Research Article

TumorBiology

Tumor Biol 2008;29:83–92
DOI: 10.1159/000135688

Received: October 30, 2007
Accepted: April 2, 2008
Published online: June 2, 2008

Genetic Variants of the Copy Number Polymorphic β-Defensin Locus Are Associated with Sporadic Prostate Cancer

Klaus Huse[a] Stefan Taudien[a] Marco Groth[a, c] Philip Rosenstiel[d] Karol Szafranski[a] Michael Hiller[f] Jochen Hampe[d] Kerstin Junker[b] Jörg Schubert[b] Stefan Schreiber[d] Gerd Birkenmeier[c] Michael Krawczak[e] Matthias Platzer[a]

[a]Genome Analysis, Leibniz Institute for Age Research, Fritz Lipmann Institute and [b]Department of Urology, Friedrich Schiller University, Jena, [c]Institute of Biochemistry, University of Leipzig, Leipzig, [d]Institute for Clinical Molecular Biology and [e]Institute of Medical Informatics and Statistics, Christian Albrechts University, Kiel, and [f]Institute of Computer Science, Bioinformatics Group, Albert Ludwigs University Freiburg, Freiburg, Germany

Key Words
Prostate cancer · DNA amplification · Genomic DNA · Polymorphism · Copy number variation · Haplotype · Innate immunity · Antibacterial defense · Defensin

Abstract
Background/Aims: Prostate cancer represents the cancer with the highest worldwide prevalence in men. Chromosome 8p23 has shown suggestive genetic linkage to early-onset familial prostate cancer and is frequently deleted in cancer cells of the urogenital tract. Within this locus some β-defensin genes (among them *DEFB4*, *DEFB103*, *DEFB104*) are localized, which are arranged in a gene cluster shown to exhibit an extensive copy number variation in the population. This structural variation considerably hampers genetic studies. In a new approach considering both sequence as well as copy number variations we aimed to compare the defensin locus at 8p23 in prostate cancer patients and controls. ***Methods:*** We apply PCR/cloning-based haplotyping and high-throughput copy number determination methods which allow assessment of both individual haplotypes and gene copy numbers not accessible to conventional SNP-based genotyping. ***Results:*** We demonstrate association of four common *DEFB104* haplotypes with the risk of prostate cancer in two independent patient cohorts. Moreover, we show that high copy numbers (>9) of the defensin gene cluster are significantly underrepresented in both patient samples. ***Conclusions:*** Our findings imply a role of the antibacterial defensins in prostate cancerogenesis qualifying distinct gene variants and copy numbers as potential tumor markers.
Copyright © 2008 S. Karger AG, Basel

Introduction

The genetic etiology of prostate cancer (PC), which is a major health burden worldwide [1], is complex and the interplay of disease genes with environmental trigger factors is still incompletely understood. A suggestive genetic linkage to chromosome 8p22–23 was reported for hereditary PC [2–4]. The region contains in a 2-Mb interval several clusters of defensin (DEF) genes [5]. Defensins are biocidal peptides acting against bacteria, fungi, and enveloped viruses with effector function in innate and

KARGER

Fax +41 61 306 12 34
E-Mail karger@karger.ch
www.karger.com

© 2008 S. Karger AG, Basel
1010–4283/08/0292–0083$24.50/0

Accessible online at:
www.karger.com/tbi

Matthias Platzer
Genome Analysis, Leibniz Institute for Age Research, Fritz Lipmann Institute
Beutenbergstrasse 11
DE–07745 Jena (Germany)
Tel. +49 3641 656 241, Fax +49 3641 656 255, E-Mail mplatzer@fli-leibniz.de

adaptive immunity [6, 7]. They have been implicated in the pathophysiology of a host of diseases with microbiological components including HIV infection [8, 9], chronic inflammatory bowel diseases (i.e. ulcerative colitis and Crohn's disease) [10, 11] and psoriasis [12]. The DEF locus exhibits extensive copy number variation (CNV) including the β-defensin (DEFB) genes *DEFB4*, *DEFB103*, and *DEFB104* [5, 11, 13–15] encoding the proteins hBD-2, 3 and 4, respectively. Copy numbers of these genes influence the mRNA level [14], and low copy numbers have been suggested to predispose to Crohn's disease [11], whereas increased numbers are associated with psoriasis [12].

Defensins have attracted attention in cancer research because an inflammatory link to cancer appears likely [16, 17]. In a murine model, *Defb29* initiates vasculogenesis in tumors. A recruitment of dendritic cell precursors through Ccr6, which are transformed to endothelial-like cells by Vegf-A [18], thus may represent a pivotal step in the progression of malignant disorders. In the same study, upregulation of *DEFB4* mRNA in human ovarian cancer and cancer-specific expression of hBD-2 and hBD-3 were observed. Very recently, it has been shown that at nanomolar concentrations hBD-3 is a ligand of MC1R [19]. Genetic variants of this receptor are associated with skin cancer. Loss of DEF genes from the 8p23 locus and/or deficient expression of the *DEFB1*-encoded protein are associated with the progression of renal cell carcinoma [20–22].

In families with hereditary pancreatic carcinoma, loss of the DEF cluster from the 8p22–23 region is associated with higher disease susceptibility [3, 23, 24]. Similar to renal cell carcinoma, PC tissues show no or only faint *DEFB1* expression [20]. However, systematic genetic studies of the DEFB locus in sporadic, common PC are still lacking. One of the reasons is that large-scale association studies to find genetic variants conferring a low to modest disease risk are not feasible using conventional genotyping technologies [5, 14, 15, 25, 26]. Only recently methods with the potential for high-throughput copy number quantification were published [13, 27]. In view of the linkage to the short arm of chromosome 8 [3, 20–24] and the fact that exposure to defensin-sensitive infectious agents contributes to prostate carcinogenesis [28], DEF genes need to be evaluated as attractive candidates for association studies in PC.

We have recently refined the genomic structure of the segmentally duplicated DEF locus and established a PCR/cloning-based haplotyping method that allows assessment of both individual haplotypes and gene copy numbers [5]. We found that an individual's haplotype variation is partially independent from copy number polymorphisms, thus adding another level of complexity to the understanding of CNV. This association study therefore considers both copy number and sequence variation as possible phenotype determinants. Here, we report that the frequency of four haplotypes of *DEFB104* significantly varies between PC patients and controls. A novel pyrosequencing approach for copy number estimation of the DEFB cluster revealed that individuals with a high copy number (>9) are underrepresented among PC patients.

