Human Molecular Genetics, 2016, Vol. 25, No. 6 1203–1214

doi: 10.1093/hmg/ddv492
Advance Access Publication Date: 4 January 2016
Association Studies Article

ASSOCIATION STUDIES ARTICLE

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

Jonine D. Figueroa[1,2,†,*], Candace D. Middlebrooks[1,†], A. Rouf Banday[1,†], Yuanqing Ye[3,†], Montserrat Garcia-Closas[1,5,†], Nilanjan Chatterjee[1,†], Stella Koutros[1], Lambertus A. Kiemeney[6], Thorunn Rafnar[7], Timothy Bishop[8], Helena Furberg[10], Giuseppe Matullo[14,15], Klaus Golka[16], Manuela Gago-Dominguez[17], Jack A. Taylor[18,19], Tony Fletcher[20], Afshan Siddiq[21], Victoria K. Cortessis[23,24,25], Charles Kooperberg[26], Olivier Cussenot[27,30,31], Simone Benhamou[32,33], Jennifer Prescott[34,35], Stefano Porru[38], Colin P. Dinney[4], Núria Malats[39], Dalsu Baris[1], Mark P. Purdue[1], Eric J. Jacobs[40], Demetrius Albanes[1], Zhaoming Wang[41], Charles C. Chung[1,4], Sita H. Vermeulen[6], Katja K. Aben[6], Tessel E. Galesloot[6], Gudmar Thorleifsson[7], Patrick Sulem[7], Kari Stefansson[7,42], Anne E. Kiltie[43], Mark Harland[8], Mark Teo[9], Kenneth Offit[11], Joseph Vijai[11], Dean Bajorin[12], Ryan Kopp[13], Giovanni Fiorito[14,15], Simonetta Guarrera[14,15], Carlotta Sacerdote[44], Silvia Selinski[16], Jan G. Hengstler[16], Holger Gerullis[45,46], Daniel Ovsiannikov[47], Meinolf Blaszkewicz[16], Jose Esteban Castelao[48], Manuel Calaza[17,49], Maria Elena Martinez[50], Patricia Cordeiro[51], Zongli Xu[18], Vijayalakshmi Panduri[18,19], Rajiv Kumar[52], Eugene Gurzau[53], Kvetoslava Koppova[54], H. Bas Bueno-De-Mesquita[21,55,56], Börje Ljungberg[57], Françoise Clavel-Chapelon[58,59,60], Elisabete Weiderpass[61,62,63,64], Vittorio Krogh[65], Miren Dorronsoro[66,67], Ruth C. Travis[68], Anne Tjønneland[69], Paul Brennan[70], Jenny Chang-Claude[52], Elio Riboli[21], David Conti[21,24], Marianna C. Stern[21,24], Malcolm C. Pike[8], David Van Den Berg[21,24], Jian-Min Yuan[71],

[†]These authors contributed equally to this work.
[‡]These authors jointly directed this work.

Received: July 31, 2015. **Revised:** October 12, 2015. **Accepted:** November 26, 2015

Chancellor Hohensee[26], Rebecca P. Jeppson[26], Geraldine Cancel-Tassin[30,31], Morgan Roupret[28,30,31], Eva Comperat[29,30,31], Constance Turman[35], Immaculata De Vivo[25,34], Edward Giovannucci[34,35,36], David J. Hunter[34,35,36,72], Peter Kraft[35,37], Sara Lindstrom[35], Angela Carta[38], Sofia Pavanello[73], Cecilia Arici[38], Giuseppe Mastrangelo[73], Ashish M. Kamat[4], Liren Zhang[3], Yilei Gong[3], Xia Pu[3], Amy Hutchinson[41], Laurie Burdett[41], William A. Wheeler[74], Margaret R. Karagas[75], Alison Johnson[76], Alan Schned[75], G. M. Monawar Hosain[77], Molly Schwenn[78], Manolis Kogevinas[67,79,80,81], Adonina Tardón[67,82], Consol Serra[67,80,83], Alfredo Carrato[84], Reina García-Closas[85], Josep Lloreta[67], Gerald Andriole Jr[86], Robert Grubb III[86], Amanda Black[1], W. Ryan Diver[40], Susan M. Gapstur[40], Stephanie Weinstein[1], Jarmo Virtamo[87], Christopher A. Haiman[24], Maria Teresa Landi[1], Neil E. Caporaso[1], Joseph F. Fraumeni Jr[1], Paolo Vineis[15,22,‡], Xifeng Wu[3,‡], Stephen J. Chanock[1,‡], Debra T. Silverman[1,‡], Ludmila Prokunina-Olsson[1,‡] and Nathaniel Rothman[1,‡]

