Identification and replication of the interplay of four genetic high risk variants for urinary bladder cancer

Carcinogenesis, 2017, Vol. 38, No. 12, 1167–1179

doi:10.1093/carcin/bgx102
Advance Access publication September 27, 2017
Original Article

ORIGINAL ARTICLE

Identification and replication of the interplay of four genetic high-risk variants for urinary bladder cancer

Silvia Selinski[1,*], Meinolf Blaszkewicz[1], Katja Ickstadt[2], Holger Gerullis[3,4], Thomas Otto[3], Emanuel Roth[5], Frank Volkert[5], Daniel Ovsiannikov[6,7], Oliver Moormann[6], Gergely Banfi[8], Peter Nyirady[8], Sita H. Vermeulen[9], Montserrat Garcia-Closas[10], Jonine D. Figueroa[11], Alison Johnson[12], Margaret R. Karagas[13], Manolis Kogevinas[14,15,16,17], Nuria Malats[18], Molly Schwenn[19], Debra T. Silverman[10], Stella Koutros[10], Nathaniel Rothman[10], Lambertus A. Kiemeney[9], Jan G. Hengstler[1,†] and Klaus Golka[1,†]

[1]Systems Toxicology, Leibniz-Institut für Arbeitsforschung an der TU Dortmund, Leibniz Research Centre for Working Environment and Human Factors (IfADo), Dortmund, Germany, [2]Faculty of Statistics, TU Dortmund University, Dortmund, Germany, [3]Department of Urology, Lukasklinik Neuss, Neuss, Germany, [4]University Hospital for Urology, Klinikum Oldenburg, School of Medicine and Health Sciences, Carl von Ossietzky University Oldenburg, Oldenburg, Germany, [5]Department of Urology, Evangelic Hospital, Paul Gerhardt Foundation, Lutherstadt Wittenberg, Germany, [6]Department of Urology, St.-Josefs-Hospital, Dortmund-Hoerde, Germany, [7]Department of Urology and Pediatric Urology, Kemperhof Hospital, Koblenz, Germany, [8]Department of Urology, Semmelweis University Budapest, Budapest, Hungary, [9]Department for Health Evidence (133 HEV) and Department of Urology (659 URO), Radboud University Medical Center (Radboudumc), Nijmegen, The Netherlands, [10]Division of Cancer Epidemiology and Genetics, Department of Health and Human Services, National Cancer Institute (NCI), National Institutes of Health (NIH), Bethesda, MD 20892, USA, [11]Usher Institute of Population Health Sciences and Informatics, CRUK Edinburgh Centre, University of Edinburgh, Edinburgh, UK, [12]Vermont Department of Health, Vermont Cancer Registry, Burlington, VT 05401, USA, [13]Department of Epidemiology, Geisel School of Medicine at Dartmouth, Hanover, NH 03756, USA, [14]Cancer Program, ISGlobal, Centre for Research in Environmental Epidemiology (CREAL), Barcelona, Spain, [15]CIBER Epidemiology and Public Health (CIBER-ESP), Health Research Institute Carlos III, Madrid, Spain, [16]Hospital del Mar Medical Research Institute, Barcelona, Spain, [17]University Pompeu Fabra (UPF), Barcelona, Spain, [18]Genetic and Molecular Epidemiology Group, Spanish National Cancer Research Center (CNIO), Madrid, and CIBERONC, Spain and [19]Maine Department of Health and Human Services, Maine Cancer Registry, Augusta, ME 04333, USA

*To whom correspondence should be addressed. Tel: +49 231 1084 216; Fax: +49 231 1084 343; Email: Selinski@ifado.de
†These authors contributed equally to this work.

Abstract

Little is known whether genetic variants identified in genome-wide association studies interact to increase bladder cancer risk. Recently, we identified two- and three-variant combinations associated with a particular increase of bladder cancer risk in a urinary bladder cancer case–control series (Leibniz Research Centre for Working Environment and Human Factors at TU Dortmund (IfADo), 1501 cases, 1565 controls). In an independent case–control series (Nijmegen Bladder Cancer Study, NBCS, 1468 cases, 1720 controls) we confirmed these two- and three-variant combinations. Pooled analysis of the two studies as discovery group (IfADo-NBCS) resulted in sufficient statistical power to test up to four-variant combinations by a logistic regression approach. The New England and Spanish Bladder Cancer Studies (2080 cases and 2167 controls) were used as a replication series. Twelve previously identified risk variants were considered. The strongest four-variant combination was obtained in never smokers. The combination of rs1014971[AA] near *apolipoprotein B mRNA editing enzyme,*

Received: August 26, 2016; **Revised:** August 18, 2017; **Accepted:** September 18, 2017

catalytic polypeptide-like 3A (*APOBEC3A*) and *chromobox homolog 6* (*CBX6*), *solute carrier family 1s4* (*urea transporter*), *member 1* (*Kidd blood group*) (*SLC14A1*) exon single nucleotide polymorphism (SNP) rs1058396[AG, GG], *UDP glucuronosyltransferase 1 family, polypeptide A complex locus* (*UGT1A*) intron SNP rs11892031[AA] and rs8102137[CC, CT] near *cyclin E1* (*CCNE1*) resulted in an unadjusted odds ratio (OR) of 2.59 (95% CI = 1.93–3.47; $P = 1.87 \times 10^{-10}$), while the individual variant ORs ranged only between 1.11 and 1.30. The combination replicated in the New England and Spanish Bladder Cancer Studies ($OR_{unadjusted}$ = 1.60, 95% CI = 1.10–2.33; $P = 0.013$). The four-variant combination is relatively frequent, with 25% in never smoking cases and 11% in never smoking controls (total study group: 19% cases, 14% controls). In conclusion, we show that four high-risk variants can statistically interact to confer increased bladder cancer risk particularly in never smokers.

Abbreviations

APOBEC3A	*apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like 3A*
CBX6	*chromobox homolog 6*
CCNE1	*cyclin E1*
CI	confidence interval
FGFR3	*fibroblast growth factor receptor 3*
GSTM1	*glutathione S-transferase M1*
IfADo	Leibniz Research Centre for Working Environment and Human Factors at TU Dortmund
LR	likelihood ratio
MYC	*v-myc avian myelocytomatosis viral oncogene homolog*
NAT2	*N-acetyltransferase 2*
NBCS	Nijmegen Bladder Cancer Study
NBS	Nijmegen Biomedical Study
OR	odds ratio
PSCA	*prostate stem cell antigen*
SLC14A1	*solute carrier family 14* (*urea transporter*), *member 1* (*Kidd blood group*)
SNP	single nucleotide polymorphism
TACC3	*transforming, acidic coiled-coil containing protein 3*
TP63	*tumor protein p63*
UBC	urinary bladder cancer.
UGT1A	*UDP glucuronosyltransferase 1 family, polypeptide A complex locus*

Introduction

Twin studies have suggested that approximately 30% of urinary bladder carcinomas (UBC) can be attributed to genetic predisposition (1). Recently, genome-wide association studies have identified genetic variants at 15 locations associated with UBC risk (2). However, the identified variants show very small odds ratios (ORs) ranging between 1.11 and 1.24 (3). An exception is the homozygous deletion variant of *glutathione S-transferase M1* (*GSTM1*) with an OR ranging between 1.28 and 1.70 in large studies with a mean OR of about 1.43 in Caucasians (4–8).

Relatively little is known whether the identified genetic variants interact to modulate UBC risk (9). An open question is to which extent the common occurrence of several risk variants in an individual enhances risk. Recently, two large studies used a weighted allele score (or polygenic risk score) for each individual to model the consequences of common occurrence of risk alleles (10, 11). In the approach of these two studies, each analysed variant was assigned a score from zero to two risk alleles, which were summed to an overall weighted score for each individual. Weights were based on the estimated OR of each variant. Subsequently, ORs of the score quartiles were estimated. Analysis of 12 variants in the study group of García-Closas *et al.* by this technique resulted in an OR of 2.94 for the highest scores (>75% quantile) compared with the lowest (<25% quantile) (11). Similarly, analysis of seven single nucleotide polymorphism (SNPs) in a Chinese study group resulted in an OR of 2.58 (>75 versus <25% quantile) (10). A limitation of this weighted allele score approach is that it does not identify the specific genetic variants that in combination enhances UBC risk as it does not model particular interactions of SNPs. Therefore, in 2012, our group applied a logistic regression approach in which all possible combinations of seven high-risk variants were considered (12, 13). This approach identified specific three-variant combinations, where carriers of three high-risk alleles had higher ORs than carriers of only one of the alleles (12).

A limitation of this study in the Leibniz Research Centre for Working Environment and Human Factors at TU Dortmund (IfADo) case–control series in 2012 was that an independent study group for replication was not available. Also, higher order than three-variant combinations could not be studied because the case numbers in individual risk combination subgroups became smaller than 100, which did not allow analysis with sufficient statistical power. Meanwhile, the Nijmegen Bladder Cancer Study (NBCS), an independent ongoing case–control series comprising 1468 cases and 1720 controls (in the present analysis), has become available for this purpose (14–18). In the present study, we used this independent group to replicate the three-variant combinations previously identified in the total IfADo case–control series. In addition, the IfADo and NBCS case-control series were combined to achieve sufficiently high case numbers to identify four-variant combinations. These four-way combinations were further explored in the New England and Spanish Bladder Cancer Studies (6, 19–21).

Materials and methods

We used 2969 cases and 3285 controls from two case–control series in Germany (the multicentric IfADo case–control series, 1501 cases/1565 controls) and the Netherlands (NBCS, 1468 cases/1720 controls) with complete genotype data for the 12 investigated genetic variants as a discovery group and 2080 cases and 2167 controls from the New England and Spanish Bladder Cancer Studies with complete genotype data as a replication series (Table 1, Supplementary Table 1, is available at *Carcinogenesis* Online).

