A genome-wide association study of bladder cancer identifies a new susceptibility locus within \textit{SLC14A1}, a urea transporter gene on chromosome 18q12.3

Montserrat Garcia-Closas\textsuperscript{1,2,*},†, Yuanqing Ye\textsuperscript{3,†}, Nathaniel Rothman\textsuperscript{1,†}, Jonine D. Figueroa\textsuperscript{1,†}, Núria Malats\textsuperscript{4,†}, Colin P. Dinney\textsuperscript{5,†}, Nilanjan Chatterjee\textsuperscript{1,†}, Ludmila Prokunina-Olsson\textsuperscript{1,†}, Zhaoming Wang\textsuperscript{6,†}, Jie Lin\textsuperscript{3,†}, Francisco X. Real\textsuperscript{4,7}, Kevin B. Jacobs\textsuperscript{6}, Dalsu Baris\textsuperscript{1}, Michael Thun\textsuperscript{8}, Immaculata De Vivo\textsuperscript{9}, Demetrius Albanes\textsuperscript{1}, Mark P. Purdie\textsuperscript{1}, Manolis Kogevinas\textsuperscript{10,11,12,13}, Ashish M. Kamat\textsuperscript{5}, Seth P. Lerner\textsuperscript{14}, H. Barton Grossman\textsuperscript{5}, Jian Gu\textsuperscript{3}, Xia Pu\textsuperscript{3}, Amy Hutchinson\textsuperscript{6}, Yi-Ping Fu\textsuperscript{1}, Laurie Burdett\textsuperscript{6}, Meredith Yeager\textsuperscript{6}, Wei Tang\textsuperscript{1}, Adonina Tardón\textsuperscript{12,15}, Consol Serra\textsuperscript{7}, Alfredo Carrato\textsuperscript{16}, Reina García-Closas\textsuperscript{17}, Josep Lloreta\textsuperscript{7,11}, Alison Johnson\textsuperscript{18}, Molly Schwenn\textsuperscript{19}, Margaret R. Karagas\textsuperscript{20}, Alan Schned\textsuperscript{22}, Gerald Andriole Jr\textsuperscript{21}, Robert Grubb III\textsuperscript{21}, Amanda Black\textsuperscript{1}, Eric J. Jacobs\textsuperscript{8}, W. Ryan Diver\textsuperscript{8}, Susan M. Gapstur\textsuperscript{8}, Stephanie J. Weinstein\textsuperscript{23}, Jarmo Virtamo\textsuperscript{22}, David J. Hunter\textsuperscript{23}, Neil Caporaso\textsuperscript{1}, Maria Teresa Landi\textsuperscript{1}, Joseph F. Fraumeni Jr\textsuperscript{1}, Debra T. Silverman\textsuperscript{1,†}, Stephen J. Chanock\textsuperscript{1,†} and Xifeng Wu\textsuperscript{3,†}

\textsuperscript{1}Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD, USA, \textsuperscript{2}Division of Genetics and Epidemiology, Institute of Cancer Research, London, UK, \textsuperscript{3}Department of Epidemiology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA, \textsuperscript{4}Spanish National Cancer Research Centre, Madrid, Spain, \textsuperscript{5}Department of Urology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA, \textsuperscript{6}Core Genotype Facility, SAIC-Frederick, Inc., National Cancer Institute-Frederick, Frederick, MD, USA, \textsuperscript{7}Departament de Ciències Experimentals i de la Salut, Universitat Pompeu Fabra, Barcelona, Spain, \textsuperscript{8}Epidemiology Research Program, American Cancer Society, Atlanta, GA, USA, \textsuperscript{9}Channing Laboratory, Department of Medicine, Brigham and Women’s Hospital, Boston, MA, USA, \textsuperscript{10}Centre de Recerca en Epidemiologia Ambiental (CREAL), Barcelona, Spain, \textsuperscript{11}Parc de Salut Mar - Institut Municipal d’Investigació’ Medica (IMIM), Barcelona, Spain, \textsuperscript{12}CIBER of Epidemiology and Public Health (CIBERESP), Barcelona, Spain, \textsuperscript{13}National School of Public Health, Athens, Greece, \textsuperscript{14}Scott Department of Urology, Baylor College of Medicine, Houston, TX, USA, \textsuperscript{15}Universidad de Oviedo, Oviedo, Spain, \textsuperscript{16}Ramón y Cajal University Hospital, Madrid, Spain, \textsuperscript{17}Unidad de Investigación, Hospital Universitario de Canarias, Las Lagunas, Spain, \textsuperscript{18}Vermont Cancer Registry, Burlington, VT, USA, \textsuperscript{19}Maine Cancer Registry, Augusta, ME, USA, \textsuperscript{20}Dartmouth Medical School, Hanover, NH, USA, \textsuperscript{21}Department of Urology, Washington University School of Medicine, St Louis, MO, USA, \textsuperscript{22}National Institute for Health and Welfare, Helsinki, Finland and \textsuperscript{23}Department of Epidemiology, Program in Molecular and Genetic Epidemiology, Harvard School of Public Health, Boston, MA, USA

