

A multi-stage genome-wide association study of bladder cancer identifies multiple susceptibility loci

A full list of authors and affiliations appears at the end of the article.

Abstract

We conducted a multi-stage, genome-wide association study (GWAS) of bladder cancer with a primary scan of 589,299 single nucleotide polymorphisms (SNPs) in 3,532 cases and 5,120 controls of European descent (5 studies) followed by a replication strategy, which included 8,381 cases and 48,275 controls (16 studies). In a combined analysis, we identified three new regions associated with bladder cancer on chromosomes 22q13.1, 19q12 and 2q37.1; rs1014971, ($P=8\times 10^{-12}$) maps to a non-genic region of chromosome 22q13.1; rs8102137 ($P=2\times 10^{-11}$) on 19q12 maps to *CCNE1*; and rs11892031 ($P=1\times 10^{-7}$) maps to the *UGT1A* cluster on 2q37.1. We confirmed four previous GWAS associations on chromosomes 3q28, 4p16.3, 8q24.21 and 8q24.3, validated previous candidate associations for the *GSTM1* deletion ($P=4\times 10^{-11}$) and a tag SNP for *NAT2* acetylation status ($P=4\times 10^{-11}$), as well as demonstrated smoking interactions with both regions. Our findings on common variants associated with bladder cancer risk should provide new insights into mechanisms of carcinogenesis.

Bladder cancer is the fourth most common incident cancer in men¹ and its frequent recurrence requires regular screening and interventions. Cigarette smoking and occupational exposure to aromatic amines have been strongly linked to bladder cancer risk.¹ A family history of bladder cancer is associated with an approximately two-fold increase in risk; however, multiple-cancer families are rare and no high-penetrance genes have been

Users may view, print, copy, download and text and data-mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use: http://www.nature.com/authors/editorial_policies/license.html#terms

Correspondence should be addressed to: Montserrat Garcia-Closas, Section of Epidemiology, Institute of Cancer Research, 15 Cotswold Rd, Belmont, Sutton, Surrey SM2 5NG, United Kingdom, Montse.GarciaClosas@icr.ac.uk.

⁶⁷These authors contributed equally to this work.

⁶⁸These authors are the designated representatives of the participating studies.

⁶⁹These authors jointly directed this work.

Author Contributions

N.R., M.G-C., N.C., J.F., D.T.S., and S.J.C. organized and designed the study.

S.J.C., K.B.J., A.H., Z.W., Y-P.F., J.P-O., L.B., X.W., M.A.T.H., M.C., D.V.D.B., S.G., S.P., R.R.M., I.D.V., T.R., D.T.B., G.C-T., J.G.H., R.K., S.C.E.B., and A.G. conducted and supervised genotyping of samples.

M.G-C., N.C., N.R., K.B.J., M.Y., N.M., D.T.S., and S.J.C. contributed to the design and execution of statistical analysis.

M.G-C., N.R., N.C., N.M., J.F., F.X.R., J.F.F., D.T.S., and S.J.C. wrote the first draft of the manuscript.

N.R., M.G-C., N.M., X.W., J.F., D.V.D.B., F.X.R., G.M., D.B., M.T., L.A.K., P.V., I.D.V., D.A., M.P.P., T.R., M.A.T.H., A.E.K., O.C., K.G., R.K., J.A.T., J.I.M., M.K., A.T., C.S., A.C., R.G-C., J.L., A.J., M.S., M.R.K., A.S., G.A., R.G., A.B., E.J.J., W.R.D.,

identified to date²⁻⁴. Large meta-analyses of candidate gene studies have provided support for associations between NAT2 slow acetylation phenotype⁵ (defined by NAT2 haplotypes) and a common gene deletion of GSTM16 with bladder cancer risk^{7,8}. Further, gene-environment interactions have been shown for smoking and NAT2 acetylation, with an increased risk in slow acetylators, apparent only among cigarette smokers^{7,8}.

Previous genome-wide association studies (GWAS) in bladder cancer have identified common variants in four genomic regions on chromosomes 3q289 (TP63), 4p16.3 (TMEM129, TACC3-FGFR3)¹⁰, 8q24.219, and 8q24.311 (PSCA) that are associated with risk. Interestingly, the variants on 8q24.21 map to a region centromeric to MYC that has been identified in GWAS of breast, colorectal and prostate cancers, as well as chronic lymphocytic leukemia¹²⁻¹⁸. Also, in follow-up analyses, an association with bladder cancer risk has been suggested for variants near the TERT-CLPTMIL locus on chromosome 5p15.33, which has also been associated by GWAS with risk for basal cell carcinoma, cutaneous melanoma, lung, brain and pancreatic cancers¹⁹⁻²³. However, the previously reported association with bladder cancer did not achieve genome-wide significance.

We conducted a multi-stage GWAS involving 3,532 cases and 5,120 controls of self-described European descent in stage I, and followed up the most notable signals in two stages of replication (stages IIa/b and III) totaling 8,381 cases and 48,275 controls (Figure 1 and Online Methods). Individuals with scan data in stage I were participants in two case-control studies carried out in Spain and the USA (Maine and Vermont component of the New England Bladder Cancer Study) and three prospective cohort studies in the USA and Finland (see Supplementary Table 1 online for details). Replication analyses in stage II were carried out using existing scan data from two earlier studies. First, we evaluated the 100 most significant SNPs (excluding previously reported loci and SNPs with pairwise $r^2 > 0.8$) in 969 cases and 957 controls from the Texas Bladder Cancer study in the USA (stage IIa)¹¹. Five of these SNPs were further evaluated in a second scan of 1,274 cases and 1,832 controls in The Netherlands (stage IIb)⁹. Three of the five SNPs were included or tagged at a pair-wise $r^2 > 0.8$ in the Dutch scan, and risk associations were confirmed for all three. In stage III, the three SNPs plus a tagging SNP for the NAT2 acetylation status were evaluated in 6,141 cases and 45,486 controls from 11 case-control and 3 prospective cohort studies in the USA and Europe (see Figure 1 and Supplementary Table 1).

After quality control analysis of genotypes, we combined the data sets in stage I resulting in 589,299 SNPs available for analysis (based on the common SNPs called from both the Illumina Human1M and Human 610-Quad) in 3,532 cases and 5,120 controls (**Online Methods**). A logistic regression model was fit for genotype trend effects (1 d.f.) adjusted for study center, age, sex, smoking status (current, former or never) and DNA source (blood/buccal). The quantile-quantile (Q-Q) plot showed little evidence for inflation of the test statistics as compared to the expected distribution (corrected λ_{1000} subjects=1.021), which minimizes the likelihood of substantial hidden population substructure or differential genotype calling between cases and controls²⁴ (**Online Methods and Supplementary Figure 1**). A Manhattan plot displays the results of the combined GWAS in stage I (Supplementary Figure 2).