IfADo case–control series

In total, 1501 confirmed UBC cases and 1565 controls without malignant disease of European descent of four case–control series from Germany and Hungary were collected by the Leibniz Research Centre for Working Environment and Human Factors at TU Dortmund (IfADo). All participants provided written informed consent. Details are given in the Supplementary Materials and Methods, available at *Carcinogenesis* Online, and elsewhere (22).

Genotypes of rs11892031[A/C], rs1495741[A/G], rs1058396[A/G], rs17674580[C/T], rs2294008[C/T], rs2978974[A/G], rs1014971[A/G], rs710521[A/G], rs798766[C/T], rs8102137[C/T] and rs9642880[G/T] were detected via TaqMan® Assay (23). The homozygous *GSTM1* deletion was detected by the amplification of the *GSTM1* DNA sequence segment with 218 bp by means of PCR (24–26). Details are given in the Supplementary Materials and Methods, available at *Carcinogenesis* Online.

Table 1. Study group characteristics of the discovery and replication series

	Discovery series						Replication series	
	IfADo		NBCS		Combined		New England and Spanish Bladder Cancer Studies	
	Cases (%)	Controls (%)	Cases (%)	Controls (%)	Cases (%)	Controls (%)	Cases (%)	Controls (%)
Gender								
Female	305 (0.20)	570 (0.36)	265 (0.18)	864 (0.50)	570 (0.19)	1434 (0.44)	353 (0.17)	407 (0.19)
Male	1196 (0.80)	995 (0.64)	1189 (0.81)	843 (0.49)	2385 (0.80)	1838 (0.56)	1727 (0.83)	1760 (0.81)
Missing	0 (0.00)	0 (0.00)	14 (0.01)	13 (0.01)	14 (0.01)	13 (0.00)	0 (0.00)	0 (0.00)
Total	1501	1565	1468	1720	2969	3285	2080	2167
Smoking habits								
Ever	1109 (0.74)	890 (0.57)	960 (0.65)	1246 (0.72)	2069 (0.70)	2136 (0.65)	1696 (0.82)	1360 (0.63)
Current	404 (0.27)	297 (0.19)	272 (0.19)	362 (0.21)	676 (0.23)	659 (0.20)	776 (0.37)	428 (0.20)
Former	705 (0.47)	593 (0.38)	688 (0.47)	884 (0.51)	1393 (0.47)	1477 (0.45)	920 (0.44)	932 (0.43)
Never	300 (0.20)	658 (0.42)	106 (0.07)	457 (0.27)	406 (0.14)	1115 (0.34)	305 (0.15)	688 (0.32)
Missing	92 (0.06)	17 (0.01)	402 (0.27)	17 (0.01)	494 (0.17)	34 (0.01)	79 (0.04)	119 (0.06)
Age								
Min-max	20–95	20–100	25–93	27–79	20–95	20–100	22–77	20–76
Median (IQR)	68 (16)	67 (18)	64 (13)	63 (15)	66 (14)	65 (16)	67 (14)	66 (14)
Mean (SD)	66.9 (11.29)	63.13 (15.49)	62.1 (9.97)	61.51 (10.33)	64.76 (10.99)	62.28 (13.08)	65.4 (10)	64.4 (10.3)
20–55 years	246 (0.16)	386 (0.25)	289 (0.20)	476 (0.28)	535 (0.18)	862 (0.26)	369 (0.18)	434 (0.22)
56–64 years	366 (0.24)	303 (0.19)	358 (0.24)	501 (0.29)	724 (0.24)	804 (0.24)	457 (0.22)	477 (0.22)
65–71 years	355 (0.24)	402 (0.26)	363 (0.25)	453 (0.26)	718 (0.24)	855 (0.26)	580 (0.28)	643 (0.30)
72+ years	529 (0.35)	473 (0.30)	189 (0.13)	277 (0.16)	718 (0.24)	750 (0.23)	674 (0.32)	613 (0.28)

For detailed information on the analysed variants, see Supplementary Table 1, available at *Carcinogenesis* Online.
IQR, interquartile range (Q75–Q25%); SD, standard deviation.

NBCS

In the current study, we used data of 1468 cases with primary UBC from the NBCS and 1720 controls from the Nijmegen Biomedical Study (NBS). The combined NBS-NBCS served as the Dutch discovery population in previous UBC genome-wide association studies (14–18).

Genotypes of rs11892031[A/C], rs1495741[A/G], rs1058396[A/G], rs2294008[C/T], rs2978974[A/G], rs1014971[A/G], rs710521[A/G], rs798766[C/T], rs8102137[C/T] and rs9642880[G/T] were determined using the Illumina HumanCNV370 BeadChip as described elsewhere (14). Genotypes of rs17674580[C/T] were imputed using the IMPUTE v2.1 software as described elsewhere (14). *GSTM1* copy number variation status was determined by an Applied Biosystems TaqMan Copy Number assay (Assay ID: Hs02575461_cn).

New England and Spanish Bladder Cancer Studies

This published case–control series includes 2080 cases and 2167 controls as described previously (6, 19–21). Information on 10 genotypes determined in the IfADo and NBCS was also available from the New England and Spanish Bladder Cancer Studies. In case of the *solute carrier family 14 (urea transporter), member 1 (Kidd blood group)* (*SLC14A1*) SNPs rs1058396 and rs17674580, the SNPs rs10775480 and rs10853535 were used as proxies ($r^2 = 0.75$ and $r^2 = 0.66$; data from 1000 Genomes Project, CEU (phase 3), ensemble genome browser version 87). A summary of study group characteristics is given in Table 1.

Statistical analysis

Analyses were performed in the discovery group (total study group: IfADo and NBCS combined) and stratified for subgroups defined by smoking habits: never (less than 100 cigarettes/lifetime), former (stopped smoking before first diagnosis of UBC/recruitment), current (still smoking at UBC diagnosis/recruitment or just stopped smoking, for details, see Supplementary Materials and Methods, available at *Carcinogenesis* Online) and ever smokers (former and current smokers combined). 'Unadjusted' analyses were performed using asymptotic chi-squared tests, ORs and 95% confidence intervals (95% CIs). Logistic regression and the Wald test adjusted for age, gender, smoking habits (if applicable) and study site (if applicable) were used for 'adjusted' analyses. Unadjusted and adjusted effects of 'single variants' on UBC risk were analysed in the complete study group, in the subgroups defined by smoking habits and separately in the IfADo and in the Nijmegen study group. SNPs were analysed assuming a recessive, dominant and additive mode of inheritance of the minor allele.

We used the NBCS to replicate the results of a previous study in the IfADo case–control series on seven variants (*GSTM1*, rs11892031[A/C], rs710521[A/G], rs1495741[A/G], rs9642880[G/T], rs8102137[C/T] and rs1014971[A/G]) (12). The frequency of the combinations in NBCS, unadjusted *P* values, ORs and 95% CIs were estimated for the 10 best two- and three-variant combinations found in the total IfADo study group and in the subgroups defined by smoking habits. To avoid spurious findings, we considered results as relevant if combinations were present in at least 100 cases and 100 controls from NBCS.

To investigate relevant combinations of 12 variants in the 'combined IfADo and Nijmegen study group', we generated four binary variables for each of the 11 SNPs as described previously (12). These variables coded either for a dominant or a recessive mode of inheritance of the minor allele or the respective complements which are necessary to define all combinations of dominant and recessive genotypes. So, we defined for an SNP with major [A] and minor [B] alleles the risk and reference genotypes:

(i) dominant [B]: AB and BB (risk) versus AA (reference),
(ii) complement of dominant [B]: AA (risk) versus AB and BB (reference),
(iii) recessive [B]: BB (risk) versus AA and AB (reference) and
(iv) complement of recessive [B]: AA and AB (risk) versus BB (reference).

Two binary variables encoded either the *GSTM1 null* or the *GSTM1 present* genotype as risk factor. We used these 46 binary variables to define variant combinations. The 'genotype at risk' of a combination was defined as presence versus absence (reference) of a particular variant combination. *P* values (unadjusted) of the asymptotic chi-squared tests were computed for all two-, three- and four-variant combinations and used as an ordering criterion. The 10 lowest nominal *P* values identified the 10 best one- to four-variant combinations in the total study group and in the smoking habits subgroups. The analyses were restricted to combinations with a

minimum frequency of 100 cases and 100 controls in the risk and in the reference group.

Two separate analyses were performed for the total study group as well as for the smoking habits subgroups: 'Analysis I' comprised a detailed analysis of the best one- to four-variant combinations. 'Analysis II' of the 10 best combinations was used to confirm the results of 'Analysis I' using a set of combinations with similarly low *P* values.

Analysis I

For the 'best (lowest unadjusted *P* value) individual variants and two-, three- and four-variant combinations' unadjusted and adjusted *P* values, ORs and 95% CIs were estimated in the discovery group (IfADo and NBCS combined) as well as in the independent case–control series of the New England and Spanish Bladder Cancer Studies.

The stability of the ORs of the best combinations was investigated by bootstrap sampling in the discovery group as described previously (12). Likelihood ratio (LR) tests were used to check whether the effect of the variant combinations was due to multiplicative interaction or due to main effects of the single variants in the combination. To test whether the unadjusted ORs of the *k*-variant combinations increase significantly with increasing number *k* of combined variants, LR tests were used adding successively the best single variant, the best two-, three- and four-variant combinations to a logistic regression model. Similarly, we used LR tests to check whether this increase in ORs could be achieved alone by adding the single variants, which were present in the best combinations, as main effects. Finally, we checked the relevance of addition of a main effect and an interaction term (multiplicative) in one step by LR tests.

