cell death processes, an up-regulation could be a tumorigenic strategy to prevent apoptosis. On the other hand, it could also reflect a simple co-regulation with *PARG*, which shows similar expressional dependence on tissue type, T2E fusion status and even rs10993994 genotype. In the latter context, *TIMM23* shows just as *PARG*, no significant correlation with the risk allele T. In contrast, *TIMM23B* is significantly correlated with the number of risk alleles in normal tissue, and to some extent also in tumor tissue. However, a closer look at the three-genotype expression box plots reveals that *TIMM23B* shows a similar U-shaped pattern like *TIMM23* and *PARG*, but the specimens with CC genotype are slightly higher than those with TT mimicking a 'linear effect'.

Overall, this expressional situation is somewhat surprising. The genetic map of 10q11, as schematically displayed in Figure 21, was used as a basis for the present work, though there are apparent inconsistencies among gene annotations of different genome browsers. One report from literature demonstrated, that *PARG* and *TIMM23* share a bidirectional promoter, which is more active in the *TIMM23* direction, but no distinction was made between the two genomic copies [103]. I assumed that the authors have considered the copy designated as *TIMM23B* in the present work, but therefore I would have expected a stronger co-regulation between *PARG* and *TIMM23B*, not *TIMM23* as observed in the correlation analyses. On the other hand, *TIMM23* and the *PARG* pseudogene *LOC100506785* could possibly share their promoter as well, which might function in a similar way. However, the *PARG* qRT-PCR assay that was used in the present work was designed not to target *LOC100506785.*

Regarding the paired expressional changes between tumor and normal tissues, no clear trend of a correlation with genotypes and T2E fusion status was observed. An involvement of *TIMM23* in modulating DNA repair capacity and susceptibility for T2E fusion is therefore rather unlikely. Furthermore, there is no obvious way how an altered *TIMM23* expression could possibly repress the micronucleus frequencies after irradiation or shorten the mitotic delay in peripheral lymphocytes.

To test the influence of *TIMM23* on T2E fusion formation using the *TMPRSS2-ERG* induction assay, solely siRNA against *TIMM23* was available. Though the knockdown efficiency was very good, no significant change in fusion induction in LNCaP was observed. Now, the question arises if *TIMM23* really exerts no effect on fusion formation, or if the *TIMM23B* activity is sufficient to maintain the cellular protein level. There is no absolute certainty of course, but LNCaP cells showed considerably decreased growth after treatment with *TIMM23* siRNA, and thus, were severely affected by the depletion. The apparent cell mortality was not proven by viability tests, but has been in fact previously described. According to Goemans and colleagues, *TIMM23*-depleted HeLa cells showed a 40 % reduced proliferation, but interestingly no apoptosis [52]. Thus, considering that the Tim23 protein is critical for mitochondrial transport and a certain cellular level would probably be required, the present experiment may actually provide evidence against an involvement of *TIMM23* in the formation of somatic fusion events.

4.3.1.4 *PARG* is the most distant, but most plausible candidate gene in 10q11

Even though its distance to rs10993994 the poly (ADP-ribose) glycohydrolase gene *PARG* was considered as the most intriguing candidate on 10q11 with respect to its manifold involvement in cellular processes. PARG is the catabolic counterpart of the PARP (poly (ADP-ribose) polymerase) protein family (17 members). As implied by the name, most of them synthesize ADP-ribose polymers (PAR), an important posttranslational modification that is involved in various cellular processes like replication, methylation, DNA repair and apoptosis. Acceptor molecules are many nuclear proteins, for example histones, DNA repair enzymes, transcription factors and last but not least PARPs themselves (automodification). The most important representative of this group is PARP-1, accounting for ~ 90 % of synthesized PARs (see [82] for review).

DNA damage entails an enhanced PARP-1 mediated PARylation of effector molecules. As PAR synthesis requires NAD^+ withdrawal, an extensive damage leads to apoptosis or even necrosis caused by the progressive energy deficit. In case of a moderate damage, the recruitment of different repair enzymes paves the way for DNA damage repair. The half-life of the polymers is only short; within minutes, they are detached and degraded again by PARG. PAR homeostasis, regulated by its assembly and disassembly is regarded as a dynamic equilibrium essential for cellular maintenance. PAR dysregulation, interfered by risk genotypes in *PARG*, could be a potential risk-mediating mechanism at 10q11.

The present expression study found PARG, similar to *TIMM23*, strongly up-regulated in tumor versus normal prostate tissue. And likewise, even to a stronger extent, overexpression was more significant in T2E positive cases. As already mentioned before, the genes could possibly be co-expressed by a bidirectional promoter [103]. Interestingly, an ETS binding motif has been found in their promoter region [165]. Thus, it is plausible, and in line with the present expression data, that both *TIMM23* and *PARG* could up-regulated especially in T2E positive tumor via the presence ERG itself. Such scenario of a positive feedback loop does not (yet) explain the role of risk genotypes in 10q11 and is possibly independent from the germline predisposition specifically to T2E positive PrCa. In fact, expression of *PARG* (and also of *TIMM23*) did not depend on the rs10993994 genotype regarding static levels in normal and tumor tissues. Nevertheless, considering the extent of up-regulation from normal to matched tumor samples, there was a trend of genotype-dependency for *PARG* (but not *TIMM23*). Surprisingly at the first glance, this finding was evident exclusively in the T2E negative PrCa, where *PARG* is not much up-regulated at all, and where association of rs10993994 with disease risk was found to be of limited relevance. Explaining this obscurity would require a more complex scenario, involving *PARG* up-regulation and oncogene fusion as interdependent processes in tumor formation: Mediated by the germline risk variation, *PARG* up-regulation may play a preceding, perhaps even predisposing role for T2E formation. Once the oncogene is present, and ERG becomes abundant, the mentioned positive feedback loop via the ETS responsive element in the *PARG* promoter leads to a drastic boost of PARG expression in the T2E positive tumor type. The present experimental data are in line with the plausibility of this hypothesis, as in fusion the negative state, *PARG* up-regulation is 1.5-fold in wild-type cases, 8-fold heterozygous and 40-fold in homozygous risk allele carriers,

while all fusion positive cases have roughly 50-fold *PARG* levels, regardless of genotype. Fine regulatory effects emerging by SNP genotypes are unlikely detectable in ETS driven tumors.

One remaining question is by which mechanism the germline risk variation in 10q11 could influence *PARG* transcription. As mentioned above, *PARG* is located quite far from the PrCa risk SNP rs10993994 and was not previously considered as a candidate gene on 10q11. As not investigated to date, also long-range interactions analogous to the situations on 17q24 and 8q24 (sections 4.3.2 and 4.3.3) would be possible. There is also a putative alternative transcription start site spanning the risk locus (180 kb telomeric of the SNP, UCSC browser, Figure 21). This hypothetical isoform could not be experimentally verified in the present work; however, technical sensitivity was limited, as the alternate exon 1 is only 45 bp in size and thus restricted the space for optimal primer design. *PARG* also produces several internal splice variants. This aspect should be taken into consideration with respect to possibility that only one particular isoform could have been relevant concerning PrCa risk, but was not adequately represented in the applied qRT-PCR method. According to Meyer-Ficca and her colleagues, isoforms differ in size and in cellular localization: The full length protein (111 kDa) localizes predominantly to the nucleus, the 102 kDa and 99 kDa proteins, which lack exon 1 including a nuclear leading signal, are cytoplasmic [105]. Later, the same research group provided evidence for a two further mitochondria-associated PARG isoforms (55 kDa and 60 kDa) [104]. It is quite conceivable that such isoforms could be differentially affected by expressional modulation by genetic variants.

The strongest evidence for an involvement of *PARG* in ETS positive PrCa was provided by the *TMPRSS2-ERG* induction assay. Depletion of *PARG* led to significant 3-fold increase of the fusion formation in LNCaP cells after irradiation. With respect to the observation in tumor tissue, where fusion positive PrCa is linked to *PARG* overexpression, *PARG* down-regulation in the context of the T2E induction assay is technically not optimal. A naïve view on this could give rise to the question if siRNA against *PARG*, which is otherwise overexpressed in tumors, would prevent LNCaP cells for developing rearrangements. However, substantial literature suggests that impaired PAR homeostasis is the crucial factor deficient double strand break repair, regardless of the direction and the factors that pull or push at either side. Many studies have addressed the involvement of PARG in

damage response and repair. The above mentioned progressive PAR-automodification of PARP1 decreases its affinity for the DNA strand and requires PARG for regeneration [134]. The importance of the PARG/PARP balance for the cell becomes even clearer considering that *PARG* knockdown in HeLa cells was found to co-repress *PARP1* expression [166]. To a certain degree, the cell can correct the imbalance. Interestingly, another HeLa study demonstrated that *PARG* depleted cells showed a repair delay and increased mitotic aberrations [8]. By this means, an enhanced formation of T2E fusions in the *TMPRSS2-ERG* induction assay would be plausible. Other studies found decreased proliferation and increased apoptosis induction in cancer cells treated with *PARG* inhibitory substances [75,186] and increased radiation-induced G2/M arrest and checkpoint activation [109]. The observation of decreased micronucleus frequency and shorter mitotic delay after radiation induction in lymphocytes with the rs10993994 T allele is not contradictive, since the T allele probably goes in line with a *PARG* up-regulation. Conceivably, slightly more *PARG* might lead to a more efficient clearance of *PARP1* automodification and thus to a faster binding on DNA damage sites. Furthermore, PARP binding is apparently important for replication fork stalling providing the cell with the time for repair [138]. A side effect of a higher throughput could possibly be a lower accuracy and thus also an enhanced susceptibility to develop genomic rearrangements such as the T2E fusion.

Taken together, *PARG* turned out as a strong candidate gene for the susceptibility to T2E positive PrCa. Further experiments should be designed to test this hypothesis. In particular, it would be interesting whether an overexpression of *PARG* in fact also leads to increased fusion formation. Since LNCaP are not well suited for plasmid transfection, the knockout of the main *PARG* counterpart *PARP1* by using siRNA or inhibitors on the enzymatic level would also be feasible for this purpose.

4.3.1.5 The *TMPRSS2-ERG* induction assay - chances and limitations

Actually, the T2E induction assay represents a DNA damage and repair test, which is focused very explicitly on the genome integrity issue specifically relevant for PrCa tumorigenesis. Although its perfect match with the scope of the present thesis, the assay has some crucial limitations. First, it is not suitable for analyzing germline variations (e.g. risk alleles) directly, because the implemented cell line has its intrinsic genotypes at risk loci which cannot be modified by reasonable efforts. Second, as the assay is time- and

cost-intensive, screening a larger number of candidate genes is also beyond practicability at present. Finally, when addressing a particular candidate gene, the various components and the endpoint measure of experimental set-up require prior considerations also taking into account the candidate gene's expected function. Trivially, for example, the assay could not be used for gaining mechanistic insights for *MYC* on 8q24, because the gene locus was found associated with T2E negative carcinoma. *SOX9* on 17q24 was also not investigated by the induction assay, although the candidate gene locus was found implicated in T2E positive PrCa. The reason for considering *SOX9* as an unsuitable assay target was its interference with the androgenic pathway as an inducer of the androgen receptor [174]. A steady androgen impulse is a requirement for the T2E induction assay, and therefore, it would have been difficult to distinguish between effects on fusion formation and technical impairment after *SOX9* knock-down.

In contrast, the *TMPRSS2-ERG* induction assay was considered as a suitable tool to elucidate the puzzling evidential situation on PrCa risk region 10q11. The rs10993994 variant showed correlations with DNA repair capacity on the one hand and nominal significant association with T2E fusion positive PrCa on the other. Subsequent expressional analyses of candidate genes on mRNA level failed to unambiguously identify the causal gene. The performance and results of the T2E induction assay for the candidate genes *MSMB*, *PARG*, *NCOA4* and *TIMM23* will now be discussed critically in more detail.

Efficient knockdown has been considered as a crucial prerequisite when testing genes in the context of in vitro T2E induction. As monitored by the transcript abundance of target genes, the knockdown efficiencies of *PARG*, *NCOA4* and *TIMM23* siRNAs were shown to be sufficient over a range from 24h to 120h. Solely *MSMB* was delayed. At 24h, *MSMB* expression was even increased above the basal level. However, as native LNCaP express only low levels of *MSMB*, an 88 % knockdown efficiency at time of irradiation (at 72 h) was considered acceptable. The knockdown efficiency of the positive control *ESCO1* was merely 76 %, and nevertheless led to the expected result. Low abundant transcripts are generally known to be prone to fluctuations, possibly explaining the increased expression of *MSMB* at 24h. Anyway, whether the knockdown efficiency of a hardly expressed gene ever plays a role remains debatable. Notably, all transcripts were already suppressed at the time of irradiation (= 72 h) and at least till the cell harvest (= 120 h). Of course, extended half-life of protein levels can lead to residual activities of the candidate under

study. An interesting investigation on global mammalian expression levels from Schwanhäusser et al. demonstrated a median half-life for proteins of 46 h [140]. The half-life of Tim23 is ~ 38 h according to their raw data [140], meaning it should be largely removed from the system at the time of damage induction (= 72 h). Unfortunately, no data were available for the other genes. For the NcoA4 homologue NcoA2, a half-life of ~ 16 h was reported, for PARP1 ~ 60 h [140].

The most striking result of this analysis was the significant *PARG* knockdown-mediated enhancement of T2E formation (3 fold) compared to cells treated with control siRNA. As already discussed in the section above, *PARG* represents an important global player which plausibly affects DNA repair, because homeostasis of PAR levels is crucial. Apart from *PARG*, however, all siRNAs led to weakly increased fusion formation, for *MSMB* and *NCOA4* even marginally significant. There are several possible explanations for this phenomenon. Possibly, this observation could reflect a general side effect of the siRNA mediated silencing pathway. The Argonaute (AGO) proteins, which represent the main functional component of the RISC complex, were shown to be involved in various other cellular processes, including double strand break repair (see [101] for review). Wei and colleagues showed strong irradiation-induced upregulation of AGO2 in *Arabidopsis thaliana* as well as a decreased repair rate in *ago2* mutants and suggested a role in recruitment of repair complexes on damaged DNA sites [177]. It would be conceivable that transfection with siRNAs lead to an AGO2 withdrawal resulting in less efficient double strand repair. The negative control siRNA would not lead to such effects since it is not complementary to any known gene and thus not initiate the formation of the RISC complex. If this scenario was true, it would be difficult to detect preventive effects of silencing particular genes in formation of damage-induced rearrangements.

Another possible explanation for the slight but consistent fusion increase after *MSMB*, *NCOA4* and *TIMM23* knock-down could be related to the genes' proximity in the same genomic region. The interval is especially dense in transcriptionally active elements, and long-range transcripts cannot be ruled out. If such long transcripts exists, it might be targeted by every siRNA used here, consequently resulting in cross-reaction of particular knockdown experiments. In this scenario, siRNAs targeting *MSMB*, *NCOA4* and *TIMM23* could additionally suppress *PARG* expression by some extent, which in turn is responsible for fusion formation.

4.3.2 *SOX9* on 17q24 as a putative *TMPRSS2-ERG* fusion promoting gene

The association study between common PrCa risk variants and the somatic T2E fusion revealed the risk allele (G) of rs1759962 on 17q24 significantly overrepresented in T2E positive tumors. The question arose, how this association might be mediated mechanistically. The genomic region 17q24 is an expanded gene desert and rs1859962 is approximately one megabase of distance to the nearest protein coding genes in both directions (*KCNJ2* centromeric and *SOX9* telomeric). The SNP is located in a ~130 kb spanning LD block [189] that contains the long non-coding RNA *CASC17* (cancer susceptibility candidate 17), annotated as *BC039327* in Figure 31, with yet unknown function. *CASC17* is not expressed in prostate as well as in various other tissues, except for testis [57,66]. It is not yet known if the expression of this RNA is affected by rs1859962, but if so, an indirect influence on prostate tissue must be assumed. This could be an endocrine loop, as in case, *CASC17* might be able to modulate testicular hormone production. Apparently, androgens play a crucial role in T2E positive tumor type, not only because AR is involved in fusion formation and maintenance of the activity of the oncogene promoter [90,98]. High androgen levels have been postulated to be a habitual characteristic of T2E positive PrCa, since this subtype was found associated with PSA value [135], a parameter that in turn depends on androgen levels. In the present study, association of T2E status with PSA at diagnosis could be replicated, and a slight influence of PSA at diagnosis on the association between rs1859962 and the T2E fusion was also present. However, previous studies found no evidence of an association between the variant and androgen levels [162] or PSA concentration [64,119]. Overall, there is only tentative support for the “cancer susceptibility candidate 17” as the most compelling candidate near rs1859962.

The present work focused on the transcription factor *SOX9* as a candidate for T2E fusion positive PrCa, since previous studies have accumulated evidence for a participation of this gene in PrCa tumorigenesis [71,136,157,158,173,174,189]. The transcription factor SOX9 is critically involved in various developmental processes, especially in chondrogenesis and sex determination (see [126] for review). Heterozygous germline mutations in *SOX9* cause Campomelic dysplasia [172] due to haploinsufficiency. *SOX9* plays also an important role in prostate development and maintenance [158,159,173]. Wang and colleagues demonstrated that *SOX9* expression is present in the normal prostate exclusively in basal

cells, but also in a subset of primary PrCa and more pronounced in recurrent tumors [174]. Overexpression in xenografts led to an increased tumor growth, invasion and angiogenesis, while downregulation inhibited the growth [173]. Furthermore, the androgen receptor was shown as a downstream target by posttranslational regulation: A knockdown of *SOX9* led to decreased androgen receptor protein levels [174]. An additional interesting aspect concerning fusion positive PrCa was provided by Cai et al., who reported a correlation between the *SOX9* expression and the presence of the T2E fusion. The authors described *SOX9* as a downstream effector of ERG in ERG-overexpressing cells by opening a cryptic androgen-regulated enhancer element [23]. Recently, Zhang and colleagues claimed to have found the missing link between *SOX9* and the PrCa risk variant rs1859962: An enhancer element within the rs1859962 LD block that accomplishes a long-range cis-interaction with the *SOX9* gene mediated by a chromatin loop. Moreover, a reporter gene assay has shown that variant alleles of two SNPs within the enhancer region lead to increased *SOX9* expression [189]. However, to my knowledge, no *SOX9* expression analyses provided evidence for a dependency on rs1859962 genotypes to date.

In the present study, mRNA expression analyses revealed an up-regulation of *SOX9* in prostate tumor compared to normal tissue. In line with Cai and colleagues [23], this effect appeared mainly caused by the fraction of T2E fusion positive cases, which fits very well with the thesis of an involvement of *SOX9* in the establishment of this somatic subtype. It is also conceivable, that this effect is promoted by the enhanced androgen receptor levels associated with *SOX9* overexpression (as mentioned above, [174]), which could lead to an enhanced colocalisation of *TMPRSS2* and *ERG* followed by increased fusion formation. Against all expectations, no correlation between the number of rs1859962 risk alleles (G) and *SOX9* expression was observed, neither splitted by normal/tumor tissue alone, nor by additionally stratifying on T2E fusion status. Also, considering the paired expressional changes between normal and matched tumor tissue revealed no genotype-dependent up-regulation of *SOX9*. Nevertheless, a trend of slight risk allele dependence in direction of an increased *SOX9* expression was noticed in tumor tissue of the subgroup of T2E positive cases, which is encouraging to pursue a larger study sample for this hypothesis. It is also important to consider the mentioned cell type specificity of *SOX9* expression as a complicating factor for the detection of genotype-dependent nuances. While the

expressional pattern observed in normal tissue might reflect the activity in the basal cells, the tumor tissue is characterized by an expressional mixture of luminal with a fraction of normal basal cells (most of the prostate tumors are acinar-type adenocarcinomas, characterized by the absence of basal cells [72]). However, the increased *SOX9* expression by long-range interaction of the rs1859962 associated enhancer element apparently depends to some extent on the androgen receptor [189], which is weakly expressed in normal basal cells. It is very likely, that the genotype-dependent effect contributes to the formation of T2E fusion in the course of tumorigenesis in early PrCa cells, but may be no longer detectable in an already existing, fusion positive tumor. There, *SOX9* expression is also influenced by various other factors, not least by ERG itself [23]. Despite all, the increased *SOX9* expression in T2E positive versus negative tumors clearly indicates its relevance for fusion positive PrCa, where *SOX9*, ETS factors and the androgen receptor operate inter-dependently on tumor cell growth.

4.3.3 The implication of 8q24 variants in *TMPRSS2-ERG* negative prostate cancer - The role of *MYC*

The 8q24 region is the most prominent and mysterious common risk locus of PrCa and other cancers. Many studies have attempted to elucidate causal variants and mechanisms within this expanded gene desert. Comprehensive fine-mapping identified at least five separate LD regions that are independently associated with PrCa [4]. In the present study, rs1447295 on region 1, rs16901979 on region 2 and rs6983267 on region 3 were included in the association analyses between common variants and DNA repair capacity as well with the somatic T2E fusion, to gain possible functional insights how these SNPs mediate PrCa risk. Subsequent expressional mRNA analyses in histological normal and tumor tissue of PrCa patients dependent on the 8q24 genotype and fusion status were done to examine a postulated long-range connection with the *MYC* gene.