Data from the first stage confirm the associations reported with tag SNPs in the four previously identified genomic regions on chromosomes 3q28 (rs710521)⁹, 8q24.21 (rs9642880)⁹, 8q24.3 (rs2294008)¹¹ and 4p16.3 (rs798766)¹⁰ as well as a suggested region in 5p15.33 (rs401681; a neighboring SNP, rs2736098, was also reported but data were not available in our study)¹⁹ (Table 1 and Supplementary Figure 3). Consistent with prior reports^{9,10}, rs9642880 on 8q24.21 and rs798766 on 4p16.3 were most strongly associated with tumors of low grade/low risk of progression (Supplementary Table 2). A stronger association with low grade/low risk disease was also suggested for rs401681 on 5p15.33 (Supplementary Table 2). In addition, we used a copy number variation TaqMan assay⁷ to assess the presence of *GSTM1* on 1p13.3 to genotype stage I samples, and confirmed an association with increased bladder cancer risk (Table 1).

In a combined analysis based on case/control counts by genotype and study, we estimated odds ratios (ORs) using logistic regression analyses adjusted for study center. Meta-analyses of estimated ORs adjusted for age, sex, smoking status and DNA source produced comparable point estimates (Supplementary Table 3). Our combined analysis of stages I, II and III identified three novel genomic regions on chromosomes 22q13.1, 19q12 and 2q37.1 that were associated with bladder cancer risk below the threshold for genome-wide significance ($P < 5 \times 10^{-7}$)²⁵ (Table 2 and Supplementary Figure 4 for study and stage specific estimates, Figure 2). We also confirmed a signal below genome-wide significance for rs1495741, which tags the NAT2 acetylator status²⁶ previously reported as a bladder cancer susceptibility locus on 8p227.8. The new SNP is located approximately 10kb of the 3' end of the gene.

The locus on chromosome 22q13.1, rs1014971 ($P_{trend} = 8.4 \times 10^{-12}$; OR per C allele = 0.88, 95%CI 0.85-0.91), was primarily associated with high-risk tumors (Supplementary Table 2). The locus is located in a non-genic region, approximately 25 kb centromeric of the catalytic polypeptide-like 3A (*APOBEC3A*) and 64 kb telomeric of the chromobox homolog 6 (*CBX6*). *APOBEC3A* is an apolipoprotein B mRNA editing enzyme that belongs to the cytidine deaminase gene family, which can play a role in the initiation of tumorigenesis by deamination of cytosine (C) to uracil (U)²⁷. *CBX6* is a component of the chromatin – associated polycomb complex involved in transcriptional repression.

In the combined analysis, we observed an association with rs8102137 on chromosome 19q12 ($P_{trend} = 1.7 \times 10^{-11}$; OR per C allele = 1.13, 95%CI 1.09-1.17), which maps to the cyclin E1 gene (*CCNE1*). *CCNE1* is a key member of the cyclin/cyclin-dependent kinase (Cdk)/retinoblastoma protein (pRB) pathway which determines the rates of cell cycle transition from G1 to S phase, and is commonly altered in bladder cancer and other tumors²⁸. Cyclin E1 expression in bladder cancer has been associated with high grade or muscle invasive tumors and poor clinical outcome²⁹. Consistently, rs8102137 was most strongly associated with risk of high grade/high risk tumors (Supplementary Table 2).

A third locus is marked by rs11892031 ($P = 1.0 \times 10^{-7}$; OR per C allele = 0.84, 95%CI 0.79-0.89) on chromosome 2q37.1 and resides in an intronic region of the UDP-glucuronosyltransferase (UGT) 1A gene locus, which encodes the UGT1A family of proteins. Glucuronidation by UGTs facilitates solubility and removal of substrates such as

endo- and xenobiotics (including carcinogens in tobacco smoke) via bile or urine³⁰. Genetic variation in *UGT1A* has been associated with predisposition to severe gastrointestinal toxicity of the anticancer drug irinotecan³¹. The *UGT1A* locus is represented by at least nine highly homologous transcripts, collectively known as UGTs, generated by alternative splicing. Tissue-specific loss or decreased expression of UGTs has been associated with several gastrointestinal cancers and bladder cancer³²⁻³⁴, as well as experimentally induced bladder cancer in animal models³⁵.

Previously, a promising signal in the *CLPTMIL-TERT* locus on chromosome 5p15.33 was reported in a region in which common variants have been associated with multiple cancers in recent GWAS¹⁹⁻²³. In addition, rare mutations in *TERT* have been linked to dyskeratosis congenita (a bone marrow failure syndrome), idiopathic pulmonary fibrosis, acute myelogenous leukemia and chronic lymphocytic leukemia³⁶⁻³⁹. In the first stage of this GWAS, we observed a moderately significant effect for rs401681 ($P = 2.9 \times 10^{-3}$), which was at genome-wide significance when combined with the Rafnar et al. data ($P = 5.0 \times 10^{-7}$; OR per C allele 1.11, 95% CI 1.07-1.16) (Table 1, Supplementary Figure 3).

The risk associated with *GSTM1* and *NAT2* varied in strength across categories of cigarette smoking, whereas genotype risk associations by smoking categories were of similar magnitude for the eight susceptibility loci identified by GWAS (Supplementary Table 4). In a combined analysis, the risk association with *GSTM1* deletion was strongest in never smokers (OR=1.75, 95%CI=1.44-2.13), and progressively weaker in former (OR=1.55, 95%CI=1.35-1.78) and current smokers (OR=1.25, 95%CI=1.07-1.46; $P_{\text{interaction}} = 0.008$ for current vs. never smokers; Table 3). The stronger association of the *GSTM1* deletion among non-smokers is a novel observation that was not evident in previous case-only meta-analyses⁷. rs1495741 located on the 3' end of *NAT2* is a marker of the *NAT2* phenotype associated with bladder cancer risk²⁶. The rs1495741 GG genotype marking the slow acetylation phenotype, compared to the combined AG/AA genotypes corresponding to the intermediate/rapid acetylation phenotypes, showed a highly significant ($P=5.5 \times 10^{-7}$) association with increased bladder cancer risk that was limited to cigarette smokers (OR=1.24, 95% CI=1.16-1.32 $P=4.3 \times 10^{-11}$; $P_{\text{interaction}}=6.3 \times 10^{-5}$) (Supplementary Figure 5 and Supplementary Table 3). This interaction is consistent with the role of *NAT2* in the detoxification of bladder carcinogens such as aromatic amines from tobacco smoke.