Methods

DNA Samples, Cell Lines and Oligonucleotides

Two independent cohorts of Caucasian patients who underwent radical prostatectomy were sampled [95 individuals each, labeled cohort 1 (University Hospital Leipzig; mean age: 62.5 years, median age: 63 years, range: 50–77 years) and cohort 2 (University Hospital Jena; mean age: 62.9 years, median age: 64 years, range: 46–69 years), respectively]. Genomic DNA was obtained from EDTA blood drawn at surgery (QIAamp DNA Mini Kit). Genomic DNA samples from 95 male Caucasian control individuals (mean age: 48.8 years, median age: 50 years, range: 31–71 years) were obtained from saliva and kindly provided by G.H. Scholz (Altenburg, Germany). The studies were approved by the Ethics Committees of the University of Leipzig (Vote No. 857) and Friedrich Schiller University Jena (Vote No. 1983-01/07). Written informed consent was obtained from all participants. The funding sources of the study played no role in study design, data collection, data analysis, data interpretation or writing of the report.

A pool of ~100 human genomic DNA samples was obtained from Roche. LNCaP (CRL-1740™), PC3 (CRL-1435™) and DU145 (HTB-81™) cell lines were obtained from ATCC. All primers were synthesized by Metabion.

Determination of Haplotypes by the Analysis of Cloned PCR Products

Sequencing of individual clones obtained by subcloning of PCR products served to identify haplotypes at the *DEFB104* locus. The primers TTCTGTAGCCCCAACACCTC and GGTGCCAAGGACATCTAGGA amplify a 500-bp fragment spanning four single nucleotide variations (rs17843871, rs2680507, rs17843872, rs4259430). PCR reactions were performed in 25-μl volume containing 5–30 ng genomic DNA and 10 pmol of each primer, using ReadyToGo PCR beads (Amersham). PCR cycling conditions were as follows: 94°C for 4 min; followed by 30 cycles at 94°C for 30 s, 60°C for 30 s, and 72°C for 30 s, with a final extension at 72°C for 10 min. PCR products were sequenced (Sanger), with PCR primer GGTGCCAAGGACATCTAGGA using Dye Terminator 3.1 chemistry and a 3730xl DNA Analyzer (Applied Biosystems). Results were visually inspected with GAP4 [29].

Equal amounts of PCR products from PC patients and controls were pooled, respectively, and aliquoted to 25 μl. Aliquots of

pooled PCR products as well as PCR products obtained from pooled genomic DNA and cell lines were supplemented with 1 μl of formamide, 2 μl of 7.5 mM ammonium acetate and 85 μl of ethanol (6%, v/v) and stored at 4°C for 1 h. The DNA precipitates were collected by centrifugation, washed twice with 100 μl of ethanol (70%, v/v) and redissolved in 5 μl of water. This solution was used for topoisomerase cloning into pCR2.1-TOPO with the TOPO TA Cloning Kit (Invitrogen) according to the manufacturer's instructions. Well-isolated colonies were transferred and grown in LB broth supplemented with ampicillin. Plasmid DNA was isolated from the cultures by BioRobot 8000 and MagAttract 96 Miniprep Core Kit (Qiagen) and inserts were sequenced in both directions using M13 universal primers. Haplotypes were called by visual inspection of the sequence traces.

Estimation of Single Nucleotide Variation Frequencies by Pyrosequencing

Relative allele frequencies in the DNA pools of variants rs17843871, rs2680507, rs17843872, and rs4259430 were estimated as described previously [25].

Estimation of Copy Numbers

For individual copy number assessment of the DEFB cluster, we performed a modified paralogue ratio test originally developed by Deutsch et al. [27] and applied by Armour et al. [13] to copy numbers of the DEFB cluster. We performed PCR using primers GTCCAAGGCTGGAATGCA and TTGGTGGGACAAGTGATG, specific for a retroposed *HSPD1* pseudogene *(HSPDP)* within the DEFB cluster (hg18; chr8:7786283–7786422 and chr8:7264985–7265124, respectively) and five paralogous single copy loci (hg18; chr2:198061320–198061459; chr3:36783978–36784117; chr4:145987444–145987583; chr5:21919673–21919812; chr12:55192269–55192407). Amplification was carried out as described above with 10 ng genomic DNA in 96-well PCR plates (ABgene) using Deep Vent DNA polymerase (New England Biolabs) and its respective buffer. Amplicons were precipitated and ligated overnight at room temperature with T7 DNA ligase into *Sma*I-digested pUC19 in a volume of 10 μl. To carry out independent pyrosequencing analyses for each orientation of the amplicon within the vector [for further details, see 25], two amplification reactions were performed using ReadyToGo PCR beads (Amersham), 2 μl of the ligated sample, 5′-biotinylated vector primers bt-CAGGAAACAGCTATGAC or bt-GTAAAACGACGGCCAG, respectively, and pseudogene sequencing primer ACAGCAGCTCTTGTAGCATTAA (PCR conditions: 94°C for 2 min; followed by 30 cycles at 94°C for 30 s, 58°C for 30 s, and 72°C for 30 s with a final extension at 72°C for 10 min). Biotin-labeled PCR products were immobilized on Streptavidin Sepharose™ (GE Healthcare) by mixing 20 μl of the PCR product with 6 μl Streptavidin Sepharose suspension, 10 μl water and 40 μl 1× binding buffer (Biotage) followed by shaking at room temperature for at least 10 min. To remove the unbiotinylated DNA strand, the samples were sequentially washed with 70% ethanol and 0.5 M NaOH using the PyroMark Vacuum Prep Tool (Biotage). Immobilized ssDNA was then washed with 1× washing buffer for 10 s, transferred to 40 μl 1× annealing buffer plus 4 μl target specific sequencing primer (20 pmol/μl in water), and kept at 80°C under continuous shaking for 10 min. After equilibration to room temperature, the sequencing reaction was performed using the Pyro Gold Reagent Kit (Biotage) in the PSQ 96MA Pyrosequencing™ instrument according to the manufacturer's instructions. Iterative addition of nucleotides was in the order AGTGTCATCTGAGT to discriminate the *HSPDP* locus at chromosome 8 (sequence downstream of the primer is GTGC) from the five calibrator loci (GGGC). Copy number of the DEFB locus in a diploid genome was estimated by 10(1 – f)/f, where f is the fraction of the calibrator allele G at the second position downstream of the sequencing primer. These calculations were based upon the assumption that (1) each DEFB cluster contains exactly one *HSPDP* locus, (2) all *HSPDP* loci in DEFB clusters contain the T allele at the diagnostic position, (3) five single copy *HSPDP* loci are located elsewhere in the genome, all containing the G allele at the diagnostic position and (4) the estimated T frequency (1 – f) is proportional to the DEFB cluster number. The first two assumptions are supported by BLAST alignments of the *HSPDP* sequence to all available human clone sequences containing 8p23.1 β-defensin cluster genes or parts of them. This includes not only clones implemented into the hg18 assembly but also those which cannot be unambiguously aligned to any of the two β-defensin clusters in hg18 and therefore may represent additional copies. Exclusively the T allele of *HSPDP* was detected and no hints were found for the existence of more than one copy of *HSPDP* in clones harboring the flanking *DEFB4* and *DEFB103* genes. Assumption 3 was supported by an analogous BLAST approach to the entire human genome. In addition, since identical primers are used for the amplification of relatively small DNA segments of equal size, a PCR bias is avoided and a proportional propagation is very likely. In our experiments, copy number estimation was confined to those cases in which the results of two independent Pyrosequencing assays did not differ by more than 10%, and the average G frequencies obtained were used in the respective calculations.