The multicenter association study between common variants and the T2E fusion revealed the rs16901979 risk allele (A) 1.89-fold overrepresented in T2E negative cases with global significance. Interestingly, the risk allele (A) of the independent SNP rs1447295 showed the same tendency with a 1.43-fold overrepresentation. The enrichment of 8q24 risk alleles in fusion negative cases implicate a pathomechanism independent from ETS driver activity.

The most probable candidate gene is *MYC*, around 600kb from rs16901979 and 260kb from rs1447295 in distance. The transcription factor c-Myc plays a substantial role in the regulation of human genes involved in many cellular processes including cell growth, apoptosis and cell cycle progression [29]. There is evidence for long-range interactions between all cancer associated LD regions with *MYC* [2,145]. However, several efforts to demonstrate a genotype-dependent expression pattern of *MYC* failed to prove this suspected link unequivocally [122,123]. With respect to the newly identified association between 8q24 and T2E negative PrCa, the hypothesis arose that accounting for somatic subgroups could reveal possible correlations between genotypes and *MYC* expression.

The present expression analysis could replicate the overexpression of *MYC* in tumor tissue, an observation which has been consistently described in literature since decades [22,47]. Also, in line with latest reports [20,62], *MYC* up-regulation was not restricted to T2E negative tumors. This finding is not surprising, given the fact that oncogenic overexpression of *MYC* plays a role in pathogenesis of many cancers (see [113] for review). Its activity can arise in different stages of tumor development [58], and various possible mechanisms are known, such as amplification of the *MYC* locus [77]. Importantly, *MYC* has recently been found induced by ETS factors, a notion that could well explain *MYC* abundance in T2E positive tumors. Since ETS positive tumors obviously acquire *MYC* expression via ETS itself [20,62], the most tempting question has been, if T2E negative tumors, in particular, up-regulate *MYC* by means of germline risk variation. This hypothesis would perfectly match the exclusive role of risk loci rs1447295 and rs16901979 in T2E negative PrCa, and could also explain why it has ever been so hard to prove *MYC* levels depending on genotype. Unfortunately, the present expression analysis, which especially focused on this issue, had some obstacles. Due to the low risk allele frequencies, the subgroups with variant genotypes comprised only a few samples. This fact was tightened by the additional splitting according to the fusion status in 3 x 2 subgroups resulting in a substantially decreased statistical power. Despite these unfavorable circumstances, the subtyping in fact revealed an increased *MYC* expression dependent on 8q24 risk alleles in T2E negative normal tissue, in particular for the rs1447295 SNP and borderline significant also for both SNPs pooled. These results represent a first indication of a role of 8q24 variants for *MYC* expression, selectively in

T2E negative cases, but larger sample sizes with T2E status will be needed for confirmation.

While rs1447295 (region 1) and rs16901979 (region 2) exhibited compelling similarity with respect to an association with fusion negative PrCa, rs6983267 (region 3) appeared to differ from region 1 and 2 in subtype specificity and in DNA damage response. Neither of the two variants in risk regions 1 and 2 showed an association with DNA damage response, as measured by the micronucleus frequency or the mitotic delay index after irradiation. In contrast, the region 3 SNP rs6983267 pointed out in the mitotic delay test, since carriers of the risk genotype GG showed a nominal significantly decreased (0.88 fold) mitotic delay after γ-irradiation compared to those with TT and TG genotypes. It should be noted, that all repair test results for 8q24 must be handled with caution as 1) the correlation of region 3 with mitotic delay did not withstand correction for multiple testing, and 2) unsuspicious findings for regions 1 and 2 could be false negative results consequence of the limited power related to infrequent alleles and 3) there is possibly a tissue specificity of the genotype-mediated effects [2]. Interestingly, for all risk variants in 8q24, also for rs6983267 [122,145], *MYC* is suggested as the most plausible candidate gene, a curiosity that raises the question why rs6983267 is not associated with T2E negative PrCa. Plausible explanations may be found in diverging co-factors, or any temporal and spatial (e.g. cell type) difference between the particular involvements of risk variants, though targeting the same gene. For rs6983267, Wasserman and colleagues demonstrated enhancer activity for *MYC* that occurs during development, as early as prostatic gland maturation [176]. The predisposition mediated by this SNP could thus become manifested long before tumorigenesis [176], and might therefore be independent of somatic subtypes. Furthermore, tissue specificity is obviously also a diverging feature for regions 1 and 2 versus 3. The latter is, besides with PrCa, also associated with colon- and ovarian cancer risk, suggesting a rather generalized role in cancer susceptibility mediated by the rs6983267 variant, while regions 1 and 2 appear to be restrictive for PrCa [50].

4.4 Conclusions and prospects

The present study demonstrates, in principle, that reducing heterogeneity via subgrouping represents a promising strategy for the identification of functional evidence behind common PrCa risk variants in order to facilitate specified candidate gene analyses.

Functional screening suggested 10q11 as a risk region, where unusual DNA damage response could predispose to T2E formation via error prone double strand break repair. Previous candidate gene studies, which were solely based on proximity to the risk SNP, have failed to consider *PARG* so far, the most intriguing candidate for fulfilling the now proposed pathomechanism at 10q11. While in the present study *PARG* has already passed experimental proof to be capable for T2E fusion induction, further studies will be needed, in particular, to uncover mechanisms by which risk genotypes could influence *PARG* expression. Moreover, the present finding may suggests PAR homeostasis as a general mechanism for DNA repair dependent PrCa susceptibility, and thus could involve members of the PARP family at other genetic loci as further suspicious genes.

Two genomic regions were differentially associated with the T2E fusion, 8q24 with fusion negative and 17q24 with fusion positive PrCa, underlining the distinctness of the two entities. Moreover, as both loci were mapped to gene deserts, deeper insights in the underlying pathomechanisms would be highly appreciated. Subtype specific analyses could provide such clues for candidate gene analyses. In particular, the *MYC* gene on 8q24 was repeatedly investigated in the past for expressional changes dependent on genotypes in different settings, leading to negative or inconsistent results. The present results strongly suggest that *MYC* studies could be much more informative if genotype dependencies were investigated with focus on the fusion negative tumor type, where 8q24 risk variants apparently play a special role. While the fusion positive tumor type probably up-regulates the *MYC* oncogene in an ETS-driven manner, fusion negative might instead make use of different mechanisms of *MYC* up-regulation, such as gene amplifications or 8q24 risk genotypes.

Noteworthy, the study highlighted the limitations of expression analyses for candidate gene examinations. Especially when multiple subgrouping is performed, on top of high inter-individual variation in gene expression, a large number of samples are needed to obtain statistical significant differences. As the presented detailed analyses according to SNP genotypes and T2E fusion status were obviously limited in power, the suggested differential expressions of *MYC* and *SOX9* depending on genotypes are currently under examination in a larger sample set.

Finally, elucidating subtype specific effects for germline variants is a versatile approach to refine genetic epidemiologic aspects of PrCa such as risk contributions and prediction models. Further studies comprising larger sample sizes are warranted to identify whole patterns of SNPs characteristic for the distinct tumor entities of PrCa.

5. Summary

Despite prostate cancer (PrCa) has the highest heritability among common cancer types, little is known about PrCa specific susceptibility genes. While large genome wide association studies have identified an increasing number of common germline variants that are moderately associated with PrCa, the underlying risk-mediating mechanisms and causal genes remain elusive. The aim of the present study was the investigation of common variants for functional involvement in DNA repair, a pathway crucially involved in cancer predisposition, and for somatic subtype specificity, in order to facilitate context specific candidate gene research.

Two different strategies were pursued: The association of common variants (1) with the results of two different tests on DNA repair capacity in peripheral blood lymphocytes (PBL) of 128 healthy probands and (2) with the presence of the somatic *TMPRSS2-ERG* (T2E) fusion in a multicenter study involving ~ 1,200 PrCa cases. The T2E fusion is found in ~ 50 % of all prostate tumors and could represent a somatic outcome of defective DNA repair. Striking loci were then analyzed for candidate gene expression dependent on genotype and T2E fusion status in histologically normal and tumor tissue pairs of 87 PrCa cases. The functional involvement of candidate genes in T2E fusion formation was further investigated by the *TMPRSS2-ERG* induction assay in a cell culture model.

Replication analysis of the initially identified 25 common variants yielded nine loci significantly associated ($p < 0.05$) in the German case-control series comprising 708 cases and 509 controls with odds ratios (OR) between ~ 1.2 and 1.6. A simple cumulative model for a set of 12 selected loci identified a fraction of 5 % of men who are at extraordinary risk (OR = 5.44, $p = 0.003$) of PrCa, attributable to an excessive load of multiple common risk alleles. The multifactorial burden could thus resemble the predisposing effect of a high-risk gene mutation.

The most striking result of functional investigations was observed for rs10993994 on 10q11: PrCa risk allele (T) carriers seemed to cope more efficiently with radiation-induced DNA double strand breaks (lower micronucleus frequency, $p = 0.0003$) in a shorter time (shorter mitotic delay, $p = 0.0353$). The performance of risk allele carriers is challenging with respect to the quality of fixed DNA lesions, and could indicate preferential usage of error-prone non homologous end joining rather than the conservative homologous

recombination repair. Intriguingly, risk alleles at 10q11 were found enriched in T2E positive (T2E+) versus T2E negative (T2E-) cases, in line with the notion that genomic rearrangements are promoted by error-prone repair pathways. In consequence, the candidate list for 10q11, previously including *MSMB*, *NCOA4* and *TIMM23*, was extended by *PARG*, a gene with a high prior plausibility since poly (ADP-ribose) glycohydrolase is evidently involved in DNA repair. While expression analyses yielded inconsistent results, *PARG* knockdown resulted in a 3-fold increased T2E fusion induction after irradiation in LNCaP ($p = 0.0015$), supporting the thesis of *PARG* as a PrCa risk gene.

Moreover, subtype-specific association analyses identified differentially associated variants on 17q24 with T2E+ PrCa (rs1859962: OR = 1.30, $p = 0.0016$), and 8q24 with T2E- PrCa (rs16901979: OR = 0.53, $p = 0.0007$). Noteworthy, a second independent SNP on 8q24 (rs1447295) was also (nominally significantly) associated with the fusion negative subgroup, supporting the specificity of 8q24 for T2E- PrCa. Both SNP regions, 8q24 and 17q24, are expanded gene deserts, where the knowledge of subtype involvement could provide novel useful clues. At 17q24, the most plausible risk mechanism is the previously reported long-range interaction of a genetic variation near rs1859962 with the distant *SOX9* gene. Although no clear genotype-dependent expression was observed in the present work, *SOX9* was, in line with literature, up-regulated in tumor versus normal tissue explicitly in T2E+ cases. *MYC*, the classical oncogene on 8q24, was up-regulated in both, T2E+ and T2E- cases, but evidently increased by 8q24 risk alleles predominantly in T2E- cases. *MYC* overexpression is a general tumor driver. Since ETS transcription factors drive *MYC* expression, ETS negative tumors may need an alternate pathway for up-regulation, possibly mediated by 8q24 risk SNPs.

In conclusion, categorizing of risk loci for mechanistic contexts, such as for DNA repair and fusion gene formation, has proven successful to diminish disease heterogeneity and to allow incorporation of more specific functional knowledge in approaching candidate genes near distinctly associated variants. In particular, the present study identified *PARG* as a novel PrCa candidate gene on the genomic region 10q11 that probably acts as a modulator of double strand break repair and thus predisposes also for the ETS positive subtype. Two further loci, 8q24 and 17q24 were found differentially associated with the T2E fusion, highlighting the distinctness of ETS positive versus negative PrCa on the germline level.

6. Literature

1. Agalliu I, Salinas CA, Hansten PD, Ostrander EA, and Stanford JL: Statin use and risk of prostate cancer: results from a population-based epidemiologic study. Am.J.Epidemiol. 168: 250-260 (2008)

2. Ahmadiyeh N, Pomerantz MM, Grisanzio C, Herman P, Jia L, Almendro V, He HH, Brown M, Liu XS, Davis M, Caswell JL, Beckwith CA, Hills A, Macconaill L, Coetzee GA, Regan MM, and Freedman ML: 8q24 prostate, breast, and colon cancer risk loci show tissue-specific long-range interaction with MYC. Proc.Natl.Acad.Sci.U.S.A 107: 9742-9746 (2010)

3. Akamatsu S, Takata R, Haiman CA, Takahashi A, Inoue T, Kubo M, Furihata M, Kamatani N, Inazawa J, Chen GK, Le Marchand L, Kolonel LN, Katoh T, Yamano Y, Yamakado M, Takahashi H, Yamada H, Egawa S, Fujioka T, Henderson BE, Habuchi T, Ogawa O, Nakamura Y, and Nakagawa H: Common variants at 11q12, 10q26 and 3p11.2 are associated with prostate cancer susceptibility in Japanese. Nat.Genet. 44: 426-9, S1 (2012)

4. Al Olama AA, Kote-Jarai Z, Giles GG, Guy M, Morrison J, Severi G, Leongamornlert DA, Tymrakiewicz M, Jhavar S, Saunders E, Hopper JL, Southey MC, Muir KR, English DR, Dearnaley DP, Ardern-Jones AT, Hall AL, O'Brien LT, Wilkinson RA, Sawyer E, Lophatananon A, Horwich A, Huddart RA, Khoo VS, Parker CC, Woodhouse CJ, Thompson A, Christmas T, Ogden C, Cooper C, Donovan JL, Hamdy FC, Neal DE, Eeles RA, and Easton DF: Multiple loci on 8q24 associated with prostate cancer susceptibility. Nat.Genet. 41: 1058-1060 (2009)

5. Albadine R, Latour M, Toubaji A, Haffner M, Isaacs WB, Platz A, Meeker AK, Demarzo AM, Epstein JI, and Netto GJ: TMPRSS2-ERG gene fusion status in minute (minimal) prostatic adenocarcinoma. Mod.Pathol. 22: 1415-1422 (2009)

6. Alen P, Claessens F, Schoenmakers E, Swinnen JV, Verhoeven G, Rombauts W, and Peeters B: Interaction of the putative androgen receptor-specific coactivator ARA70/ELE1alpha with multiple steroid receptors and identification of an internally deleted ELE1beta isoform. Mol.Endocrinol. 13: 117-128 (1999)

7. Allan CA and McLachlan RI: Androgens and obesity. Curr.Opin.Endocrinol.Diabetes Obes. 17: 224-232 (2010)

8. Ame JC, Fouquerel E, Gauthier LR, Biard D, Boussin FD, Dantzer F, de Murcia G, and Schreiber V: Radiation-induced mitotic catastrophe in PARG-deficient cells. J.Cell Sci. 122: 1990-2002 (2009)

9. Amin Al Olama A, Benlloch S, Antoniou AC, Giles GG, Severi G, Neal DE, Hamdy FC, Donovan JL, Muir K, Schleutker J, Henderson BE, Haiman CA, Schumacher F, Pashayan N, Pharoah P, Ostrander EA, Stanford JL, Batra J, Clements JA, Champers S, Weischer M, Nordestgaard BG, Ingles SA, Sorensen KD, Orntoft T, Park JY, Cybulski C, Maier C, Doerk T, Dickinson JL, Cannon-Albright L, Brenner H, Rebbeck T, Habuchi T, Thibodeau SN, Cooney K, Chappuis P, Hutter P, Kaneva R, Foulkes WD, Zeegers MP, Lu YJ, Zhang HW, Stephenson R, Cox A, Southey MC, Spurdle A, FitzGerald LM, Leongamornlert DA, Saunders EJ, Tymrakiewicz M, Guy M, Dadaev T, Little SJ, Govindasami K, Sawyer EJ, Wilkinson RA, Herkommer K, Morrison J, Hopper J, Lophatonanon A, Rinckleb A, The UK Genetic Prostate Cancer Study Collaborators/British Association of Urological Surgeons' Section of Oncology, The UK ProtecT Study Collaborators, The PRACTICAL Consortium, Kote-Jarai Z, Eeles RA, Easton D. Risk Analysis of Prostate Cancer in PRACTICAL, a Multinational Consortium, Using 25 Known Prostate Cancer Susceptibility Loci. (Publication in Preparation)

10. Amundadottir LT, Sulem P, Gudmundsson J, Helgason A, Baker A, Agnarsson BA, Sigurdsson A, Benediktsdottir KR, Cazier JB, Sainz J, Jakobsdottir M, Kostic J, Magnusdottir DN, Ghosh S, Agnarsson K, Birgisdottir B, Le Roux L, Olafsdottir A, Blondal T, Andresdottir M, Gretarsdottir OS, Bergthorsson JT, Gudbjartsson D, Gylfason A, Thorleifsson G, Manolescu A, Kristjansson K, Geirsson G, Isaksson H, Douglas J, Johansson JE, Balter K, Wiklund F, Montie JE, Yu X, Suarez BK, Ober C, Cooney KA,

Gronberg H, Catalona WJ, Einarsson GV, Barkardottir RB, Gulcher JR, Kong A, Thorsteinsdottir U, and Stefansson K: A common variant associated with prostate cancer in European and African populations. Nat.Genet. 38: 652-658 (2006)

11. Aymard F, Bugler B, Schmidt CK, Guillou E, Caron P, Briois S, Iacovoni JS, Daburon V, Miller KM, Jackson SP, and Legube G: Transcriptionally active chromatin recruits homologous recombination at DNA double-strand breaks. Nat.Struct.Mol.Biol. (2014)

12. Baca SC, Prandi D, Lawrence MS, Mosquera JM, Romanel A, Drier Y, Park K, Kitabayashi N, Macdonald TY, Ghandi M, Van Allen E, Kryukov GV, Sboner A, Theurillat JP, Soong TD, Nickerson E, Auclair D, Tewari A, Beltran H, Onofrio RC, Boysen G, Guiducci C, Barbieri CE, Cibulskis K, Sivachenko A, Carter SL, Saksena G, Voet D, Ramos AH, Winckler W, Cipicchio M, Ardlie K, Kantoff PW, Berger MF, Gabriel SB, Golub TR, Meyerson M, Lander ES, Elemento O, Getz G, Demichelis F, Rubin MA, and Garraway LA: Punctuated evolution of prostate cancer genomes. Cell 153: 666-677 (2013)

13. Baeyens A, Thierens H, Claes K, Poppe B, Messiaen L, De Ridder L, and Vral A: Chromosomal radiosensitivity in breast cancer patients with a known or putative genetic predisposition. Br.J.Cancer 87: 1379-1385 (2002)

14. Ban S, Konomi C, Iwakawa M, Yamada S, Ohno T, Tsuji H, Noda S, Matui Y, Harada Y, Cologne JB, and Imai T: Radiosensitivity of peripheral blood lymphocytes obtained from patients with cancers of the breast, head and neck or cervix as determined with a micronucleus assay. J.Radiat.Res. 45: 535-541 (2004)

15. Barros-Silva JD, Paulo P, Bakken AC, Cerveira N, Lovf M, Henrique R, Jeronimo C, Lothe RA, Skotheim RI, and Teixeira MR: Novel 5' Fusion Partners of ETV1 and ETV4 in Prostate Cancer. Neoplasia. 15: 720-726 (2013)

16. Bartkova J, Horejsi Z, Koed K, Kramer A, Tort F, Zieger K, Guldberg P, Sehested M, Nesland JM, Lukas C, Orntoft T, Lukas J, and Bartek J: DNA damage response as a candidate anti-cancer barrier in early human tumorigenesis. Nature 434: 864-870 (2005)

17. Beke L, Nuytten M, Van Eynde A, Beullens M, and Bollen M: The gene encoding the prostatic tumor suppressor PSP94 is a target for repression by the Polycomb group protein EZH2. Oncogene 26: 4590-4595 (2007)

18. Beucher A, Birraux J, Tchouandong L, Barton O, Shibata A, Conrad S, Goodarzi AA, Krempler A, Jeggo PA, and Lobrich M: ATM and Artemis promote homologous recombination of radiation-induced DNA double-strand breaks in G2. EMBO J. 28: 3413-3427 (2009)

19. Bongarzone I, Butti MG, Coronelli S, Borrello MG, Santoro M, Mondellini P, Pilotti S, Fusco A, Della PG, and Pierotti MA: Frequent activation of ret protooncogene by fusion with a new activating gene in papillary thyroid carcinomas. Cancer Res. 54: 2979-2985 (1994)