Our three-stage study had adequate power to detect variants of moderate effect sizes over a range of common allele frequencies. For the newly discovered SNP markers, the power to detect the observed associations at a level of genome-wide significance was at 54%, 30%, 30% and 6% for rs104971, rs1495741, rs8102137 and rs11892031, respectively. In light of the limited power to discover SNPs with modest effect sizes, additional loci with similar effect sizes will likely be identified with larger scale GWAS. Based on a recent estimator⁴⁰ that incorporates novel and previously reported loci together, we estimate that approximately two dozen additional bladder cancer susceptibility SNP markers of similar magnitude and frequencies might be discovered. Future studies should be powered with adequate sample size to detect additional variants.

With the exception of the *GSTM1* deletion, relative risk estimates for novel loci are based on associations using tag SNPs, which most likely underestimate the association with biologically important alleles. Accordingly, further studies are needed to define the functional variants and the clinical utility of risk models that combine genetic markers with epidemiologic risk factors for bladder cancer (i.e. smoking, occupational and environmental exposures, family history). Our combined analysis of 12,254 individuals with bladder cancer and 53,395 controls has uncovered three new genomic regions associated with bladder cancer risk. Fine-mapping studies of these three regions are needed to identify candidate variants for functional studies that should shed light into biological mechanisms for the associations reported through GWAS. This knowledge could establish the foundation for developing improved preventive, diagnostic and/or therapeutic approaches.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Authors

Nathaniel Rothman^{1,67}, Montserrat Garcia-Closas^{1,67}, Nilanjan Chatterjee^{1,67}, Nuria Malats^{2,67}, Xifeng Wu^{3,67}, Jonine Figueroa^{1,68}, Francisco X Real^{2,4,68}, David Van Den Berg^{5,68}, Giuseppe Matullo^{6,7,68}, Dalsu Baris^{1,68}, Michael Thun^{8,68}, Lambertus A Kiemeneij^{9,10,11,68}, Paolo Vineis^{12,7,68}, Immaculata De Vivo^{13,68}, Demetrius Albanes^{1,68}, Mark P Purdue^{1,68}, Thorunn Rafnar^{14,68}, Michelle A T Hildebrandt^{3,68}, Anne E Kiltie^{15,68}, Olivier Cussenot^{16,17,68}, Klaus Golka^{18,68}, Rajiv Kumar^{19,68}, Jack A Taylor^{20,68}, Jose I Mayordomo^{21,68}, Kevin B Jacobs^{22,68}, Manolis Kogevinas^{23,24,25,26,68}, Amy Hutchinson²², Zhaoming Wang²², Yi-Ping Fu¹, Ludmila Prokunina-Olsson¹, Laurie Burdette²², Meredith Yeager²², William Wheeler²⁷, Adonina Tardón^{25,28}, Consol Serra²⁹, Alfredo Carrato³⁰, Reina García-Closas³¹, Josep Lloreta³², Alison Johnson³³, Molly Schwenn³⁴, Margaret R Karagas³⁵, Alan Schned³⁵, Gerald Andriole Jr³⁶, Robert Grubb III³⁶, Amanda Black¹, Eric J Jacobs⁸, W Ryan Diver⁸, Susan M Gapstur⁸, Stephanie J Weinstein¹, Jarmo Virtamo³⁷, Victoria K Cortessis⁴, Manuela Gago-Dominguez⁴, Malcolm C Pike^{4,38}, Mariana C Stern⁴, Jian-Min Yuan³⁹, David Hunter⁴⁰, Monica McGrath⁴⁰, Colin P Dinney⁴¹, Bogdan Czerniak⁴², Meng Chen³, Hushan Yang³, Sita H Vermeulen^{43,9}, Katja K Aben^{9,10}, J Alfred Witjes¹¹, Remco R Makkinje⁴³, Patrick Sulem¹⁴, Soren Besenbacher¹⁴, Kari Stefansson^{14,44}, Elio Riboli¹², Paul Brennan⁴⁵, Salvatore Panico⁴⁶, Carmen Navarro^{47,25}, Naomi E Allen⁴⁸, H Bas Bueno-de-Mesquita⁴⁹, Dimitrios Trichopoulos^{50,51}, Neil Caporaso¹, Maria Teresa Landi¹, Federico Canzian⁵², Borje Ljungberg⁵³, Anne Tjonneland⁵⁴, Francoise Clavel-Chapelon⁵⁵, David T Bishop⁵⁶, Mark T W Teo⁵⁶, Margaret A Knowles⁵⁶, Simonetta Guarrera⁷, Silvia Polidoro⁷, Fulvio Ricceri^{6,7}, Carlotta Sacerdote^{7,57}, Alessandra Allione⁷, Geraldine Cancel-Tassin¹⁷, Silvia Selinski¹⁸, Jan G Hengstler¹⁸, Holger Dietrich⁵⁸, Tony Fletcher⁵⁹, Peter Rudnai⁶⁰, Eugen Gurzau⁶¹, Kvetoslava Koppova⁶², Sophia C E Bolick²⁰, Ashley Godfrey²⁰, Zongli Xu²⁰, José I Sanz-Velez⁶³, María D García-Prats⁶³, Manuel Sanchez²¹, Gabriel Valdivia²¹,

Stefano Porru⁶⁴, Simone Benhamou^{65,66}, Robert N Hoover¹, Joseph F Fraumeni Jr.¹, Debra T Silverman^{1,69}, and Stephen J Chanock^{1,69}