Analysis II

The 'ten best two-, three- and four-variant combinations' were used to compare subgroup (never versus former and current, never versus ever smokers) differences regarding the frequency of the single variants in the *k*-variant combinations. Exact chi-squared tests were used to compare never, former and current smokers and Fisher's exact tests for pairwise comparisons. Increase of the ORs with increasing number *k* of combined variants was tested using the Tukey test.

All calculations were performed using the software R, version 3.0.1 (R Development Core Team, 2014) and SAS/STAT, versions 9.3 and 9.4 (SAS Institute Inc., Cary, NC). Details of the statistical analysis are given in the Supplementary Materials and Methods.

Results

Replication of previously reported two- and three-variant combinations in the NBCS

In a previous study, we reported ORs of seven genetic variants individually, as well as of two- and three-variant combinations in the IfADo study group (12, 13). In a first step of the present study, we performed similar analyses in the NBCS and estimated the individual and combination effects. Unadjusted ORs of the individual variants with respect to UBC risk were similar between both study groups (Supplementary Tables 1B, 2). In both study groups, IfADo and NBCS, the lowest *P* values were obtained for the *GSTM1* deletion variant and for rs9642880[TT]. Moreover, the previously reported strongest two- and three-variant combinations identified in the IfADo study group (12, 13) resulted in similar unadjusted ORs in the NBCS (Table 3). The top 10 two-variant combinations identified in the IfADo study group were all significant also in the NBCS and the ORs obtained from both independent groups never differed by more than 0.14 (Table 3). Also, the top 10 three-variant combinations identified in the IfADo study group were all significant in the NBCS (Table 3). The analyses of the individual variants as well as two- and three-variant combinations were repeated in the subgroups of ever, current, former and never smokers (Supplementary Tables 3 and 4, available at *Carcinogenesis* Online). Unadjusted ORs of the individual variants were also quite similar between both study groups stratifying for smoking habits (Supplementary Table 3). In current and never smokers, case numbers were too low ($N < 100$ per risk group in cases and in controls) to allow a comparison of the NBCS and the IfADo study groups (Supplementary Table 4, available at *Carcinogenesis* Online). However, in ever and former smokers when case and control numbers exceeded $N = 100$, similar ORs were obtained in both study groups (Supplementary Table 4, available at *Carcinogenesis* Online). Age and sex distributions were similar between both study groups, while the NBCS contained less never smokers compared with the IfADo case–control series (Table 1). In conclusion, both study groups were similar with respect to the influence of the individual variants as well as previously identified two- and three-variant combinations. Therefore, a combination of the two study groups with the aim to identify possible four-variant combinations seemed justified.

Interplay of high-risk genetic variants in the combined IfADo-Nijmegen case–control series: relevance of four-variant combinations

To study the possible interactions between genetic variants in a larger case–control series as previously possible (12), the aforementioned NBCS and IfADo study groups were combined resulting in a series of 2969 UBC cases and 3285 controls (Table 1). Data of 12 genetic variants were available in both study groups (Supplementary Table 1B). Five additional variants were analysed compared with the previous study (12). All analysed variants have been reported to be individually associated with UBC risk. In the present study, all but rs2978974 of the individually analysed variants were significantly associated in the unadjusted analysis (additive genetic model, Supplementary Table 1B). After adjustment for age, gender and smoking habits, 10 of them remained significant while rs2978974 ($OR_{additive} = 1.01$) and rs149571 ($OR_{additive} = 1.08$) were not significant assuming an additive, recessive or dominant mode of inheritance (Supplementary Tables 1B and 2, available at *Carcinogenesis* Online). The significant SNPs all showed ORs between 1.09 and 1.27 (additive model; significant recessive: 1.30–1.38; significant dominant 1.13–1.35; Supplementary Table 2).

All possible two-, three- and four-variant combinations were tested, amounting to a total of 118 888 analysed effects. The resulting best combinations (Analysis I) showed a continuous increase in unadjusted ORs for increasing variant numbers between one and four (Figure 1A and B; Table 3). A remarkable difference was obtained for smokers and never smokers. Variant combinations resulted in higher ORs for the never smokers. Ever smokers showed the lowest ORs, while former and current smokers ranged in between (Figure 1A and B). The total group (Table 3) showed similar results as the ever smokers, which can be explained by the fact that most of the cases (70%) were current or former smokers. Higher than four-variant combinations were not tested because of small case numbers and thus statistical power limitations. For up to four-variant combinations, the case numbers in all subgroups were higher than $N = 100$ (Supplementary Table 8, available at *Carcinogenesis* Online).

Next, we tested whether an increase in the number of combined variants resulted in significantly increased/decreased risk (Figure 2A, Supplementary Table 6A, available at *Carcinogenesis* Online). The best one- to four-variant combinations were added successively to the model, to test whether the resulting increases in the LR statistics are significant. For never smokers, significant increases in unadjusted ORs were obtained for all one- to four-variant combinations added in a stepwise manner. The strongest

Table 2. The previously reported top ten two- and three-variant combinations in the IfADo study group (12, 13) showed similar unadjusted ORs in the NBCS

	Schwender *et al.* (12)				NBCS									
		95% CI				95% CI			Variant combination		Reference			
Variant combination	OR	Lower	Upper	*P* value	OR	Lower	Upper	*P* value	Cases	Controls	Cases	Controls	Min. freq.	OR diff.
Ten best two-variant combinations														
rs11892031[AA] × *GSTM1 null*	1.42	1.23	1.64	**9.50×10^{-07}**	1.29	1.12	1.49	**0.0003**	725	740	743	980	725	0.13
rs9642880[GG, GT] × *GSTM1 present*	0.70	0.60	0.81	**1.57×10^{-06}**	0.69	0.60	0.80	**1.30×10^{-06}**	443	660	1025	1060	443	0.00
rs710521[AA, AG] × *GSTM1 null*	1.41	1.23	1.63	**1.67×10^{-06}**	1.32	1.15	1.52	**8.67×10^{-05}**	804	822	664	898	664	0.09
rs8102137[CC, CT] × *GSTM1 null*	1.43	1.23	1.66	**3.73×10^{-06}**	1.31	1.13	1.52	**0.0005**	500	487	968	1233	487	0.12
rs9642880[GT, TT] × *GSTM1 null*	1.39	1.21	1.61	**5.41×10^{-06}**	1.34	1.16	1.55	**4.66×10^{-05}**	662	653	806	1067	653	0.05
rs710521[AA, AG] × *GSTM1 present*	0.73	0.63	0.84	**1.45×10^{-05}**	0.81	0.70	0.93	**0.0031**	580	769	888	951	580	-0.08
rs11892031[AA, AC] × *GSTM1 null*	1.36	1.19	1.57	**1.46×10^{-05}**	1.30	1.13	1.49	**0.0003**	847	882	621	838	621	0.07
rs9642880[TT] × rs710521[AA, AG]	1.43	1.21	1.70	**2.92×10^{-05}**	1.36	1.15	1.61	**0.0003**	374	345	1094	1375	345	0.07
rs710521[AA] × *GSTM1 null*	1.39	1.19	1.62	**3.42×10^{-05}**	1.36	1.16	1.58	**9.52×10^{-05}**	483	457	985	1263	457	0.03
rs11892031[AA, AC] × *GSTM1 present*	0.74	0.65	0.86	**3.77×10^{-05}**	0.77	0.67	0.88	**0.0002**	616	835	852	885	616	-0.02
Ten best three-variant combinations														
rs710521[AA, AG] × rs11892031[AA] × *GSTM1 null*	1.48	1.28	1.70	**8.22×10^{-08}**	1.31	1.14	1.51	**0.0002**	687	691	781	1029	687	0.17
rs9642880[GG, GT] × rs710521[AA, AG] × *GSTM1 present*	0.67	0.58	0.78	**1.98×10^{-07}**	0.73	0.63	0.85	**3.97×10^{-05}**	417	606	1051	1114	417	-0.06
rs8102137[CC, CT] × rs11892031[AA] × *GSTM1 null*	1.49	1.27	1.74	**7.02×10^{-07}**	1.30	1.11	1.52	**0.0012**	427	413	1041	1307	413	0.19
rs710521[AA, AG] × rs8102137[CC, CT] × *GSTM1 null*	1.47	1.26	1.72	**1.19×10^{-06}**	1.36	1.17	1.59	**7.91×10^{-05}**	475	447	993	1273	447	0.11
rs710521[AA, AG] × rs11892031[AA, AC] × *GSTM1 null*	1.41	1.23	1.63	**1.68×10^{-06}**	1.31	1.14	1.51	**0.0001**	801	821	667	899	667	0.10
rs9642880[GT, TT] × rs11892031[AA] × *GSTM1 null*	1.42	1.23	1.65	**2.21×10^{-06}**	1.37	1.19	1.59	**2.15×10^{-05}**	571	545	897	1175	545	0.05
rs9642880[GG, GT] × rs11892031[AA, AC] × *GSTM1 present*	0.70	0.60	0.81	**2.38×10^{-06}**	0.69	0.60	0.80	**1.02×10^{-06}**	441	659	1027	1061	441	0.01
rs9642880[GT, TT] × rs710521[AA, AG] × *GSTM1 null*	1.42	1.22	1.64	**3.23×10^{-06}**	1.37	1.18	1.58	**1.85×10^{-05}**	627	607	841	1113	607	0.05
rs9642880[GG, GT] × rs1495741[AA, AG] × *GSTM1 present*	0.70	0.61	0.82	**3.42×10^{-06}**	0.70	0.60	0.82	**3.60×10^{-06}**	423	629	1045	1091	423	0.00
rs8102137[CC, CT] × rs11892031[AA, AC] × *GSTM1 null*	1.43	1.23	1.66	**4.20×10^{-06}**	1.31	1.13	1.52	**0.0004**	500	486	968	1234	486	0.11

The comparison of the ten best two- and three-variant combinations in the subgroups of smokers (ever, current and former) and never smokers are given in Supplementary Table 4, available at *Carcinogenesis* Online. "Reference" is the group of all genotypes except for the considered variant combination (i.e. the complement of the variant combination). For example, the reference group for variant combination "rs11892031 [AA] × *GSTM1 null*" is "rs11892031 [AC, CC] and or *GSTM1 present*".