Received April 27, 2011; Revised July 26, 2011; Accepted August 1, 2011

\textsuperscript{*}To whom correspondence should be addressed at: Section of Epidemiology, Institute of Cancer Research, 15 Cotswold Rd, Belmont, Sutton, Surrey SM2 5NG, UK. Tel: +44 2087224239; Email: montse.garciaclosas@icr.ac.uk

\textsuperscript{†}M.G.-C., Y.Y., N.R., J.D.F. are co-first authors; N.M., C.P.D., N.C., L.P.-O., Z.W. and J.L. are co-second authors; D.T.S., S.J.C. and X.W. are co-last authors.

Published by Oxford University Press 2011.
 Genome-wide and candidate-gene association studies of bladder cancer have identified 10 susceptibility loci thus far. We conducted a meta-analysis of two previously published genome-wide scans (4501 cases and 6076 controls of European background) and followed up the most significant association signals [17 single nucleotide polymorphisms (SNPs) in 10 genomic regions] in 1382 cases and 2201 controls from four studies. A combined analysis adjusted for study center, age, sex, and smoking status identified a novel susceptibility locus that mapped to a region of 18q12.3, marked by rs7238033 ($P = 8.7 \times 10^{-9}$; allelic odds ratio 1.20 with 95% CI: 1.13–1.28) and two highly correlated SNPs, rs10775480/rs10853535 ($r^2 = 1.00$; $P = 8.9 \times 10^{-5}$; allelic odds ratio 1.16 with 95% CI: 1.10–1.22). The signal localizes to the solute carrier family 14 member 1 gene, SLC14A1, a urea transporter that regulates cellular osmotic pressure. In the kidney, SLC14A1 regulates urine volume and concentration whereas in erythrocytes it determines the Kidd blood groups. Our findings suggest that genetic variation in SLC14A1 could provide new etiological insights into bladder carcinogenesis.

INTRODUCTION

The risk for developing urinary bladder cancer, the fourth most common incident cancer in men, is strongly related to cigarette smoking and occupational exposure to aromatic amines (1). Family history is associated with an ∼2-fold risk in cancer (2), and the genetic contribution of common and uncommon alleles to bladder cancer risk has been pursued in candidate-gene studies and more recently, genome-wide association studies (GWAS). Together, the two approaches have yielded 10 distinct loci. It is notable that three of the discovered loci contain carcinogen-metabolizing genes: the N-acetyltransferase 2 (NAT2) (3), a common gene deletion of glutathione S-transferase Mu 1 (GSTM1) (4–7) and the UDP-glucuronosyltransferase UGT1A gene locus on chromosome 2q37.1 discovered by a recent GWAS (7). Cigarette smoking modifies risk associations with NAT2 and GSTM1. While the NAT2 slow acetylation status alone appears to increase the risk of bladder cancer in cigarette smokers, the GSTM1 null genotype shows stronger associations with the risk in never smokers (7).