Statistics

Haplotype frequencies among cases and controls were tested for statistical significance using both a global and haplotypewise χ^2 statistics. However, since all analyzed clones were derived from a fixed number of individuals, the corresponding haplotypes did not represent statistically independent observations. This implies that the distribution of the calculated χ^2 statistics could not be assumed to follow a χ^2 distribution. Instead, p values were obtained by stepwise simulation. To this end, an individual's diploid copy number was first simulated according to the data generated in the experiments described above. Next, each copy was assigned a haplotype at random, using the combined haplotype distribution as observed in cases and controls. (Note that, under the null hypothesis of equal haplotype distributions in the two groups, the pooled observations represent the best estimate of the joint distribution.) From the simulated genotypes, comprising the same number of individuals as the original data, a haplotype distribution for subsequent clone sampling was derived giving each individual (not copy) the same weight. In this way, the simulation took post-PCR pooling into account. Finally, the same number of clones was sampled from each group-specific distribution as in the original experiment. This process was repeated one million times. From the simulated data, χ^2 statistics were calculated in each round and the relative number of times by which the simulated χ^2 statistics exceeded the original one was taken as its p value ($p = 0$ was conservatively interpreted as $p < 10^{-4}$). Since an omnibus χ^2 test was performed prior to the haplotypewise tests, no correction for

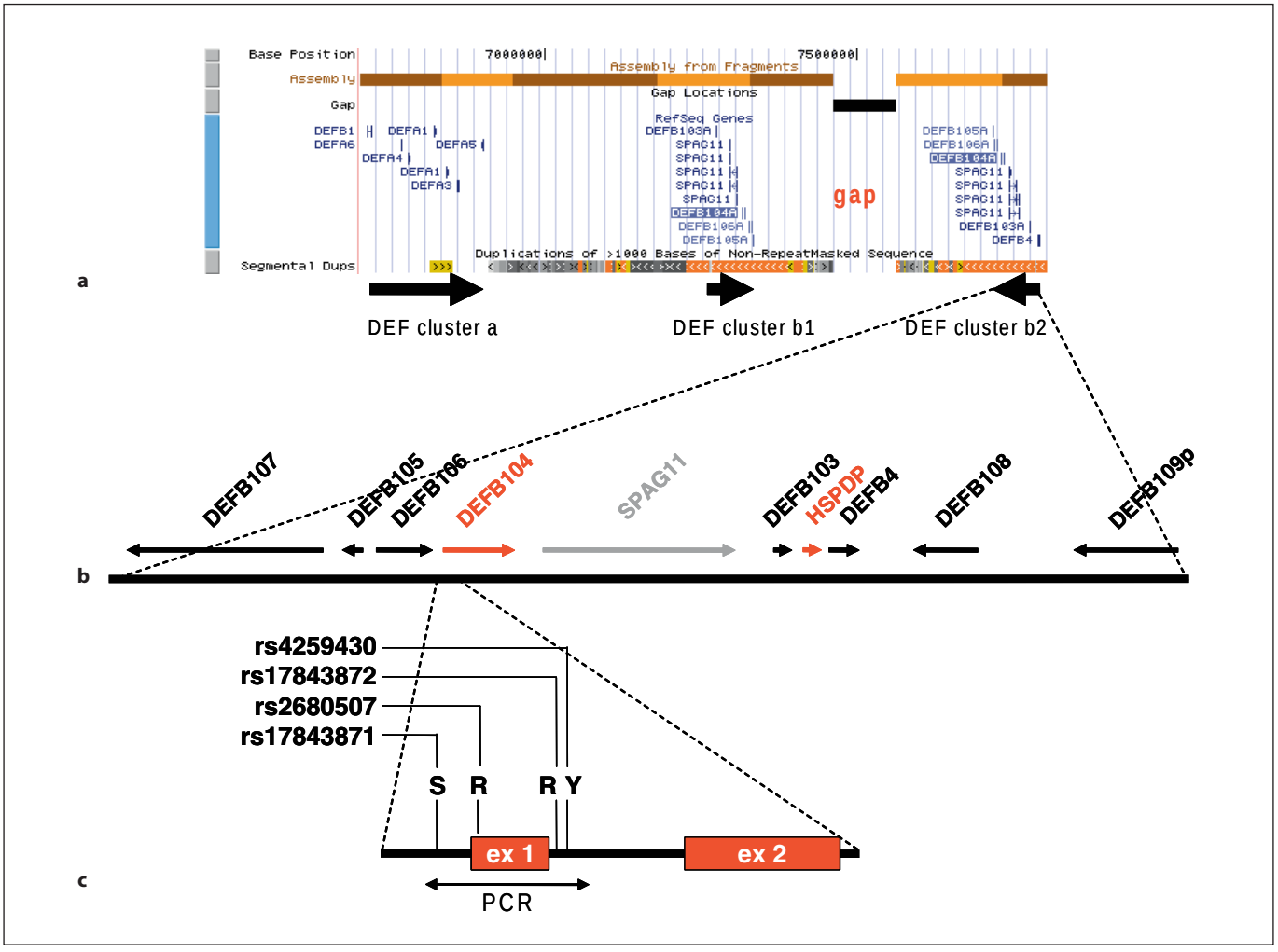

Fig. 1. Human DEF gene locus at 8p23. **a** Representation in the UCSC genome browser, hg18. **b** Gene organization of the DEFB cluster with *DEFB104* and *HSPDP* in red. **c** Amplified *DEFB104* region with the four MSV used for haplotyping.

multiple testing was deemed necessary for the latter. Wilcoxon rank sum and Fisher's exact tests of DEFB copy number differences were performed using the R statistical software package (http://www.rproject.org).

Results

Assessment of DEFB104 *Haplotypes*

We amplified a 500-bp fragment in the region of *DEFB104*, located within the DEFB cluster on chromosome 8p23 and harboring four single nucleotide variations (fig. 1). Direct sequencing of the PCR products from individual DNA samples revealed a variation of the peak heights in the electropherograms, indicating differences of both copy number and allele distribution within these copies (fig. 2). Due to the complex structure of the locus, it cannot be decided whether the electropherograms reflect SNPs (single nucleotide polymorphisms), PSVs (paralogous sequence variations), or MSVs (multisite variations) [30]. We will henceforth use the terms 'MSV' for convenience and 'haplotype' for the physical combinations of their respective alleles.