Affiliations

¹Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, Maryland, USA ²Spanish National Cancer Research Centre, Madrid, Spain ³Department of Epidemiology, The University of Texas M.D. Anderson Cancer Center, Houston, Texas, USA ⁴Departament de Ciències Experimentals i de la Salut, Universitat Pompeu Fabra, Barcelona, Spain ⁵Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, California, USA ⁶Department of Genetics, Biology and Biochemistry, University of Torino, Torino, Italy ⁷Human Genetics Foundation (HuGeF), Torino, Italy ⁸Epidemiology Research Program, American Cancer Society, Atlanta, Georgia, USA ⁹Department of Epidemiology, Biostatistics and Health Technology Assessment, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands ¹⁰Comprehensive Cancer Center East, Nijmegen, The Netherlands ¹¹Department of Urology, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands ¹²Imperial College London, London, UK ¹³Channing Laboratory, Department of Medicine, Brigham and Women's Hospital, Boston, Massachusetts, USA ¹⁴deCODE Genetics, Reykjavik, Iceland ¹⁵Gray Institute for Radiation Oncology and Biology, University of Oxford, Oxford, UK ¹⁶Department of Urology, Assistance Publique-Hôpitaux de Paris, Tenon Hospital, Paris, France ¹⁷Centre de Recherche sur les Pathologies Prostatiques, Paris, France ¹⁸Leibniz Research Centre for Working Environment and Human Factors, Dortmund, Germany ¹⁹Division of Molecular Genetic Epidemiology, German Cancer Research Center, Heidelberg, Germany ²⁰National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina, USA ²¹University of Zaragoza, Zaragoza, Spain ²²Core Genotype Facility, SAIC-Frederick, Inc., NCI-Frederick, Frederick, Maryland, USA ²³Centre for Research in Environmental Epidemiology (CREAL), Barcelona, Spain ²⁴Municipal Institute of Medical Research, Barcelona, Spain ²⁵CIBER Epidemiología y Salud Pública (CIBERESP), Barcelona, Spain ²⁶National School of Public Health, Athens, Greece ²⁷Information Management Services, Inc., Rockville, Maryland, USA ²⁸Instituto Universitario de Oncología, Universidad de Oviedo, Oviedo, Spain ²⁹Universitat Pompeu Fabra, Barcelona, Spain ³⁰Ramón y Cajal University Hospital, Madrid, Spain ³¹Unidad de Investigación, Hospital Universitario de Canarias, La Laguna, Spain ³²Hospital del Mar-IMIM, Univesitat Pompeu Fabra, Barcelona, Spain ³³Vermont Cancer Registry, Burlington, Vermont, USA ³⁴Maine Cancer Registry, Augusta, Maine, USA ³⁵Dartmouth Medical School, Hanover, New Hampshire, USA ³⁶Department of Urology, Washington University School of Medicine, St. Louis, Missouri, USA ³⁷National Institute for Health and Welfare, Helsinki, Finland ³⁸Department of Epidemiology and Biostatistics, Memorial Sloan-Kettering Cancer Center, New York, USA ³⁹Masonic Cancer Center, University of Minnesota, Minneapolis, Minnesota, USA ⁴⁰Department of Epidemiology, Program in Molecular and Genetic Epidemiology, Harvard School of Public Health, Boston, Massachusetts, USA ⁴¹Department of Urology, University of

Texas M.D. Anderson Cancer Center, Houston, Texas, USA ⁴²Department of Pathology, University of Texas M.D. Anderson Cancer Center, Houston, Texas, USA ⁴³Department of Human Genetics, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands ⁴⁴Faculty of Medicine, University of Iceland, Reykjavik, Iceland ⁴⁵International Agency for Research on Cancer, Lyon, France ⁴⁶Department of Clinical and Experimental Medicine, Federico II University, Naples, Italy ⁴⁷Department of Epidemiology, Murcia Regional Health Council, Murcia, Spain ⁴⁸Cancer Epidemiology Unit, Nuffield Department of Clinical Medicine University of Oxford, Oxford, UK ⁴⁹National Institute for Public Health and the Environment (RIVM), Bilthoven, The Netherlands ⁵⁰Department of Epidemiology, Harvard School of Public Health, Boston, Massachusetts, USA ⁵¹Bureau of Epidemiologic Research, Academy of Athens, Greece ⁵²Genomic Epidemiology Group, German Cancer Research Center, Heidelberg, Germany ⁵³Department of Surgical and Perioperative Sciences, University of Umea, Umea, Sweden ⁵⁴Danish Cancer Society, Institute of Cancer Epidemiology, Copenhagen, Denmark ⁵⁵Centre for Research in Epidemiology and Population Health, Institute Gustave-Roussy, Paris South University, Villejuif, France ⁵⁶Leeds Institute of Molecular Medicine, St James's University Hospital, University of Leeds, Leeds, UK ⁵⁷Unit of Cancer Epidemiology, University of Torino, Torino, Italy ⁵⁸Department of Urology, Paul Gerhardt Foundation, Wittenberg, Germany ⁵⁹Institute of Hygiene and Tropical Medicine, London, UK ⁶⁰National Institute of Environmental Health, Budapest, Hungary ⁶¹Babes Bolyai University, Environmental Health Center, Cluj-Napoca, Romania ⁶²State Health Institute, Banská Bystrica, Slovakia ⁶³San Jorge University Hospital, Huesca, Spain ⁶⁴Department of Experimental and Applied Medicine, Section of Occupational Medicine and Industrial Hygiene, University of Brescia, Italy ⁶⁵Institut national de la santé et de la recherche médicale, U946, Fondation Jean Dausset-CEPH, Paris, France ⁶⁶Centre national de la recherche scientifique, UMR8200, Institut Gustave-Roussy, Villejuif, France

Acknowledgments

The bladder cancer GWAS was supported by the intramural research program of the National Institutes of Health, National Cancer Institute.

This project has been funded in part with federal funds from the National Cancer Institute, National Institutes of Health, under Contract No. HHSN261200800001E. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government.

Appendix

Online Methods

Study Participants

Participants were drawn from 21 studies (Supplementary Table 1). For stage I, cases were defined as histologically confirmed primary carcinoma of the urinary bladder including carcinoma *in situ* (ICD-0-2 topography codes C67.0-C67.9 or ICD9 codes 188.1-188.9).

Each participating study obtained informed consent from study participants and approval from its Institutional Review Board (IRB) for this study. For stage I only, participating studies obtained institutional certification permitting data sharing in accordance with the NIH Policy for Sharing of Data Obtained in NIH Supported or Conducted Genome-Wide Association Studies (GWAS).

Genotyping and Quality Control

For stage I, genome-wide genotyping was conducted using three chips, SBCS (HumanHap 1 Million), NEBC-ME/VT (Human Hap 610-Quad), ATBC, CPS-II and PLCO (cases) (Human Hap 610) and controls from CGEMS/GEI for PLCO (Human Hap 550-r equivalent). DNA samples were selected for genotyping based on pre-genotyping quality control measures performed for GWAS at the Core Genotyping Facility of the NCI 4,089 blood samples and 2,813 buccal samples were analyzed. Repeat genotyping was performed on 38 blood samples (19 cases and 19 controls) and 10 buccal samples (2 cases and 8 controls) on Illumina 1M chips after suitable metrics identified performance issues. Cancer free controls (N=2003) were previously scanned in CGEMS18 and a lung cancer GWAS21.