The remaining seven susceptibility loci for bladder cancer are scattered across the genome and are currently under active investigation to understand the biological basis of the contribution of common genetic variants to bladder cancer risk. These include common single nucleotide polymorphism (SNP) markers in regions that harbor plausible candidate genes for further study: 3q28 (8) (TP63), 4p16.3 (TMEM129, TACC3-FGFR3) (9), 8q24.21 (8), 8q24.3 (10) (PSCA), 5p15.33 (TERT-CLPTM1L) (7,11), 22q13.1 (7) and 19q12 (7) (CCNE1). GWAS of bladder cancer have included primarily cases with urothelial (transitional cell) carcinomas, which represent ∼95% of malignant bladder tumors occurring in industrialized countries (1). Most (80%) urothelial carcinomas are low grade and non-invasive at presentation (TaG1/TaG2) but have high recurrence rate, thus requiring regular screening and interventions. These tumors are clinically and molecularly distinct from the more aggressive high-grade, non-muscle invasive (TaG3/T1G2/T1G3) and muscle invasive (T2-3) tumors. Distinct tumor types are postulated to develop through different pathways, suggesting heterogeneous etiologic factors, both genetic and environmental (12,13). Smoking and occupational exposure to aromatic amines has similar associations with tumors of different grade and stage (14); however, some loci discovered though GWAS have been reported to be differentially associated according to stage and grade. While loci on chromosomes 8q24.21, 4p16.3 (TMEM129, TACC3-FGFR3) and 5p15.33 (TERT-CLPTM1L) are more strongly associated with tumors of low grade/low risk of progression (7–9), loci on chromosomes 22q13.1 and 19q12 (CCNE1) are more strongly associated with the risk of high-grade/high-risk tumors (7).

Meta-analysis of existing GWAS data offers the opportunity to discover additional loci based on current projections for the number of independent regions harboring common variants associated with bladder cancer risk (15). In this study, we conducted a meta-analysis of two previously published GWAS, followed by a validation stage in further studies to discover additional susceptibility loci.

RESULTS

The study design for the meta-analysis (stage I) and follow up (stage II) is summarized in Figure 1, and study populations are described in Supplementary Material, Table S1. In stage I, a meta-analysis of genome-wide scan data from two previously published GWAS (7,10) was conducted in 4501 cases and 6076 controls of European background using 555,912 SNPs (common to all subjects scanned on the HumanHap 500, 610-Quad and 1M Illumina Infinium arrays). The quantile–quantile (Q–Q) plot showed minimal evidence for inflation of the test statistics when compared with the expected distribution (corrected $\lambda_{1000}$ subjects = 1.002), which suggests that there is no substantial hidden population substructure or differential genotype calling between cases and controls (16) (Supplementary Material, Fig. S1). A Manhattan plot displays the results of the combined GWAS meta-analysis in stage I (Supplementary Material, Fig. S2). Seven distinct genomic loci were notable for $P$-value < $1.5 \times 10^{-5}$ in meta-analyses of allelic odds ratio estimates derived from logistic regression analysis adjusted for study center, age, sex and smoking status. We repeated the meta-analyses among current smokers only and identified three additional loci with $P$-value < $2 \times 10^{-6}$ that were
advanced in the replication effort, resulting on a total of 10 genomic regions being evaluated in stage II.

Based on results from stage I meta-analyses, SNPs in linkage disequilibrium with the 10 known loci or those with minor allele frequency (MAF) <5% in the controls were excluded from further analysis. Seventeen SNPs were selected based on rank $P$-value for follow up in stage II, which included highly correlated SNPs. These included the most significant SNPs in the most promising seven regions, namely 7p12.1 with two SNPs, 18q12.3 with three SNPs, 19q13.33 with three SNPs, 6q23.2 with three SNPs, as well as one SNP in each of the following regions: 4p16.1, 10q25.3 and 20p13. We also selected three additional variants that were highly significant in current smokers (6p21.32, 6p24.3, 9q12.3). The 17 SNPs were genotyped in each of the following regions: 1382 cases and 2201 controls from two case–control studies and two prospective cohorts in the USA (Fig. 1).