Min. freq., lowest frequency of the risk combination and the reference genotypes, respectively, observed in cases and controls. OR diff., difference between the published OR (12) and the OR in the NBCS.

$P \leq 0.05$ are printed bold.

Table 3. Different variant combinations are relevant in smokers and never smokers

Subgroup	Number of variants	Variant combinations	Cases		Controls				95% CI		
			N	%	*N*	%	OR	1/OR	Lower	Upper	*P* value
A. Best combinations in the Discovery series (combined IfADo case–control series and NBCS)											
All	1	*GSTM1 present*	1252	0.42	1622	0.49	0.75	**1.34**	0.68	0.83	1.16×10^{-08}
2969 cases	2	*GSTM1 present* × rs9642880[GG, GT]	913	0.31	1283	0.39	0.69	**1.44**	0.62	0.77	6.95×10^{-12}
3285 controls	3	*GSTM1 present* × rs9642880[GG, GT] × rs17674580[CC, CT]	767	0.26	1129	0.34	0.67	**1.50**	0.60	0.74	2.59×10^{-13}
	4	*GSTM1 present* × rs9642880[GG, GT] × rs17674580[CC, CT] × rs2294008[CC, CT]	586	0.20	920	0.28	0.63	**1.58**	0.56	0.71	2.83×10^{-14}
Ever	1	*GSTM1 present*	876	0.42	1038	0.49	0.78	**1.29**	0.69	0.88	4.70×10^{-05}
2069 cases	2	*GSTM1 present* × rs9642880[GG, GT]	640	0.31	835	0.39	0.70	**1.43**	0.61	0.79	3.15×10^{-08}
2136 controls	3	*GSTM1 present* × rs9642880[GG, GT] × rs2294008[CC, CT]	490	0.24	684	0.32	0.66	**1.52**	0.57	0.75	1.86×10^{-09}
	4	*GSTM1 present* × rs9642880[GG, GT] × rs2294008[CC, CT] × rs798766[CC, CT]	464	0.22	660	0.31	0.65	**1.55**	0.56	0.74	6.17×10^{-10}
Current	1	*GSTM1 null*	402	0.59	349	0.53	**1.30**	0.77	1.05	1.62	1.66×10^{-02}
676 cases	2	*GSTM1 present* × rs710521[AG, GG]	108	0.16	153	0.23	0.63	**1.59**	0.48	0.83	9.03×10^{-04}
659 controls	3	*GSTM1 null* × rs11892031[AA] × rs1058396[AG, GG]	303	0.45	226	0.34	**1.56**	0.64	1.25	1.94	8.75×10^{-05}
	4	*GSTM1 null* × rs11892031[AA] × rs1058396[AG, GG] × rs1014971[AA, AG]	272	0.40	194	0.29	**1.61**	0.62	1.29	2.03	3.70×10^{-05}
Former	1	rs9642880[TT]	394	0.28	303	0.21	**1.53**	0.65	1.29	1.81	1.33×10^{-06}
1393 cases	2	rs9642880[GG, GT] × rs17674580[CC, CT]	831	0.60	1029	0.70	0.64	**1.55**	0.55	0.75	2.15×10^{-08}
1477 controls	3	rs9642880[TT] × rs8102137[CC, CT] × rs1495741[AA, AG]	229	0.16	135	0.09	**1.96**	0.51	1.56	2.45	6.67×10^{-09}
	4	rs9642880[TT] × rs8102137[CC, CT] × rs1495741[AA, AG] × rs17674580[CC, CT]	204	0.15	110	0.07	**2.13**	0.47	1.67	2.72	1.29×10^{-09}
Never	1	rs1014971[AA]	207	0.51	452	0.41	**1.53**	0.66	1.22	1.92	2.88×10^{-04}
406 cases	2	rs1014971[AA] × rs17674580[CT, TT]	142	0.35	252	0.23	**1.84**	0.54	1.44	2.36	1.33×10^{-06}
1115 controls	3	rs1014971[AA] × rs11892031[AA] × rs1058396[AG, GG]	161	0.40	276	0.25	**2.00**	0.50	1.57	2.54	1.81×10^{-08}
	4	rs1014971[AA] × rs11892031[AA] × rs1058396[AG, GG] × rs8102137[CC, CT]	100	0.25	125	0.11	**2.59**	0.39	1.93	3.47	1.87×10^{-10}
B. Replication series: New England and Spanish Bladder Cancer Studies											
All	1	*GSTM1 present*	823	0.40	1033	0.48	0.70	**1.42**	0.62	0.79	2.56×10^{-08}
2080 cases	2	*GSTM1 present* × rs9642880[GG, GT]	598	0.29	793	0.37	0.70	**1.43**	0.62	0.80	7.89×10^{-08}
2167 controls	3	*GSTM1 present* × rs9642880[GG, GT] × rs17674580[CC, CT][a]	482	0.23	646	0.30	0.71	**1.41**	0.62	0.82	1.42×10^{-06}
	4	*GSTM1 present* × rs9642880[GG, GT] × rs17674580[CC, CT][a] × rs2294008[CC, CT]	383	0.18	517	0.24	0.72	**1.39**	0.62	0.84	1.88×10^{-05}
Ever	1	*GSTM1 present*	642	0.39	606	0.46	0.76	**1.32**	0.65	0.88	1.79×10^{-04}
1654 cases	2	*GSTM1 present* × rs9642880[GG, GT]	506	0.31	502	0.38	0.72	**1.39**	0.62	0.84	2.98×10^{-05}

Table 3. *Continued*

Subgroup	Number of variants	Variant combinations	Cases		Controls				95% CI		
			N	%	*N*	%	OR	1/OR	Lower	Upper	*P* value
1325 controls	3	*GSTM1 present* × rs9642880[GG, GT] × rs2294008[CC, CT]	397	0.24	401	0.30	0.73	**1.37**	0.62	0.86	1.29×10^{-04}
	4	*GSTM1 present* × rs9642880[GG, GT] × rs2294008[CC, CT] × rs798766[CC, CT]	385	0.23	391	0.30	0.73	**1.37**	0.62	0.85	1.21×10^{-04}
Current	1	*GSTM1 null*	458	0.59	238	0.56	1.17	0.85	0.92	1.49	0.20
776 cases	2	*GSTM1 present* × rs710521[AG, GG]	115	0.15	86	0.20	0.70	**1.43**	0.51	0.95	0.02
428 controls	3	*GSTM1 null* × rs11892031[AA] × rs1058396[AG, GG][b]	290	0.37	118	0.28	**1.58**	0.63	1.22	2.05	4.96×10^{-04}
	4	*GSTM1 null* × rs11892031[AA] × rs1058396[AG, GG][b] × rs1014971[AA, AG]	270	0.35	114	0.27	**1.49**	0.67	1.14	1.93	3.03×10^{-03}
Former	1	rs9642880[TT]	198	0.22	188	0.20	1.09	0.92	0.87	1.36	0.47
920 cases	2	rs9642880[GG, GT] × rs17674580[CC, CT][a]	580	0.63	605	0.65	0.93	1.08	0.76	1.12	0.42
932 controls	3	rs9642880[TT] × rs8102137[CC, CT] × rs1495741[AA, AG]	113	0.12	94	0.10	1.25	0.80	0.94	1.67	0.13
	4	rs9642880[TT] × rs8102137[CC, CT] × rs1495741[AA, AG] × rs17674580[CC, CT][a]	80	0.09	74	0.08	1.11	0.90	0.80	1.54	0.55
Never	1	rs1014971[AA]	155	0.51	303	0.44	**1.32**	0.76	1.01	1.72	0.048
305 cases	2	rs1014971[AA] × rs17674580[CT, TT][a]	110	0.36	213	0.31	1.26	0.79	0.95	1.67	0.11
688 controls	3	rs1014971[AA] × rs11892031[AA] × rs1058396[AG, GG][b]	93	0.30	176	0.26	1.27	0.79	0.94	1.71	0.11
	4	rs1014971[AA] × rs11892031[AA] × rs1058396[AG, GG][b] × rs8102137[CC, CT]	54	0.18	81	0.12	**1.60**	0.63	1.10	2.33	0.013

The best (lowest *P* value, minimum frequency 100) single polymorphisms and the best two- to four-variant combinations discovered in the combined IfADo and NBCS, unadjusted OR, 95% CI and *P* values are shown (**A**). Furthermore, the unadjusted OR of the reference group (complement of the variant combination) compared with the variant combination is given (1/OR). Both, OR and 1/OR, are highlighted grey and are printed bold in case of exceeding one significantly. The best combinations found in current and never smokers could be replicated in the New England and Spanish Bladder Cancer Studies (**B**) using rs10775480 and rs10853535 as proxies for rs1058396 and rs17674580 ($r^2 = 0.75$ and $r^2 = 0.66$) (19). Results for the IfADo and Nijmegen study groups as well as age, gender, smoking habits and study group adjusted OR, 95% CIs and *P* values of the Wald test are given in Supplementary Table 5, available at *Carcinogenesis* Online. Performance of the best combinations in the other subgroups is shown in Supplementary Table 10, available at *Carcinogenesis* Online.

[a]rs10853535 was used as a proxy for rs17674580.