Genotype clusters were estimated with samples by study with preliminary completion rates greater than 98% per individual study (namely SBCS, NEBC-ME/VT, PLCO, ATBC and CPS-II). Genotypes for the analytical build were based on study specific clustering. SNP assays with locus call rates lower than 90% were excluded.

SNPs with extreme departures from Hardy-Weinberg proportions ($P < 1 \times 10^{-7}$) were excluded from the association analysis due to the increased likelihood of spurious associations due to problematic assays or genotyping calling.⁴² Additional participants were excluded based on: 1) completion rates lower than 94-96% (n=203 samples); 2) heterozygosity of less than 22% or >35% (n=12); 3) inter-study unexpected duplicates (n=5); 4) phenotype exclusions (due to ineligibility or incomplete information) (n=94).

Assessment of population structure of study participants was performed with STRUCTURE⁴³ by seeding the analysis with founder genotypes from three HapMap populations (Phase I and II build 26).⁴⁴ A set of 12,898 SNPs with extremely low pair-wise correlation ($r^2 < 0.004$) was selected for this analysis.⁴⁵⁻⁴⁷ A total of 55 participants (43 cases and 12 controls) were estimated to have less than 85% HapMap CEU admixture (Supplementary Figure 6). Principal component analysis (PCA) of scanned subjects (excluding inferred sib and half-sib pairs) was performed with GLU (a similar procedure to EIGENSTRAT)^{45,46} and did not reveal notable eigenvectors. Consequently, a study-specific indicator was used for the stage I analysis⁴⁶.

We estimated 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: $\lambda(\text{corrected}) = 1 + (\lambda - 1) \times [n_{\text{case}}^{-1} + n_{\text{cont}}^{-1}] / [2 \times 10^{-3}]$.⁴⁸ The corrected estimated λ_{1000} is 1.021 while the uncorrected λ is 1.086 (Supplementary Figure 1).

Twenty participant pairs were identified as potential relatives based on genotyping sharing in excess of theoretical expectations. A set of 4,546 SNPs were selected (with completion

rates >95%, MAF>0.3 and $r^2 < 0.01$ in the three HapMap populations) and used to run PREST49 to formally test for cryptic relatedness. 19 unexpected full-sib and 1 parent-child pairs were identified and excluded from PCA (but included in the association analysis). 243 expected duplicates (including 6 triplicates in ATBC) were evaluated and yielded a concordance rate of 99.99%.

The final participant count for stage I analysis was 3,532 cases and 5,120 controls (Supplementary Table 1). The number of SNPs available for association analysis in all studies but SBCS was 589,299. In the SBCS, genotyped with the Infinium HumanHap 1 M chip, after quality control metrics were applied, 1,002,634 SNPs were available and 571,643 overlapped exactly with the 610Quad/550k data.

TaqMan custom genotyping assays (ABI, Foster City, CA) were designed and optimized for 4 SNPs, including the tag SNP for *NAT2*. In an analysis of 1,107 samples from three studies, the comparison of the Illumina calls with the TaqMan assays showed an average concordance rate of 99.4% (range 99.2-99.8%); no shifts from wild type to homozygotes were observed. The Illumina Infinium cluster plots for the four novel associations, rs1014971, rs8102137, rs11892031 and rs1495741 are shown in Supplementary Figure 7.

Association Analysis

Association analyses for stage I were conducted using logistic regression, adjusted for age (in five-year categories), sex, smoking (current, former or never), DNA source (buccal/ blood) and study. 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.⁵⁰ A score test with one degree of freedom was performed on all genetic parameters in each model to determine statistical significance. We assessed heterogeneity in genetic effects across studies using the I^2 statistic. For the inclusion of stage II and III data, we used genotype counts by case-control status and study, and conducted a fixed effects meta-analysis. We also conducted a meta-analysis based on estimates of allelic odds ratio adjusted by age, sex, smoking status, DNA source and study; the estimates did not materially differ from the fixed-effects meta-analysis (Supplementary Table 3).

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. Models for tumor grade constrained the effect size to increase linearly across levels. Genotype-smoking interactions were assessed using logistic regression for grouped data adjusted by study and including interaction terms. Forest plots by smoking, including summary estimates from fixed effects meta-analyses, are also shown for rs1495741.

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.

Estimate of Recombination Hotspots

SequenceLDhot51 that uses an approximate marginal likelihood method⁵² 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 descent (by pooling 5 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 pooled controls.

Data Access

The CGEMS data portal provides access to individual level data for investigators from certified scientific institutions after approval of their submitted Data Access Request.

URLs:

CGEMS portal: <http://cgems.cancer.gov/>

CGF: <http://cgf.nci.nih.gov/>

GLU: <http://code.google.com/p/glu-genetics/>

EIGENSTRAT: <http://genepath.med.harvard.edu/~reich/EIGENSTRAT.htm>

SNP500Cancer: <http://snp500cancer.nci.nih.gov/>

STRUCTURE: <http://pritch.bsd.uchicago.edu/structure.html>

Tagzilla: <http://tagzilla.nci.nih.gov/>

References

1. Silverman, D.; SS, D.; LE, M.; N, R. Bladder Cancer. In: D, S.; Fraumeni, JF., Jr., editors. *Cancer Epidemiology and Prevention*. Oxford University Press; New York, NY: 2006. p. 1101-1127.
2. Kantor AF, Hartge P, Hoover RN, Fraumeni JF Jr. Familial and environmental interactions in bladder cancer risk. *Int J Cancer*. 1985; 35:703–6. [PubMed: 4008097]
3. Murta-Nascimento C, et al. Risk of bladder cancer associated with family history of cancer: do low-penetrance polymorphisms account for the increase in risk? *Cancer Epidemiol Biomarkers Prev*. 2007; 16:1595–600. [PubMed: 17684133]
4. Aben KK, et al. Segregation analysis of urothelial cell carcinoma. *Eur J Cancer*. 2006; 42:1428–33. [PubMed: 16737809]
5. Lower GM Jr. et al. N-acetyltransferase phenotype and risk in urinary bladder cancer: approaches in molecular epidemiology. Preliminary results in Sweden and Denmark. *Environ Health Perspect*. 1979; 29:71–9. [PubMed: 510245]
6. Bell DA, et al. 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*. 1993; 85:1159–64. [PubMed: 8320745]
7. Garcia-Closas M, et al. NAT2 slow acetylation, GSTM1 null genotype, and risk of bladder cancer: results from the Spanish Bladder Cancer Study and meta-analyses. *Lancet*. 2005; 366:649–59. [PubMed: 16112301]
8. Rothman N, Garcia-Closas M, Hein DW. Commentary: Reflections on G. M. Lower and colleagues' 1979 study associating slow acetylator phenotype with urinary bladder cancer: meta-analysis,