We used fixed effects meta-analyses based on estimates of allelic odds ratios for each study, adjusted by study center, age, sex and smoking status (current, former or never), to obtain combined (stages I and II) and stage-specific estimates. Analyses of combined estimates identified one locus on chromosome 18q12.3 reaching genome-wide level of significance ($P < 5 \times 10^{-8}$, Table 1, Supplementary Material, Table S2). The strongest signal in the 18q12.3 locus ($P = 8.9 \times 10^{-9}$) was for rs10775480 tagged by rs10853535 (pairwise $r^2 = 1.0$ in HapMap CEU parents). We used data on rs10775480 for all studies, except for Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial (PLCO) and Spanish Bladder Cancer Study (SBCS) in which rs10853535 was used because rs10775480 was not successfully called in the PLCO and was not included on the HumanHap 1 M array used in SBCS. Three other SNPs in the region showed associations with risk: rs7238033 had a similar signal as rs10775480/rs10853535, but it was genotyped in 8 of 10 studies included in the primary GWAS, whereas two surrogate SNPs rs11082469/rs11877720 (pairwise $r^2 = 1.0$) had a weaker signal. These associations were not independent of rs10775480/rs10853535, therefore further analyses of the 18q12.3 locus by tumor characteristics, gene–gene and gene–smoking interactions (shown below) focused on rs10775480/rs10853535. In an analysis of tumor subtypes, data suggested a possible association of the 18q12.3 locus with tumors of higher grade but the finding is not statistically significant ($P_{interaction} = 0.071$) (Supplementary Material, Table S3). We found no evidence of modification of the 18q12.3 locus risk association by smoking status (Supplementary Material, Table S4).

We explored whether the allelic relative risk for the new locus on 18q12.3 varied by genotypes for each of the 10 previously identified loci (Supplementary Material, Table S5). Genotype-specific estimates and $P$-values for multiplicative
gene × gene interactions were obtained by including main effects and an interaction term between the 18q12.3 locus and each locus in a series of multivariate logistic regression models adjusted by study center, sex, age and smoking status. Our preliminary analyses suggest that the association at 18q12.3 is limited to subjects with the AA genotype (87%) of the UGTA1 at 18q12.3 is limited to subjects with the AA genotype and each locus in a series of multivariate logistic regression effects and an interaction term between the 18q12.3 locus and member 2 (SLC14A2).