20. Borno ST, Fischer A, Kerick M, Falth M, Laible M, Brase JC, Kuner R, Dahl A, Grimm C, Sayanjali B, Isau M, Rohr C, Wunderlich A, Timmermann B, Claus R, Plass C, Graefen M, Simon R, Demichelis F, Rubin MA, Sauter G, Schlomm T, Sultmann H, Lehrach H, and Schweiger MR: Genome-wide DNA methylation events in TMPRSS2-ERG fusion-negative prostate cancers implicate an EZH2-dependent mechanism with miR-26a hypermethylation. Cancer Discov. 2: 1024-1035 (2012)

21. Brenner JC, Ateeq B, Li Y, Yocum AK, Cao Q, Asangani IA, Patel S, Wang X, Liang H, Yu J, Palanisamy N, Siddiqui J, Yan W, Cao X, Mehra R, Sabolch A, Basrur V, Lonigro RJ, Yang J, Tomlins SA, Maher CA, Elenitoba-Johnson KS, Hussain M, Navone NM, Pienta KJ, Varambally S, Feng FY, and Chinnaiyan AM: Mechanistic rationale for inhibition of poly(ADP-ribose) polymerase in ETS gene fusion-positive prostate cancer. Cancer Cell 19: 664-678 (2011)

22. Buttyan R, Sawczuk IS, Benson MC, Siegal JD, and Olsson CA: Enhanced expression of the c-myc protooncogene in high-grade human prostate cancers. Prostate 11: 327-337 (1987)

23. Cai C, Wang H, He HH, Chen S, He L, Ma F, Mucci L, Wang Q, Fiore C, Sowalsky AG, Loda M, Liu XS, Brown M, Balk SP, and Yuan X: ERG induces androgen receptor-mediated regulation of SOX9 in prostate cancer. J.Clin.Invest 123: 1109-1122 (2013)

24. Carpten J, Nupponen N, Isaacs S, Sood R, Robbins C, Xu J, Faruque M, Moses T, Ewing C, Gillanders E, Hu P, Bujnovszky P, Makalowska I, Baffoe-Bonnie A, Faith D, Smith J, Stephan D, Wiley K, Brownstein M, Gildea D, Kelly B, Jenkins R, Hostetter G, Matikainen M, Schleutker J, Klinger K, Connors T, Xiang Y, Wang Z, De Marzo A, Papadopoulos N, Kallioniemi OP, Burk R, Meyers D, Gronberg H, Meltzer P, Silverman R, Bailey-Wilson J, Walsh P, Isaacs W, and Trent J: Germline mutations in the ribonuclease L gene in families showing linkage with HPC1. Nat.Genet. 30: 181-184 (2002)

25. Chan PS, Chan LW, Xuan JW, Chin JL, Choi HL, and Chan FL: In situ hybridization study of PSP94 (prostatic secretory protein of 94 amino acids) expression in human prostates. Prostate 41: 99-109 (1999)

26. Chang BL, Cramer SD, Wiklund F, Isaacs SD, Stevens VL, Sun J, Smith S, Pruett K, Romero LM, Wiley KE, Kim ST, Zhu Y, Zhang Z, Hsu FC, Turner AR, Adolfsson J, Liu W, Kim JW, Duggan D, Carpten J, Zheng SL, Rodriguez C, Isaacs WB, Gronberg H, and Xu J: Fine mapping association study and functional analysis implicate a SNP in MSMB at 10q11 as a causal variant for prostate cancer risk. Hum.Mol.Genet. 18: 1368-1375 (2009)

27. Chen B, Zhao J, Zhang S, Wu W, and Qi R: Aspirin inhibits the production of reactive oxygen species by downregulating Nox4 and inducible nitric oxide synthase in human endothelial cells exposed to oxidized low-density lipoprotein. J.Cardiovasc.Pharmacol. 59: 405-412 (2012)

28. Clark JP and Cooper CS: ETS gene fusions in prostate cancer. Nat.Rev.Urol. 6: 429-439 (2009)

29. Dang CV, O'Donnell KA, Zeller KI, Nguyen T, Osthus RC, and Li F: The c-Myc target gene network. Semin.Cancer Biol. 16: 253-264 (2006)

30. Darnel AD, LaFargue CJ, Vollmer RT, Corcos J, and Bismar TA: TMPRSS2-ERG fusion is frequently observed in Gleason pattern 3 prostate cancer in a Canadian cohort. Cancer Biol.Ther. 8: 125-130 (2009)

31. Demichelis F, Fall K, Perner S, Andren O, Schmidt F, Setlur SR, Hoshida Y, Mosquera JM, Pawitan Y, Lee C, Adami HO, Mucci LA, Kantoff PW, Andersson SO, Chinnaiyan AM, Johansson JE, and Rubin MA: TMPRSS2:ERG gene fusion associated with lethal prostate cancer in a watchful waiting cohort. Oncogene 26: 4596-4599 (2007)

32. Demichelis F, Garraway LA, and Rubin MA: Molecular archeology: unearthing androgen-induced structural rearrangements in prostate cancer genomes. Cancer Cell 23: 133-135 (2013)

33. DGU: Interdisziplinäre Leitlinie der Qualität S3 zur Früherkennung, Diagnose und Therapie der verschiedenen Stadien des Prostatakarzinoms 1.03 (2011)

34. Dong X, Wang L, Taniguchi K, Wang X, Cunningham JM, McDonnell SK, Qian C, Marks AF, Slager SL, Peterson BJ, Smith DI, Cheville JC, Blute ML, Jacobsen SJ, Schaid DJ, Tindall DJ, Thibodeau SN, and Liu W: Mutations in CHEK2 associated with prostate cancer risk. Am.J.Hum.Genet. 72: 270-280 (2003)

35. Duggan D, Zheng SL, Knowlton M, Benitez D, Dimitrov L, Wiklund F, Robbins C, Isaacs SD, Cheng Y, Li G, Sun J, Chang BL, Marovich L, Wiley KE, Balter K, Stattin P, Adami HO, Gielzak M, Yan G, Sauvageot J, Liu W, Kim JW, Bleecker ER, Meyers DA, Trock BJ, Partin AW, Walsh PC, Isaacs WB, Gronberg H, Xu J, and Carpten JD: Two genome-wide association studies of aggressive prostate cancer implicate putative prostate tumor suppressor gene DAB2IP. J.Natl.Cancer Inst. 99: 1836-1844 (2007)

36. Eeles RA, Kote-Jarai Z, Al Olama AA, Giles GG, Guy M, Severi G, Muir K, Hopper JL, Henderson BE, Haiman CA, Schleutker J, Hamdy FC, Neal DE, Donovan JL, Stanford JL, Ostrander EA, Ingles SA, John EM, Thibodeau SN, Schaid D, Park JY, Spurdle A, Clements J, Dickinson JL, Maier C, Vogel W, Dork T, Rebbeck TR, Cooney KA, Cannon-Albright L, Chappuis PO, Hutter P, Zeegers M, Kaneva R, Zhang HW,

Lu YJ, Foulkes WD, English DR, Leongamornlert DA, Tymrakiewicz M, Morrison J, Ardern-Jones AT, Hall AL, O'Brien LT, Wilkinson RA, Saunders EJ, Page EC, Sawyer EJ, Edwards SM, Dearnaley DP, Horwich A, Huddart RA, Khoo VS, Parker CC, Van As N, Woodhouse CJ, Thompson A, Christmas T, Ogden C, Cooper CS, Southey MC, Lophatananon A, Liu JF, Kolonel LN, Le Marchand L, Wahlfors T, Tammela TL, Auvinen A, Lewis SJ, Cox A, FitzGerald LM, Koopmeiners JS, Karyadi DM, Kwon EM, Stern MC, Corral R, Joshi AD, Shahabi A, McDonnell SK, Sellers TA, Pow-Sang J, Chambers S, Aitken J, Gardiner RA, Batra J, Kedda MA, Lose F, Polanowski A, Patterson B, Serth J, Meyer A, Luedeke M, Stefflova K, Ray AM, Lange EM, Farnham J, Khan H, Slavov C, Mitkova A, Cao G, and Easton DF: Identification of seven new prostate cancer susceptibility loci through a genome-wide association study. Nat.Genet. 41: 1116-1121 (2009)

37. Eeles RA, Kote-Jarai Z, Giles GG, Olama AA, Guy M, Jugurnauth SK, Mulholland S, Leongamornlert DA, Edwards SM, Morrison J, Field HI, Southey MC, Severi G, Donovan JL, Hamdy FC, Dearnaley DP, Muir KR, Smith C, Bagnato M, Ardern-Jones AT, Hall AL, O'Brien LT, Gehr-Swain BN, Wilkinson RA, Cox A, Lewis S, Brown PM, Jhavar SG, Tymrakiewicz M, Lophatananon A, Bryant SL, Horwich A, Huddart RA, Khoo VS, Parker CC, Woodhouse CJ, Thompson A, Christmas T, Ogden C, Fisher C, Jamieson C, Cooper CS, English DR, Hopper JL, Neal DE, and Easton DF: Multiple newly identified loci associated with prostate cancer susceptibility. Nat.Genet. 40: 316-321 (2008)

38. Eeles RA, Olama AA, Benlloch S, Saunders EJ, Leongamornlert DA, Tymrakiewicz M, Ghoussaini M, Luccarini C, Dennis J, Jugurnauth-Little S, Dadaev T, Neal DE, Hamdy FC, Donovan JL, Muir K, Giles GG, Severi G, Wiklund F, Gronberg H, Haiman CA, Schumacher F, Henderson BE, Le Marchand L, Lindstrom S, Kraft P, Hunter DJ, Gapstur S, Chanock SJ, Berndt SI, Albanes D, Andriole G, Schleutker J, Weischer M, Canzian F, Riboli E, Key TJ, Travis RC, Campa D, Ingles SA, John EM, Hayes RB, Pharoah PD, Pashayan N, Khaw KT, Stanford JL, Ostrander EA, Signorello LB, Thibodeau SN, Schaid D, Maier C, Vogel W, Kibel AS, Cybulski C, Lubinski J, Cannon-Albright L, Brenner H, Park JY, Kaneva R, Batra J, Spurdle AB, Clements JA, Teixeira MR, Dicks E, Lee A, Dunning AM, Baynes C, Conroy D, Maranian MJ, Ahmed S, Govindasami K, Guy M, Wilkinson RA, Sawyer EJ, Morgan A, Dearnaley DP, Horwich A, Huddart RA, Khoo VS, Parker CC, Van As NJ, Woodhouse CJ, Thompson A, Dudderidge T, Ogden C, Cooper CS, Lophatananon A, Cox A, Southey MC, Hopper JL, English DR, Aly M, Adolfsson J, Xu J, Zheng SL, Yeager M, Kaaks R, Diver WR, Gaudet MM, Stern MC, Corral R, Joshi AD, Shahabi A, Wahlfors T, Tammela TL, Auvinen A, Virtamo J, Klarskov P, Nordestgaard BG, Roder MA, Nielsen SF, Bojesen SE, Siddiq A, FitzGerald LM, Kolb S, Kwon EM, Karyadi DM, Blot WJ, Zheng W, Cai Q, McDonnell SK, Rinckleb AE, Drake B, Colditz G, Wokolorczyk D, Stephenson RA, Teerlink C, Muller H, Rothenbacher D, Sellers TA, Lin HY, Slavov C, Mitev V, Lose F, Srinivasan S, Maia S, Paulo P, Lange E, Cooney KA, Antoniou AC, Vincent D, Bacot F, Tessier DC, Kote-Jarai Z, and Easton DF: Identification of 23 new prostate cancer susceptibility loci using the iCOGS custom genotyping array. Nat.Genet. 45: 385-391 (2013)

39. Egber L, Luedeke M, Rinckleb A, Wright JL, Maier C, Neuhouser ML, and Stanford JL: Obesity and the risk of prostate cancer according to *TMPRSS2:ERG* fusion (Publication in preparation)

40. Epstein JI, Allsbrook WC, Jr., Amin MB, and Egevad LL: The 2005 International Society of Urological Pathology (ISUP) Consensus Conference on Gleason Grading of Prostatic Carcinoma. Am.J.Surg.Pathol. 29: 1228-1242 (2005)

41. Erkko H, Xia B, Nikkila J, Schleutker J, Syrjakoski K, Mannermaa A, Kallioniemi A, Pylkas K, Karppinen SM, Rapakko K, Miron A, Sheng Q, Li G, Mattila H, Bell DW, Haber DA, Grip M, Reiman M, Jukkola-Vuorinen A, Mustonen A, Kere J, Aaltonen LA, Kosma VM, Kataja V, Soini Y, Drapkin RI, Livingston DM, and Winqvist R: A recurrent mutation in PALB2 in Finnish cancer families. Nature 446: 316-319 (2007)

42. Esgueva R, Perner S, LaFargue J, Scheble V, Stephan C, Lein M, Fritzsche FR, Dietel M, Kristiansen G, and Rubin MA: Prevalence of TMPRSS2-ERG and SLC45A3-ERG gene fusions in a large prostatectomy cohort. Mod.Pathol. 23: 539-546 (2010)

43. Ewing CM, Ray AM, Lange EM, Zuhlke KA, Robbins CM, Tembe WD, Wiley KE, Isaacs SD, Johng D, Wang Y, Bizon C, Yan G, Gielzak M, Partin AW, Shanmugam V, Izatt T, Sinari S, Craig DW, Zheng SL,

Walsh PC, Montie JE, Xu J, Carpten JD, Isaacs WB, and Cooney KA: Germline mutations in HOXB13 and prostate-cancer risk. N.Engl.J.Med. 366: 141-149 (2012)

44. Fattah F, Lee EH, Weisensel N, Wang Y, Lichter N, and Hendrickson EA: Ku regulates the non-homologous end joining pathway choice of DNA double-strand break repair in human somatic cells. PLoS.Genet. 6: e1000855 (2010)

45. Fenech M: Cytokinesis-block micronucleus cytome assay. Nat.Protoc. 2: 1084-1104 (2007)

46. FitzGerald LM, Zhang X, Kolb S, Kwon EM, Liew YC, Hurtado-Coll A, Knudsen BS, Ostrander EA, and Stanford JL: Investigation of the relationship between prostate cancer and MSMB and NCOA4 genetic variants and protein expression. Hum.Mutat. 34: 149-156 (2013)

47. Fleming WH, Hamel A, MacDonald R, Ramsey E, Pettigrew NM, Johnston B, Dodd JG, and Matusik RJ: Expression of the c-myc protooncogene in human prostatic carcinoma and benign prostatic hyperplasia. Cancer Res. 46: 1535-1538 (1986)

48. Freedman ML, Monteiro AN, Gayther SA, Coetzee GA, Risch A, Plass C, Casey G, De Biasi M, Carlson C, Duggan D, James M, Liu P, Tichelaar JW, Vikis HG, You M, and Mills IG: Principles for the post-GWAS functional characterization of cancer risk loci. Nat.Genet. 43: 513-518 (2011)

49. Furusato B, Gao CL, Ravindranath L, Chen Y, Cullen J, McLeod DG, Dobi A, Srivastava S, Petrovics G, and Sesterhenn IA: Mapping of TMPRSS2-ERG fusions in the context of multi-focal prostate cancer. Mod.Pathol. 21: 67-75 (2008)

50. Ghoussaini M, Song H, Koessler T, Al Olama AA, Kote-Jarai Z, Driver KE, Pooley KA, Ramus SJ, Kjaer SK, Hogdall E, DiCioccio RA, Whittemore AS, Gayther SA, Giles GG, Guy M, Edwards SM, Morrison J, Donovan JL, Hamdy FC, Dearnaley DP, Ardern-Jones AT, Hall AL, O'Brien LT, Gehr-Swain BN, Wilkinson RA, Brown PM, Hopper JL, Neal DE, Pharoah PD, Ponder BA, Eeles RA, Easton DF, and Dunning AM: Multiple loci with different cancer specificities within the 8q24 gene desert. J.Natl.Cancer Inst. 100: 962-966 (2008)

51. Gillanders EM, Xu J, Chang BL, Lange EM, Wiklund F, Bailey-Wilson JE, Baffoe-Bonnie A, Jones M, Gildea D, Riedesel E, Albertus J, Isaacs SD, Wiley KE, Mohai CE, Matikainen MP, Tammela TL, Zheng SL, Brown WM, Rokman A, Carpten JD, Meyers DA, Walsh PC, Schleutker J, Gronberg H, Cooney KA, Isaacs WB, and Trent JM: Combined genome-wide scan for prostate cancer susceptibility genes. J.Natl.Cancer Inst. 96: 1240-1247 (2004)

52. Goemans CG, Boya P, Skirrow CJ, and Tolkovsky AM: Intra-mitochondrial degradation of Tim23 curtails the survival of cells rescued from apoptosis by caspase inhibitors. Cell Death.Differ. 15: 545-554 (2008)

53. Green DR and Kroemer G: The pathophysiology of mitochondrial cell death. Science 305: 626-629 (2004)

54. Gudmundsson J, Sulem P, Gudbjartsson DF, Blondal T, Gylfason A, Agnarsson BA, Benediktsdottir KR, Magnusdottir DN, Orlygsdottir G, Jakobsdottir M, Stacey SN, Sigurdsson A, Wahlfors T, Tammela T, Breyer JP, McReynolds KM, Bradley KM, Saez B, Godino J, Navarrete S, Fuertes F, Murillo L, Polo E, Aben KK, van Oort IM, Suarez BK, Helfand BT, Kan D, Zanon C, Frigge ML, Kristjansson K, Gulcher JR, Einarsson GV, Jonsson E, Catalona WJ, Mayordomo JI, Kiemeney LA, Smith JR, Schleutker J, Barkardottir RB, Kong A, Thorsteinsdottir U, Rafnar T, and Stefansson K: Genome-wide association and replication studies identify four variants associated with prostate cancer susceptibility. Nat.Genet. 41: 1122-1126 (2009)

55. Gudmundsson J, Sulem P, Manolescu A, Amundadottir LT, Gudbjartsson D, Helgason A, Rafnar T, Bergthorsson JT, Agnarsson BA, Baker A, Sigurdsson A, Benediktsdottir KR, Jakobsdottir M, Xu J, Blondal T, Kostic J, Sun J, Ghosh S, Stacey SN, Mouy M, Saemundsdottir J, Backman VM, Kristjansson K, Tres A, Partin AW, Albers-Akkers MT, Godino-Ivan MJ, Walsh PC, Swinkels DW, Navarrete S, Isaacs SD, Aben KK, Graif T, Cashy J, Ruiz-Echarri M, Wiley KE, Suarez BK, Witjes JA, Frigge M, Ober C,

Jonsson E, Einarsson GV, Mayordomo JI, Kiemeney LA, Isaacs WB, Catalona WJ, Barkardottir RB, Gulcher JR, Thorsteinsdottir U, Kong A, and Stefansson K: Genome-wide association study identifies a second prostate cancer susceptibility variant at 8q24. Nat.Genet. 39: 631-637 (2007)

56. Gudmundsson J, Sulem P, Rafnar T, Bergthorsson JT, Manolescu A, Gudbjartsson D, Agnarsson BA, Sigurdsson A, Benediktsdottir KR, Blondal T, Jakobsdottir M, Stacey SN, Kostic J, Kristinsson KT, Birgisdottir B, Ghosh S, Magnusdottir DN, Thorlacius S, Thorleifsson G, Zheng SL, Sun J, Chang BL, Elmore JB, Breyer JP, McReynolds KM, Bradley KM, Yaspan BL, Wiklund F, Stattin P, Lindstrom S, Adami HO, McDonnell SK, Schaid DJ, Cunningham JM, Wang L, Cerhan JR, St Sauver JL, Isaacs SD, Wiley KE, Partin AW, Walsh PC, Polo S, Ruiz-Echarri M, Navarrete S, Fuertes F, Saez B, Godino J, Weijerman PC, Swinkels DW, Aben KK, Witjes JA, Suarez BK, Helfand BT, Frigge ML, Kristjansson K, Ober C, Jonsson E, Einarsson GV, Xu J, Gronberg H, Smith JR, Thibodeau SN, Isaacs WB, Catalona WJ, Mayordomo JI, Kiemeney LA, Barkardottir RB, Gulcher JR, Thorsteinsdottir U, Kong A, and Stefansson K: Common sequence variants on 2p15 and Xp11.22 confer susceptibility to prostate cancer. Nat.Genet. 40: 281-283 (2008)