[1]Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD, USA, [2]Usher Institute of Population Health Sciences and Informatics, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK, [3]Department of Epidemiology, MD Anderson Cancer Center, Houston, TX, USA, [4]Department of Urology, MD Anderson Cancer Center, Houston, TX, USA, [5]Division of Genetics and Epidemiology, Institute of Cancer Research, London, UK, [6]Radboud Institute for Health Sciences, Radboud University Medical Center, Nijmegen, The Netherlands, [7]deCODE Genetics/Amgen, Inc., Reykjavik, Iceland, [8]Section of Epidemiology and Biostatistics, [9]Radiotherapy Research Group, Leeds Institute of Cancer and Pathology, University of Leeds, Leeds LS9 7TF, UK, [10]Department of Epidemiology and Biostatistics, [11]Department of Medicine, [12]Genitourinary Oncology Service, Division of Solid Tumor Oncology, Department of Medicine, [13]Urology Service, Department of Surgery, Memorial Sloan Kettering Cancer Center, New York, NY, USA, [14]Department of Medical Sciences, University of Turin, Turin, Italy, [15]Human Genetics Foundation, Turin, Italy, [16]Leibniz Research Centre for Working Environment and Human Factors, Dortmund, Germany, [17]Genomic Medicine Group, Galician Foundation of Genomic Medicine, Servicio Galego de Saude (SERGAS), Instituto de Investigación Sanitaria de Santiago (IDIS), Santiago de Compostela, Spain, [18]Epidemiology Branch, National Institute of Environmental Health Sciences (NIEHS), [19]Epigenetic and Stem Cell Biology Laboratory, National Institute of Environmental Health Sciences (NIEHS), National Institutes of Health (NIH), Research Triangle Park, NC, USA, [20]London School of Hygiene and Tropical Medicine, London, UK, [21]School of Public Health, [22]MRC-PHE Centre for Environment and Health, School of Public Health, Imperial College London, London, UK, [23]Department of Preventive Medicine, USC Keck School of Medicine, [24]Department of Obstetrics and Gynecology, [25]Norris Comprehensive Cancer Center, USC Keck School of Medicine, University of Southern California, Los Angeles, CA, USA, [26]Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA, USA, [27]Department of Urology, Tenon, [28]Department of Urology, Pitié-Salpétrière, [29]Department of Pathology, Pitié-Salpétrière, Assistance-Publique Hôpitaux de Paris (APHP), Paris, France, [30]Centre de Recherche sur les Pathologies Prostatiques, Paris, France, [31]UPMC Univ Paris 06, GRC n°5, ONCOTYPE-URO, Paris, France, [32]Institut national de la sante et de la recherche medicale, U946, Foundation Jean Dausset Centre d'Etude du Polymorphisme Humain (CEPH), Paris, France, [33]Centre National de la Receherche Scientifique, UMR8200, Institut Gustave-Roussy, Villejuif, France, [34]Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA, [35]Department of Epidemiology, [36]Department of Nutrition, [37]Department of Biostatistics, Harvard School of Public Health, Boston, MA, USA, [38]Department of Medical and Surgical Specialties, Radiological Sciences and Public Health, University of Brescia, Brescia, Italy, [39]Genetic and Molecular Epidemiology Group, Spanish National Cancer Research Centre (CNIO), Madrid, Spain, [40]Epidemiology Research Program, American Cancer Society, Atlanta, GA, USA, [41]Cancer Genomics Research Laboratory, Division of Cancer