[b]rs10775480 was used as a proxy for rs1058396.

individual variant for never smokers, rs1014971[AG, GG], was significant in a univariate logistic regression model with an LR statistic of LR = 13.16, *P* = 0.0003 (Figures 1C and 2A; Supplementary Table 6A, available at *Carcinogenesis* Online). Addition of the best two-variant combination (rs1014971, rs17674580) to the logistic regression model with the best variant (rs1014971) alone increased the LR statistic significantly by LR = 9.92, *P* = 0.0016. Adding the best three-variant combination (rs1014971, rs1058396, rs11892031) to the model with the main (individual) effect and the two-variant combination further improved the LR statistic by LR = 9.75, *P* = 0.0018. Further inclusion of the best four-variant combination (rs1014971, rs1058396, rs11892031 and rs8102137), which now has to compete with the best individual, two- and three-variant combinations led to additionally improved significance (LR = 12.01, *P* = 0.0005). Similar constellations were obtained for the former smokers and for the total group (Figure 2A). For the ever and for the current smokers, significant improvements for the LR statistics were obtained for up to three-variant combinations but not for a further added four-variant combination.

To analyse whether the improvement of the model fit was due to addition of further individual variants present in the combinations or rather than interaction effects, we repeated the analysis adding successively the new individual variants in the two- to four-variant combinations as main effects in the logistic regression model (Supplementary Figure 1A and Table 6B, available at *Carcinogenesis* Online). Furthermore, we added both—interactions and corresponding individual variants—in each step to the model (Supplementary Figure 1B and Table 6C, available at *Carcinogenesis* Online). For the improvement of the model fit, the interactions were more relevant than the corresponding main effects. Additionally, we tested the significance of each two-, three- and four-way interaction in the presence of the corresponding individual variants as main effects separately for each top combination (Supplementary Table 7). Remarkably, the four-way interactions were still significant in a logistic regression model that also contained the corresponding main effects. For instance, in never smokers, the interaction term rs1014971[AA] × rs1058396[AG, GG] × rs11892031[AA] × rs8102137[CC, CT] (*P* = 0.0003) was significant in a common model containing also rs1014971[AA] (*P* = 0.4318), rs1058396[AG, GG] (*P* = 0.1236), rs11892031[AA] (*P* = 0.1207) and rs8102137[CC, CT] (*P* = 0.8005), which were not significant in presence of the interaction term (Supplementary Table 7).

The analyses of Figures 1 and 2A focussed on a single top performing combination. To study if the described patterns remain stable for further relevant combinations, we included the ten best combinations for one to four variants (Analysis II) into a box plot analysis (Figure 2B, Supplementary Table 8). A similar pattern was obtained for the ten top combinations (Figure 2B) compared with the best one (Figure 1). The "top ten approach" additionally offered the advantage that the assumed increase/decrease in the unadjusted ORs by increase of the number of combined variants could be analysed by pairwise comparisons. This illustrated that the increase in mean ORs from individual effects to four-variant combinations was significant for never, former and current smokers (Figure 2B). For the total group and ever smokers, mean ORs increased significantly combining up to three variants.

Different SNP combinations are relevant in ever and never smokers

As demonstrated in the previous paragraph, genetic variants interact with respect to UBC risk. Next, we analysed if the same or different variants were relevant in ever and never smokers. The variant with the lowest (unadjusted) *P* value in ever smokers

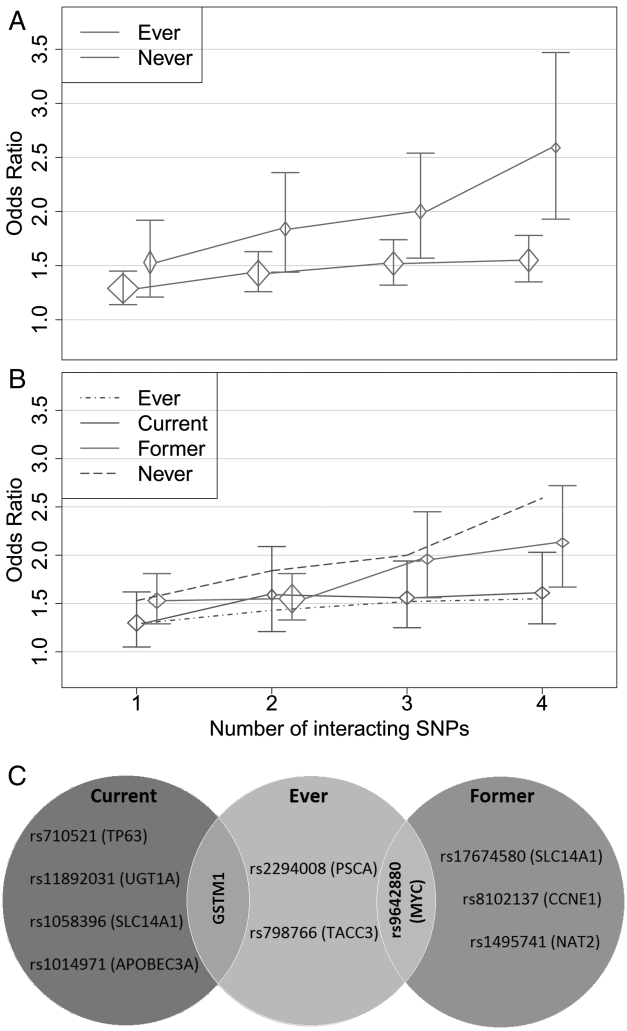

Figure 1. The best combinations of up to four risk variants showed a continuous increase in unadjusted ORs and a different composition in smoker subgroups. Combinations of 12 variants were analysed in the combined IfADo and Nijmegen case–control series. Unadjusted ORs of the best single variants, two-, three- and four-variant combinations in never and ever smokers (A) and in current and formers smokers (B, transparent lines indicate results of never and ever smokers as reference) are given. The height and width of the diamonds correspond to the square root of the combination frequency of controls and cases in the subgroup. Vertical bars indicate the 95% CIs. The overlap of polymorphisms in the top one- to four-variant combinations in the subgroup analyses of ever, current and former smokers indicated that *GSTM1* is more relevant for current smokers in contrast to rs9642880 (*MYC*) that seems to be more relevant in former smokers (C). Associated genes are given in parenthesis.

was the *GSTM1* deletion (Table 3). For two-variant combinations, SNP rs9642880 near *v-myc avian myelocytomatosis viral oncogene homolog* (*MYC*) together with *GSTM1* deletion resulted in the lowest *P* value. Next, the 5' UTR *prostate stem cell antigen* (*PSCA*) SNP rs2294008 and the *transforming, acidic coiled-coil containing protein 3* (*TACC3*) intron SNP rs798766 amended the most significant combinations in ever smokers.

A completely different sequence of best variant combinations was obtained for the never smokers (Table 3). The most significant individual variant was rs1014971 near *apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like 3A* (*APOBEC3A*). Next, rs1014971 was amended by the *SLC14A1* intron SNP rs17674580 to form the strongest two-variant combination. The three-variant combination included again rs1014971 together