57. Gudmundsson J, Sulem P, Steinthorsdottir V, Bergthorsson JT, Thorleifsson G, Manolescu A, Rafnar T, Gudbjartsson D, Agnarsson BA, Baker A, Sigurdsson A, Benediktsdottir KR, Jakobsdottir M, Blondal T, Stacey SN, Helgason A, Gunnarsdottir S, Olafsdottir A, Kristinsson KT, Birgisdottir B, Ghosh S, Thorlacius S, Magnusdottir D, Stefansdottir G, Kristjansson K, Bagger Y, Wilensky RL, Reilly MP, Morris AD, Kimber CH, Adeyemo A, Chen Y, Zhou J, So WY, Tong PC, Ng MC, Hansen T, Andersen G, Borch-Johnsen K, Jorgensen T, Tres A, Fuertes F, Ruiz-Echarri M, Asin L, Saez B, van Boven E, Klaver S, Swinkels DW, Aben KK, Graif T, Cashy J, Suarez BK, van Vierssen TO, Frigge ML, Ober C, Hofker MH, Wijmenga C, Christiansen C, Rader DJ, Palmer CN, Rotimi C, Chan JC, Pedersen O, Sigurdsson G, Benediktsson R, Jonsson E, Einarsson GV, Mayordomo JI, Catalona WJ, Kiemeney LA, Barkardottir RB, Gulcher JR, Thorsteinsdottir U, Kong A, and Stefansson K: Two variants on chromosome 17 confer prostate cancer risk, and the one in TCF2 protects against type 2 diabetes. Nat.Genet. 39: 977-983 (2007)

58. Gurel B, Iwata T, Koh CM, Jenkins RB, Lan F, Van Dang C, Hicks JL, Morgan J, Cornish TC, Sutcliffe S, Isaacs WB, Luo J, and De Marzo AM: Nuclear MYC protein overexpression is an early alteration in human prostate carcinogenesis. Mod.Pathol. 21: 1156-1167 (2008)

59. Haffner MC, Aryee MJ, Toubaji A, Esopi DM, Albadine R, Gurel B, Isaacs WB, Bova GS, Liu W, Xu J, Meeker AK, Netto G, De Marzo AM, Nelson WG, and Yegnasubramanian S: Androgen-induced TOP2B-mediated double-strand breaks and prostate cancer gene rearrangements. Nat.Genet. 42: 668-675 (2010)

60. Haiman CA, Chen GK, Blot WJ, Strom SS, Berndt SI, Kittles RA, Rybicki BA, Isaacs WB, Ingles SA, Stanford JL, Diver WR, Witte JS, Hsing AW, Nemesure B, Rebbeck TR, Cooney KA, Xu J, Kibel AS, Hu JJ, John EM, Gueye SM, Watya S, Signorello LB, Hayes RB, Wang Z, Yeboah E, Tettey Y, Cai Q, Kolb S, Ostrander EA, Zeigler-Johnson C, Yamamura Y, Neslund-Dudas C, Haslag-Minoff J, Wu W, Thomas V, Allen GO, Murphy A, Chang BL, Zheng SL, Leske MC, Wu SY, Ray AM, Hennis AJ, Thun MJ, Carpten J, Casey G, Carter EN, Duarte ER, Xia LY, Sheng X, Wan P, Pooler LC, Cheng I, Monroe KR, Schumacher F, Le Marchand L, Kolonel LN, Chanock SJ, Van Den BD, Stram DO, and Henderson BE: Genome-wide association study of prostate cancer in men of African ancestry identifies a susceptibility locus at 17q21. Nat.Genet. 43: 570-573 (2011)

61. Han S, Brenner JC, Sabolch A, Jackson W, Speers C, Wilder-Romans K, Knudsen KE, Lawrence TS, Chinnaiyan AM, and Feng FY: Targeted radiosensitization of ETS fusion-positive prostate cancer through PARP1 inhibition. Neoplasia. 15: 1207-1217 (2013)

62. Hawksworth D, Ravindranath L, Chen Y, Furusato B, Sesterhenn IA, McLeod DG, Srivastava S, and Petrovics G: Overexpression of C-MYC oncogene in prostate cancer predicts biochemical recurrence. Prostate Cancer Prostatic.Dis. 13: 311-315 (2010)

63. Helfand BT, Fought AJ, Loeb S, Meeks JJ, Kan D, and Catalona WJ: Genetic prostate cancer risk assessment: common variants in 9 genomic regions are associated with cumulative risk. J.Urol. 184: 501-505 (2010)

64. Helfand BT, Loeb S, Meeks JJ, Fought AJ, Kan D, and Catalona WJ: Pathological outcomes associated with the 17q prostate cancer risk variants. J.Urol. 181: 2502-2507 (2009)

65. Helgeson BE, Tomlins SA, Shah N, Laxman B, Cao Q, Prensner JR, Cao X, Singla N, Montie JE, Varambally S, Mehra R, and Chinnaiyan AM: Characterization of TMPRSS2:ETV5 and SLC45A3:ETV5 gene fusions in prostate cancer. Cancer Res. 68: 73-80 (2008)

66. Hill-Harfe KL, Kaplan L, Stalker HJ, Zori RT, Pop R, Scherer G, and Wallace MR: Fine mapping of chromosome 17 translocation breakpoints > or = 900 Kb upstream of SOX9 in acampomelic campomelic dysplasia and a mild, familial skeletal dysplasia. Am.J.Hum.Genet. 76: 663-671 (2005)

67. Hofer MD, Kuefer R, Maier C, Herkommer K, Perner S, Demichelis F, Paiss T, Vogel W, Rubin MA, and Hoegel J: Genome-wide linkage analysis of TMPRSS2-ERG fusion in familial prostate cancer. Cancer Res. 69: 640-646 (2009)

68. Hooper JD, Clements JA, Quigley JP, and Antalis TM: Type II transmembrane serine proteases. Insights into an emerging class of cell surface proteolytic enzymes. J.Biol.Chem. 276: 857-860 (2001)

69. Hsu T, Trojanowska M, and Watson DK: Ets proteins in biological control and cancer. J.Cell Biochem. 91: 896-903 (2004)

70. Hu YC, Yeh S, Yeh SD, Sampson ER, Huang J, Li P, Hsu CL, Ting HJ, Lin HK, Wang L, Kim E, Ni J, and Chang C: Functional domain and motif analyses of androgen receptor coregulator ARA70 and its differential expression in prostate cancer. J.Biol.Chem. 279: 33438-33446 (2004)

71. Huang Z, Hurley PJ, Simons BW, Marchionni L, Berman DM, Ross AE, and Schaeffer EM: Sox9 is required for prostate development and prostate cancer initiation. Oncotarget. 3: 651-663 (2012)

72. Humphrey PA: Histological variants of prostatic carcinoma and their significance. Histopathology 60: 59-74 (2012)

73. Jin G, Lu L, Cooney KA, Ray AM, Zuhlke KA, Lange EM, Cannon-Albright LA, Camp NJ, Teerlink CC, FitzGerald LM, Stanford JL, Wiley KE, Isaacs SD, Walsh PC, Foulkes WD, Giles GG, Hopper JL, Severi G, Eeles R, Easton D, Kote-Jarai Z, Guy M, Rinckleb A, Maier C, Vogel W, Cancel-Tassin G, Egrot C, Cussenot O, Thibodeau SN, McDonnell SK, Schaid DJ, Wiklund F, Gronberg H, Emanuelsson M, Whittemore AS, Oakley-Girvan I, Hsieh CL, Wahlfors T, Tammela T, Schleutker J, Catalona WJ, Zheng SL, Ostrander EA, Isaacs WB, and Xu J: Validation of prostate cancer risk-related loci identified from genome-wide association studies using family-based association analysis: evidence from the International Consortium for Prostate Cancer Genetics (ICPCG). Hum.Genet. 131: 1095-1103 (2012)

74. Johns LE and Houlston RS: A systematic review and meta-analysis of familial prostate cancer risk. BJU.Int. 91: 789-794 (2003)

75. Keil C, Petermann E, and Oei SL: Tannins elevate the level of poly(ADP-ribose) in HeLa cell extracts. Arch.Biochem.Biophys. 425: 115-121 (2004)

76. Kicinski M, Vangronsveld J, and Nawrot TS: An epidemiological reappraisal of the familial aggregation of prostate cancer: a meta-analysis. PLoS.One. 6: e27130 (2011)

77. Koh CM, Bieberich CJ, Dang CV, Nelson WG, Yegnasubramanian S, and De Marzo AM: MYC and Prostate Cancer. Genes Cancer 1: 617-628 (2010)

78. Kollara A and Brown TJ: Expression and function of nuclear receptor co-activator 4: evidence of a potential role independent of co-activator activity. Cell Mol.Life Sci. (2012)

79. Korenchuk S, Lehr JE, MClean L, Lee YG, Whitney S, Vessella R, Lin DL, and Pienta KJ: VCaP, a cell-based model system of human prostate cancer. In Vivo 15: 163-168 (2001)

80. Kote-Jarai Z, Easton DF, Stanford JL, Ostrander EA, Schleutker J, Ingles SA, Schaid D, Thibodeau S, Dork T, Neal D, Donovan J, Hamdy F, Cox A, Maier C, Vogel W, Guy M, Muir K, Lophatananon A, Kedda MA, Spurdle A, Steginga S, John EM, Giles G, Hopper J, Chappuis PO, Hutter P, Foulkes WD, Hamel N, Salinas CA, Koopmeiners JS, Karyadi DM, Johanneson B, Wahlfors T, Tammela TL, Stern MC, Corral R, McDonnell SK, Schurmann P, Meyer A, Kuefer R, Leongamornlert DA, Tymrakiewicz M, Liu JF, O'Mara T, Gardiner RA, Aitken J, Joshi AD, Severi G, English DR, Southey M, Edwards SM, Al Olama AA, and Eeles RA: Multiple novel prostate cancer predisposition loci confirmed by an international study: the PRACTICAL Consortium. Cancer Epidemiol.Biomarkers Prev. 17: 2052-2061 (2008)

81. Kote-Jarai Z, Olama AA, Giles GG, Severi G, Schleutker J, Weischer M, Campa D, Riboli E, Key T, Gronberg H, Hunter DJ, Kraft P, Thun MJ, Ingles S, Chanock S, Albanes D, Hayes RB, Neal DE, Hamdy FC, Donovan JL, Pharoah P, Schumacher F, Henderson BE, Stanford JL, Ostrander EA, Sorensen KD, Dork T, Andriole G, Dickinson JL, Cybulski C, Lubinski J, Spurdle A, Clements JA, Chambers S, Aitken J, Gardiner RA, Thibodeau SN, Schaid D, John EM, Maier C, Vogel W, Cooney KA, Park JY, Cannon-Albright L, Brenner H, Habuchi T, Zhang HW, Lu YJ, Kaneva R, Muir K, Benlloch S, Leongamornlert DA, Saunders EJ, Tymrakiewicz M, Mahmud N, Guy M, O'Brien LT, Wilkinson RA, Hall AL, Sawyer EJ, Dadaev T, Morrison J, Dearnaley DP, Horwich A, Huddart RA, Khoo VS, Parker CC, Van As N, Woodhouse CJ, Thompson A, Christmas T, Ogden C, Cooper CS, Lophatonanon A, Southey MC, Hopper JL, English DR, Wahlfors T, Tammela TL, Klarskov P, Nordestgaard BG, Roder MA, Tybjaerg-Hansen A, Bojesen SE, Travis R, Canzian F, Kaaks R, Wiklund F, Aly M, Lindstrom S, Diver WR, Gapstur S, Stern MC, Corral R, Virtamo J, Cox A, Haiman CA, Le Marchand L, Fitzgerald L, Kolb S, Kwon EM, Karyadi DM, Orntoft TF, Borre M, Meyer A, Serth J, Yeager M, Berndt SI, Marthick JR, Patterson B, Wokolorczyk D, Batra J, Lose F, McDonnell SK, Joshi AD, Shahabi A, Rinckleb AE, Ray A, Sellers TA, Lin HY, Stephenson RA, Farnham J, Muller H, Rothenbacher D, Tsuchiya N, Narita S, Cao GW, Slavov C, Mitev V, Easton DF, and Eeles RA: Seven prostate cancer susceptibility loci identified by a multi-stage genome-wide association study. Nat.Genet. 43: 785-791 (2011)

82. Krishnakumar R and Kraus WL: The PARP side of the nucleus: molecular actions, physiological outcomes, and clinical targets. Mol.Cell 39: 8-24 (2010)

83. Laxman B, Morris DS, Yu J, Siddiqui J, Cao J, Mehra R, Lonigro RJ, Tsodikov A, Wei JT, Tomlins SA, and Chinnaiyan AM: A first-generation multiplex biomarker analysis of urine for the early detection of prostate cancer. Cancer Res. 68: 645-649 (2008)

84. Leongamornlert D, Saunders E, Dadaev T, Tymrakiewicz M, Goh C, Jugurnauth-Little S, Kozarewa I, Fenwick K, Assiotis I, Barrowdale D, Govindasami K, Guy M, Sawyer E, Wilkinson R, Antoniou AC, Eeles R, and Kote-Jarai Z: Frequent germline deleterious mutations in DNA repair genes in familial prostate cancer cases are associated with advanced disease. Br.J.Cancer 110: 1663-1672 (2014)

85. Li P, Yu X, Ge K, Melamed J, Roeder RG, and Wang Z: Heterogeneous expression and functions of androgen receptor co-factors in primary prostate cancer. Am.J.Pathol. 161: 1467-1474 (2002)

86. Lichtenstein P, Holm NV, Verkasalo PK, Iliadou A, Kaprio J, Koskenvuo M, Pukkala E, Skytthe A, and Hemminki K: Environmental and heritable factors in the causation of cancer--analyses of cohorts of twins from Sweden, Denmark, and Finland. N.Engl.J.Med. 343: 78-85 (2000)

87. Lieber MR: The mechanism of human nonhomologous DNA end joining. J.Biol.Chem. 283: 1-5 (2008)

88. Ligr M, Li Y, Zou X, Daniels G, Melamed J, Peng Y, Wang W, Wang J, Ostrer H, Pagano M, Wang Z, Garabedian MJ, and Lee P: Tumor suppressor function of androgen receptor coactivator ARA70alpha in prostate cancer. Am.J.Pathol. 176: 1891-1900 (2010)

89. Lilja H and Abrahamsson PA: Three predominant proteins secreted by the human prostate gland. Prostate 12: 29-38 (1988)

90. Lin C, Yang L, Tanasa B, Hutt K, Ju BG, Ohgi K, Zhang J, Rose DW, Fu XD, Glass CK, and Rosenfeld MG: Nuclear receptor-induced chromosomal proximity and DNA breaks underlie specific translocations in cancer. Cell 139: 1069-1083 (2009)

91. Little J, Wilson B, Carter R, Walker K, Santaguida P, Tomiak E, Beyene J, and Raina P: Multigene panels in prostate cancer risk assessment. Evid.Rep.Technol.Assess.(Full.Rep.): 1-166 (2012)

92. Livak KJ and Schmittgen TD: Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods 25: 402-408 (2001)

93. Lou H, Yeager M, Li H, Bosquet JG, Hayes RB, Orr N, Yu K, Hutchinson A, Jacobs KB, Kraft P, Wacholder S, Chatterjee N, Feigelson HS, Thun MJ, Diver WR, Albanes D, Virtamo J, Weinstein S, Ma J, Gaziano JM, Stampfer M, Schumacher FR, Giovannucci E, Cancel-Tassin G, Cussenot O, Valeri A, Andriole GL, Crawford ED, Anderson SK, Tucker M, Hoover RN, Fraumeni JF, Jr., Thomas G, Hunter DJ, Dean M, and Chanock SJ: Fine mapping and functional analysis of a common variant in MSMB on chromosome 10q11.2 associated with prostate cancer susceptibility. Proc.Natl.Acad.Sci.U.S.A 106: 7933-7938 (2009)

94. Lüdeke M: Predisposition for TMPRSS2-ERG fusion in prostate cancer by variants in DNA repair genes. Dissertation (Dr. rer.nat.), Universität Ulm (2012)

95. Luedeke M, Linnert CM, Hofer MD, Surowy HM, Rinckleb AE, Hoegel J, Kuefer R, Rubin MA, Vogel W, and Maier C: Predisposition for TMPRSS2-ERG fusion in prostate cancer by variants in DNA repair genes. Cancer Epidemiol.Biomarkers Prev. 18: 3030-3035 (2009)

96. Maher CA, Palanisamy N, Brenner JC, Cao X, Kalyana-Sundaram S, Luo S, Khrebtukova I, Barrette TR, Grasso C, Yu J, Lonigro RJ, Schroth G, Kumar-Sinha C, and Chinnaiyan AM: Chimeric transcript discovery by paired-end transcriptome sequencing. Proc.Natl.Acad.Sci.U.S.A 106: 12353-12358 (2009)

97. Maier C, Vesovic Z, Bachmann N, Herkommer K, Braun AK, Surowy HM, Assum G, Paiss T, and Vogel W: Germline mutations of the MSR1 gene in prostate cancer families from Germany. Hum.Mutat. 27: 98-102 (2006)

98. Mani RS, Tomlins SA, Callahan K, Ghosh A, Nyati MK, Varambally S, Palanisamy N, and Chinnaiyan AM: Induced chromosomal proximity and gene fusions in prostate cancer. Science 326: 1230 (2009)

99. Mao Z, Bozzella M, Seluanov A, and Gorbunova V: Comparison of nonhomologous end joining and homologous recombination in human cells. DNA Repair (Amst) 7: 1765-1771 (2008)

100. Mehra R, Tomlins SA, Shen R, Nadeem O, Wang L, Wei JT, Pienta KJ, Ghosh D, Rubin MA, Chinnaiyan AM, and Shah RB: Comprehensive assessment of TMPRSS2 and ETS family gene aberrations in clinically localized prostate cancer. Mod.Pathol. 20: 538-544 (2007)

101. Meister G: Argonaute proteins: functional insights and emerging roles. Nat.Rev.Genet. 14: 447-459 (2013)

102. Mestayer C, Blanchere M, Jaubert F, Dufour B, and Mowszowicz I: Expression of androgen receptor coactivators in normal and cancer prostate tissues and cultured cell lines. Prostate 56: 192-200 (2003)

103. Meyer RG, Meyer-Ficca ML, Jacobson EL, and Jacobson MK: Human poly(ADP-ribose) glycohydrolase (PARG) gene and the common promoter sequence it shares with inner mitochondrial membrane translocase 23 (TIM23). Gene 314: 181-190 (2003)

104. Meyer RG, Meyer-Ficca ML, Whatcott CJ, Jacobson EL, and Jacobson MK: Two small enzyme isoforms mediate mammalian mitochondrial poly(ADP-ribose) glycohydrolase (PARG) activity. Exp.Cell Res. 313: 2920-2936 (2007)

105. Meyer-Ficca ML, Meyer RG, Coyle DL, Jacobson EL, and Jacobson MK: Human poly(ADP-ribose) glycohydrolase is expressed in alternative splice variants yielding isoforms that localize to different cell compartments. Exp.Cell Res. 297: 521-532 (2004)

106. Miller SA, Dykes DD, and Polesky HF: A simple salting out procedure for extracting DNA from human nucleated cells. Nucleic Acids Res. 16: 1215 (1988)

107. Mosquera JM, Mehra R, Regan MM, Perner S, Genega EM, Bueti G, Shah RB, Gaston S, Tomlins SA, Wei JT, Kearney MC, Johnson LA, Tang JM, Chinnaiyan AM, Rubin MA, and Sanda MG: Prevalence of TMPRSS2-ERG fusion prostate cancer among men undergoing prostate biopsy in the United States. Clin.Cancer Res. 15: 4706-4711 (2009)

108. Mwamukonda K, Chen Y, Ravindranath L, Furusato B, Hu Y, Sterbis J, Osborn D, Rosner I, Sesterhenn IA, McLeod DG, Srivastava S, and Petrovics G: Quantitative expression of TMPRSS2 transcript in prostate tumor cells reflects TMPRSS2-ERG fusion status. Prostate Cancer Prostatic.Dis. 13: 47-51 (2010)

109. Nakadate Y, Kodera Y, Kitamura Y, Tachibana T, Tamura T, and Koizumi F: Silencing of poly(ADP-ribose) glycohydrolase sensitizes lung cancer cells to radiation through the abrogation of DNA damage checkpoint. Biochem.Biophys.Res.Commun. 441: 793-798 (2013)

110. Nam RK, Sugar L, Wang Z, Yang W, Kitching R, Klotz LH, Venkateswaran V, Narod SA, and Seth A: Expression of TMPRSS2:ERG gene fusion in prostate cancer cells is an important prognostic factor for cancer progression. Cancer Biol.Ther. 6: 40-45 (2007)

111. Nam RK, Sugar L, Yang W, Srivastava S, Klotz LH, Yang LY, Stanimirovic A, Encioiu E, Neill M, Loblaw DA, Trachtenberg J, Narod SA, and Seth A: Expression of the TMPRSS2:ERG fusion gene predicts cancer recurrence after surgery for localised prostate cancer. Br.J.Cancer 97: 1690-1695 (2007)

112. Nature Collections: iCOGS. Nature collections (2013)

113. Nesbit CE, Tersak JM, and Prochownik EV: MYC oncogenes and human neoplastic disease. Oncogene 18: 3004-3016 (1999)

114. Norppa H and Falck GC: What do human micronuclei contain? Mutagenesis 18: 221-233 (2003)

115. Norris JD, Chang CY, Wittmann BM, Kunder RS, Cui H, Fan D, Joseph JD, and McDonnell DP: The homeodomain protein HOXB13 regulates the cellular response to androgens. Mol.Cell 36: 405-416 (2009)