**Tables**

### Table 1. Association of SNPs in 18q12.3 (SLC14A1) region with the risk for urinary bladder cancer

<table>
<thead>
<tr>
<th>SNP</th>
<th>Stage</th>
<th>n</th>
<th>Cases</th>
<th>Controls</th>
<th>MAF</th>
<th>Allelic OR</th>
<th>95% CI</th>
<th>P-value</th>
<th>$I^2$</th>
<th>P-het</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs7238033[C/T]</td>
<td>Combined</td>
<td>8</td>
<td>3992</td>
<td>4977</td>
<td>0.43</td>
<td>1.20</td>
<td>1.13, 1.28</td>
<td>8.72E-09</td>
<td>0.0</td>
<td>0.829</td>
</tr>
<tr>
<td></td>
<td>Stage I</td>
<td>4</td>
<td>2685</td>
<td>3150</td>
<td>0.43</td>
<td>1.19</td>
<td>1.11, 1.29</td>
<td>6.55E-06</td>
<td>0.0</td>
<td>0.603</td>
</tr>
<tr>
<td></td>
<td>Stage II</td>
<td>4</td>
<td>1307</td>
<td>1827</td>
<td>0.44</td>
<td>1.22</td>
<td>1.09, 1.36</td>
<td>3.31E-04</td>
<td>0.0</td>
<td>0.641</td>
</tr>
<tr>
<td>rs10775480[C/T]</td>
<td>Combined</td>
<td>10</td>
<td>5801</td>
<td>7894</td>
<td>0.43</td>
<td>1.16</td>
<td>1.10, 1.22</td>
<td>8.95E-09</td>
<td>0.0</td>
<td>0.535</td>
</tr>
<tr>
<td>rs10853535[T/C]</td>
<td><strong>Stage I</strong></td>
<td>6</td>
<td>4499</td>
<td>6068</td>
<td>0.43</td>
<td>1.15</td>
<td>1.08, 1.22</td>
<td>4.41E-06</td>
<td>0.0</td>
<td>0.453</td>
</tr>
<tr>
<td></td>
<td><strong>Stage II</strong></td>
<td>4</td>
<td>1302</td>
<td>1826</td>
<td>0.44</td>
<td>1.22</td>
<td>1.09, 1.36</td>
<td>3.16E-04</td>
<td>0.0</td>
<td>0.519</td>
</tr>
<tr>
<td>rs11082469[A/G]</td>
<td>Combined</td>
<td>10</td>
<td>5792</td>
<td>7821</td>
<td>0.49</td>
<td>0.89</td>
<td>0.85, 0.94</td>
<td>1.84E-05</td>
<td>33.5</td>
<td>0.150</td>
</tr>
<tr>
<td>rs11877720[A/G]</td>
<td><strong>Stage I</strong></td>
<td>6</td>
<td>4487</td>
<td>5987</td>
<td>0.49</td>
<td>0.91</td>
<td>0.86, 0.96</td>
<td>1.11E-03</td>
<td>45.0</td>
<td>0.105</td>
</tr>
<tr>
<td></td>
<td><strong>Stage II</strong></td>
<td>4</td>
<td>1305</td>
<td>1834</td>
<td>0.49</td>
<td>0.85</td>
<td>0.76, 0.95</td>
<td>3.00E-03</td>
<td>0.0</td>
<td>0.399</td>
</tr>
</tbody>
</table>

Results from meta-analyses of allelic odds ratios adjusted by study center, age, gender and smoking status, obtained from up to 10 case-control and cohort studies (Stage I: SBCS, PLCO, NEBCS-ME, VT, CPSII, ATBC, TXBCS1; Stage II: NBCS-NH, HPFS, NHS, TXBCS2). Data on rs7238033 was not available in PLCO and SBCS. The effect estimates for rs7238033 and rs10775480 in the subset of three studies with data on both SNPs were very similar ($P = 3.4 \times 10^{-4}$ for rs7238033 based on 2193 cases and 1716 controls and $P = 2.6 \times 10^{-4}$ for rs10775480 based on 2184 cases and 1705 controls). Inclusion of both SNP simultaneously in the same model resulted in loss of significance for both SNPs. These SNPs were used as surrogates in PLCO and SBCS because of no availability of data on rs10775480 and rs11082469 from these studies. The pairwise $r^2$ values based on HapMap data (release 28) for the surrogate pairs are $r^2 = 1.00$. The pairwise $r^2$ values between rs10775480/rs10853535 and rs11082469/rs11877720 are $r^2 = 0.66$.

**DISCUSSION**

Our combined analysis of 5883 individuals with bladder cancer and 8277 controls has identified a new genomic region (rs11892031) encoding a UGTA1 family of proteins that facilitate solubility and removal of carcinogens through urine ($P_{interaction} = 0.018$; Table 2); however, given the limited power to detect gene × gene interactions and the possibility that this could represent a false positive finding, further work is needed to confirm this interaction. We did not observe notable departures from multiplicative joint effects for other susceptibility regions.

second non-synonymous variant, rs2298720 (Glu44Lys) defines a weaker version of the JkA group, Jk(a)W (18). The three variants analyzed in our meta-analysis and follow-up samples adequately tag the Kidd blood groups as shown in haplotype analyses presented in Supplementary Material, Figure S3: the risk allele corresponds to the Jk(A) group, whereas the protective allele corresponds to the Jk(B) and Jk(A)W groups.