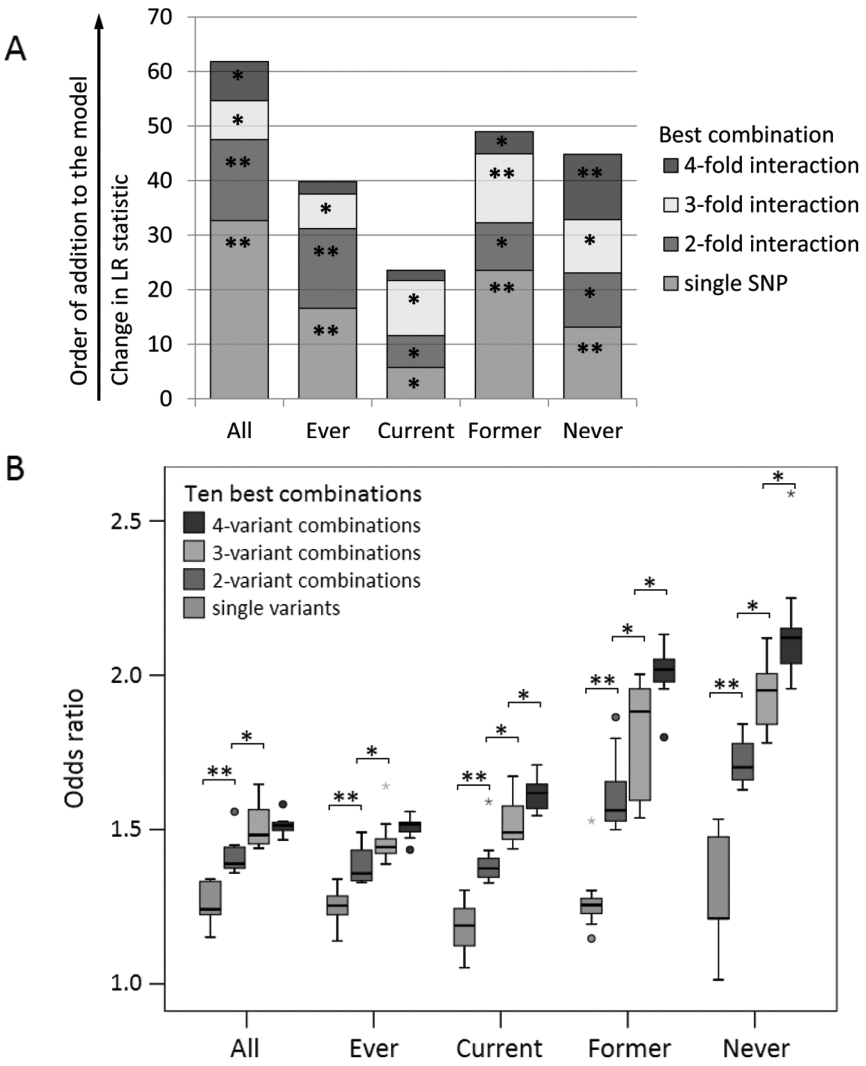

Figure 2. The increase of unadjusted ORs of up to four-variant combinations was significant for the best and the ten best combinations. (**A**) Impact of best one- to four-variant combinations on UBC risk in a common logistic regression model (without adjustment for further covariates) in the combined IfADo-Nijmegen study group ("All") and stratified by smoking habits ("Ever", "Current", "Former" and "Never" smokers). LR tests indicated that the best one- to three- (ever and current smokers) or four-variant combinations (all combined, former and never smokers) had a significant impact on UBC risk in a common logistic regression model. For the total study group as well as for all subgroups of ever (current and former smokers combined), current, former and never smokers, the LR statistic for addition of the respective best variant combination to a logistic regression model that contained all lower order best combinations was plotted successively, i.e. the LR statistic for the best main effect compared with the null model, the LR statistic for the model containing the main effect and the best two-variant combination compared with the model with the main effect only, etc. Significant combinations are indicated by * in case of $P < 0.05$ and by ** in case of $P < 0.001$. The best combinations are given in Table 3, the LR statistics and corresponding *P* values are given in Supplementary Table 6A. (B) The ten best unadjusted ORs of single variants, two-, three- and four-variant combinations show a significant increase for increasing numbers of combined variants in smokers and never smokers. Box plots of the ten best unadjusted ORs (i.e. having the lowest unadjusted *P* value) of one- to four-variant combinations with at least 100 cases and 100 controls in the combined study group ("All") or subgroup of ever, current, former and never smokers were plotted. Pairwise comparisons of the ORs of the ten best risk variants and two- to four-variant combinations were performed using the Tukey test. Significant increases were obtained for up to three-way (all combined, ever smokers) or four-way (current, former and never smokers) combinations, respectively. Significant variants and combinations are indicated by * in case of $P < 0.05$ and by ** in case of $P < 0.001$. The ten best one- to four-variant combinations are given in Supplementary Table 8A–E, available at *Carcinogenesis* Online.

with the second *SLC14A1* SNP (rs1058396 instead of rs17674580 as in the two-variant combination) and the *UDP glucuronosyltransferase 1 family, polypeptide A complex locus* (*UGT1A*) intron SNP rs11892031. The four-variant combination with the lowest *P* value additionally included rs8102137 near *cyclin E1* (*CCNE1*). In conclusion, considering the most significant combinations up to four variants, there was no overlap between ever and never smokers (Table 3).

Next, we further analysed the group of ever smokers (2069 cases, 2136 controls), which consisted of former (1393 cases, 1477 controls) and current (676 cases, 659 controls) smokers (Supplementary Figure 2). For both, ever and current smokers, the deletion of phase II detoxifying *GSTM1* was identified as the most significant variant (Table 3). In contrast, *GSTM1* was not among the best four-variant combinations in former smokers. Instead former and ever smokers overlapped in rs9642880 near the *MYC* oncogene. This suggests that quitting cigarette smoking may lead to a shift in relevance for the population at risk from detoxifying *GSTM1* to *MYC*, a gene which is known to act as a proto-oncogene. To identify the most characteristic differences between ever smokers (considering also former and current smokers) and never smokers, we analysed the top ten combinations of two, three and four variants in these subgroups (Table 4, Supplementary Tables 8 and 9, available at *Carcinogenesis* Online).

Table 4. Variants in the ten best four-variant combinations differ between smokers and never smokers

Variants	Nearest gene	All	Ever smokers	Current smokers	Former smokers	Never smokers	*P* value	*P C* versus *F*	*P C* versus *N*	*P F* versus *N*	*P E* versus *N*
GSTM1	GSTM1	9	8	10	0	0	**0.0005**	**1.08×10^{-05}**	**1.08×10^{-05}**	1.0000	**0.0007**
rs11892031[A/C]	UGT1A	3	2	6	4	8	0.2569	0.2105	**0.0031**	**1.08×10^{-05}**	**1.08×10^{-05}**
rs1495741[A/G]	NAT2	1	1	2	6	3	0.2374	0.1698	1.0000	0.0573	**0.0055**
rs1058396[A/G]	SLC14A1	2	2	8	4	9	0.0770	0.6563	0.6285	0.1698	**0.0230**
rs17674580[C/T]	SLC14A1	4	3	0	5	2	**0.0420**	0.1698	1.0000	0.3698	0.5820
rs2294008[C/T]	PSCA	6	8	0	1	1	1.0000	**0.0325**	0.4737	0.3498	1.0000
rs2978974[A/G]	PSCA	1	1	2	0	0	0.3198	1.0000	1.0000	1.0000	**0.0055**
rs1014971[A/G]	CBX6-APOBEC3A	0	0	3	0	10	**0.0005**	0.4737	0.4737	1.0000	1.0000
rs710521[A/G]	TP63	2	1	1	2	1	1.0000	1.0000	1.0000	1.0000	1.0000
rs798766[C/T]	TACC3	2	2	1	0	2	0.7266	1.0000	1.0000	0.4737	1.0000
rs8102137[C/T]	CCNE1	0	2	5	8	4	0.2689	0.3498	1.0000	0.1698	0.6285
rs9642880[G/T]	MYC	10	10	2	10	0	**0.0005**	**0.0007**	0.4737	**1.08×10^{-05}**	**1.08×10^{-05}**

Subgroups are compared regarding the occurrence of each variant in the ten best combinations by chi-squared or fishers exact tests. Unadjusted *P* values, ORs and 95% CIs of all ten best single variants, two-, three and four-variant combinations are given in Supplementary Table 8A-E, available at *Carcinogenesis* Online. Tests for the ten best two- and three-variant combinations are given in Supplementary Table 9A, B, available at *Carcinogenesis* Online.
P value: *P* value of the exact chi-squared test of homogeneity of the variant frequency in the ten best four-variant combinations in current, former and never smokers. The *P* value of Fisher's exact test of homogeneity of the variant frequency in the ten best four-variant combinations is given for current versus former smokers (*P C* versus *F*), current versus never smokers (*P C* versus *N*), former versus never smokers (*P F* versus N) and ever versus never smokers (*P E* versus *N*).
UGT1A, UDP glucuronosyltransferase 1 family, polypeptide A complex locus, NAT2, *N*-acetyltransferase 2, CBX6: chromobox homolog 6; TP63, tumor protein p63.
$P \leq 0.05$ are printed bold.

The frequency of the specific variant among the top ten combinations was analysed, and significant differences between the four groups were analysed by a chi-squared test. This led to the observations that (i) *GSTM1 null* was significantly more frequent in ever and current smokers than in never smokers and more frequent in current compared with former smokers, supporting the observations in Table 3 and Figure 1C that *GSTM1* lost its relevance when smoking had been ceased. So, *GSTM1* is a typical "current smoker variant", (ii) rs1014971 near *APOBEC3A* was more frequent in never smokers, compared with ever, current and former smokers indicating that this variant is a typical "never smoker SNP" and (iii) rs9642880 near *MYC* was more frequent in ever and former smokers compared with never smokers and more frequent in former compared with current smokers. Therefore, rs9642880 is a typical "former smoker SNP" and seems to be relevant for cigarette smoke exposed individuals only if smoking occurred in the past with no current exposure.

Replication of the four-variant combinations in independent case–control series

Finally, we tested whether the four-variant combinations, particularly the combination rs1014971[AA] × rs1058396[AG, GG] × rs11892031[AA] × rs8102137[CC, CT] in never smokers, identified in the combined IfADo-NBCS case–control series, could be confirmed in independent study groups. For this purpose, corresponding SNP data from 2080 additional bladder cancer cases and 2167 controls (Table 1) were available from the published New England and Spanish Bladder Cancer Studies (6, 19–21). The rs1014971[AA] × rs1058396[AG, GG] × rs1 1892031[AA] × rs8102137[CC, CT] combination in never smokers was confirmed in the New England and Spanish Bladder Cancer Studies resulting in increasing (unadjusted) ORs of 1.32 for the single variant to 1.60 for the four-variant combinations, respectively (Table 3B). Similarly, the four-variant combination in current smokers (*GSTM1 null* × rs11892031 [AA] × rs1058396[AG, GG] × rs1014971[AA, AG]) was confirmed (Table 3B). In contrast, the four-variant combination in former smokers (rs9642880[TT] × rs8102137[CC, CT] × rs1495741[AA, AG] × rs17674580[CC, CT]) could not be replicated, possibly due to the fact that former smokers are much more heterogeneous compared with never or current smokers. The adjusted logistic regression (Supplementary Table 5C) led to similar results as the unadjusted analysis shown in Table 3B.

Discussion

An important question is whether genetic variants can interact leading to higher ORs for combined high-risk alleles than the combination of individual variants alone. In a previous study, we identified three-variant combinations of seven confirmed UBC risk variants in the IfADo case–control series and obtained remarkable differences between ever and never smokers (12, 13). In the present study, we used the NBCS to confirm the previous results. Importantly, all frequent three-variant combinations could be replicated. In particular, we confirmed the results for the total study group and the subgroups of ever and former smokers. Three-variant combinations identified in current and never smokers were not significant, possibly because of their low frequency ($N < 100$ in cases and in controls).

As results and study group characteristics were comparable, we combined both case–control series to analyse up to four-variant combinations with sufficient power. Moreover, five further UBC risk variants were added to the combination analysis. We restricted the analysis to combinations present in at least 100 cases and controls to obtain robust results. We also restricted the analysis to four-variant combinations to avoid a bias towards frequent variants in higher fold combinations.