116. Oikawa T and Yamada T: Molecular biology of the Ets family of transcription factors. Gene 303: 11-34 (2003)

117. Paulo P, Barros-Silva JD, Ribeiro FR, Ramalho-Carvalho J, Jeronimo C, Henrique R, Lind GE, Skotheim RI, Lothe RA, and Teixeira MR: FLI1 is a novel ETS transcription factor involved in gene fusions in prostate cancer. Genes Chromosomes.Cancer 51: 240-249 (2012)

118. Peng Y, Chiriboga L, Yee H, Pei Z, Wang Z, and Lee P: Androgen receptor coactivator ARA70alpha and ARA70beta isoform-specific antibodies: new tools for studies of expression and immunohistochemical localization. Appl.Immunohistochem.Mol.Morphol. 16: 7-12 (2008)

119. Penney KL, Salinas CA, Pomerantz M, Schumacher FR, Beckwith CA, Lee GS, Oh WK, Sartor O, Ostrander EA, Kurth T, Ma J, Mucci L, Stanford JL, Kantoff PW, Hunter DJ, Stampfer MJ, and Freedman ML: Evaluation of 8q24 and 17q risk loci and prostate cancer mortality. Clin.Cancer Res. 15: 3223-3230 (2009)

120. Perner S, Mosquera JM, Demichelis F, Hofer MD, Paris PL, Simko J, Collins C, Bismar TA, Chinnaiyan AM, De Marzo AM, and Rubin MA: TMPRSS2-ERG fusion prostate cancer: an early molecular event associated with invasion. Am.J.Surg.Pathol. 31: 882-888 (2007)

121. Pflueger D, Rickman DS, Sboner A, Perner S, LaFargue CJ, Svensson MA, Moss BJ, Kitabayashi N, Pan Y, de la TA, Kuefer R, Tewari AK, Demichelis F, Chee MS, Gerstein MB, and Rubin MA: N-myc

downstream regulated gene 1 (NDRG1) is fused to ERG in prostate cancer. Neoplasia. 11: 804-811 (2009)

122. Pomerantz MM, Ahmadiyeh N, Jia L, Herman P, Verzi MP, Doddapaneni H, Beckwith CA, Chan JA, Hills A, Davis M, Yao K, Kehoe SM, Lenz HJ, Haiman CA, Yan C, Henderson BE, Frenkel B, Barretina J, Bass A, Tabernero J, Baselga J, Regan MM, Manak JR, Shivdasani R, Coetzee GA, and Freedman ML: The 8q24 cancer risk variant rs6983267 shows long-range interaction with MYC in colorectal cancer. Nat.Genet. 41: 882-884 (2009)

123. Pomerantz MM, Beckwith CA, Regan MM, Wyman SK, Petrovics G, Chen Y, Hawksworth DJ, Schumacher FR, Mucci L, Penney KL, Stampfer MJ, Chan JA, Ardlie KG, Fritz BR, Parkin RK, Lin DW, Dyke M, Herman P, Lee S, Oh WK, Kantoff PW, Tewari M, McLeod DG, Srivastava S, and Freedman ML: Evaluation of the 8q24 prostate cancer risk locus and MYC expression. Cancer Res. 69: 5568-5574 (2009)

124. Pomerantz MM, Shrestha Y, Flavin RJ, Regan MM, Penney KL, Mucci LA, Stampfer MJ, Hunter DJ, Chanock SJ, Schafer EJ, Chan JA, Tabernero J, Baselga J, Richardson AL, Loda M, Oh WK, Kantoff PW, Hahn WC, and Freedman ML: Analysis of the 10q11 cancer risk locus implicates MSMB and NCOA4 in human prostate tumorigenesis. PLoS.Genet. 6: e1001204 (2010)

125. Porter MP and Stanford JL: Obesity and the risk of prostate cancer. Prostate 62: 316-321 (2005)

126. Pritchett J, Athwal V, Roberts N, Hanley NA, and Hanley KP: Understanding the role of SOX9 in acquired diseases: lessons from development. Trends Mol.Med. 17: 166-174 (2011)

127. Rinckleb AE, Luedeke M, FitzGerald LM, Stanford JL, Schleutker J, Eeles RA, Teixeira MR, Cannon-Albright L, Ostrander EA, Weikert S., Hoegel J, Herkommer K, Wahlfors T, Visakorpi T, Leinonen KA, Tammela TL, Cooper CS, Kote-Jarai Z, Edwards S, Goh CL, McCarthy F., Parker C., Flohr P, Paulo P, Jerónimo C., Henrique R, Krause H., Schrader M, Vogel W, Kuefer R, Hofer MD, Perner S, Rubin MA, Agarwal A.M., Easton DF, Amin Al Olama A., Benlloch S, The PRACTICAL (Prostate Cancer Association Group to Investigate Cancer-Associated Alterations in the Genome) consortium, and Maier C: Prostate cancer risk regions in 8q24 and 17q24 are differentially associated with somatic TMPRSS2-ERG fusion status (Publication in preparation) (2014)

128. Rinckleb AE, Surowy HM, Luedeke M, Varga D, Schrader M, Hoegel J, Vogel W, and Maier C: The prostate cancer risk locus at 10q11 is associated with DNA repair capacity. DNA Repair (Amst) 11: 693-701 (2012)

129. Robert Koch-Institut: Krebs in Deutschland 2009/2010. 9. Ausgabe. Robert Koch-Institut (Hrsg) und die Gesellschaft der epidemiologischen Krebsregister in Deutschland e.V. (Hrsg). Berlin, (2013)

130. Rokman A, Ikonen T, Seppala EH, Nupponen N, Autio V, Mononen N, Bailey-Wilson J, Trent J, Carpten J, Matikainen MP, Koivisto PA, Tammela TL, Kallioniemi OP, and Schleutker J: Germline alterations of the RNASEL gene, a candidate HPC1 gene at 1q25, in patients and families with prostate cancer. Am.J.Hum.Genet. 70: 1299-1304 (2002)

131. Rouzier C, Haudebourg J, Carpentier X, Valerio L, Amiel J, Michiels JF, and Pedeutour F: Detection of the TMPRSS2-ETS fusion gene in prostate carcinomas: retrospective analysis of 55 formalin-fixed and paraffin-embedded samples with clinical data. Cancer Genet.Cytogenet. 183: 21-27 (2008)

132. San Filippo J, Sung P, and Klein H: Mechanism of eukaryotic homologous recombination. Annu.Rev.Biochem. 77: 229-257 (2008)

133. Saramaki OR, Harjula AE, Martikainen PM, Vessella RL, Tammela TL, and Visakorpi T: TMPRSS2:ERG fusion identifies a subgroup of prostate cancers with a favorable prognosis. Clin.Cancer Res. 14: 3395-3400 (2008)

134. Satoh MS and Lindahl T: Role of poly(ADP-ribose) formation in DNA repair. Nature 356: 356-358 (1992)

135. Schaefer G, Mosquera JM, Ramoner R, Park K, Romanel A, Steiner E, Horninger W, Bektic J, Ladurner-Rennau M, Rubin MA, Demichelis F, and Klocker H: Distinct ERG rearrangement prevalence in prostate cancer: higher frequency in young age and in low PSA prostate cancer. Prostate Cancer Prostatic.Dis. 16: 132-138 (2013)

136. Schaeffer EM, Marchionni L, Huang Z, Simons B, Blackman A, Yu W, Parmigiani G, and Berman DM: Androgen-induced programs for prostate epithelial growth and invasion arise in embryogenesis and are reactivated in cancer. Oncogene 27: 7180-7191 (2008)

137. Schaid DJ, McDonnell SK, Zarfas KE, Cunningham JM, Hebbring S, Thibodeau SN, Eeles RA, Easton DF, Foulkes WD, Simard J, Giles GG, Hopper JL, Mahle L, Moller P, Badzioch M, Bishop DT, Evans C, Edwards S, Meitz J, Bullock S, Hope Q, Guy M, Hsieh CL, Halpern J, Balise RR, Oakley-Girvan I, Whittemore AS, Xu J, Dimitrov L, Chang BL, Adams TS, Turner AR, Meyers DA, Friedrichsen DM, Deutsch K, Kolb S, Janer M, Hood L, Ostrander EA, Stanford JL, Ewing CM, Gielzak M, Isaacs SD, Walsh PC, Wiley KE, Isaacs WB, Lange EM, Ho LA, Beebe-Dimmer JL, Wood DP, Cooney KA, Seminara D, Ikonen T, Baffoe-Bonnie A, Fredriksson H, Matikainen MP, Tammela TL, Bailey-Wilson J, Schleutker J, Maier C, Herkommer K, Hoegel JJ, Vogel W, Paiss T, Wiklund F, Emanuelsson M, Stenman E, Jonsson BA, Gronberg H, Camp NJ, Farnham J, Cannon-Albright LA, Catalona WJ, Suarez BK, and Roehl KA: Pooled genome linkage scan of aggressive prostate cancer: results from the International Consortium for Prostate Cancer Genetics. Hum.Genet. 120: 471-485 (2006)

138. Schreiber V, Hunting D, Trucco C, Gowans B, Grunwald D, de Murcia G, and De Murcia JM: A dominant-negative mutant of human poly(ADP-ribose) polymerase affects cell recovery, apoptosis, and sister chromatid exchange following DNA damage. Proc.Natl.Acad.Sci.U.S.A 92: 4753-4757 (1995)

139. Schumacher FR, Berndt SI, Siddiq A, Jacobs KB, Wang Z, Lindstrom S, Stevens VL, Chen C, Mondul AM, Travis RC, Stram DO, Eeles RA, Easton DF, Giles G, Hopper JL, Neal DE, Hamdy FC, Donovan JL, Muir K, Al Olama AA, Kote-Jarai Z, Guy M, Severi G, Gronberg H, Isaacs WB, Karlsson R, Wiklund F, Xu J, Allen NE, Andriole GL, Barricarte A, Boeing H, Bueno-de-Mesquita HB, Crawford ED, Diver WR, Gonzalez CA, Gaziano JM, Giovannucci EL, Johansson M, Le Marchand L, Ma J, Sieri S, Stattin P, Stampfer MJ, Tjonneland A, Vineis P, Virtamo J, Vogel U, Weinstein SJ, Yeager M, Thun MJ, Kolonel LN, Henderson BE, Albanes D, Hayes RB, Feigelson HS, Riboli E, Hunter DJ, Chanock SJ, Haiman CA, and Kraft P: Genome-wide association study identifies new prostate cancer susceptibility loci. Hum.Mol.Genet. 20: 3867-3875 (2011)

140. Schwanhausser B, Busse D, Li N, Dittmar G, Schuchhardt J, Wolf J, Chen W, and Selbach M: Global quantification of mammalian gene expression control. Nature 473: 337-342 (2011)

141. Seppala EH, Ikonen T, Mononen N, Autio V, Rokman A, Matikainen MP, Tammela TL, and Schleutker J: CHEK2 variants associate with hereditary prostate cancer. Br.J.Cancer 89: 1966-1970 (2003)

142. Shibata A, Conrad S, Birraux J, Geuting V, Barton O, Ismail A, Kakarougkas A, Meek K, Taucher-Scholz G, Lobrich M, and Jeggo PA: Factors determining DNA double-strand break repair pathway choice in G2 phase. EMBO J. 30: 1079-1092 (2011)

143. Sobin L.H., Gospodarowicz M.K., and Wittekind C.: TNM Classification of Malignant Tumours 7th (2009)

144. Sole X, Hernandez P, de Heredia ML, Armengol L, Rodriguez-Santiago B, Gomez L, Maxwell CA, Aguilo F, Condom E, Abril J, Perez-Jurado L, Estivill X, Nunes V, Capella G, Gruber SB, Moreno V, and Pujana MA: Genetic and genomic analysis modeling of germline c-MYC overexpression and cancer susceptibility. BMC.Genomics 9: 12 (2008)

145. Sotelo J, Esposito D, Duhagon MA, Banfield K, Mehalko J, Liao H, Stephens RM, Harris TJ, Munroe DJ, and Wu X: Long-range enhancers on 8q24 regulate c-Myc. Proc.Natl.Acad.Sci.U.S.A 107: 3001-3005 (2010)

146. Stanford JL, Wicklund KG, McKnight B, Daling JR, and Brawer MK: Vasectomy and risk of prostate cancer. Cancer Epidemiol.Biomarkers Prev. 8: 881-886 (1999)

147. Summersgill B, Clark J, and Shipley J: Fluorescence and chromogenic in situ hybridization to detect genetic aberrations in formalin-fixed paraffin embedded material, including tissue microarrays. Nat.Protoc. 3: 220-234 (2008)

148. Sun C, Dobi A, Mohamed A, Li H, Thangapazham RL, Furusato B, Shaheduzzaman S, Tan SH, Vaidyanathan G, Whitman E, Hawksworth DJ, Chen Y, Nau M, Patel V, Vahey M, Gutkind JS, Sreenath T, Petrovics G, Sesterhenn IA, McLeod DG, and Srivastava S: TMPRSS2-ERG fusion, a common genomic alteration in prostate cancer activates C-MYC and abrogates prostate epithelial differentiation. Oncogene 27: 5348-5353 (2008)

149. Sun J, Chang BL, Isaacs SD, Wiley KE, Wiklund F, Stattin P, Duggan D, Carpten JD, Trock BJ, Partin AW, Walsh PC, Gronberg H, Xu J, Isaacs WB, and Zheng SL: Cumulative effect of five genetic variants on prostate cancer risk in multiple study populations. Prostate 68: 1257-1262 (2008)

150. Sun J, Zheng SL, Wiklund F, Isaacs SD, Purcell LD, Gao Z, Hsu FC, Kim ST, Liu W, Zhu Y, Stattin P, Adami HO, Wiley KE, Dimitrov L, Sun J, Li T, Turner AR, Adams TS, Adolfsson J, Johansson JE, Lowey J, Trock BJ, Partin AW, Walsh PC, Trent JM, Duggan D, Carpten J, Chang BL, Gronberg H, Isaacs WB, and Xu J: Evidence for two independent prostate cancer risk-associated loci in the HNF1B gene at 17q12. Nat.Genet. 40: 1153-1155 (2008)

151. Surowy H: Characterization of cellular DNA repair capacity by molecular and cytogenic approaches: Suitability as intermediate phenotype of cancer. Dissertation (Dr.biol.hum), Universität Ulm (2011)

152. Takata R, Akamatsu S, Kubo M, Takahashi A, Hosono N, Kawaguchi T, Tsunoda T, Inazawa J, Kamatani N, Ogawa O, Fujioka T, Nakamura Y, and Nakagawa H: Genome-wide association study identifies five new susceptibility loci for prostate cancer in the Japanese population. Nat.Genet. 42: 751-754 (2010)

153. Tavtigian SV, Simard J, Teng DH, Abtin V, Baumgard M, Beck A, Camp NJ, Carillo AR, Chen Y, Dayananth P, Desrochers M, Dumont M, Farnham JM, Frank D, Frye C, Ghaffari S, Gupte JS, Hu R, Iliev D, Janecki T, Kort EN, Laity KE, Leavitt A, Leblanc G, McArthur-Morrison J, Pederson A, Penn B, Peterson KT, Reid JE, Richards S, Schroeder M, Smith R, Snyder SC, Swedlund B, Swensen J, Thomas A, Tranchant M, Woodland AM, Labrie F, Skolnick MH, Neuhausen S, Rommens J, and Cannon-Albright LA: A candidate prostate cancer susceptibility gene at chromosome 17p. Nat.Genet. 27: 172-180 (2001)

154. Teerlink CC, Thibodeau SN, McDonnell SK, Schaid DJ, Rinckleb A, Maier C, Vogel W, Cancel-Tassin G, Egrot C, Cussenot O, Foulkes WD, Giles GG, Hopper JL, Severi G, Eeles R, Easton D, Kote-Jarai Z, Guy M, Cooney KA, Ray AM, Zuhlke KA, Lange EM, FitzGerald LM, Stanford JL, Ostrander EA, Wiley KE, Isaacs SD, Walsh PC, Isaacs WB, Wahlfors T, Tammela T, Schleutker J, Wiklund F, Gronberg H, Emanuelsson M, Carpten J, Bailey-Wilson J, Whittemore AS, Oakley-Girvan I, Hsieh CL, Catalona WJ, Zheng SL, Jin G, Lu L, Xu J, Camp NJ, and Cannon-Albright LA: Association analysis of 9,560 prostate cancer cases from the International Consortium of Prostate Cancer Genetics confirms the role of reported prostate cancer associated SNPs for familial disease. Hum.Genet. (2013)

155. The Breast Cancer Linkage Consortium: Cancer risks in BRCA2 mutation carriers. J.Natl.Cancer Inst. 91: 1310-1316 (1999)

156. Thomas G, Jacobs KB, Yeager M, Kraft P, Wacholder S, Orr N, Yu K, Chatterjee N, Welch R, Hutchinson A, Crenshaw A, Cancel-Tassin G, Staats BJ, Wang Z, Gonzalez-Bosquet J, Fang J, Deng X, Berndt SI, Calle EE, Feigelson HS, Thun MJ, Rodriguez C, Albanes D, Virtamo J, Weinstein S, Schumacher FR, Giovannucci E, Willett WC, Cussenot O, Valeri A, Andriole GL, Crawford ED, Tucker M, Gerhard DS, Fraumeni JF, Jr., Hoover R, Hayes RB, Hunter DJ, and Chanock SJ: Multiple loci identified in a genome-wide association study of prostate cancer. Nat.Genet. 40: 310-315 (2008)

157. Thomsen MK, Ambroisine L, Wynn S, Cheah KS, Foster CS, Fisher G, Berney DM, Moller H, Reuter VE, Scardino P, Cuzick J, Ragavan N, Singh PB, Martin FL, Butler CM, Cooper CS, and Swain A: SOX9

elevation in the prostate promotes proliferation and cooperates with PTEN loss to drive tumor formation. Cancer Res. 70: 979-987 (2010)

158. Thomsen MK, Butler CM, Shen MM, and Swain A: Sox9 is required for prostate development. Dev.Biol. 316: 302-311 (2008)

159. Thomsen MK, Francis JC, and Swain A: The role of Sox9 in prostate development. Differentiation 76: 728-735 (2008)

160. Tomlins SA, Mehra R, Rhodes DR, Smith LR, Roulston D, Helgeson BE, Cao X, Wei JT, Rubin MA, Shah RB, and Chinnaiyan AM: TMPRSS2:ETV4 gene fusions define a third molecular subtype of prostate cancer. Cancer Res. 66: 3396-3400 (2006)

161. Tomlins SA, Rhodes DR, Perner S, Dhanasekaran SM, Mehra R, Sun XW, Varambally S, Cao X, Tchinda J, Kuefer R, Lee C, Montie JE, Shah RB, Pienta KJ, Rubin MA, and Chinnaiyan AM: Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. Science 310: 644-648 (2005)

162. Tsilidis KK, Travis RC, Appleby PN, Allen NE, Lindstrom S, Schumacher FR, Cox D, Hsing AW, Ma J, Severi G, Albanes D, Virtamo J, Boeing H, Bueno-de-Mesquita HB, Johansson M, Quiros JR, Riboli E, Siddiq A, Tjonneland A, Trichopoulos D, Tumino R, Gaziano JM, Giovannucci E, Hunter DJ, Kraft P, Stampfer MJ, Giles GG, Andriole GL, Berndt SI, Chanock SJ, Hayes RB, and Key TJ: Interactions between genome-wide significant genetic variants and circulating concentrations of insulin-like growth factor 1, sex hormones, and binding proteins in relation to prostate cancer risk in the National Cancer Institute Breast and Prostate Cancer Cohort Consortium. Am.J.Epidemiol. 175: 926-935 (2012)

163. Tsurusaki T, Koji T, Sakai H, Kanetake H, Nakane PK, and Saito Y: Cellular expression of beta-microseminoprotein (beta-MSP) mRNA and its protein in untreated prostate cancer. Prostate 35: 109-116 (1998)

164. Tu JJ, Rohan S, Kao J, Kitabayashi N, Mathew S, and Chen YT: Gene fusions between TMPRSS2 and ETS family genes in prostate cancer: frequency and transcript variant analysis by RT-PCR and FISH on paraffin-embedded tissues. Mod.Pathol. 20: 921-928 (2007)

165. Uchiumi F, Sakakibara G, Sato J, and Tanuma S: Characterization of the promoter region of the human PARG gene and its response to PU.1 during differentiation of HL-60 cells. Genes Cells 13: 1229-1247 (2008)