- historical refinements of the hypothesis, and lessons learned. *Int J Epidemiol.* 2007; 36:23–8. [PubMed: 17510073]
9. Kiemeny LA, et al. Sequence variant on 8q24 confers susceptibility to urinary bladder cancer. *Nat Genet.* 2008; 40:1307–12. [PubMed: 18794855]
 10. Kiemeny LA, et al. A sequence variant at 4p16.3 confers susceptibility to urinary bladder cancer. *Nat Genet.* 2010; 42:415–9. [PubMed: 20348956]
 11. Wu X, et al. Genetic variation in the prostate stem cell antigen gene PSCA confers susceptibility to urinary bladder cancer. *Nat Genet.* 2009; 41:991–5. [PubMed: 19648920]
 12. Eeles RA, et al. Identification of seven new prostate cancer susceptibility loci through a genome-wide association study. *Nat Genet.* 2009; 41:1116–21. [PubMed: 19767753]
 13. Yeager M, et al. Identification of a new prostate cancer susceptibility locus on chromosome 8q24. *Nat Genet.* 2009; 41:1055–7. [PubMed: 19767755]
 14. Crowther-Swanepoel D, et al. Common variants at 2q37.3, 8q24.21, 15q21.3 and 16q24.1 influence chronic lymphocytic leukemia risk. *Nat Genet.* 2010; 42:132–6. [PubMed: 20062064]
 15. Tomlinson IP, et al. A genome-wide association study identifies colorectal cancer susceptibility loci on chromosomes 10p14 and 8q23.3. *Nat Genet.* 2008; 40:623–30. [PubMed: 18372905]
 16. Easton DF, et al. Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature.* 2007; 447:1087–93. [PubMed: 17529967]
 17. Zanke BW, et al. Genome-wide association scan identifies a colorectal cancer susceptibility locus on chromosome 8q24. *Nat Genet.* 2007; 39:989–94. [PubMed: 17618283]
 18. Yeager M, et al. Genome-wide association study of prostate cancer identifies a second risk locus at 8q24. *Nat Genet.* 2007; 39:645–9. [PubMed: 17401363]
 19. Rafnar T, et al. Sequence variants at the TERT-CLPTM1L locus associate with many cancer types. *Nat Genet.* 2009; 41:221–7. [PubMed: 19151717]
 20. Stacey SN, et al. New common variants affecting susceptibility to basal cell carcinoma. *Nat Genet.* 2009
 21. Landi MT, et al. A genome-wide association study of lung cancer identifies a region of chromosome 5p15 associated with risk for adenocarcinoma. *Am J Hum Genet.* 2009; 85:679–91. [PubMed: 19836008]
 22. Shete S, et al. Genome-wide association study identifies five susceptibility loci for glioma. *Nat Genet.* 2009; 41:899–904. [PubMed: 19578367]
 23. Petersen GM, et al. A genome-wide association study identifies pancreatic cancer susceptibility loci on chromosomes 13q22.1, 1q32.1 and 5p15.33. *Nat Genet.* 2010; 42:224–8. [PubMed: 20101243]
 24. Freedman ML, et al. Assessing the impact of population stratification on genetic association studies. *Nat Genet.* 2004; 36:388–93. [PubMed: 15052270]
 25. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature.* 2007; 447:661–78. [PubMed: 17554300]
 26. García-Closas M, et al. A single nucleotide polymorphism identified in a genome-wide scan tags variation in the N-acetyltransferase 2 phenotype in populations of European background. *Pharmacogenet Genomics.* 2010
 27. Conticello SG. The AID/APOBEC family of nucleic acid mutators. *Genome Biol.* 2008; 9:229. [PubMed: 18598372]
 28. Malumbres M, Barbacid M. To cycle or not to cycle: a critical decision in cancer. *Nat Rev Cancer.* 2001; 1:222–31. [PubMed: 11902577]
 29. Richter J, et al. High-throughput tissue microarray analysis of cyclin E gene amplification and overexpression in urinary bladder cancer. *Am J Pathol.* 2000; 157:787–94. [PubMed: 10980118]
 30. Strassburg CP, Lankisch TO, Manns MP, Ehmer U. Family 1 uridine-5'-diphosphate glucuronosyltransferases (UGT1A): from Gilbert's syndrome to genetic organization and variability. *Arch Toxicol.* 2008; 82:415–33. [PubMed: 18491077]
 31. Ando Y, et al. Polymorphisms of UDP-glucuronosyltransferase gene and irinotecan toxicity: a pharmacogenetic analysis. *Cancer Res.* 2000; 60:6921–6. [PubMed: 11156391]