Inspection of the four MSVs in 285 individuals (2 × 95 PC patients and 95 control individuals) did not reveal any sample in which all positions appeared to be homozygous. In accordance with differences in copy number and allele distribution within copies, numerous asym-

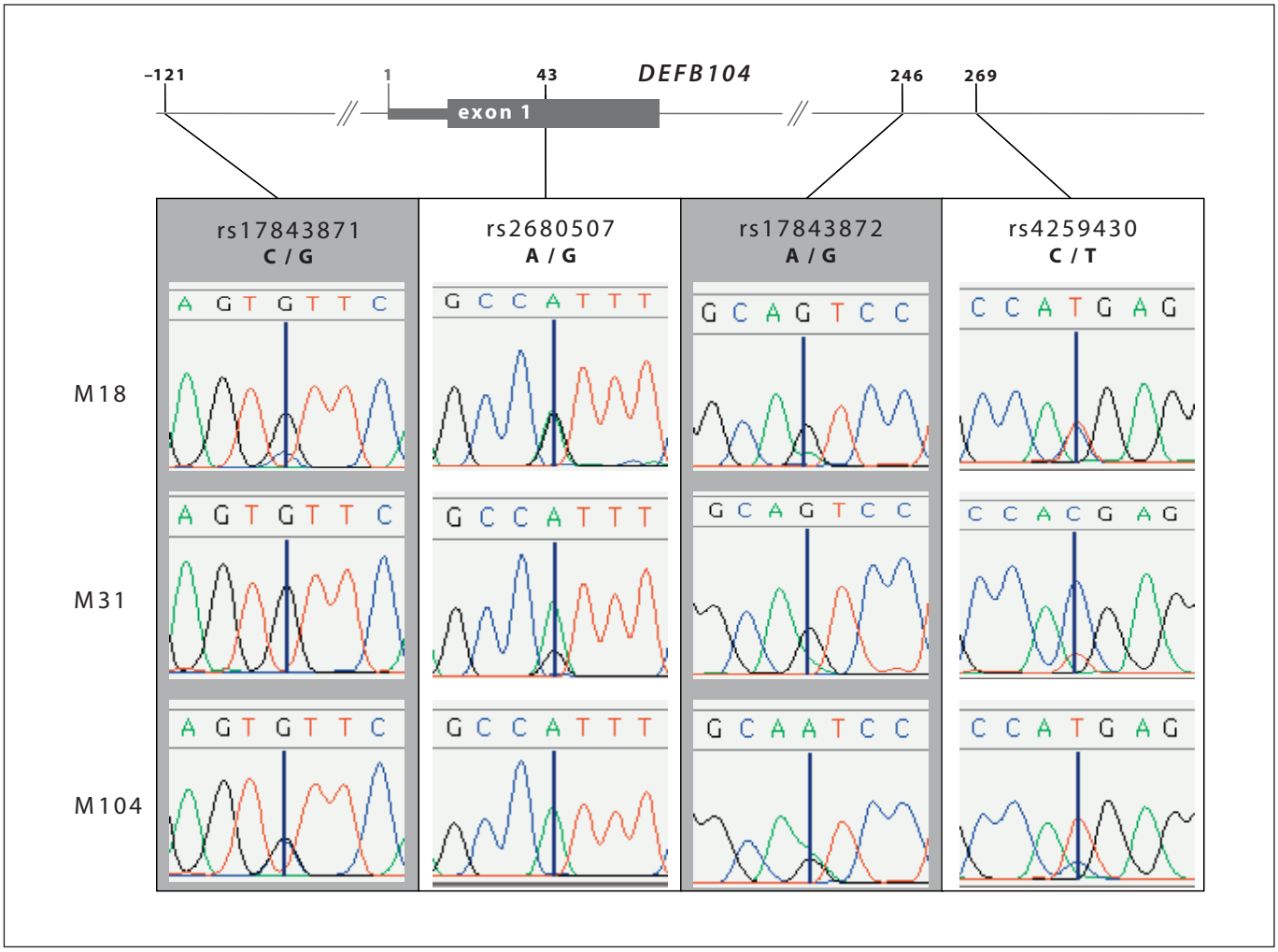

Fig. 2. Traces from the sequencing of PCR products derived from three individuals (M18, M31, M104). Variation of the peak heights for the different alleles at the MSV positions indicates differences in copy numbers as well as in the distribution of genetic variants within these copies.

metric or skewed heterozygote MSV patterns were also observed.

Haplotype Frequencies Differ Significantly between PC Patients and Controls

To analyze the potential association between particular MSV patterns and PC predisposition, we applied a PCR/cloning-based haplotyping method [5]. We performed PCR on the individual samples and pooled the resulting amplicons separately for PC patients (cohort 1) and controls, respectively. We then subcloned the two PCR product pools as well as individual amplicons from three PC cell lines and assessed the haplotype frequencies of the four MSVs within the amplified region (table 1). The most common haplotype in both groups, GAGC, is the only one for which GenBank and TraceDB entries support the existence in great apes (chimpanzee and baboon). Therefore, it likely represents the ancestral haplotype.

Prior to further quantitative analyses, it must be taken into account that post-PCR pooling leads to an overrepresentation of alleles derived from patients with low DEFB cluster copy number, whereas alleles from genomes with a high copy number are underrepresented. If there is a preference of certain haplotypes to be present in high- or low-copy-number genomes, the outcome of post-PCR pooling would be biased. Cross-validation of post- versus pre-PCR pooling could not be performed on DNA samples collected from PC patients and controls. Although

Table 1. *DEFB104* haplotype distribution in two independent cohorts of PC patients, control individuals and PC cell lines

HT	PC cohort 1		PC cohort 2		Control		Genomic pool		HTf	DU145		LNCaP		PC3		PC cohort vs. control (P)	
	n	f	n	f	n	f	n	f		n	f	n	f	n	f	cohort 1	cohort 2
GAGC	369	0.515	263	0.411	219	0.397	82	0.446	0.442	–	–	43	0.672	13	0.464	0.005	n.s.
GGGC	154	**0.215**	116	**0.181**	179	**0.325**	61	0.331	0.263	–	–	21	0.328	15	0.536	**0.003**	**$<10^{-4}$**
GAAT	104	**0.145**	81	**0.127**	36	**0.065**	9	0.049	0.096	–	–	–	–	–	–	**0.002**	**0.02**
CAGT	44	0.061	30	0.047	27	0.049	2	0.011	0.042	–	–	–	–	–	–	n.s.	n.s.
GAAC	26	**0.036**	18	**0.028**	–	–	1	0.005	0.017	34	0.479	–	–	–	–	**0.001**	**0.003**
GAGT	10	0.014	38	0.059	15	0.027	5	0.027	0.032	–	–	–	–	–	–	n.s.	n.s.
CAGC	8	0.011	32	0.050	12	0.022	8	0.044	0.032	–	–	–	–	–	–	n.s.	n.s.
CAAT	1	**0.001**	34	**0.053**	63	**0.114**	16	0.087	0.064	–	–	–	–	–	–	**$<10^{-4}$**	**0.008**
GGAT	–	–	18	0.028	–	–	–	–	0.007	–	–	–	–	–	–	–	} $<10^{-4}$
GGAC	–	–	5	0.008	–	–	–	–	0.002	37	0.521	–	–	–	–	–	
GGGT	–	–	5	0.008	–	–	–	–	0.002	–	–	–	–	–	–	–	