In <0.1% of the population, erythrocytes in individuals lacking both Jk(A) and Jk(B) antigens are resistant to lysis by 2M urea (19, 20). These individuals are unable to adequately concentrate urine, producing larger volumes of diluted urine (21). A similar phenotype observed in a knock-out mouse model of SLC14A1, demonstrated ~50% reduction in urea concentrating capacity and 1.5-fold greater daily output of urine of lower osmolality (22). There is no clear evidence of different renal function in individuals with common Jk(A), Jk(B) and Jk(A)W blood groups, but this should be explored further in bladder cancer patients and controls. Variations in urine volumes and concentration could modify exposure of bladder epithelium to carcinogens in the urine. In this regard, our study raises the possibility that common genetic variation could alter the function or expression of the Kidd blood group antigen and contribute to the risk for urinary bladder cancer. Fine-mapping studies of this region should provide an optimal set of variants for functional studies to pursue the role of the Kidd blood group antigen and contribute to the risk for urinary bladder cancer. It is also possible that the common variants act through a different mechanism yet to be determined and not related to the Kidd blood group antigen.

The addition of the 18q12.3 locus to the catalog of conclusively associated regions with urinary bladder cancer represents a small, but important step in defining a comprehensive set of variants that could be used in generating a risk model in combination with other risk factors for bladder cancer (i.e. smoking, occupational and environmental exposures and family history). At the same time, the discovery of novel genomic regions provides the foundation for biological insights into the genetics of urinary bladder cancer.
untreated incident bladder cancer cases recruited from the University of Texas MD Anderson Cancer Center and Baylor College of Medicine between 1999 and 2007. There were no age, gender, cancer-stage restrictions. All included subjects were self-identified as European background. Informed consent was obtained from all study participants before the collection of epidemiological data and blood samples by trained M.D. Anderson staff interviewers. The study was approved by the IRB of MD Anderson Cancer Center, Baylor College of Medicine and the Kelsey-Seybold Clinic.

Stage II follow up

Cases and control in stage II were derived from two case–control studies [Texas Bladder Cancer Study 2 (TXCBS 2), the New Hampshire component of the New England Bladder Cancer Study (NEBCS-NH)], and two prospective cohorts [Nurse’s Health Study (NHS) and Health Professionals Follow up Study (HPFS)]. Each participating study obtained informed consent from study participants and approval from its IRB for this study. There were no age, gender, ethnicity or cancer-stage restrictions for TXCBS1, and analyses were restricted to individuals of European background as determined by self-report for NEBCS-NH, NHS and HPFS.

Genotyping and quality control

Genotyping, quality control and assessment of population structure for studies in stage I has been previously described [see Rothman et al. for NCI GWAS (7) and Wu et al. for MD Anderson GWAS (10)]. The meta-analysis was performed on the common set of SNPs called in both the MD Anderson and NCI GWAS sets. All samples had been scanned with Illumina Infinium Arrays.

We estimate the inflation of the test statistic, λ, adjusted to a sample size of 1000 cases/1000 controls as per the method of de Bakker et al. (24): λ(corrected) = 1 + (λ - 1) × \([n_{\text{case}} + n_{\text{cont}}] / 2 \times 10^{-3}\). The corrected estimated \(λ_{1000}\) is 1.002 while the uncorrected \(λ\) is 1.014 (Supplementary Material, Fig. S1).

The final participant count for stage I analysis was 4501 cases and 6076 controls. The number of SNPs available for association analysis in all studies after quality control metrics applied was 556 429 for MD Anderson and 591 637 for NCI. Of these SNPs, 555 912 overlapped exactly between the Illumina HumanHap 1 M chip and 610Quad/550K data and were used for this meta-analysis.

TagMan custom genotyping assays (ABI, Foster City, CA, USA) were designed and optimized for the 17 SNPs genotyped in stage II studies. In an analysis of 1000 samples, the comparison of the Illumina calls with the TaqMan assays showed a concordance rate of 100%. The Illumina Infinium cluster plots for the two surrogate SNPs in the novel region in 18q12.3 (rs10775480 and rs10853535) are shown in Supplementary Material, Figure S4.