166. Uchiumi F, Watanabe T, Ohta R, Abe H, and Tanuma S: PARP1 gene expression is downregulated by knockdown of PARG gene. Oncol.Rep. 29: 1683-1688 (2013)

167. van Gils CH, Bostick RM, Stern MC, and Taylor JA: Differences in base excision repair capacity may modulate the effect of dietary antioxidant intake on prostate cancer risk: an example of polymorphisms in the XRCC1 gene. Cancer Epidemiol.Biomarkers Prev. 11: 1279-1284 (2002)

168. Vanaja DK, Cheville JC, Iturria SJ, and Young CY: Transcriptional silencing of zinc finger protein 185 identified by expression profiling is associated with prostate cancer progression. Cancer Res. 63: 3877-3882 (2003)

169. Varambally S, Dhanasekaran SM, Zhou M, Barrette TR, Kumar-Sinha C, Sanda MG, Ghosh D, Pienta KJ, Sewalt RG, Otte AP, Rubin MA, and Chinnaiyan AM: The polycomb group protein EZH2 is involved in progression of prostate cancer. Nature 419: 624-629 (2002)

170. Varga D, Hoegel J, Maier C, Jainta S, Hoehne M, Patino-Garcia B, Michel I, Schwarz-Boeger U, Kiechle M, Kreienberg R, and Vogel W: On the difference of micronucleus frequencies in peripheral blood lymphocytes between breast cancer patients and controls. Mutagenesis 21: 313-320 (2006)

171. Varga D, Michel I, Patino-Garcia B, Paiss T, Vogel W, and Maier C: Radiosensitivity detected by the micronucleus test is not generally increased in sporadic prostate cancer patients. Cytogenet.Genome Res. 111: 41-45 (2005)

172. Wagner T, Wirth J, Meyer J, Zabel B, Held M, Zimmer J, Pasantes J, Bricarelli FD, Keutel J, Hustert E, Wolf U, Tommerup N, Schempp W, and Scherer G: Autosomal sex reversal and campomelic dysplasia are caused by mutations in and around the SRY-related gene SOX9. Cell 79: 1111-1120 (1994)

173. Wang H, Leav I, Ibaragi S, Wegner M, Hu GF, Lu ML, Balk SP, and Yuan X: SOX9 is expressed in human fetal prostate epithelium and enhances prostate cancer invasion. Cancer Res. 68: 1625-1630 (2008)

174. Wang H, McKnight NC, Zhang T, Lu ML, Balk SP, and Yuan X: SOX9 is expressed in normal prostate basal cells and regulates androgen receptor expression in prostate cancer cells. Cancer Res. 67: 528-536 (2007)

175. Wang L, McDonnell SK, Elkins DA, Slager SL, Christensen E, Marks AF, Cunningham JM, Peterson BJ, Jacobsen SJ, Cerhan JR, Blute ML, Schaid DJ, and Thibodeau SN: Role of HPC2/ELAC2 in hereditary prostate cancer. Cancer Res. 61: 6494-6499 (2001)

176. Wasserman NF, Aneas I, and Nobrega MA: An 8q24 gene desert variant associated with prostate cancer risk confers differential in vivo activity to a MYC enhancer. Genome Res. 20: 1191-1197 (2010)

177. Wei W, Ba Z, Gao M, Wu Y, Ma Y, Amiard S, White CI, Rendtlew Danielsen JM, Yang YG, and Qi Y: A role for small RNAs in DNA double-strand break repair. Cell 149: 101-112 (2012)

178. Weischenfeldt J, Simon R, Feuerbach L, Schlangen K, Weichenhan D, Minner S, Wuttig D, Warnatz HJ, Stehr H, Rausch T, Jager N, Gu L, Bogatyrova O, Stutz AM, Claus R, Eils J, Eils R, Gerhauser C, Huang PH, Hutter B, Kabbe R, Lawerenz C, Radomski S, Bartholomae CC, Falth M, Gade S, Schmidt M, Amschler N, Hass T, Galal R, Gjoni J, Kuner R, Baer C, Masser S, von Kalle C, Zichner T, Benes V, Raeder B, Mader M, Amstislavskiy V, Avci M, Lehrach H, Parkhomchuk D, Sultan M, Burkhardt L, Graefen M, Huland H, Kluth M, Krohn A, Sirma H, Stumm L, Steurer S, Grupp K, Sultmann H, Sauter G, Plass C, Brors B, Yaspo ML, Korbel JO, and Schlomm T: Integrative genomic analyses reveal an androgen-driven somatic alteration landscape in early-onset prostate cancer. Cancer Cell 23: 159-170 (2013)

179. Wright JL, Chéry L, Holt S, Lin DW, Luedeke M, Rinckleb AE, Maier C, Stanford JL. Aspirin use is associated with a reduced risk of *TMPRSS2:ERG* fusion positive prostate cancer. (Publicaton in preparation)

180. Xu J, Dimitrov L, Chang BL, Adams TS, Turner AR, Meyers DA, Eeles RA, Easton DF, Foulkes WD, Simard J, Giles GG, Hopper JL, Mahle L, Moller P, Bishop T, Evans C, Edwards S, Meitz J, Bullock S, Hope Q, Hsieh CL, Halpern J, Balise RN, Oakley-Girvan I, Whittemore AS, Ewing CM, Gielzak M, Isaacs SD, Walsh PC, Wiley KE, Isaacs WB, Thibodeau SN, McDonnell SK, Cunningham JM, Zarfas KE, Hebbring S, Schaid DJ, Friedrichsen DM, Deutsch K, Kolb S, Badzioch M, Jarvik GP, Janer M, Hood L, Ostrander EA, Stanford JL, Lange EM, Beebe-Dimmer JL, Mohai CE, Cooney KA, Ikonen T, Baffoe-Bonnie A, Fredriksson H, Matikainen MP, Tammela TL, Bailey-Wilson J, Schleutker J, Maier C, Herkommer K, Hoegel JJ, Vogel W, Paiss T, Wiklund F, Emanuelsson M, Stenman E, Jonsson BA, Gronberg H, Camp NJ, Farnham J, Cannon-Albright LA, and Seminara D: A combined genomewide linkage scan of 1,233 families for prostate cancer-susceptibility genes conducted by the international consortium for prostate cancer genetics. Am.J.Hum.Genet. 77: 219-229 (2005)

181. Xu J, Lange EM, Lu L, Zheng SL, Wang Z, Thibodeau SN, Cannon-Albright LA, Teerlink CC, Camp NJ, Johnson AM, Zuhlke KA, Stanford JL, Ostrander EA, Wiley KE, Isaacs SD, Walsh PC, Maier C, Luedeke M, Vogel W, Schleutker J, Wahlfors T, Tammela T, Schaid D, McDonnell SK, DeRycke MS, Cancel-Tassin G, Cussenot O, Wiklund F, Gronberg H, Eeles R, Easton D, Kote-Jarai Z, Whittemore AS, Hsieh CL, Giles GG, Hopper JL, Severi G, Catalona WJ, Mandal D, Ledet E, Foulkes WD, Hamel N, Mahle L, Moller P, Powell I, Bailey-Wilson JE, Carpten JD, Seminara D, Cooney KA, and Isaacs WB: HOXB13 is a

susceptibility gene for prostate cancer: results from the International Consortium for Prostate Cancer Genetics (ICPCG). Hum.Genet. 132: 5-14 (2013)

182. Xu J, Mo Z, Ye D, Wang M, Liu F, Jin G, Xu C, Wang X, Shao Q, Chen Z, Tao Z, Qi J, Zhou F, Wang Z, Fu Y, He D, Wei Q, Guo J, Wu D, Gao X, Yuan J, Wang G, Xu Y, Wang G, Yao H, Dong P, Jiao Y, Shen M, Yang J, Ou-Yang J, Jiang H, Zhu Y, Ren S, Zhang Z, Yin C, Gao X, Dai B, Hu Z, Yang Y, Wu Q, Chen H, Peng P, Zheng Y, Zheng X, Xiang Y, Long J, Gong J, Na R, Lin X, Yu H, Wang Z, Tao S, Feng J, Sun J, Liu W, Hsing A, Rao J, Ding Q, Wiklund F, Gronberg H, Shu XO, Zheng W, Shen H, Jin L, Shi R, Lu D, Zhang X, Sun J, Zheng SL, and Sun Y: Genome-wide association study in Chinese men identifies two new prostate cancer risk loci at 9q31.2 and 19q13.4. Nat.Genet. 44: 1231-1235 (2012)

183. Xu J, Zheng SL, Komiya A, Mychaleckyj JC, Isaacs SD, Chang B, Turner AR, Ewing CM, Wiley KE, Hawkins GA, Bleecker ER, Walsh PC, Meyers DA, and Isaacs WB: Common sequence variants of the macrophage scavenger receptor 1 gene are associated with prostate cancer risk. Am.J.Hum.Genet. 72: 208-212 (2003)

184. Xu J, Zheng SL, Turner A, Isaacs SD, Wiley KE, Hawkins GA, Chang BL, Bleecker ER, Walsh PC, Meyers DA, and Isaacs WB: Associations between hOGG1 sequence variants and prostate cancer susceptibility. Cancer Res. 62: 2253-2257 (2002)

185. Xu X, Valtonen-Andre C, Savblom C, Hallden C, Lilja H, and Klein RJ: Polymorphisms at the Microseminoprotein-beta locus associated with physiologic variation in beta-microseminoprotein and prostate-specific antigen levels. Cancer Epidemiol.Biomarkers Prev. 19: 2035-2042 (2010)

186. Yang GY, Liao J, Kim K, Yurkow EJ, and Yang CS: Inhibition of growth and induction of apoptosis in human cancer cell lines by tea polyphenols. Carcinogenesis 19: 611-616 (1998)

187. Yeager M, Orr N, Hayes RB, Jacobs KB, Kraft P, Wacholder S, Minichiello MJ, Fearnhead P, Yu K, Chatterjee N, Wang Z, Welch R, Staats BJ, Calle EE, Feigelson HS, Thun MJ, Rodriguez C, Albanes D, Virtamo J, Weinstein S, Schumacher FR, Giovannucci E, Willett WC, Cancel-Tassin G, Cussenot O, Valeri A, Andriole GL, Gelmann EP, Tucker M, Gerhard DS, Fraumeni JF, Jr., Hoover R, Hunter DJ, Chanock SJ, and Thomas G: Genome-wide association study of prostate cancer identifies a second risk locus at 8q24. Nat.Genet. 39: 645-649 (2007)

188. Yeh S and Chang C: Cloning and characterization of a specific coactivator, ARA70, for the androgen receptor in human prostate cells. Proc.Natl.Acad.Sci.U.S.A 93: 5517-5521 (1996)

189. Zhang X, Cowper-Sal lR, Bailey SD, Moore JH, and Lupien M: Integrative functional genomics identifies an enhancer looping to the SOX9 gene disrupted by the 17q24.3 prostate cancer risk locus. Genome Res. 22: 1437-1446 (2012)

190. Zheng SL, Sun J, Wiklund F, Smith S, Stattin P, Li G, Adami HO, Hsu FC, Zhu Y, Balter K, Kader AK, Turner AR, Liu W, Bleecker ER, Meyers DA, Duggan D, Carpten JD, Chang BL, Isaacs WB, Xu J, and Gronberg H: Cumulative association of five genetic variants with prostate cancer. N.Engl.J.Med. 358: 910-919 (2008)

Appendix

Appendix A: Gene symbols according to the HUGO Gene Nomenclature Committee

ACSL3	acyl-CoA synthetase long-chain family member 3
AFM	afamin
AGO2	argonaute RISC catalytic component 2
ALAS1	aminolevulinate, delta-, synthase 1
AR	androgen receptor
ARMC2	armadillo repeat containing 2
ATM	ataxia telangiectasia mutated
ATR	ataxia telangiectasia and Rad3 related
BOD1	biorientation of chromosomes in cell division 1
BRCA1	breast cancer 1, early onset
BRCA2	breast cancer 2, early onset
C2orf43	chromosome 2 open reading frame 43
CABLES2	Cdk5 and Abl enzyme substrate 2
CANT1	calcium activated nucleotidase 1
CHEK2	checkpoint kinase 2
CCHCR1	coiled-coil alpha-helical rod protein 1
CLDN11	claudin 11
CTBP2	C-terminal binding protein 2
DDX5	DEAD (Asp-Glu-Ala-Asp) box helicase 5
EBF2	early B-cell factor 2
EEFSEC	eukaryotic elongation factor, selenocysteine-tRNA-specific
EHBP1	EH domain binding protein 1
ELAC2	elaC ribonuclease Z 2
ERG	v-ets avian erythroblastosis virus E26 oncogene homolog
ESCO1	establishment of sister chromatid cohesion N-acetyltransferase 1
ETS1	v-ets avian erythroblastosis virus E26 oncogene homolog 1
ETV1	ets variant 2
ETV4	ets variant 4
ETV5	ets variant 5
EZH2	enhancer of zeste homolog 2 (Drosophila)
FAM57A	family with sequence similarity 57, member A
*FAM111*A	family with sequence similarity 111, member A
FARP2	FERM, RhoGEF and pleckstrin domain protein 2
FERMT2	fermitin family member 2
FGF10	fibroblast growth factor 10
FOXP1	forkhead box P1
FOXP4	forkhead box P4
GATA5	GATA binding protein 5
G6PD	glucose-6-phosphate dehydrogenase
GGCX	gamma-glutamyl carboxylase
GOLM1	golgi membrane protein 1
GPRC6A	G protein-coupled receptor, family C, group 6, member A
GRHL1	grainyhead-like 1 (Drosophila)
GSPT2	G1 to S phase transition 2
HERPUD1	homocysteine-inducible, endoplasmic reticulum stress-inducible, ubiquitin-like domain member 1
HMGN2P46	high mobility group nucleosomal binding domain 2 pseudogene 46
HNF1B	HNF1 homeobox B
HNRPA2B1	heterogeneous nuclear ribonucleoprotein A2/B1
HOXB13	homeobox B13
HPN	hepsin
IRX4	iroquois homeobox 4
ITGA6	integrin, alpha 6
JAZF1	JAZF zinc finger 1
KLF4	Kruppel-like factor 4 (gut)
KLK2	kallikrein-related peptidase 2
KLK3	kallikrein-related peptidase 3
KRT8	keratin 8
LILRA3	leukocyte immunoglobulin-like receptor, subfamily A (without TM domain), member 3
LMTK2	lemur tyrosine kinase 2
MAGED1	melanoma antigen family D, 1
MDM4	Mdm4 p53 binding protein homolog (mouse)

MLH1	mutL homolog 1
MLPH	melanophilin
MMP7	matrix metallopeptidase 7 (matrilysin, uterine)
MSH2	mutS homolog 2
MSH6	mutS homolog 6
MSMB	microseminoprotein, beta-
MSR1	macrophage scavenger receptor 1
MYC	v-myc avian myelocytomatosis viral oncogene homolog
NCOA4	nuclear receptor coactivator 4
NDRG1	N-myc downstream regulated 1
NKX3-1	NK3 homeobox 1
NOTCH4	notch 4
NUDT10	nudix (nucleoside diphosphate linked moiety X)-type motif 10
NUDT11	nudix (nucleoside diphosphate linked moiety X)-type motif 11
OGG1	8-oxoguanine DNA glycosylase
OR51E2	olfactory receptor, family 51, subfamily E, member 2
PARG	poly (ADP-ribose) glycohydrolase
PARP1	poly (ADP-ribose) polymerase 1
PCA3	prostate cancer antigen 3 (non-protein coding)
PDLIM5	PDZ and LIM domain 5
PMS2	PMS2 postmeiotic segregation increased 2 (S. cerevisiae)
PRAC1	prostate cancer susceptibility candidate 1
PRLH	prolactin releasing hormone
PRPH	peripherin
PTEN	phosphatase and tensin homolog
RAB17	RAB17, member RAS oncogene family
RAD23B	*RAD23* homolog B (S. cerevisiae)
RAD51B	RAD51 paralog B
RASSF6	Ras association (RalGDS/AF-6) domain family member 6
RFX6	regulatory factor X, 6
RNASEL	ribonuclease L (2',5'-oligoisoadenylate synthetase-dependent)
RNF181	ring finger protein 181
SALL3	spalt-like transcription factor 3
SESN1	sestrin 1
SHROOM2	shroom family member 2
SIDT1	SID1 transmembrane family, member 1
SKIL	SKI-like oncogene
SLC22A3	solute carrier family 22 (organic cation transporter), member 3
SLC25A37	solute carrier family 25 (mitochondrial iron transporter), member 37
SLC45A3	solute carrier family 45, member 3
SOX9	SRY (sex determining region Y)-box 9
SP8	Sp8 transcription factor
SPINK1	serine peptidase inhibitor, Kazal type 1
SPOP	speckle-type POZ protein
TAF1B	TATA box binding protein (TBP)-associated factor, RNA polymerase I, B, 63kDa
TBX5	T-box 5
TERT	telomerase reverse transcriptase
TET2	tet methylcytosine dioxygenase 2
THADA	thyroid adenoma associated
TIMM23	translocase of inner mitochondrial membrane 23 homolog (yeast)
TIMM23B	translocase of inner mitochondrial membrane 23 homolog B (yeast)
TMPRSS2	Transmembranprotease Serine 2
TRIM8	tripartite motif containing 8
TTLL1	tubulin tyrosine ligase-like family, member 1
TUBA1C	tubulin, alpha 1c
UBTF	upstream binding transcription factor, RNA polymerase I
VAMP5	vesicle-associated membrane protein 5
VAMP8	vesicle-associated membrane protein 8
VPS53	vacuolar protein sorting 53 homolog (S. cerevisiae)
XRCC4	X-ray repair complementing defective repair in Chinese hamster cells 4
ZBTB38	zinc finger and BTB domain containing 38
ZGPAT	zinc finger, CCCH-type with G patch domain
ZNF652	zinc finger protein 652
ZRANB1	zinc finger, RAN-binding domain containing 1

Appendix B: The classifier for counting binucleated lymphocytes and micronuclei for the micronucleus assay using Metafer MSearch 4 software, according to Harald Surowy [151].