32. Strassburg CP, Manns MP, Tukey RH. Differential down-regulation of the UDP-glucuronosyltransferase 1A locus is an early event in human liver and biliary cancer. *Cancer Res.* 1997; 57:2979–85. [PubMed: 9230212]
33. Strassburg CP, Nguyen N, Manns MP, Tukey RH. Polymorphic expression of the UDP-glucuronosyltransferase UGT1A gene locus in human gastric epithelium. *Mol Pharmacol.* 1998; 54:647–54. [PubMed: 9765507]
34. Giuliani L, et al. Can down-regulation of UDP-glucuronosyltransferases in the urinary bladder tissue impact the risk of chemical carcinogenesis? *Int J Cancer.* 2001; 91:141–3. [PubMed: 11149414]
35. Iida K, et al. Suppression of AhR signaling pathway is associated with the down-regulation of UDP-glucuronosyltransferases during BBN-induced urinary bladder carcinogenesis in mice. *J Biochem.* 2009; 147:353–60. [PubMed: 19880377]
36. Calado RT, Young NS. Telomere maintenance and human bone marrow failure. *Blood.* 2008; 111:4446–55. [PubMed: 18239083]
37. Armanios MY, et al. Telomerase mutations in families with idiopathic pulmonary fibrosis. *N Engl J Med.* 2007; 356:1317–26. [PubMed: 17392301]
38. Tsakiri KD, et al. Adult-onset pulmonary fibrosis caused by mutations in telomerase. *Proc Natl Acad Sci U S A.* 2007; 104:7552–7. [PubMed: 17460043]
39. Calado RT, et al. Constitutional hypomorphic telomerase mutations in patients with acute myeloid leukemia. *Proc Natl Acad Sci U S A.* 2009; 106:1187–92. [PubMed: 19147845]
40. Park JH, et al. Estimation of effect size distribution from genome-wide association studies and implications for future discoveries. *Nat Genet.* 2010; 42:570–5. [PubMed: 20562874]
41. Hein DW. N-acetyltransferase 2 genetic polymorphism: effects of carcinogen and haplotype on urinary bladder cancer risk. *Oncogene.* 2006; 25:1649–58. [PubMed: 16550165]
42. Wigginton JE, Cutler DJ, Abecasis GR. A note on exact tests of Hardy-Weinberg equilibrium. *Am J Hum Genet.* 2005; 76:887–93. [PubMed: 15789306]
43. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. *Genetics.* 2000; 155:945–59. [PubMed: 10835412]
44. Frazer KA, et al. A second generation human haplotype map of over 3.1 million SNPs. *Nature.* 2007; 449:851–61. [PubMed: 17943122]
45. Price AL, et al. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet.* 2006; 38:904–9. [PubMed: 16862161]
46. Patterson N, Price AL, Reich D. Population structure and eigenanalysis. *PLoS Genet.* 2006; 2:e190. [PubMed: 17194218]
47. Yu K, et al. Population substructure and control selection in genome-wide association studies. *PLoS One.* 2008; 3:e2551. [PubMed: 18596976]
48. de Bakker PI, et al. Practical aspects of imputation-driven meta-analysis of genome-wide association studies. *Hum Mol Genet.* 2008; 17:R122–8. [PubMed: 18852200]
49. Sun L, Wilder K, McPeck MS. Enhanced pedigree error detection. *Hum Hered.* 2002; 54:99–110. [PubMed: 12566741]
50. Clayton D. Testing for association on the X chromosome. *Biostatistics.* 2008; 9:593–600. [PubMed: 18441336]
51. Fearnhead P. SequenceLDhot: detecting recombination hotspots. *Bioinformatics.* 2006; 22:3061–6. [PubMed: 17060358]
52. Fearnhead P, Harding RM, Schneider JA, Myers S, Donnelly P. Application of coalescent methods to reveal fine-scale rate variation and recombination hotspots. *Genetics.* 2004; 167:2067–81. [PubMed: 15342541]

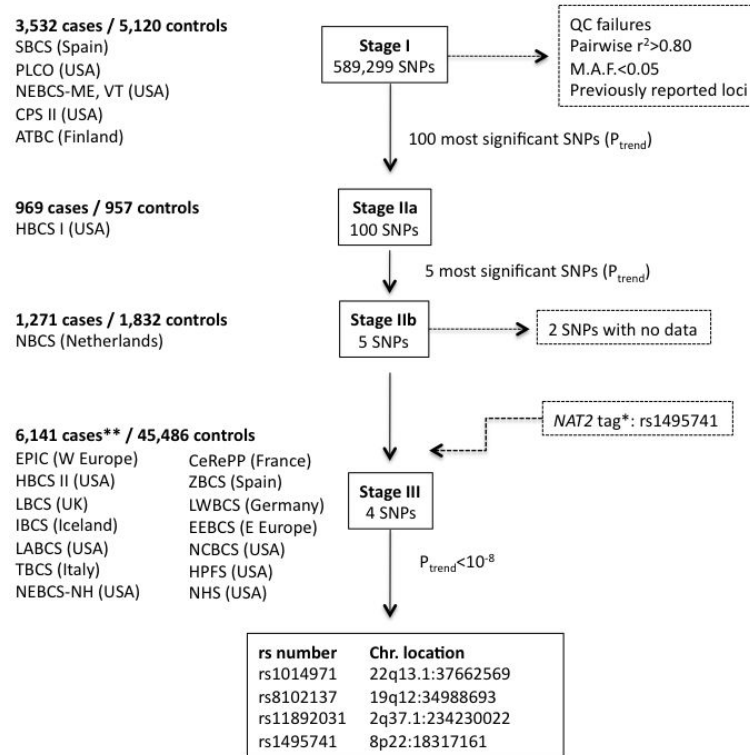


Figure 1. Study design of multi-stage GWAS of bladder cancer

See **Online Methods** and Supplementary Table 1 for details of study designs and sample sizes. *The tag SNP, rs1495741 located 3' of NAT241 was genotyped in subjects in stage II and III studies as well as on the Illumina bead chips used in stage I. **Includes 338 additional cases from NBCS that were added to the final combined analyses.

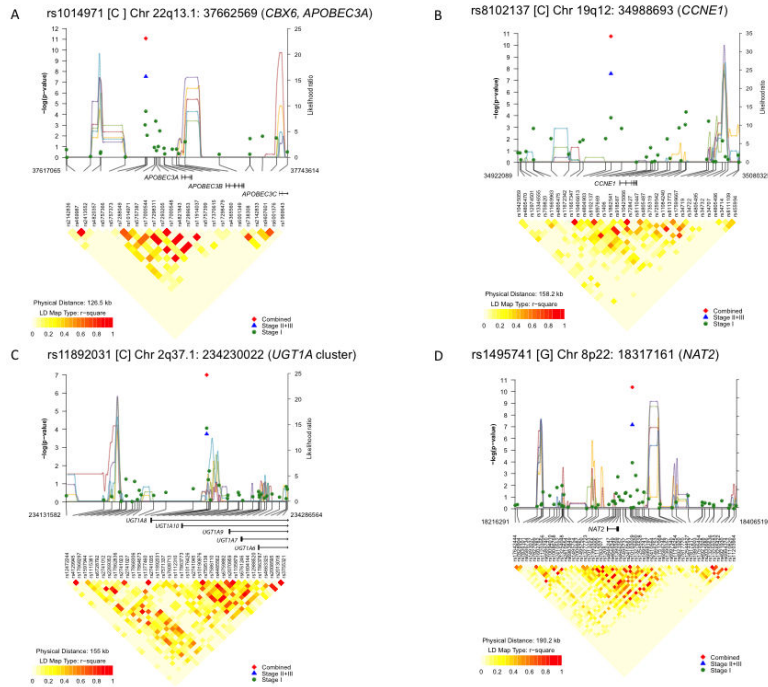


Figure 2. Association results, recombination and linkage disequilibrium plots for four regions on chromosomes 22q13.1, 19q12, 2q37.1 and 8p22

Results of stage I (green circles), combined stages II and III (blue diamonds) and combined data from the three stages (red diamonds) with P-values for log-additive association results with recombination rates (cm/Mb) based on HapMap phase II data. Pairwise r^2 values based on control populations are displayed at the bottom for all SNPs included in the GWAS analysis. **Panel A** depicts chromosome 22q13.1 region (37,617,065 to 37,743,614). **Panel B** depicts the region of chromosome 19q12 (34,922,089 to 35,080,325). **Panel C** depicts the region of 2q37.1 (234,131,582 to 234,286,564). **Panel D** depicts the region of 8p22 (18,216,291 to 18,406,519). Genomic coordinates are based on NCBI Human Genome Build 36.3.