Statistical analysis

Meta-analyses in stage I were based on allelic odds ratio estimates derived from logistic regression models adjusted for population structure as determined by population structure analyses with STRUCTURE (23). There were no age, gender, cancer-stage restrictions. Each participating study obtained informed consent from study participants and approval from its Institutional Review Board (IRB) for this study. Participating studies obtained institutional certification permitting data sharing in accordance with the NIH Policy for Sharing of Data Obtained in NIH Supported or Conducted GWAS.

Stage I meta-analyses

NCI GWAS. Case and controls in the National Cancer Institute (NCI) GWAS were derived from two case–control studies [SBCS and the Maine and Vermont components of the New England Bladder Cancer Study (NEBCS-ME, VT), and three prospective cohorts [PLCO, The American Cancer Society Cancer Prevention Study II Nutrition Cohort (CPSII), ATBC, TXCBS1; Stage II: NCBS-NH, HPFS, NHS, TXCBS2].

Surrogate SNPs in PLCO and SBCS: rs10775480 and rs10853535; \(r^2 = 1.08\). \(∗\)P-value from test for interaction with rs10775480[T]/rs10853535[C].

MATERIAL AND METHODS

Study populations

Participants were drawn from 10 prospective cohort and case–control studies (Supplementary Material, Table S1).

Stage I meta-analyses

NCI GWAS. Case and controls in the National Cancer Institute (NCI) GWAS were derived from two case–control studies [SBCS and the Maine and Vermont components of the New England Bladder Cancer Study (NEBCS-ME, VT), and three prospective cohorts [PLCO, The American Cancer Society Cancer Prevention Study II Nutrition Cohort (CPSII), ATBC, TXCBS1; Stage II: NCBS-NH, HPFS, NHS, TXCBS2]. Surrogate SNPs in PLCO and SBCS: rs10775480 and rs10853535; \(r^2 = 1.08\). *P-value from test for interaction with rs10775480[T]/rs10853535[C].

Table 2. Association of rs10775480[T]/rs10853535[C] in 18q12.3 (SLC14A1) with the risk for urinary bladder cancer, stratified by rs11892031 (UGT1A1) and rs1495741 (NAT2) genotypes

<table>
<thead>
<tr>
<th>Genotypes</th>
<th>Cases</th>
<th>Controls</th>
<th>OR</th>
<th>95% CI</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs11892031 (UGT1A1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AA</td>
<td>5004</td>
<td>6607</td>
<td>1.19</td>
<td>1.12</td>
<td>1.25</td>
</tr>
<tr>
<td>AC</td>
<td>753</td>
<td>1200</td>
<td>1.02</td>
<td>0.89</td>
<td>1.16</td>
</tr>
<tr>
<td>CC</td>
<td>22</td>
<td>57</td>
<td>0.82</td>
<td>0.42</td>
<td>1.60</td>
</tr>
<tr>
<td>Total</td>
<td>5779</td>
<td>7864</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs1495741 (NAT2)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AA</td>
<td>3649</td>
<td>4652</td>
<td>1.13</td>
<td>1.06</td>
<td>1.21</td>
</tr>
<tr>
<td>AG</td>
<td>1863</td>
<td>2749</td>
<td>1.20</td>
<td>1.10</td>
<td>1.31</td>
</tr>
<tr>
<td>GG</td>
<td>241</td>
<td>374</td>
<td>1.33</td>
<td>1.04</td>
<td>1.70</td>
</tr>
<tr>
<td>GG/AG</td>
<td>775</td>
<td>1257</td>
<td>1.01</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>5753</td>
<td>7775</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Results from logistic regression models to estimate allelic odds ratios adjusted by study center, age, gender and smoking status, including data from 10 case–control and cohort studies (Stage I: SBCS, PLCO, NEBCS-ME, VT, CPSII, ATBC, TXCBS1; Stage II: NCBS-NH, HPFS, NHS, TXCBS2). *Surrogate SNPs in PLCO and SBCS: rs10775480 and rs10853535; \(r^2 = 1.08\). **P-value from test for interaction with rs10775480[T]/rs10853535[C].