Classifier Name:	STC
Classifier Description:	short-time whole-blood cultures

Nuclei parameters

Nuclei: Image Processing Operations:	Sharpen(3,4)
Nuclei: Object Threshold (in %):	18
Nuclei: Minimum Area (in 1/100 µm²):	6000
Nuclei: Maximum Area (in 1/100 µm²):	50000
Nuclei: Maximum Relative Concavity Depth (in 1/1000):	90
Nuclei: Maximum Aspect Ratio (in 1/1000):	1500
Nuclei: Maximum Distance (in 1/10 µm):	250
Nuclei: Maximum Area Asymmetry (in %):	90
Nuclei: Region of Interest Radius (in 1/10 µm):	300
Nuclei: Maximum Object Area in ROI (in 1/100 µm²):	5000

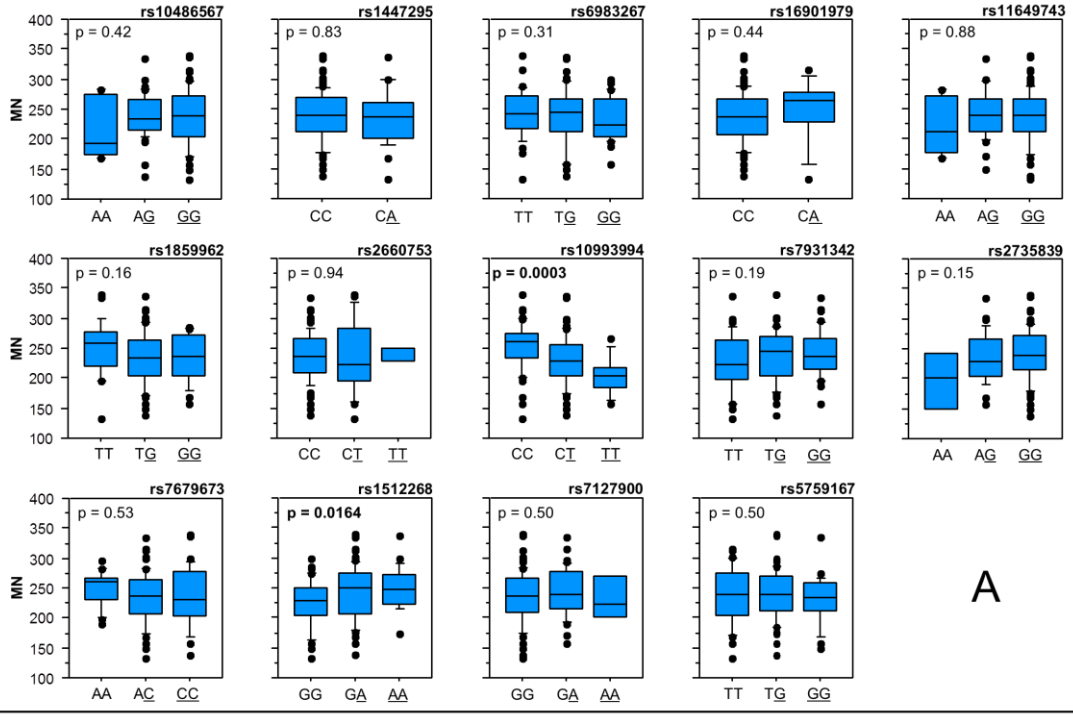
Micronuclei parameters

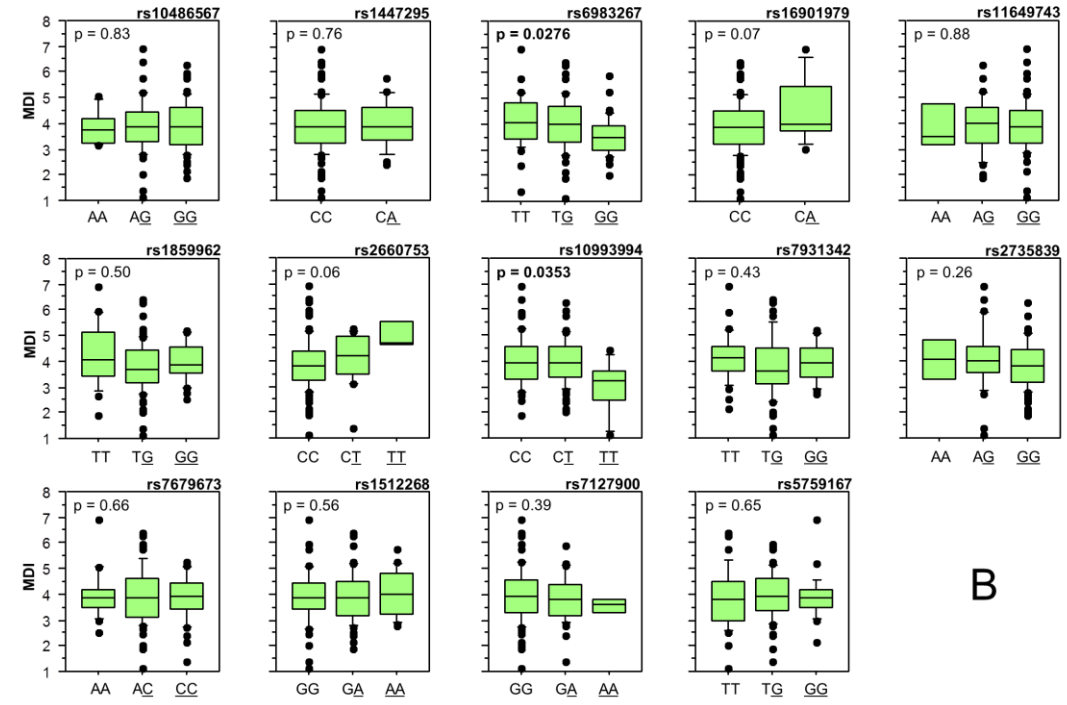
Micronuclei: Image Processing Operations:	MedianV(3) MedianH(3) Average(3,1) Sharpen(5,5)
Micronuclei: Object Threshold (in %):	5
Micronuclei: Minimum Area (in 1/100 µm²):	400
Micronuclei: Maximum Area (in 1/100 µm²):	6000
Micronuclei: Maximum Relative Concavity Depth (in 1/1000):	1000
Micronuclei: Maximum Aspect Ratio (in 1/1000):	1720
Micronuclei: Maximum Distance (in 1/10 µm):	400

Appendix C: Cumulative number of risk alleles (RA) of the twelve investigated SNPs among the prostate cancer case and control groups.

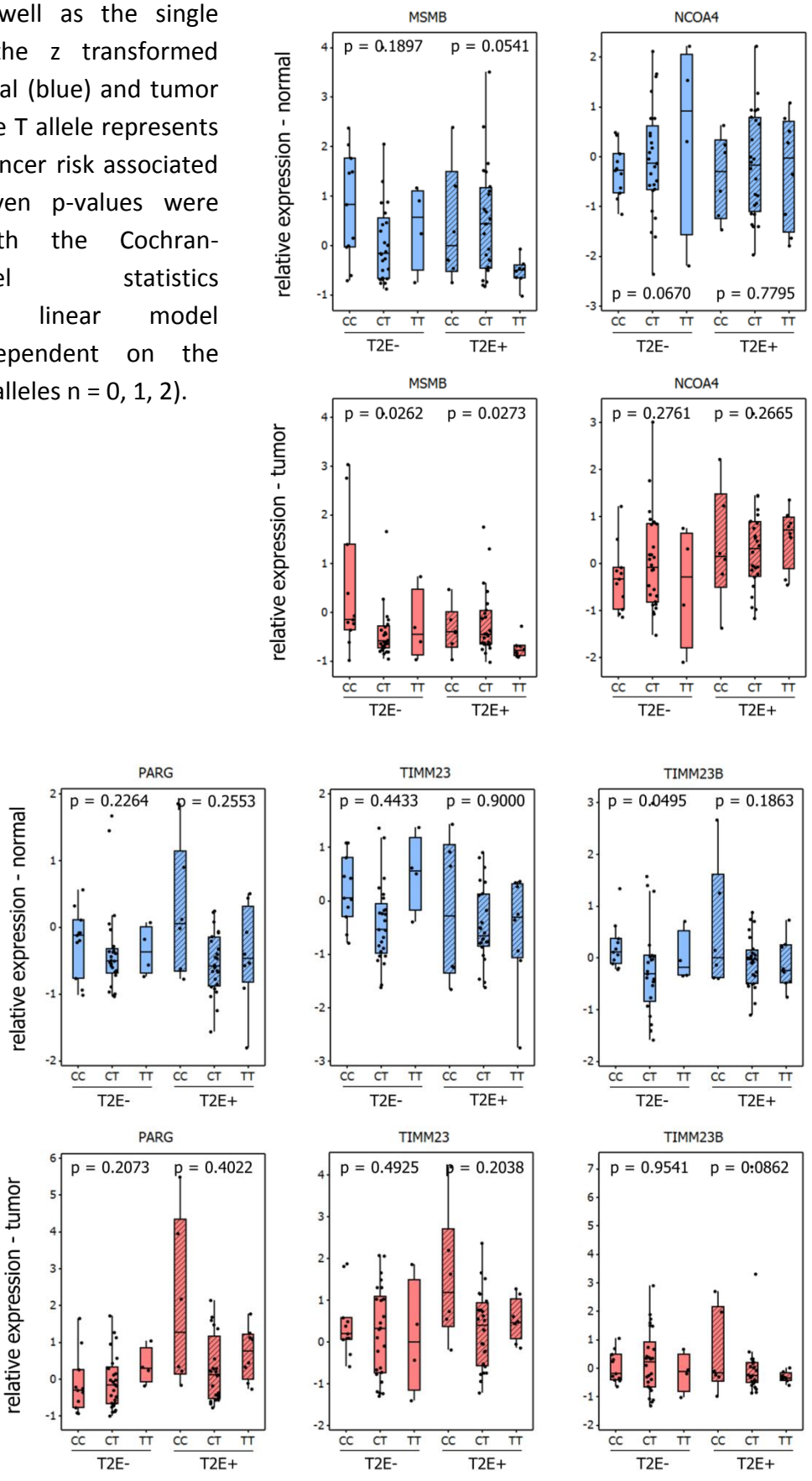
Number of risk alleles	Controls (n = 497)	All cases (n =679)	Familiar cases (n = 371)	Sporadic cases (n = 308)
0	0	0	0	0
1	0	0	0	0
2	0	0	0	0
3	0	0	0	0
4	1	2	1	1
5	3	2	2	0
6	15	4	2	2
7	37	16	9	7
8	63	62	27	35
9	92	89	48	41
10	101	99	54	45
11	81	130	76	54
12	57	121	56	65
13	28	74	50	24
14	13	48	29	19
15	4	21	10	11
16	0	9	5	4
17	1	2	2	0
18	0	0	0	0
19	0	0	0	0
20	1	0	0	0
21	0	0	0	0
22	0	0	0	0
23	0	0	0	0
24	0	0	0	0

Appendix D: Boxplots of the micronucleus test (A) and mitotic delay assay (B) results. (modified from [128]) Given are the micronucleus (MN) frequencies after damage induction with γ-irradiation and the mitotic delay indices (MDI) in n = 128 healthy probands dependent on the genotypes of 14 common prostate cancer risk variants. The particular risk alleles are underlined. The p-values were calculated under an additive model with linear regression.

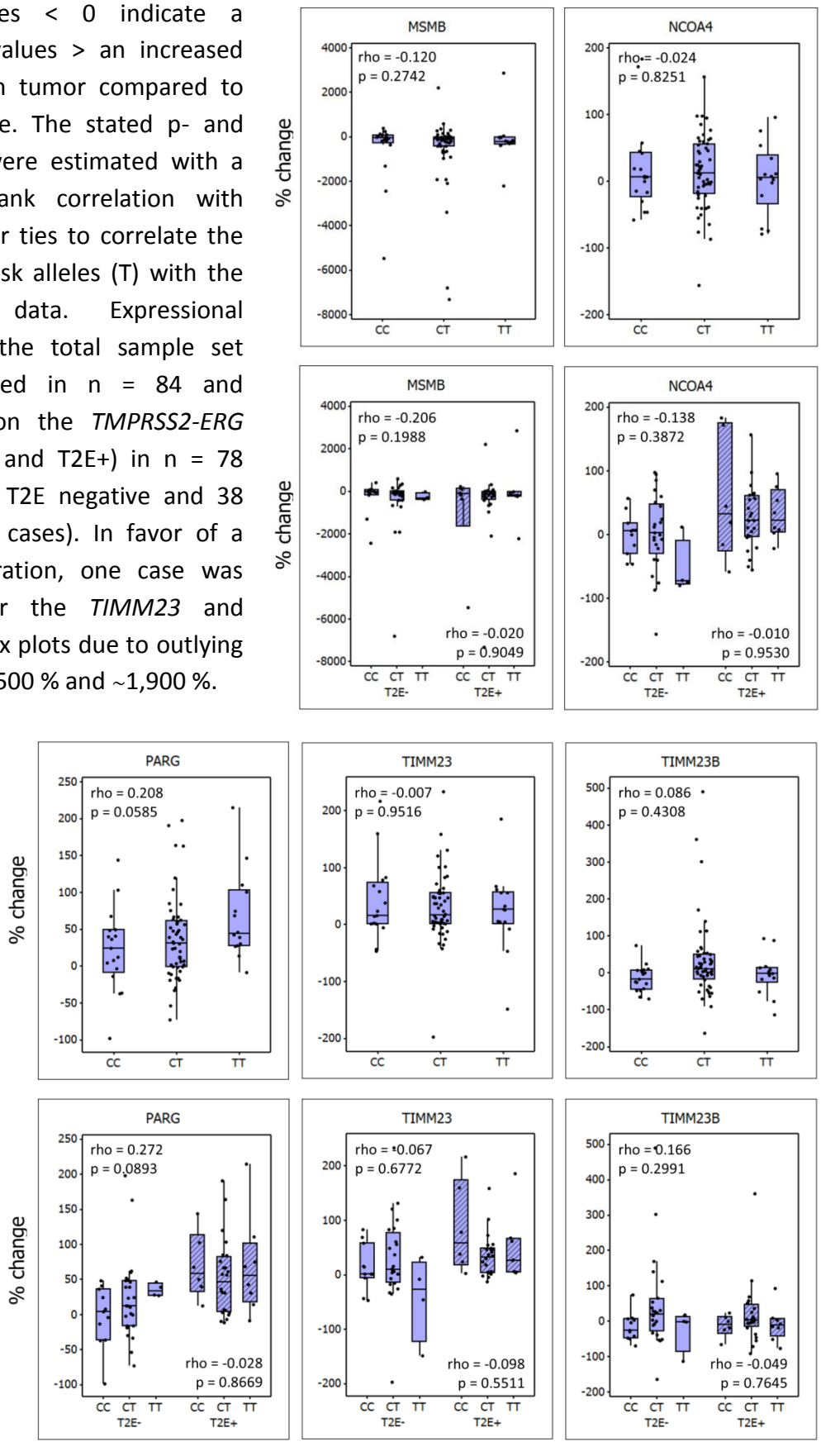




Appendix E: Relative expression of the 10q11 region genes dependent on the *TMPRSS2-ERG* (T2E) fusion status and the rs10993994 genotype in prostate normal- and tumor tissue of 87 patients. Shown are the box plots as well as the single values from the z transformed results of normal (blue) and tumor (red) tissue. The T allele represents the prostate cancer risk associated allele. The given p-values were calculated with the Cochran-Mantel-Haenszel statistics assuming a linear model (expression dependent on the number of risk alleles n = 0, 1, 2).



Appendix F: Percentage change of the expressions of the 10q11 genes in tumor versus normal tissue dependent on rs10993994 genotype and *TMPRSS2-ERG* status. The value indicates the expressional change in %; values < 0 indicate a decreased, values > an increased expression in tumor compared to normal tissue. The stated p- and rho-values were estimated with a Spearman rank correlation with correction for ties to correlate the number of risk alleles (T) with the expression data. Expressional changes in the total sample set were assessed in n = 84 and dependent on the *TMPRSS2-ERG* status (T2E- and T2E+) in n = 78 patients (40 T2E negative and 38 T2E positive cases). In favor of a better illustration, one case was excluded for the *TIMM23* and *TIMM23B* box plots due to outlying values of ~1,500 % and ~1,900 %.



Acknowledgments

An dieser Stelle möchte ich mich ganz besonders herzlich bei meiner Doktormutter PD Dr. Christiane Maier für die Begleitung und kompetente Betreuung meiner Arbeit bedanken. Liebe Chrissie, danke für die zahlreichen Erfahrungen und Tagungen, die Du mir ermöglicht hast. Danke, dass Du Dir stets geduldig Zeit für mich genommen hast. Und danke vor allem für den Zusammenhalt und das Gefühl, auch in schwierigen Zeiten nie alleine dazustehen.

Vielen Dank an PD Dr. Dominic Varga für die Bereitschaft, das Zweitgutachten für meine Arbeit zu übernehmen.

Herrn Prof. Dr. Mark Schrader danke ich für die freundliche Aufnahme unserer Arbeitsgruppe in der Abteilung für Urologie und für die Unterstützung meiner Arbeit.

Dem ehemaligen ärztlichen Direktor der Humangenetik, Herrn Prof. Dr. Walther Vogel, möchte ich herzlich für die freundliche Aufnahme im Institut, sowie für die fachliche und finanzielle Unterstützung meiner Arbeit danken.

Danke an meinen Kollegen Dr. Manuel Lüdeke für seine Hilfsbereitschaft und sein Engagement, sowie auch an meine früheren Mit-Doktoranden Dr. Harald Surowy und Dr. Silvia Kastler für den fachlichen Input und die gute Zeit.

Herzlichen Dank an Herrn Prof. Dr. Josef Högel für die Engelsgeduld und für die Prise Humor bei zahlreichen statistischen Fragen.

Vielen Dank an Herrn Prof. Dr. Speit, der uns freundlich einen Arbeitsplatz in seiner Zellkultur zur Verfügung gestellt hat und an seine damalige TA, Regina Linsenmeyer, für ihre Hilfsbereitschaft.

Ich danke Margot Brugger, Irina Wiest, Bärbel Weber, Birgit Schmoll, und allen anderen, zum Teil ehemaligen, Mitgliedern des Instituts für Humangenetik für die schöne Zeit, an die ich mich immer gern zurückernnern werde.

Danke an Andrea Nottelmann und Heidi Schmidt für die tolle produktive Zusammenarbeit. Ich wünsche euch weiterhin ganz viel Glück und Erfolg!

Vielen Dank an das Uro-Team PD Dr. Marcus Cronauer, Wolfgang Streicher und Anke Bauer für die gute Zusammenarbeit in der Urologie. Danke vor allem auch an Michaela Eggel für die schönen TaqMan-Daten!

Meiner besseren Hälfte danke ich für die schöne Zeit abseits vom Labor sowie für die entgegengebrachte Geduld in so mancher zähen Phase des Schreibens.

Zu guter Letzt möchte ich mich herzlich bei meinen Eltern und bei meiner Schwester Katja bedanken. Ihr habt mit in jeglicher Hinsicht unterstützt. Danke für alles!

CURRICULUM VITAE

Der Lebenslauf wurde aus Gründen des Datenschutzes in der elektronischen Version entfernt

PUBLICATIONS

1. Luedeke M, Linnert CM, Hofer MD, Surowy HM, **Rinckleb AE**, Hoegel J, Kuefer R, Rubin MA, Vogel W, Maier C. Predisposition for TMPRSS2-ERG fusion in prostate cancer by variants in DNA repair genes. Cancer Epidemiology, Biomarkers & Prevention 2009; 18(11):3030-3035.

2. Surowy H, **Rinckleb A**, Luedeke M, Stuber M, Wecker A, Varga D, Maier C, Hoegel J, Vogel W. Heritability of baseline and induced micronucleus frequencies. Mutagenesis 2011; 26(1):111-117.

3. Kote-Jarai Z, Olama AA, Giles GG, Severi G, Schleutker J, Weischer M, Campa D, Riboli E, Key T, Gronberg H, Hunter DJ, Kraft P, Thun MJ, Ingles S, Chanock S, Albanes D, Hayes RB, Neal DE, Hamdy FC, Donovan JL, Pharoah P, Schumacher F, Henderson BE, Stanford JL, Ostrander EA, Sorensen KD, Dörk T, Andriole G, Dickinson JL, Cybulski C, Lubinski J, Spurdle A, Clements JA, Chambers S, Aitken J, Gardiner RA, Thibodeau SN, Schaid D, John EM, Maier C, Vogel W, Cooney KA, Park JY, Cannon-Albright L, Brenner H, Habuchi T, Zhang HW, Lu YJ, Kaneva R, Muir K, Benlloch S, Leongamornlert DA, Saunders EJ, Tymrakiewicz M, Mahmud N, Guy M, O'Brien LT, Wilkinson RA, Hall AL, Sawyer EJ, Dadaev T, Morrison J, Dearnaley DP, Horwich A, Huddart RA, Khoo VS, Parker CC, Van As N, Woodhouse CJ, Thompson A, Christmas T, Ogden C, Cooper CS, Lophatonanon A, Southey MC, Hopper JL, English DR, Wahlfors T, Tammela TL, Klarskov P, Nordestgaard BG, Røder MA, Tybjærg-Hansen A, Bojesen SE, Travis R, Canzian F, Kaaks R, Wiklund F, Aly M, Lindstrom S, Diver WR, Gapstur S, Stern MC, Corral R, Virtamo J, Cox A, Haiman CA, Le Marchand L, Fitzgerald L, Kolb S, Kwon EM, Karyadi DM, Orntoft TF, Borre M, Meyer A, Serth J, Yeager M, Berndt SI, Marthick JR, Patterson B, Wokolorczyk D, Batra J, Lose F, McDonnell SK, Joshi AD, Shahabi A, **Rinckleb AE**, Ray A, Sellers TA, Lin HY, Stephenson RA, Farnham J, Muller H, Rothenbacher D, Tsuchiya N, Narita S, Cao GW, Slavov C, Mitev V, Easton DF, Eeles RA; UK Genetic Prostate Cancer Study Collaborators/British Association of Urological Surgeons' Section of Oncology; UK ProtecT Study Collaborators, The Australian Prostate Cancer BioResource; PRACTICAL Consortium. Seven prostate cancer susceptibility loci identified by a multi-stage genome-wide association study. Nature Genetics 2011; 43(8):785-91.

4. Jin G, Lu L, Cooney KA, Ray AM, Zuhlke KA, Lange EM, Cannon-Albright LA, Camp NJ, Teerlink CC, Fitzgerald LM, Stanford JL, Wiley KE, Isaacs SD, Walsh PC, Foulkes WD, Giles GG, Hopper JL, Severi G, Eeles R, Easton D, Kote-Jarai Z, Guy M, **Rinckleb A**, Maier C, Vogel W, Cancel-Tassin G, Egrot C, Cussenot O, Thibodeau SN, McDonnell SK, Schaid DJ, Wiklund F, Grönberg H, Emanuelsson M, Whittemore AS, Oakley-Girvan I, Hsieh CL, Wahlfors T, Tammela T, Schleutker J, Catalona WJ, Zheng SL, Ostrander EA, Isaacs WB, Xu J; International Consortium for Prostate Cancer Genetics. Validation of prostate cancer risk-related loci identified from genome-wide association studies using family-based association analysis: evidence from the International Consortium for Prostate Cancer Genetics (ICPCG). Human Genetics 2012; 131(7):1095-103.

5. Luedeke M, Coinac I, Linnert CM, Bogdanova N, **Rinckleb AE**, Schrader M, Vogel W, Hoegel J, Meyer A, Dörk T, Maier C. Prostate cancer risk is not altered by TP53AIP1 germline mutations in a German case-control series. PLoS One 2012; 7(3);e34128.