The strongest four-variant combination in never smokers was obtained for the high-risk alleles of rs1014971[AA] near *APOBEC3A* and *chromobox homolog 6*, the *SLC14A1* exon SNP rs1058396[AG, GG], the *UDP glucuronosyltransferase 1 family, polypeptide A complex locus* intron SNP rs11892031[AA] and rs8102137[CT, CC] near *CCNE1*. The combination resulted in an unadjusted OR of 2.59 (95% CI = 1.93–3.47, $P = 1.87 \times 10^{-10}$; Table 3). This four-variant combination was still relatively frequent, with 25 and 11% in cases and controls, respectively. The relatively high OR of the combination (OR = 2.59) is remarkable considering that the

individual ORs were small (OR = 1.11–1.30; Supplementary Tables 3K, L). Furthermore, the individual variants were not significant ($P \geq 0.1207$) in a common logistic regression model in presence of the four-way interaction effect ($P = 0.0003$; Supplementary Table 7). LR tests also showed that the increase from two- to four-variant combinations led to a significant increase in ORs for each step. Next, we tested whether the four-variant combination in never smokers (rs1014971[AA] × rs1058396[AG, GG] × r s11892031[AA] × rs8102137[CC, CT]), identified in the combined IfADo-NBCS study group, could be confirmed in an independent case–control series, the published New England and Spanish Bladder Cancer Studies (6, 19–21). The increased risk replicated in this group ($OR_{unadjusted} = 1.60$, 95% CI = 1.10–2.33, $P = 0.0130$) with similar frequencies of the combination (18% in cases, 12% in controls, Table 3B).

The main effect in never smokers, i.e. the most important individual variant, of the four-variant combination was attributable to rs1014971[AA]. This SNP maps to an intergenic region close to the *chromobox homolog 6* and *APOBEC3A* genes (6). *Chromobox homolog* genes have been reported to be involved in regulation of heterochromatin while *APOBEC3A* seems to be associated with genetic instability (27–29). The second SNP in the four-variant combination rs1058396[AG, GG] is a *SLC14A1* exon SNP (15). SLC14A1 is a urea transporter in the bladder which influences urine concentration as measured by specific gravity (30). The third variant of the four-variant combination was rs11892031[AA], an intron SNP of the phase II metabolism gene *UDP glucuronosyltransferase 1 family, polypeptide A complex locus* involved in glucuronidation (6, 31, 32). Variant number four in the combination was rs8102137[CC, CT] near *CCNE1* (6). *CCNE1* is involved in cell cycle transition from G1 to S phase (6). In conclusion, the four strongest interacting SNPs for never smokers seem to be associated with the biological processes of chromatin modification, genetic stability, detoxification and proliferation. This leads to the question why there is an interaction effect between these variants. We speculate that the interacting variants seem to belong to completely different biological processes, which may be assigned to detoxification, proliferation and DNA stability, thereby covering a broad set of functions relevant for carcinogenesis, instead of focusing on a specific single function. However, this interpretation should be treated with caution, since little is known about the functions of the variants themselves, let alone the combinations.

Increased ORs for the strongest four-variant combinations in the combined IfADo-Nijmegen study group were obtained in the never as well as the current smokers (Table 3A). The results of both four-variant combinations were confirmed in the New England and Spanish Bladder Cancer Studies, although ORs were numerically lower (but *P* values still significant; Table 3B), which is not unusual for analysis of an independent replication series. The strongest interaction of the entire study, the four-variant combination in never smokers, is characterized by a monotonous trend, where each added variant leads to increased ORs. However, also other scenarios were obtained. The four-variant combinations in current and former smokers show a non-monotonous trend with a decreased OR by the two-variant combination but increased ORs associated with the following variants (Table 3A). Such combinations are possible because the applied approach searches for the strongest variant combination independently within all two-, three- and four-variant combinations—ignoring that a best lower fold combination has been found already. Remarkably, the previous set of combined variants usually kept stable with increasing numbers of combined variants. So, the switch from a protective two-variant combination to a risk combination of three and four variants is easily explained: Either low risk genotypes were added to the low-risk genotype of the most important variant in the subgroup (*GSTM1 positive* for current smokers, rs9642880[GG, GT] for former smokers) or vice versa high-risk genotypes were combined (Table 3A). Importantly, the four-variant combinations as well as the non-monotonous trend were confirmed in the New England and Spanish Bladder Cancer Studies (Table 3B).

The variants of the best four-variant combination of the subgroup of ever smokers showed no overlap with that of the never smokers. In smokers, the strongest variant of the four-variant combination was the *GSTM1* deletion. *GSTM1* is known to be involved in detoxification of cigarette smoke carcinogens (8, 33–35). Already in previous publications, the relevance of this polymorphism has been demonstrated particularly in smokers and individuals occupationally exposed to polycyclic aromatic hydrocarbons (8, 11, 36). Interestingly, after cessation of occupational and/or environmental exposure to polycyclic aromatic hydrocarbons, the *GSTM1* polymorphism was no longer relevant (37). The previous interaction analysis of the IfADo case–control series comprising seven of the investigated polymorphisms also identified the *GSTM1* deletion as the most important variant in smokers (12). The second variant in the four-variant ever smoker combination was rs9642880, a variant approximately 30 000 bases upstream of *MYC* which has been reported to be associated with RNA levels of the oncogene (16, 31, 38). The biological function of the third and the fourth variant in the combination, the glycoprotein *PSCA* and the microtubule regulating *TACC3* variants rs2294008 and rs798766 in relation to bladder carcinogenesis is still not fully understood (17, 31, 39–43). Interactions of the *GSTM1* deletion, rs9642880 and rs2294008 but not rs798766 with smoking habits have also been found in a large UBC study (3942 cases, 5680 controls) (11). Missense SNP rs2294008 alters the *PSCA* start codon and results in less promoter activity but more mRNA (39, 40). *PSCA* owning an androgen responsive element in its promoter region influences *PSCA* expression (44). It is hypothesized that gender-specific UBC risk might be modulated via androgen responsive element-depending *PSCA* transcription activity in presence of rs2294008. However, relevant signalling pathways remain unclear (41). The functional role of the *TACC3* intron SNP rs798766 is still unclear (17, 31). TACC3 is a centrosomal adaptor protein that is involved in spindle microtubule dynamics during cell division (42, 43). In particular, TACC3 protein complexes seem to be essential for mitotic spindle assembly and dynamics and, hence, for prevention of genomic instability (42, 43). However, the observed association of rs798766 with UBC risk might also be due to the nearby *fibroblast growth factor receptor 3* (*FGFR3*) gene. Fibroblast growth factor receptors play a key role in activation of signalling pathways, for instance, the RAS/MAPK, phospholipase C, gamma 1 (PLCγ1), phosphatidylinositol-4,5-bisphosphate 3-kinase (PI3K) and signal transducer and activator of transcription pathways that regulate proliferation, migration and differentiation (45). Point mutations in the *FGFR3* gene and increased expression of the variant gene are common in low-grade non-invasive papillary urothelial bladder carcinomas (45–47). Overexpression of wild-type *FGFR3* is more common in invasive than in non-invasive tumors (45, 47). Recently, *FGFR3-TACC3* gene fusions have been detected in UBC as well as glioblastoma patients and cell lines (46–48). The fusion seems to result in a loss of the C-terminus of *FGFR3* and an overexpression of the *FGFR3-TACC3* fusion product (47, 48). The protein seems to be highly oncogenic *in vivo* and *in vitro* and can be assumed to induce signalling via the MAPK pathway in urothelial cells (47, 48).

It should be considered that differences in variant combinations were not only observed between ever and never smokers but also between current and former smokers. The subgroup of former smokers lost the "smoker variant" *GSTM1* in their best four-variant combination. Instead, the "never smoker SNP" rs8102137 5.8 kb upstream *CCNE1* occurs in the best four-variant combination of former smokers. The SNP rs9642880 near *MYC* is the strongest variant in the subgroup of former smokers. It is not present in the best four-variant combination of current smokers, nor of never smokers. However, it should be taken into account that this four-variant combination identified in the combined IfADo-Nijmegen case–control series was not confirmed in the replication series and should therefore be discussed with caution. Reasons for the discrepant result may be the general heterogeneity of the subgroup of former smokers, differences in definition and determination of a status as a former smoker which is associated with a higher degree of uncertainty than in case of current or never smokers. Or this observation could be a chance finding.

In conclusion, the present logistic regression based approach demonstrated that specific combinations of three to four variants confer UBC risk. The highest unadjusted OR was obtained for a four-variant combination in the subgroup of never smokers (OR = 2.59). Importantly, this four-variant combination was confirmed in an independent case–control series. Moreover, different SNP combinations were obtained for current smokers (OR = 1.61) and former smokers (OR = 2.13). The dominant SNP for never smokers is rs1014971 near *APOBEC3A*. The most important "smoker variant" is *GSTM1*, while former smokers shift to rs9642880 near *MYC*. The study demonstrates the strength of logistic regression approaches for SNP interactions that identify specific interaction profiles. This present study further supports the concept, according to which individual polymorphisms add only little to overall UBC risk, but the combined presence of many individually weak SNPs leads to substantially increased ORs (3).

Supplementary Material

Supplementary data is available at *Carcinogenesis* online.

Funding

KI's work has been supported by Deutsche Forschungsgemeinschaft (DFG) within the Collaborative Research Center SFB 876 "Providing Information by Resource-Constrained Analysis", project C4. The NBS (source of controls) and the NBCS (source of the cases) were funded by the Radboud university medical center. This work was supported by the Intramural Research Program of the National Institutes of Health and the Division of Cancer Epidemiology and Genetics, National Cancer Institute (Z01 CP010187-13).

Acknowledgements

The NBS is a population-based survey conducted at the Department for Health Evidence and the Department of Laboratory Medicine of the Radboud university medical center. Principal investigators of the NBS are L.A.L.M. Kiemeney, M. den Heijer, A.L.M. Verbeek, D.W. Swinkels and B. Franke. Principal investigators of the NBCS are L.A.L.M. Kiemeney, K. Aben and S.H. Vermeulen. The two studies thank deCODE Genetics, Reykjavik, Iceland for the genotyping of all the participants. The authors thank Ms. Hannah Bürger, Ms. Claudia Brockhaus, Ms. Kirsten Liesenhoff-Henze, Ms. Katrin Linßen and Ms. Marion Page for excellent technical support.
Conflict of Interest Statement: The authors have no conflict of interest to declare.