Haplotypes refer to the order of alleles at MSVs rs17843871, rs2680507, rs17843872 and rs4259430. Figures in bold represent common haplotypes (HTf >0.01) with a significant frequency difference between the control group and both PC cohorts. P = Statistical significance of differences in the haplotype distribution between PC patients and control individuals; p values were obtained by stepwise simulation (for details, see Materials and Methods); HT = haplotype; n = clone number; f = fraction; HTf = average haplotype frequency in PC cohorts 1, 2, controls and genomic pool; n.s. = not significant; – = not observed.

Table 2. Comparison of MSV major allele frequencies obtained via subcloning/Sanger sequencing vs. Pyrosequencing

MSV	MA	PC cohort 1			Control			Genomic pool			DU145			LNCaP			PC3		
		cloning		PS	cloning		PS	cloning		PS	cloning		PS	cloning		PS	cloning		PS
		(n = 716)			(n = 551)			(n = 184)			(n = 71)			(n = 64)			(n = 28)		
rs17843871	G	663	0.93	0.92	449	0.81	0.84	158	0.86	0.84	–	–	–	–	–	–	–	–	–
rs2680507	A	562	0.78	0.79	372	0.68	0.69	123	0.67	0.71	34	0.48	0.49	43	0.67	0.64	13	0.46	0.51
rs17843872	G	585	0.81	0.81	452	0.82	0.82	158	0.86	0.83	–	–	–	–	–	–	–	–	–
rs4259430	C	557	0.78	0.78	410	0.74	0.73	152	0.83	0.79	–	–	–	–	–	–	–	–	–

MA = Major allele; n = total number of analyzed alleles; PS = Pyrosequencing; – = not polymorphic.

DNA isolation has been done by the same procedure, differences in sample quality could not be avoided (data not shown). Instead, we used a commercially available pool of genomic DNA samples derived from approximately 100 individuals as pre-PCR control. For this pool it has been shown that SNP allele frequencies are equal to those for the Caucasian population, indicating homogeneous representation of individual DNA samples. However, no information is available concerning age and gender of the included individuals. When estimated from the genomic pool, the observed haplotype distribution was not found to differ significantly from that of the post-PCR pool of control individuals (simulation p = 0.26, global χ^2 test; table 1). Though due to different DNA samples this control experiment cannot completely rule out biases introduced by postamplification pooling, it reduces the possibility that a bias is introduced by this approach.

To ensure that the analyzed clones and their numbers were not subjected to haplotype-related biases during propagation in *Escherichia coli*, we compared the allele frequencies of the four MSVs determined by subcloning with those obtained by Pyrosequencing. Results from both methods were in reasonable agreement (table 2). Moreover, for the clones with the two most abundant haplotypes (GAGC, GGGC), we assessed the insert orientation with respect to the vector and found a ratio close to 1:1. This was also the case for the combination of all other clones, indicating unbiased in vivo clone

propagation. Taken together, the described control experiments highlight the validity of our haplotyping approach.

Comparison of the haplotype distribution between PC patients (cohort 1) and control individuals revealed a highly significant difference (global $\chi^2 = 146.094$, simulation $p < 10^{-4}$). The frequencies of five of the observed haplotypes differed significantly between the two groups (simulation $p < 0.05$, haplotypewise χ^2 statistics; table 1). To verify these findings, we replicated our analysis in an independent, age-matched PC sample (cohort 2). Remarkably, four common haplotypes (allele frequency >1%) showed a significant frequency difference between the control group and both PC cohorts (highlighted in bold in table 1), with GAAC only present in the PC groups.

Haplotype Distribution in PC Cell Lines

Haplotype analysis of PC cell lines revealed that both androgen-responsive (LNCaP) and hormone-independent cells (DU145, PC3) contained only two haplotypes per genome (table 1). For LNCaP and PC3, these were the most common ones (GAGC and GGGC). DU145, in contrast, contains relatively rare haplotypes: (1) GAAC – enriched in the PC groups, and (2) GGAC – found at minor frequency in PC cohort 2, but neither in PC cohort 1 nor the controls.

DEFB CNV in PC Patients

To assess the copy number of the DEFB cluster, we adapted a paralogue ratio test [13, 27] that compares a retroposed *HSPD1* pseudogene *(HSPDP)* within the DEFB cluster (fig. 1) with five paralogues on chromosomes 2, 3, 4, 5 and 12 as calibrator loci. Quantification of one diagnostic MSV (positions within hg18 annotated DEF clusters are chr8:7265040 and chr8:7786367, respectively) by Pyrosequencing revealed that both PC cohorts lack individuals with a copy number >9 (for individual copy numbers see online supplement material www.karger.com/doi/10.1159/000135688), which is a significant difference to control individuals ($p < 0.01$ for PC cohorts 1 and 2, Fisher's exact test; fig. 3). Furthermore, both PC cohorts have a lower median number of *HSPDP* copies than controls, whereas the difference in the overall copy number distribution is statistically significant only for cohort 1 ($p < 0.001$, Wilcoxon rank sum test; fig. 3). Copy number determination in the PC cell lines seemed to be inappropriate because it is known that these cells exhibit unusual karyotypes and, consequently, an unknown number of reference copies that will deteriorate any calculation.

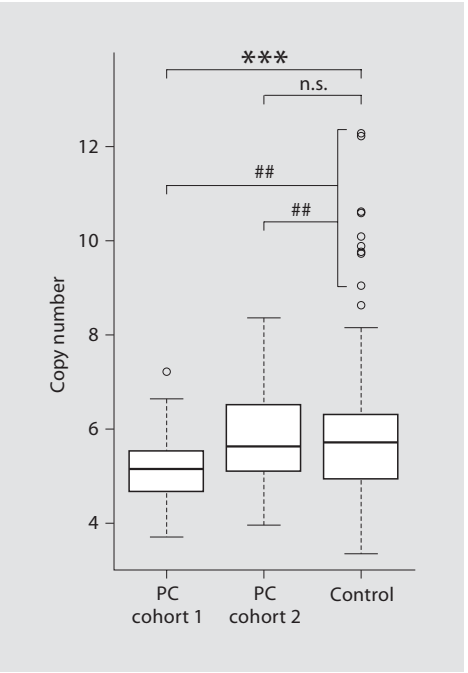

Fig. 3. Copy number distribution in PC cohort 1 and 2 and in controls shown as box plots. Boxes represent the region between the first and the third quartile (i.e. the central 50% of the data). The bold horizontal line represents the median. The ends of the dashed vertical lines mark the smallest and largest copy number located less than 1.5 times the interquartile distance from the first and third quartile, respectively. Other observations may be deemed outliers and are shown as open circles. *** $p < 0.001$ in a Wilcoxon rank sum test; ## $p < 0.01$ in a post hoc Fisher's exact test, respectively, comparing the number of individuals with copy number below and above 9. n.s. = Not significant.