MD Anderson GWAS. Cases and controls for the primary scan were derived from the Texas Bladder Cancer Study 1 (TXCBS 1), a hospital-based bladder cancer case–control study (10). Cases were defined as histologically confirmed and previously
study center, age (in 5-year categories), sex and smoking status (current, former or never). Each SNP genotype was coded as a count of minor alleles, with the exception of X-linked SNPs among men that were coded as 2 if the participant carried the minor allele and 0 if he carried the major allele (25). A score test with one degree of freedom was performed on all genetic parameters in each model to determine statistical significance. For the inclusion of stage II data, we conducted meta-analyses based on estimates of allelic odds ratio for each study, adjusted by study center, age, sex and smoking status (never, former, current).

Polytomous logistic regression was used to obtain estimates of effect for different tumor subtypes. Case-only analyses with tumor type as an outcome were used to test for differences in effect size across subtypes. Polytomous logistic regression models for tumor grade constrained the effect size to increase linearly across levels. Gene–gene interactions were assessed using logistic regression models adjusted by study center, age, sex and smoking status and including interaction terms. Genotype–smoking interactions were assessed using logistic regression for grouped data adjusted by study center and sex, and including interaction terms.

A haplotype-based association analysis was performed across the region of interest (shown in the Supplementary Material, Figure S3) using the PLINK (version 1.07) (26). Haplotype-specific odds ratio and P-values were estimated for each haplotype versus all others, adjusted for the effects of age, gender, study center and when appropriate, smoking status.

Data analysis and management was performed with GLU (Genotyping Library and Utilities version 1.0), a suite of tools available as an open-source application for management, storage and analysis of GWAS data, and STATA S.E. v.11.1 (College Station, TX, USA).

**Estimate of recombination hotspots**

SequenceL.Dhot (27) that uses an approximate marginal likelihood method (28) was used to compute likelihood ratio (LR) statistics for a set of putative hotspots across the region of interest. We sequentially analyzed subsets of 100 controls of European background (by pooling five controls from each study). We used Phasev2.1 to infer the haplotypes as well as background recombination rates. The analysis was repeated with five non-overlapping sets of 100 controls.
ACKNOWLEDGEMENTS

Leslie Carroll (Information Management Services, Silver Spring, MD, USA); Gemma Castaño-Vinyals (Institut Municipal d’Investigació Mèdica, Barcelona, Spain); Fernando Fernández (Institut Municipal d’Investigació Mèdica, Barcelona, Spain); Paul Hurwitz (Westat, Inc., Rockville, MD, USA); Charles Lawrence (Westat, Inc., Rockville, MD, USA); Marta Lopez-Brea (Marqués de Valdecilla University Hospital, Santander, Cantabria, Spain); Anna McIntosh (Westat, Inc., Rockville, MD, USA); Marta Lopez-Brea (Marqués de Valdecilla University Hospital, Santander, Cantabria, Spain); Robert Saal (Westat, Inc., Rockville, MD, USA); Angeles Panadero (Hospital Ciudad de Còrria, Còrria (Cáceres), Spain); Fernando Rivera (Marqués de Valdecilla University Hospital, Santander, Cantabria, Spain); Robert Saal (Westat, Inc., Rockville, MD, USA); Maria Sala (Institut Municipal d’Investigació Mèdica, Barcelona, Spain); Kirk Snyder (Information Management Services, Inc., Silver Spring, MD, USA); Anne Taylor (Information Management Services, Inc., Silver Spring, MD, USA); Montserrat Torà (Institut Municipal d’Investigació Mèdica, Barcelona, Spain); Jane Wang (Information Management Services, Silver Spring, MD, USA).

AUTHOR CONTRIBUTIONS

All authors contributed to the writing of the manuscript.

REFERENCES