References

1. Lichtenstein, P. *et al.* (2000) Environmental and heritable factors in the causation of cancer – analyses of cohorts of twins from Sweden, Denmark, and Finland. N. Engl. J. Med., 343, 78–85.
2. Selinski, S. (2014) Urinary bladder cancer risk variants: recent findings and new challenges of GWAS and confirmatory studies. Arch. Toxicol., 88, 1469–1475.
3. Golka, K. *et al.* (2011) Genetic variants in urinary bladder cancer: collective power of the "wimp SNPs". Arch. Toxicol., 85, 539–554.
4. Jiang, Z. *et al.* (2011) Glutathione S-transferase M1 polymorphism and bladder cancer risk: a meta-analysis involving 33 studies. Exp. Biol. Med. (Maywood)., 236, 723–728.
5. Moore, L.E. *et al.* (2011) GSTM1 null and NAT2 slow acetylation genotypes, smoking intensity and bladder cancer risk: results from the New England bladder cancer study and NAT2 meta-analysis. Carcinogenesis, 32, 182–189.
6. Rothman, N. *et al.* (2010) A multi-stage genome-wide association study of bladder cancer identifies multiple susceptibility loci. Nat. Genet., 42, 978–984.
7. García-Closas, M. *et al.* (2005) NAT2 slow acetylation, GSTM1 null genotype, and risk of bladder cancer: results from the Spanish Bladder Cancer Study and meta-analyses. Lancet, 366, 649–659.
8. Engel, L.S. *et al.* (2002) Pooled analysis and meta-analysis of glutathione S-transferase M1 and bladder cancer: a HuGE review. Am. J. Epidemiol., 156, 95–109.
9. Selinski, S. (2014) The post GWAS era: strategies to identify gene-gene and gene-environment interactions in urinary bladder cancer. EXCLI J., 13, 1198–1203.
10. Wang, M. *et al.* (2014) Cumulative effect of genome-wide association study – identified genetic variants for bladder cancer. Int. J. Cancer, 135, 2653–2660.
11. Garcia-Closas, M. *et al.* (2013) Common genetic polymorphisms modify the effect of smoking on absolute risk of bladder cancer. Cancer Res., 73, 2211–2220.
12. Schwender, H. *et al.* (2012) Distinct SNP combinations confer susceptibility to urinary bladder cancer in smokers and non-smokers. PLoS One, 7, e51880.
13. Schwender, H. *et al.* (2015) Correction: distinct SNP combinations confer susceptibility to urinary bladder cancer in smokers and non-smokers. PLoS One, 10, e0137937.
14. Rafnar, T. *et al.* (2014) Genome-wide association study yields variants at 20p12.2 that associate with urinary bladder cancer. Hum. Mol. Genet., 23, 5545–5557.
15. Rafnar, T. *et al.* (2011) European genome-wide association study identifies SLC14A1 as a new urinary bladder cancer susceptibility gene. Hum. Mol. Genet., 20, 4268–4281.
16. Kiemeney, L.A. *et al.* (2008) Sequence variant on 8q24 confers susceptibility to urinary bladder cancer. Nat. Genet., 40, 1307–1312.
17. Kiemeney, L.A. *et al.* (2010) A sequence variant at 4p16.3 confers susceptibility to urinary bladder cancer. Nat. Genet., 42, 415–419.
18. Rafnar, T. *et al.* (2009) Sequence variants at the TERT-CLPTM1L locus associate with many cancer types. Nat. Genet., 41, 221–227.
19. Garcia-Closas, M. *et al.* (2011) A genome-wide association study of bladder cancer identifies a new susceptibility locus within SLC14A1, a urea transporter gene on chromosome 18q12.3. Hum. Mol. Genet., 20, 4282–4289.
20. Figueroa, J.D. *et al.* (2016) Identification of a novel susceptibility locus at 13q34 and refinement of the 20p12.2 region as a multi-signal locus associated with bladder cancer risk in individuals of European ancestry. Hum. Mol. Genet., 25, 1203–1214.
21. Figueroa, J.D. *et al.* (2014) Genome-wide association study identifies multiple loci associated with bladder cancer risk. Hum. Mol. Genet., 23, 1387–1398.

22. Selinski, S. *et al.* (2013) Refinement of the prediction of N-acetyltransferase 2 (NAT2) phenotypes with respect to enzyme activity and urinary bladder cancer risk. Arch. Toxicol., 87, 2129–2139.
23. Saravana Devi, S. *et al.* (2008) Distribution of detoxifying genes polymorphism in Maharastrian population of central India. Chemosphere, 70, 1835–1839.
24. Arand, M. *et al.* (1996) A multiplex polymerase chain reaction protocol for the simultaneous analysis of the glutathione S-transferase GSTM1 and GSTT1 polymorphisms. Anal. Biochem., 236, 184–186.
25. Kempkes, M. *et al.* (1996) Glutathione S-transferase GSTM1 and GSTT1 null genotypes as potential risk factors for urothelial cancer of the bladder. Arch. Toxicol., 71, 123–126.
26. Krause, G. *et al.* (2004) Glutathione S-transferase T1 and M1 (GSTT1, GSTM1) (genotyping). In Angerer, J. and Müller, M. (eds.) *Analyses of hazardous substances in biological materials*. Vol 9. Special issue: Marker of susceptibility. Wiley-VCH Verlag GmbH, Weinheim, Germany, pp. 183–210.
27. Kaustov, L. *et al.* (2011) Recognition and specificity determinants of the human cbx chromodomains. J. Biol. Chem., 286, 521–529.
28. Nik-Zainal, S. *et al.* (2014) Association of a germline copy number polymorphism of APOBEC3A and APOBEC3B with burden of putative APOBEC-dependent mutations in breast cancer. Nat. Genet., 46, 487–491.
29. Vartanian, J.P. *et al.* (2008) Evidence for editing of human papillomavirus DNA by APOBEC3 in benign and precancerous lesions. Science, 320, 230–233.
30. Koutros, S. *et al.* (2013) Differential urinary specific gravity as a molecular phenotype of the bladder cancer genetic association in the urea transporter gene, SLC14A1. Int. J. Cancer, 133, 3008–3013.
31. Dudek, A.M. *et al.* (2013) Urinary bladder cancer susceptibility markers. What do we know about functional mechanisms? Int. J. Mol. Sci., 14, 12346–12366.
32. Tang, W. *et al.* (2012) Mapping of the UGT1A locus identifies an uncommon coding variant that affects mRNA expression and protects from bladder cancer. Hum. Mol. Genet., 21, 1918–1930.
33. Bolt, H.M. *et al.* (2006) Relevance of the deletion polymorphisms of the glutathione S-transferases GSTT1 and GSTM1 in pharmacology and toxicology. Curr. Drug Metab., 7, 613–628.
34. Hengstler, J.G. *et al.* (1998) Polymorphisms of N-acetyltransferases, glutathione S-transferases, microsomal epoxide hydrolase and sulfotransferases: influence on cancer susceptibility. Recent Results Cancer Res., 154, 47–85.
35. Bell, D.A. *et al.* (1993) Genetic risk and carcinogen exposure: a common inherited defect of the carcinogen-metabolism gene glutathione S-transferase M1 (GSTM1) that increases susceptibility to bladder cancer. J. Natl. Cancer Inst., 85, 1159–1164.
36. Golka, K. *et al.* (2009) Susceptibility to urinary bladder cancer: relevance of rs9642880[T], GSTM1 0/0 and occupational exposure. Pharmacogenet. Genomics, 19, 903–906.
37. Ovsiannikov, D. *et al.* (2012) Polymorphic enzymes, urinary bladder cancer risk, and structural change in the local industry. J. Toxicol. Environ. Health A, 75, 557–565.
38. Wang, M. *et al.* (2009) Common genetic variants on 8q24 contribute to susceptibility to bladder cancer in a Chinese population. Carcinogenesis, 30, 991–996.
39. Wu, X. *et al.* (2009) Genetic variation in the prostate stem cell antigen gene PSCA confers susceptibility to urinary bladder cancer. Nat. Genet., 41, 991–995.
40. Kohaar, I. *et al.* (2013) Genetic variant as a selection marker for anti-prostate stem cell antigen immunotherapy of bladder cancer. J. Natl. Cancer Inst., 105, 69–73.
41. Gakis, G. *et al.* (2013) Gender-specific differences in muscle-invasive bladder cancer: the concept of sex steroid sensitivity. World J. Urol., 31, 1059–1064.
42. Ha, G.H. *et al.* (2013) Transforming acidic coiled-coil proteins (TACCs) in human cancer. Cancer Lett., 336, 24–33.
43. Thakur, H.C. *et al.* (2013) Role of centrosomal adaptor proteins of the TACC family in the regulation of microtubule dynamics during mitotic cell division. Biol. Chem., 394, 1411–1423.
44. Zhigang, Z. *et al.* (2008) Flutamide reduced prostate cancer development and prostate stem cell antigen mRNA expression in high grade prostatic intraepithelial neoplasia. Int. J. Cancer, 122, 864–870.
45. di Martino, E. *et al.* (2012) A Decade of FGF receptor research in bladder cancer: past, present, and future challenges. Adv. Urol., 2012, 429213.
46. Cancer Genome Atlas Research Network. (2014) Comprehensive molecular characterization of urothelial bladder carcinoma. Nature, 507, 315–322.
47. Williams, S.V. *et al.* (2013) Oncogenic FGFR3 gene fusions in bladder cancer. Hum. Mol. Genet., 22, 795–803.
48. Parker, B.C. *et al.* (2013) The tumorigenic FGFR3-TACC3 gene fusion escapes miR-99a regulation in glioblastoma. J. Clin. Invest., 123, 855–865.