Discussion

Convincing epidemiological evidence supports the role of chronic infection and inflammation as a pivotal step in PC carcinogenesis [28]. A high number of sexual partners and a history of sexually transmitted infections have been described as an independent risk factor for PC [31, 32]. The presence of bacterial signatures in the prostate gland under physiological conditions [33] and the finding that variants in the innate immune receptor TLR4 [34] are associated with an increased risk for early PC development have asserted the proposed link between the defense against persisting infections and prostate carcinogenesis [35]. Increasing evidence indicates that persistent mucosal or epithelial cell colonization by microorganisms in general induces carcinogenesis via the concomitant inflammatory response [16, 17]. This qualifies

antibacterial peptides such as defensins as candidates for an important role also in cancer development and manifestation. Thereby defensins may be expressed in prostate tissue [36] and/or immunocompetent cells [7] attracted to the sites of tumorigenesis.

In the present study, we demonstrate for the first time an association between genetic variants in the DEFB locus, a major player in the innate immune defense, and the risk of PC. The investigation was only possible because new technologies were developed. Conventional genotyping assays would fail for this locus due to CNV ranging between 2 and 12 in an individual genome [14, 15] (fig. 2). In previous efforts to refine the genomic structure of the 8p23 locus [5], we determined haplotypes of *DEFB4* and *DEFB104* in four individual genomes by a newly established PCR/cloning-based method. Five different haplotypes, defined by four MSVs, were identified for *DEFB104*. In contrast, only four haplotypes were found for *DEFB4*, although haplotypes at the analyzed locus of similar size are determined by five variable positions. Due to its higher diversity and its central position within the DEFB cluster, we chose *DEFB104* for a detailed analysis to characterize new genetic markers for PC predisposition. As the PCR/cloning-based haplotyping method [5] is not applicable to large sample numbers, we performed a post-PCR pooling strategy. This approach did not introduce a significant bias as results were comparable to a pre-PCR pooled sample (table 1). Moreover, it should be noted that, in contrast to our gender-matched male PC controls, the commercially available pre-PCR pool contained genomic DNA from both males and females. The absence of a difference between the haplotype distributions of the controls and the pre-PCR pool is therefore in agreement with the expectations for the autosomal locus 8p23, and does not provide any evidence for gender-specific haplotype patterns.

The frequencies of four common haplotypes differed significantly between each of two independent, age-matched PC Caucasian patient cohorts and controls, respectively. Remarkably, haplotype GAAC is seen in the PC cohorts (3.6 and 2.8%, respectively), is absent in the controls, and represented by just one clone obtained from the pooled genomic DNA. In contrast, CAAT was found to be strikingly underrepresented in the cancer groups (0.1 and 5.3%, respectively) but to be fairly abundant among the controls (11.4%). Since haplotypes of individual clones, as used in our experiments, do not represent statistically independent observations, the significance of the observed differences between PC patients and controls had to be assessed by simulation. One crucial assumption underlying this simulation was the statistical independence of the *DEFB104* haplotype in a given DEFB cluster from both the copy number and the haplotypes present in other DEFB copies on the same chromosome. According to our current understanding of CNVs [37–40], however, the larger range of possible DEFB cluster copy numbers [5, 11, 14, 15], the large size of ~350 kb of the repeat unit, and the inverse orientation of the two most outward copies predispose the locus to frequent intra- and interchromosomal recombination, unequal crossing-over and gene conversion [5], thereby justifying the setup of our simulations. Moreover, the simulations were based upon haplotype and copy number distributions as obtained from the combined data sets of similar numbers of PC patients and control individuals. In comparison to a random control data set, this approach resulted in data pools that were highly enriched by PC patients. That significant differences were observed despite these constraints is a strong argument that our observation is valid.

Against the background of the extraordinary plasticity of the DEFB locus and its wide range of CNV, an important question concerns possible differences in copy numbers among cells of the same individual as a result of mitotic recombination that in turn may increase with age. This also challenges our assumption that copy numbers determined in blood are representative of other normal tissues/organs including prostate. There is, however, currently no evidence for somatic DEFB mosaicism [13]. Furthermore, the inheritance of copy numbers is in >99.8% compatible with Mendelian rules [40], excluding frequent de novo recombination.

Only the second MSV (rs2680507, A→G) within the haplotyped region is a nonsynonymous coding variant and entails an amino acid exchange (I→V). Among the PC risk-associated haplotypes, the one carrying the G allele (GGGC) is overrepresented in controls, whereas two of the A allele-carrying haplotypes (GAAT, GAAC) are more and the third (CAAT) is less abundant in PC patients. It should, however, be stated that currently there is no reason to regard any of the PC-associated *DEFB104* haplotypes as a causative genetic variant. The *DEFB104* locus was selected as genetic marker for the analyses simply due to methodological reasons (high MSV density). Most likely, the conspicuous haplotypes are linked to causative variations somewhere else in the *DEF* cluster, a question that can only be resolved by supplementing non-genetic studies.

We estimated the copy number of the DEFB cluster in patients and controls adapting a paralogue ratio test [27]

by calibrating a DEFB cluster-specific *HSPDP* MSV to autosomal loci via Pyrosequencing.

In parallel, a very similar high-throughput method was recently developed by others [13] using a *Hae*III fragment length polymorphism. Although earlier studies [15] suggest that the DEFB cluster is always inherited as a unit resulting in equal copy numbers of all defensin genes and *HSPDP*, this was only assumed [11, 12] but never explicitly shown. Only recently could we confirm this assumption experimentally by applying a multiplex ligation-dependent probe amplification assay [41] and therewith justifying the application of the paralogue ratio test to assess the copy number of the DEFB cluster. Remarkably, both PC cohorts consistently lacked individuals with copy numbers >9 (fig. 3). Notably, low copy numbers of the DEFB cluster give rise to inflammatory disorders like Crohn's disease [11]. This disease is associated with an increased risk of colorectal cancer [42], but an association with PC has not yet been described. It would, therefore, be of interest to analyze the appropriate DEF haplotype composition in Crohn's disease patients.