6. Bailey-Wilson JE, Childs EJ, Cropp CD, Schaid DJ, Xu J, Camp NJ, Cannon-Albright LA, Farnham JM, George A, Powell I, Carpten JD, Giles GG, Hopper JL, Severi G, English DR, Foulkes WD, Mæhle L, Møller P, Eeles R, Easton D, Guy M, Edwards S, Badzioch MD, Whittemore AS, Oakley-Girvan I, Hsieh CL, Dimitrov L, Stanford JL, Karyadi DM, Deutsch K, McIntosh L, Ostrander EA, Wiley KE, Isaacs SD, Walsh PC, Thibodeau SN, McDonnell SK, Hebbring S, Lange EM, Cooney KA, Tammela TL, Schleutker J, Maier C, Bochum S, Hoegel J, Grönberg H, Wiklund F, Emanuelsson M, Cancel-Tassin G, Valeri A, Cussenot O, Isaacs WB; **International Consortium for Prostate Cancer Genetics†**. Analysis of Xq27-28 linkage in the international consortium for prostate cancer genetics (ICPCG) families. BMC medical genetics 2012; 13(1):46. †**Rinckleb A** is mentioned as collaborator within the group authorship

7. **Rinckleb AE**, Surowy HM, Luedeke M, Varga D, Schrader M, Hoegel J, Vogel W, Maier C. The prostate cancer risk SNP at 10q11 is associated with DNA repair capacity. DNA Repair 2012. 11(8):693-701.

8. Amin Al Olama A, Kote-Jarai Z, Schumacher FR, Wiklund F, Berndt SI, Benlloch S, Giles GG, Severi G, Neal DE, Hamdy FC, Donovan JL, Hunter DJ, Henderson BE, Thun MJ, Gaziano M, Giovannucci EL, Siddiq A, Travis RC, Cox DG, Canzian F, Riboli E, Key TJ, Andriole G, Albanes D, Hayes RB, Schleutker J, Auvinen A, Tammela TL, Weischer M, Stanford JL, Ostrander EA, Cybulski C, Lubinski J, Thibodeau SN, Schaid DJ, Sorensen KD, Batra J, Clements JA, Chambers S, Aitken J, Gardiner RA, Maier C, Vogel W, Dörk T, Brenner H, Habuchi T, Ingles S, John EM, Dickinson JL, Cannon-Albright L, Teixeira MR, Kaneva R, Zhang HW, Lu YJ, Park JY, Cooney KA, Muir KR, Leongamornlert DA, Saunders E, Tymrakiewicz M, Mahmud N, Guy M, Govindasami K, O'Brien LT, Wilkinson RA, Hall AL, Sawyer EJ, Dadaev T, Morrison J, Dearnaley DP, Horwich A, Huddart RA, Khoo VS, Parker CC, Van As N, Woodhouse CJ, Thompson A, Dudderidge T, Ogden C, Cooper CS, Lophatonanon A, Southey MC, Hopper JL, English D, Virtamo J, Le Marchand L, Campa D, Kaaks R, Lindstrom S, Diver WR, Gapstur S, Yeager M, Cox A, Stern MC, Corral R, Aly M, Isaacs W, Adolfsson J, Xu J, Zheng SL, Wahlfors T, Taari K, Kujala P, Klarskov P, Nordestgaard BG, Røder MA, Frikke-Schmidt R, Bojesen SE, FitzGerald LM, Kolb S, Kwon EM, Karyadi DM, Orntoft TF, Borre M, **Rinckleb A**, Luedeke M, Herkommer K, Meyer A, Serth J, Marthick JR, Patterson B, Wokolorczyk D, Spurdle A, Lose F, McDonnell SK, Joshi AD, Shahabi A, Pinto P, Santos J, Ray A, Sellers TA, Lin HY, Stephenson RA, Teerlink C, Muller H, Rothenbacher D, Tsuchiya N, Narita S, Cao GW, Slavov C, Mitev V; UK Genetic Prostate Cancer Study Collaborators/British Association of Urological Surgeons' Section of Oncology; UK ProtecT Study Collaborators; Australian Prostate Cancer Bioresource; PRACTICAL Consortium, Chanock S, Gronberg H, Haiman CA, Kraft P, Easton DF, Eeles RA. A meta-analysis of genome-wide association studies to identify prostate cancer susceptibility loci associated with aggressive and non-aggressive disease. Human Molecular Genetics 2013; 22(2):408-15.

9. Eeles RA, Olama AA, Benlloch S, Saunders EJ, Leongamornlert DA, Tymrakiewicz M, Ghoussaini M, Luccarini C, Dennis J, Jugurnauth-Little S, Dadaev T, Neal DE, Hamdy FC, Donovan JL, Muir K, Giles GG, Severi G, Wiklund F, Gronberg H, Haiman CA, Schumacher F, Henderson BE, Le Marchand L, Lindstrom S, Kraft P, Hunter DJ, Gapstur S, Chanock SJ, Berndt SI, Albanes D, Andriole G, Schleutker J, Weischer M, Canzian F, Riboli E, Key TJ, Travis RC, Campa D, Ingles SA, John EM, Hayes RB, Pharoah PD, Pashayan N, Khaw KT, Stanford JL, Ostrander EA, Signorello LB, Thibodeau SN, Schaid D, Maier C, Vogel W, Kibel AS, Cybulski C, Lubinski J, Cannon-Albright L, Brenner H, Park JY, Kaneva R, Batra J, Spurdle AB, Clements JA, Teixeira MR, Dicks E, Lee A, Dunning AM, Baynes C, Conroy D, Maranian MJ, Ahmed S, Govindasami K, Guy M, Wilkinson RA, Sawyer EJ, Morgan A, Dearnaley DP, Horwich A, Huddart RA, Khoo VS, Parker CC, Van As NJ, Woodhouse CJ, Thompson A, Dudderidge T, Ogden C, Cooper CS, Lophatananon A, Cox A, Southey MC, Hopper JL, English DR, Aly M, Adolfsson J, Xu J, Zheng SL, Yeager M, Kaaks R, Diver WR, Gaudet MM, Stern MC, Corral R, Joshi AD, Shahabi A, Wahlfors T, Tammela TL, Auvinen A, Virtamo J, Klarskov P, Nordestgaard BG, Røder MA, Nielsen SF, Bojesen SE, Siddiq A, Fitzgerald LM, Kolb S, Kwon EM, Karyadi DM, Blot WJ, Zheng W, Cai Q, McDonnell SK, **Rinckleb AE**, Drake B, Colditz G, Wokolorczyk D, Stephenson RA, Teerlink C, Muller H, Rothenbacher D, Sellers TA, Lin HY, Slavov C, Mitev V, Lose F, Srinivasan S, Maia S, Paulo P, Lange E, Cooney KA, Antoniou AC, Vincent D, Bacot F, Tessier DC; COGS–Cancer Research UK GWAS–ELLIPSE (part of GAME-ON) Initiative; Australian Prostate Cancer Bioresource; UK Genetic Prostate Cancer Study Collaborators/British Association of Urological Surgeons' Section of Oncology; UK ProtecT (Prostate testing for cancer and Treatment) Study Collaborators; PRACTICAL (Prostate Cancer Association Group to Investigate Cancer-Associated Alterations in the Genome) Consortium, Kote-Jarai Z, Easton DF. Identification of 23 new prostate cancer susceptibility loci using the iCOGS custom genotyping array. Nature Genetics 2013; 45(4):385-91, 391e1-2.

10. Kote-Jarai Z, Saunders EJ, Leongamornlert DA, Tymrakiewicz M, Dadaev T, Jugurnauth-Little S, Ross-Adams H, Al Olama AA, Benlloch S, Halim S, Russell R, Dunning AM, Luccarini C, Dennis J, Neal DE, Hamdy FC, Donovan JL, Muir K, Giles GG, Severi G, Wiklund F, Gronberg H, Haiman CA, Schumacher F, Henderson BE, Le Marchand L, Lindstrom S, Kraft P, Hunter DJ, Gapstur S, Chanock S, Berndt SI, Albanes D, Andriole G, Schleutker J, Weischer M, Canzian F, Riboli E, Key TJ, Travis RC, Campa D, Ingles SA, John EM, Hayes RB, Pharoah P, Khaw KT, Stanford JL, Ostrander EA, Signorello LB, Thibodeau SN, Schaid D, Maier C, Vogel W, Kibel AS, Cybulski C, Lubinski J, Cannon-Albright L, Brenner H, Park JY, Kaneva R, Batra J, Spurdle A, Clements JA, Teixeira MR, Govindasami K, Guy M, Wilkinson RA, Sawyer EJ, Morgan A, Dicks E, Baynes C, Conroy D, Bojesen SE, Kaaks R, Vincent D, Bacot F, Tessier DC; COGS-CRUK GWAS-ELLIPSE (Part of GAME-ON) Initiative; UK Genetic Prostate Cancer Study Collaborators/British Association of Urological Surgeons' Section of Oncology; UK ProtecT Study Collaborators; **PRACTICAL Consortium†**, Easton DF, Eeles RA. Fine-mapping identifies multiple prostate cancer risk loci at 5p15, one of which associates with TERT expression. Human Molecular Genetics 2013; 22(12):2520-2528.
†Rinckleb A is mentioned as collaborator within the group authorship

11. Teerlink CC, Thibodeau SN, McDonnell SK, Schaid DJ, **Rinckleb A**, Maier C, Vogel W, Cancel-Tassin G, Egrot C, Cussenot O, Foulkes WD, Giles GG, Hopper JL, Severi G, Eeles R, Easton D, Kote-Jarai Z, Guy M, Cooney KA, Ray AM, Zuhlke KA, Lange EM, Fitzgerald LM, Stanford JL, Ostrander EA, Wiley KE, Isaacs SD, Walsh PC, Isaacs WB, Wahlfors T, Tammela T, Schleutker J, Wiklund F, Grönberg H, Emanuelsson M, Carpten J, Bailey-Wilson J, Whittemore AS, Oakley-Girvan I, Hsieh CL, Catalona WJ, Zheng SL, Jin G, Lu L, Xu J; International Consortium for Prostate Cancer Genetics, Camp NJ, Cannon-Albright LA. Association analysis of 9,560 prostate cancer cases from the International Consortium of Prostate Cancer Genetics confirms the role of reported prostate cancer associated SNPs for familial disease. Human Genetics 2014; 133(3):347-56.

12. Reiter R, Brosch S, Lüdeke M, Fischbein E, **Rinckleb A**, Haase S, Schwandt A, Pickhard A, Maier C, Högel J, Vogel W. Do Orofacial Clefts Represent Different Genetic Entities? The Cleft palate-craniofacial journal 2014; [Epub ahead of print]

13. Amin Al Olama A, Benlloch S, Antoniou AC, Giles GG, Severi G, Neal DE, Hamdy FC, Donovan JL, Muir K, Schleutker J, Henderson BE, Haiman CA, Schumacher F, Pashayan N, Pharoah P, Ostrander EA, Stanford JL, Batra J, Clements JA, Champers S, Weischer M, Nordestgaard BG, Ingles SA, Sorensen KD, Orntoft T, Park JY, Cybulski C, Maier C, Doerk T, Dickinson JL, Cannon-Albright L, Brenner H, Rebbeck T, Habuchi T, Thibodeau SN, Cooney K, Chappuis P, Hutter P, Kaneva R, Foulkes WD, Zeegers MP, Lu YJ, Zhang HW, Stephenson R, Cox A, Southey MC, Spurdle A, FitzGerald LM, Leongamornlert DA, Saunders EJ, Tymrakiewicz M, Guy M, Dadaev T, Little SJ, Govindasami K, Sawyer EJ, Wilkinson RA, Herkommer K, Morrison J, Hopper J, Lophatonanon A, **Rinckleb A**, The UK Genetic Prostate Cancer Study Collaborators/British Association of Urological Surgeons' Section of Oncology, The UK ProtecT Study Collaborators, The PRACTICAL Consortium, Kote-Jarai Z, Eeles RA, Easton D. Risk Analysis of Prostate Cancer in PRACTICAL, a Multinational Consortium, Using 25 Known Prostate Cancer Susceptibility Loci. [Submitted]

14. Amin Al Olama A, Kote-Jarai Z, Berndt SI, Conti D, Schumacher F, Han Y, Benlloch S, Hazelett DJ, Wang Z, Saunders E, Leongamornlert D, Lindstrom S, Jugurnauth-Little S, Dadaev T, TymrakiewiczM, Stram DO, Wan P, Stram A, Sheng X, Pooler LC, Park K, Xia L, Kolonel LN, LeMarchand L, Hoover RN, Machiela M, Yeager M, Burdette L, Chung CC, Hutchinson A, Yu K, Goh C, Ahmed M, Govindasami K, Guy M, Tammela T, Schleutker J, Wahlfors T, Xu J, Aly M, Donovan J, Travis R, Key T, Siddiq A, Canzian F, Khaw KT, Takahashi A, Kubo M, Southey M, Pharoah P, Pashayan N, Weischer M, Nordestgaard BG, Nielsen SF, Klarskov P, Roder A, Iverson P, Thibodeau S, Stanford JL, Kolb S, Holt S, Gapstur S, Diver RW, Stevens, Maier C, Luedeke M, Herkommer K, **Rinckleb A**, Strom SS, Pettaway C, Yeboah ED, Tettey Y, Biritwum RB, Adjei AA, Tay E, Truelove A, Niwa S, Chokkalingam AP, Cannon-Albright L, Cybulski C, Park J, Sellers T, Lin HY, Isaacs WB, Brenner H, Chen C, Giovannucci EL, Ma J, Stampfer M, Penney K, Mucci L, John EM, Ingles SA, Kittles RA, Murphy AB, Pandha H, Agnieszka M, Kierzek A, Blot W, Signorello L, Zheng W, Albanes D, Virtamo J, Weinstein S, Nemesure B, Carpten J, Leske C, Wu SY, Hennis A, Kibel AS, Rybicki BA, Neslund-Dudas C, Hsing AW, Chu L, Goodman P, Klein EA, Zheng L, Batra J, Clements J, Sprudle A, Teixeira MR, Paulo P, Maia S, Slavov C, Kaneva R, Witte J, Casey G, Seminara D, Riboli E, Hamdy FC, Coetzee GA, Li Q, Matthew Freedman ML, Hunter DJ, Muir K, Gronberg H, Neal DE, Giles G, Severi G, The Breast and Prostate Cancer Cohort Consortium (BPC3), The PRACTICAL (Prostate Cancer Association Group to Investigate Cancer-Associated Alterations in the Genome) Consortium, The COGS (Collaborative Oncological Gene-environment Study) Consortium, The GAME-ON/ELLIPSE Consortium, Cook MB, Nakagawa H, Wiklund F, Kraft P, Chanock SJ, Henderson BE, Easton DF, Eeles RA, Haiman CA. A multiethnic meta-analysis of 89,189 individuals reveals 23 novel susceptibility loci for prostate cancer. [Submitted]

15. Maier C, Herkommer K, Luedeke M, **Rinckleb A**, Schrader M and Vogel W. Subgroups of familial and aggressive prostate cancer with considerable frequencies of *BRCA2* mutations. [In Revision]

16. **Rinckleb AE**, Luedeke M, FitzGerald LM, Stanford JL, Schleutker J, Eeles RA, Teixeira MR, Cannon-Albright L, Ostrander EA, Weikert S., Hoegel J, Herkommer K, Wahlfors T, Visakorpi T, Leinonen KA, Tammela TL, Cooper CS, Kote-Jarai Z, Edwards S, Goh CL, McCarthy F., Parker C., Flohr P, Paulo P, Jerónimo C., Henrique R, Krause H., Schrader M, Vogel W, Kuefer R, Hofer MD, Perner S, Rubin MA, Agarwal A.M., Easton DF, Amin Al Olama A., Benlloch S, The PRACTICAL (Prostate Cancer Association Group to Investigate Cancer-Associated Alterations in the Genome) consortium, and Maier C: Prostate cancer risk regions in 8q24 and 17q24 are differentially associated with somatic *TMPRSS2-ERG* fusion status. [In Revision]

17. Egber L, Luedeke M, **Rinckleb A**, Wright JL, Maier C, Neuhouser ML, Stanford JL. Obesity and the risk of prostate cancer according to *TMPRSS2:ERG* fusion. [In Revision]

18. Wright JL, Chéry L, Holt S, Lin DW, Luedeke M, **Rinckleb AE**, Maier C, Stanford JL. Aspirin use is associated with a reduced risk of *TMPRSS2:ERG* fusion positive prostate cancer. [In Revision]

19. Surowy H, Varga D, Burwinkel B, Marmé F, Sohn C, Luedeke M, Rinckleb A, Maier C, Deissler H, Volcic M, Wiesmüller L, Hasenburg A, Klar M, Hoegel J, Vogel W. A low frequency missense variant (rs3810813) in *SLX4* is associated with early-onset breast cancer and reduced DNA repair capacity. [In Revision]

PRESENTATIONS

Oral Presentations

1. **Rinckleb AE**, Luedeke M, Surowy HM, Vogel W, Hoegel J, Herkommer K, Kote-Jarai Z, Eeles RA, Kubisch C, Schrader M, and Maier C. Kumulative Analyse risikoassoziierter "common variants" mit dem Prostatakarzinom [V4.7]. 2. Symposium "Urologische Forschung der Deutschen Gesellschaft für Urologie". Mainz, 11.-13.11.2010. Abstract in: Urologe A 2010; 50(1):99

2. **Rinckleb AE**, Luedeke M, Hoegel J, Schrader M, Vogel W, Stanford J, Schleutker J, Teixeira M, Eeles RA, and Maier C. Common prostate cancer risk variants are associated with the somatic *TMPRSS2-ERG* fusion [V11.6]. 65. Kongress der Deutschen Gesellschaft für Urologie/Meeting of the European Section for Urological Research (ESUR). Leipzig, 26.-29.09.2013.

Poster Presentations

3. **Rinckleb A**, Maier C, Vogel W, and Luedeke M. Gene blocking approach for the oncogene activator TMPRSS2 by application of peptide nucleic acids (PNA) in prostate cancer cell lines [P146]. 20. Jahrestagung der Deutschen Gesellschaft für Humangenetik (GfH). Aachen, 01.-03.04.2009. Abstract in: Medizinische Genetik 2009; 21(1):120

4. **Rinckleb A**, Kote-Jarai Z, Luedeke M, Surowy H, Easton DF, Eeles RA, Vogel W, and Maier C. Re-Analysis of 25 common risk variants for familial prostate cancer risk in Germany [P-CancG-024]. 21. Jahrestagung der Deutschen Gesellschaft für Humangenetik (GfH), gemeinsam mit der Österreichischen Gesellschaft für Humangenetik (ÖHG) und der Schweizerischen Gesellschaft für Medizinische Genetik (SGMG). Hamburg, 02.-04.03.2010. Abstract in: Medizinische Genetik 2010; 22(1):100

5. **Rinckleb AE**, Surowy HM, Schmidt H, Luedeke M, Vogel W, Schrader M, and Maier C. The prostate cancer tagging SNP at 10q11 is associated with DNA repair capacity [P2.3]. 3. Symposium "Urologische Forschung der Deutschen Gesellschaft für Urologie". Jena, 17.-19.11.2011. Abstract in: Urologe A 2011; 51(1):105

6. **Rinckleb AE**, Schmidt H, Luedeke M, Hoegel J, Schrader M, Marienfeld RB, Vogel W, and Maier C, Influence of a Prostate Cancer risk locus on 10q11 on expression profiles of flanking genes [P6.3]. 64. Kongress der Deutschen Gesellschaft für Urologie. Leipzig, 26.-29.09.2012.

7. **Rinckleb AE**, Luedeke M, Stanford JL, Schleutker J, Eeles RA, Teixeira MR, Cannon-Albright L, Ostrander EA, Weikert S., Hoegel J, Herkommer K, FitzGerald LM, Wahlfors T, Visakorpi T, Leinonen KA, Tammela TL, Cooper CS, Kote-Jarai Z, Edwards S, Goh CL, McCarthy F., Parker C., Flohr P, Paulo P, Jerónimo C., Henrique R, Krause H., Schrader M, Vogel W, Kuefer R, Hofer MD, Perner S, Rubin MA, Agarwal AM, Easton DF, Amin Al Olama A, Benlloch S, The PRACTICAL (Prostate Cancer Association Group to Investigate Cancer-Associated Alterations in the Genome) consortium, Maier C and Hoegel J. Prostate cancer risk regions in 8q24 and 17q24 are differentially associated with somatic *TMPRSS2:ERG* fusion status. 25. Jahrestagung der Deutschen Gesellschaft für Humangenetik (GfH). Essen, 19.-21.03.2014. Abstract in: Medizinische Genetik 2014/1: 75