Marker ^a , risk allele ^b , chr ^c , location ^c and gene ^d	Groups of studies	N ^e	Cases	Controls	Freq ^f	Allelic OR (95% CI) ^g	P value ^h		
Deletion Assay	Previously reported ⁱ 7,9	28	5,072	6,466	0.51	1.46	1.58	1.9E-21	
Chr 1p13.3	Stage ^j	4	2,480	3,222	0.49	1.49	1.33	1.68	3.7E-11
<i>GSTM1</i>	Combined	32	7,552	9,688	0.53	1.47	1.38	1.57	5.0E-31

^aNCBI dbSNP identifier.

^bRisk allele shown in [].

^cChromosome and NCBI Human Genome Build 36.3 location.

^dGene neighborhood closest to the most notable SNP.

^eN: number of studies

^fRisk allele frequency in control populations.

^gEstimate assuming multiplicative odds model, OR, odds ratio; CI, 95% confidence interval.

^h1 d.f. trend test.

ⁱSummary estimates differ slightly from previously published because we used a fixed effects meta-analysis.

^jData from SBCS was excluded from stage I because it had been included in the previous meta-analyses published in Garcia-Closas et al. 20057.. Data from NEBCS are reported separately (Unpublished data to appear in "GSTM1 null and NAT2 Slow Acetylation Genotypes, Smoking Intensity, and Bladder Cancer Risk: Results from the New England Bladder Cancer Case-Control Study and Meta-Analyses" by Moore LE, Barris D, Figueroa J, Garcia-Closas M, Karagas M, Schwenn M, Johnson A, Lubin J, Hein DW, Dagnall C, Colt J, Kida M, Jones M, Schmed A, Cherala S, Chanock S, Cantor K, Silverman D, Rothman N)

Table 2

Novel SNPs identified in a multi-stage GWAS of bladder cancer

Results of the meta-analysis of genotype counts included in combined stages I, II and III. The initial scan results were adjusted by age, gender, smoking status, study site and DNA source.

Marker ^a , minor allele ^b , chr ^c , location ^c and gene ^d	Studies included	N ^e	Cases	Controls	Freq. ^f	Allelic OR (95%CI) ^g	P value ^h		
rs1014971 [C]	Stage I	5	3,529	5,092	0.38	0.87	0.93	6.1E-05	
Chr 22q13.1:	Stage II and III	15	8,277	47,517	0.37	0.89	0.85	0.93	3.0E-08
37662569	Combined	20	11,806	52,609	0.38	0.88	0.85	0.91	8.4E-12
<i>CBX6, APOBEC3A</i>									
rs8102137 [C]	Stage I	5	3,530	5,114	0.33	1.13	1.06	1.21	1.6E-04
Chr 19q12:	Stage II and III	15	8,261	47,708	0.32	1.13	1.08	1.18	2.6E-08
34988693	Combined	20	11,791	52,822	0.33	1.13	1.09	1.17	1.7E-11
<i>CCNE1</i>									
rs11892031 [C]	Stage I	5	3,524	5,108	0.08	0.79	0.70	0.89	6.9E-05
Chr 2q37.1:	Stage II and III	15	8,284	47,727	0.08	0.86	0.80	0.93	1.8E-04
234230022	Combined	20	11,808	52,835	0.08	0.84	0.79	0.89	1.0E-07
<i>UGT1A cluster j</i>									
rs1495741 [G]	Stage I	5	3525	5116	0.24	0.86	0.80	0.93	1.5E-04
Chr 8p22:	Stage II and III	15	8,279	47,744	0.22	0.87	0.83	0.92	6.8E-08
18317161	Combined	20	11,804	52,860	0.24	0.87	0.83	0.91	4.2E-11
<i>NAT2</i>									

^aNCBI dbSNP identifier.

^bMinor allele shown in [].

^cChromosome and NCBI Human Genome Build 36.3 location.

^dGene neighborhood closest to the most notable SNP.

^eN: number of studies

^fMinor allele frequency in control populations.

^gEstimate assuming multiplicative odds model, OR, odds ratio; CI, 95% confidence interval.

f_1 d.f. trend test.

Table 3
Interaction of NAT2 tagSNP (rs1495741) and GSTM1 deletion with cigarette smoking in bladder cancer risk

Results from logistic regression analyses of aggregated data adjusted by study. rs1495741 genotypes were classified according to an established approach based on the phenotype of NAT2 acetylators; AA/AG were collapsed to tag rapid/intermediate acetylation, and GG tags slow acetylation.²⁶

<i>NAT2</i> (rs1495741)	Cases			Controls			P		
	AA/AG	GG	AA/AG	GG	OR	95% CI	P	Interaction	
All subjects	3,784	6,915	5,233	8,182	1.15	1.09	1.22	2.9E-07	
By smoking status									
Never smoker	760	1,202	1,679	2,758	0.95	0.85	1.06	3.3E-01 Ref.	
Ever smoker	3,024	5,713	3,554	5,424	1.24	1.16	1.32	4.3E-11 6.3E-05	
Former smoker	1,859	3,455	2,300	3,559	1.20	1.11	1.29	6.8E-06 6.7E-04	
Current smoker	1,165	2,258	1,254	1,865	1.27	1.14	1.40	7.2E-06 1.6E-04	
<i>GSTM1 Del</i>	Present	Null	Present	Null					
All subjects	1,319	1,995	1,717	1,726	1.47	1.33	1.62	4.4E-14	
By smoking status									
Never smoker	210	346	519	510	1.71	1.38	2.12	6.9E-07 Ref.	
Ever smoker	1,109	1,649	1,198	1,216	1.47	1.30	1.67	2.1E-09 1.1E-01	
Former smoker	564	961	622	653	1.62	1.39	1.89	4.7E-10 6.9E-01	
Current smoker	545	688	576	563	1.19	1.00	1.40	4.5E-02 8.1E-03	