In summary, we have performed the first disease association study of a locus, considering both copy number and sequence variation. We were able to correlate sequence variants within the DEFB cluster with PC in two independent, age-matched Caucasian patient cohorts and postulate that certain *DEFB104* haplotypes may serve as predictive markers for PC. DEFB cluster copy numbers higher than nine are significantly underrepresented in both patient samples. This adds a new level of complexity to the currently emerging research focus on the functional impact of CNVs in the human genome and on their role in complex diseases [43, 44]. Although the functional consequences of the described genetic variants are not yet resolved, the results point to a pivotal role of a perturbed antimicrobial barrier defense and/or of the defensin-triggered inflammation response in the etiopathogenesis of PC [19, 28]. Further prospective studies are necessary to show whether the respective haplotypes, in connection with copy number determination, may serve as reliable and feasible genetic markers in PC diagnostics and prognostics.

Acknowledgments

We thank Ivonne Görlich and Beate Szafranski for skillful technical assistance. This work was supported by grants from the Deutsche Forschungsgemeinschaft to K.H. and J.H. (Hu 498/3-1), from the German Ministry of Education and Research to S.S. and M.K. (01GS0426), to P.R. (NUW-S23T16) and to M.P. (01GR0504, 0313652D) as well as from the Wilhelm Sander Stiftung to M.P. (2005.045.1) and G.B. (2006.053.1).

References

1 Jemal A, Murray T, Ward E, Samuels A, Tiwari RC, Ghafoor A, Feuer EJ, Thun MJ: Cancer statistics, 2005. CA Cancer J Clin 2005;55:10–30.

2 Xu J, Zheng SL, Hawkins GA, Faith DA, Kelly B, Isaacs SD, Wiley KE, Chang B, Ewing CM, Bujnovszky P, Carpten JD, Bleecker ER, Walsh PC, Trent JM, Meyers DA, Isaacs WB: Linkage and association studies of prostate cancer susceptibility: evidence for linkage at 8p22-23. Am J Hum Genet 2001;69:341–350.

3 Wiklund F, Jonsson BA, Goransson I, Bergh A, Gronberg H: Linkage analysis of prostate cancer susceptibility: confirmation of linkage at 8p22-23. Hum Genet 2003;112:414–418.

4 Chang BL, Liu W, Sun J, Dimitrov L, Li T, Turner AR, Zheng SL, Isaacs WB, Xu J: Integration of somatic deletion analysis of prostate cancers and germline linkage analysis of prostate cancer families reveals two small consensus regions for prostate cancer genes at 8p. Cancer Res 2007;67:4098–4103.

5 Taudien S, Galgoczy P, Huse K, Reichwald K, Schilhabel M, Szafranski K, Shimizu A, Asakawa S, Frankish A, Loncarevic IF, Shimizu N, Siddiqui R, Platzer M: Polymorphic segmental duplications at 8p23.1 challenge the determination of individual defensin gene repertoires and the assembly of a contiguous human reference sequence. BMC Genomics 2004;5:92.

6 Ganz T: Defensins: antimicrobial peptides of innate immunity. Nat Rev Immunol 2003;3:710–720.

7 Oppenheim JJ, Biragyn A, Kwak LW, Yang D: Roles of antimicrobial peptides such as defensins in innate and adaptive immunity. Ann Rheum Dis 2003;62(suppl 2):ii17–ii21.

8 Quinones-Mateu ME, Lederman MM, Feng Z, Chakraborty B, Weber J, Rangel HR, Marotta ML, Mirza M, Jiang B, Kiser P, Medvik K, Sieg SF, Weinberg A: Human epithelial beta-defensins 2 and 3 inhibit HIV-1 replication. Aids 2003;17:F39–F48.

9 Zhang L, Yu W, He T, Yu J, Caffrey RE, Dalmasso EA, Fu S, Pham T, Mei J, Ho JJ, Zhang W, Lopez P, Ho DD: Contribution of human alpha-defensin 1, 2, and 3 to the anti-HIV-1 activity of CD8 antiviral factor. Science 2002;298:995–1000.

10 Fahlgren A, Hammarstrom S, Danielsson A, Hammarstrom ML: Beta-defensin-3 and -4 in intestinal epithelial cells display increased mRNA expression in ulcerative colitis. Clin Exp Immunol 2004;137:379–385.

11 Fellermann K, Stange DE, Schaeffeler E, Schmalzl H, Wehkamp J, Bevins CL, Reinisch W, Teml A, Schwab M, Lichter P, Radlwimmer B, Stange EF: A chromosome 8 gene-cluster polymorphism with low human beta-defensin 2 gene copy number predisposes to Crohn disease of the colon. Am J Hum Genet 2006;79:439–448.

12 Hollox EJ, Huffmeier U, Zeeuwen PL, Palla R, Lascorz J, Rodijk-Olthuis D, van de Kerkhof PC, Traupe H, de Jongh G, den Heijer M, Reis A, Armour JA, Schalkwijk J: Psoriasis is associated with increased beta-defensin genomic copy number. Nat Genet 2008;40:23–25.

13 Armour JA, Palla R, Zeeuwen PL, den Heijer M, Schalkwijk J, Hollox EJ: Accurate, high-throughput typing of copy number variation using paralogue ratios from dispersed repeats. Nucleic Acids Res 2007;35:e19.

14 Hollox EJ, Armour JA, Barber JC: Extensive normal copy number variation of a beta-defensin antimicrobial-gene cluster. Am J Hum Genet 2003;73:591–600.

15 Linzmeier RM, Ganz T: Human defensin gene copy number polymorphisms: comprehensive analysis of independent variation in alpha- and beta-defensin regions at 8p22-p23. Genomics 2005;86:423–430.

16 Karin M, Lawrence T, Nizet V: Innate immunity gone awry: linking microbial infections to chronic inflammation and cancer. Cell 2006;124:823–835.

17 Lu H, Ouyang W, Huang C: Inflammation, a key event in cancer development. Mol Cancer Res 2006;4:221–233.

18 Conejo-Garcia JR, Benencia F, Courreges MC, Kang E, Mohamed-Hadley A, Buckanovich RJ, Holtz DO, Jenkins A, Na H, Zhang L, Wagner DS, Katsaros D, Caroll R, Coukos G: Tumor-infiltrating dendritic cell precursors recruited by a beta-defensin contribute to vasculogenesis under the influence of Vegf-A. Nat Med 2004;10:950–958.

19 Candille SI, Kaelin CB, Cattanach BM, Yu B, Thompson DA, Nix MA, Kerns JA, Schmutz SM, Millhauser GL, Barsh GS: A beta-defensin mutation causes black coat color in domestic dogs. Science 2007;318:1418–1423.

20 Donald CD, Sun CQ, Lim SD, Macoska J, Cohen C, Amin MB, Young AN, Ganz TA, Marshall FF, Petros JA: Cancer-specific loss of beta-defensin 1 in renal and prostatic carcinomas. Lab Invest 2003;83:501–505.

21 Schullerus D, von Knobloch R, Chudek J, Herbers J, Kovacs G: Microsatellite analysis reveals deletion of a large region at chromosome 8p in conventional renal cell carcinoma. Int J Cancer 1999;80:22–24.

22 Young AN, de Oliveira Salles PG, Lim SD, Cohen C, Petros JA, Marshall FF, Neish AS, Amin MB: Beta defensin-1, parvalbumin, and vimentin: a panel of diagnostic immunohistochemical markers for renal tumors derived from gene expression profiling studies using cDNA microarrays. Am J Surg Pathol 2003;27:199–205.

23 Wu SQ, Hafez GR, Zhang J, Newton M, Chen A, Lange J, Wilding G: Identification of the prostate cancer micro-foci with chromosome 8p deletion at the tumor interface area by histopathological-FISH parallel examination. Int J Oncol 2001;19:1143–1147.

24 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, Isaacs WB: Common sequence variants of the macrophage scavenger receptor 1 gene are associated with prostate cancer risk. Am J Hum Genet 2003;72:208–212.

25 Groth M, Huse K, Reichwald K, Taudien S, Hampe J, Rosenstiel P, Birkenmeier G, Schreiber S, Platzer M: Method for preparing single-stranded DNA templates for Pyrosequencing using vector ligation and universal biotinylated primers. Anal Biochem 2006;356:194–201.

26 Mars WM, Patmasiriwat P, Maity T, Huff V, Weil MM, Saunders GF: Inheritance of unequal numbers of the genes encoding the human neutrophil defensins HP-1 and HP-3. J Biol Chem 1995;270:30371–30376.

27 Deutsch S, Choudhury U, Merla G, Howald C, Sylvan A, Antonarakis SE: Detection of aneuploidies by paralogous sequence quantification. J Med Genet 2004;41:908–915.

28 De Marzo AM, Platz EA, Sutcliffe S, Xu J, Gronberg H, Drake CG, Nakai Y, Isaacs WB, Nelson WG: Inflammation in prostate carcinogenesis. Nat Rev Cancer 2007;7:256–269.

29 Bonfield JK, Smith K, Staden R: A new DNA sequence assembly program. Nucleic Acids Res 1995;23:4992–4999.

30 Fredman D, White SJ, Potter S, Eichler EE, Den Dunnen JT, Brookes AJ: Complex SNP-related sequence variation in segmental genome duplications. Nat Genet 2004;36:861–866.

31 Andersson SO, Baron J, Bergstrom R, Lindgren C, Wolk A, Adami HO: Lifestyle factors and prostate cancer risk: a case-control study in Sweden. Cancer Epidemiol Biomarkers Prev 1996;5:509–513.

32 Rosenblatt KA, Wicklund KG, Stanford JL: Sexual factors and the risk of prostate cancer. Am J Epidemiol 2001;153:1152–1158.

33 Hochreiter WW, Duncan JL, Schaeffer AJ: Evaluation of the bacterial flora of the prostate using a 16S rRNA gene based polymerase chain reaction. J Urol 2000;163:127–130.

34 Zheng SL, Augustsson-Balter K, Chang B, Hedelin M, Li L, Adami HO, Bensen J, Li G, Johnasson JE, Turner AR, Adams TS, Meyers DA, Isaacs WB, Xu J, Gronberg H: Sequence variants of toll-like receptor 4 are associated with prostate cancer risk: results from the Cancer Prostate in Sweden Study. Cancer Res 2004;64:2918–2922.

35 Alexeyev O, Bergh J, Marklund I, Thellenberg-Karlsson C, Wiklund F, Gronberg H, Bergh A, Elgh F: Association between the presence of bacterial 16S RNA in prostate specimens taken during transurethral resection of prostate and subsequent risk of prostate cancer (Sweden). Cancer Causes Control 2006;17:1127–1133.

36 Oudes AJ, Campbell DS, Sorensen CM, Walashek LS, True LD, Liu AY: Transcriptomes of human prostate cells. BMC Genomics 2006;7:92.

37 Bailey JA, Eichler EE: Primate segmental duplications: crucibles of evolution, diversity and disease. Nat Rev Genet 2006;7:552–564.

38 Freeman JL, Perry GH, Feuk L, Redon R, McCarroll SA, Altshuler DM, Aburatani H, Jones KW, Tyler-Smith C, Hurles ME, Carter NP, Scherer SW, Lee C: Copy number variation: new insights in genome diversity. Genome Res 2006;16:949–961.

39 Lupski JR: Genome structural variation and sporadic disease traits. Nat Genet 2006;38:974–976.

40 Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, Fiegler H, Shapero MH, Carson AR, Chen W, Cho EK, Dallaire S, Freeman JL, Gonzalez JR, Gratacos M, Huang J, Kalaitzopoulos D, Komura D, MacDonald JR, Marshall CR, Mei R, Montgomery L, Nishimura K, Okamura K, Shen F, Somerville MJ, Tchinda J, Valsesia A, Woodwark C, Yang F, Zhang J, Zerjal T, Zhang J, Armengol L, Conrad DF, Estivill X, Tyler-Smith C, Carter NP, Aburatani H, Lee C, Jones KW, Scherer SW, Hurles ME: Global variation in copy number in the human genome. Nature 2006;444:444–454.

41 Groth M, Szafranski K, Taudien S, Huse K, Mueller O, Rosenstiel P, Nygren AOH, Schreiber S, Birkenmeier G, Platzer M: High-resolution mapping of the 8p23.1 beta-defensin cluster reveals strictly concordant copy number variation of all genes. Hum Mutat, 2008 May 9. Epub ahead of print.

42 Bernstein CN, Blanchard JF, Kliewer E, Wajda A: Cancer risk in patients with inflammatory bowel disease: a population-based study. Cancer 2001;91:854–862.

43 Aitman TJ, Dong R, Vyse TJ, Norsworthy PJ, Johnson MD, Smith J, Mangion J, Roberton-Lowe C, Marshall AJ, Petretto E, Hodges MD, Bhangal G, Patel SG, Sheehan-Rooney K, Duda M, Cook PR, Evans DJ, Domin J, Flint J, Boyle JJ, Pusey CD, Cook HT: Copy number polymorphism in Fcgr3 predisposes to glomerulonephritis in rats and humans. Nature 2006;439:851–855.

44 Gonzalez E, Kulkarni H, Bolivar H, Mangano A, Sanchez R, Catano G, Nibbs RJ, Freedman BI, Quinones MP, Bamshad MJ, Murthy KK, Rovin BH, Bradley W, Clark RA, Anderson SA, O'Connell RJ, Agan BK, Ahuja SS, Bologna R, Sen L, Dolan MJ, Ahuja SK: The influence of CCL3L1 gene-containing segmental duplications on HIV-1/AIDS susceptibility. Science 2005;307:1434–1440.