Epidemiology and Genetics, National Cancer Institute, Gaithersburg, MD, USA, [42]Faculty of Medicine, University of Iceland, Reykjavik, Iceland, [43]CRUK/MRC Oxford Institute for Radiation Oncology, Department of Oncology, University of Oxford, Old Road Campus Research Building, Roosevelt Drive, Headington, Oxford OX3 7DQ, UK, [44]Cancer Epidemiology, CPO Piemonte, Turin, Italy, [45]University Hospital for Urology, Klinikum Oldenburg, School of Medicine and Health Sciences, Carl von Ossietzky University Oldenburg, Oldenburg, Germany, [46]Department of Urology, Lukasklinik Neuss, Germany, [47]Department of Urology, St. Josefs Hospital, Dortmund-Hörde, Germany, [48]Oncology and Genetics Unit, Complejo Hospitalario, Instituto de Investigacion Biomedica (IBI) Orense-Pontevedra-Vigo, Xerencia de Xestion Integrada de Vigo-SERGAS, Vigo, Spain, [49]Center for Research in Molecular Medicine and Chronic Diseases (CIMUS), University of Santiago de Compostela, Galicia, Spain, [50]Department of Family Medicine and Public Health, Moores Cancer Center, University of California San Diego, San Diego, CA, USA, [51]Department of Urology, Complejo Hospitalario, University of Santiago de Compostela, Servicio Galego de Saude (SERGAS), Santiago de Compostela, Spain, [52]Division of Cancer Epidemiology, German Cancer Research Center (DKFZ), Heidelberg, Baden-Württemberg; University Cancer Center Hamburg (UCCH), University Medical Center Hamburg-Eppendorf, Hamburg, Germany, [53]Environmental Health Center, Cluj, Romania, [54]State Health Institute, Banska Bystrica, Slovakia, [55]Department for Determinants of Chronic Diseases (DCD), National Institute for Public Health and the Environment (RIVM), Bilthoven, The Netherlands, [56]Department of Social & Preventive Medicine, Faculty of Medicine, University of Malaya, Kuala Lumpur, Malaysia, [57]Department of Surgical and Perioperative Sciences, Urology and Andrology, Umea University, Umea, Sweden, [58]Inserm, Centre for research in Epidemiology and Population Health (CESP), U1018, Nutrition, Hormones and Women's Health team, Villejuif F-94805, France, [59]Université Paris Sud, UMRS 1018, Villejuif F-94805, France, [60]Institut Gustave Roussy, Villejuif F-94805, France, [61]Department of Community Medicine, Faculty of Health Sciences, University of Tromsø, The Arctic University of Norway, Tromsø, Norway, [62]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden, [63]Cancer Registry of Norway, Institute of Population-Based Cancer Research, Oslo, Norway, [64]Genetic Epidemiology Group, Folkhälsan Research Center, Helsinki, Finland, [65]Epidemiology and Prevention Unit, Fondazione IRCCS Istituto Nazionale dei Tumori, Milano, Italy, [66]Health Department, BioDonostia Research Institute, Basque Region, Spain, [67]Centro de Investigación Biomédica en Red de Epidemiología y Salud Pública (CIBERESP), Madrid, Spain, [68]Cancer Epidemiology Unit, University of Oxford, Oxford, UK, [69]Danish Cancer Society Research Center, Copenhagen, Denmark, [70]International Agency for Research on Cancer (IARC), Lyon, France, [71]University of Pittsburgh Cancer Institute, Pittsburgh, PA, USA, [72]Broad Institute of Harvard and MIT, Cambridge, MA, USA, [73]Department of Cardiac, Thoracic and Vascular Sciences, University of Padova, Padova, Italy, [74]Information Management Services, Silver Spring, MD, USA, [75]Geisel School of Medicine, Dartmouth College, Hanover, NH, USA, [76]Vermont Cancer Registry, Burlington, VT, USA, [77]New Hampshire State Cancer Registry, Concord, NH, USA, [78]Maine Cancer Registry, Augusta, ME, USA, [79]Centre for Research in Environmental Epidemiology (CREAL), Barcelona, Spain, [80]Municipal Institute of Medical Research, (IMIM—Hospital del Mar), Barcelona, Spain, [81]National School of Public Health, Athens, Greece, [82]Instituto Universitario de Oncología, Universidad de Oviedo, Oviedo, Spain, [83]Departament de Ciències Experimentals i de la Salut, Universitat Pompeu Fabra, Barcelona, Spain, [84]Ramon y Cajal University Hospital, IRYCIS, Madrid, Spain, [85]Unidad de Investigación, Hospital Universitario de Canarias, La Laguna, Spain, [86]Division of Urologic Surgery, Washington University School of Medicine, Saint Louis, MO, USA and [87]Department of Chronic Disease Prevention, National Institute for Health and Welfare, Helsinki, Finland

*To whom correspondence should be addressed at: Usher Institute of Population Health Sciences and Informatics, Medical School, University of Edinburgh, Teviot Place, Edinburgh EH8 9AG, UK. Tel: +44 (0)1316514140; Fax: +44 (0)7478467698; Email: jonine.figueroa@ed.ac.uk

Abstract

Candidate gene and genome-wide association studies (GWAS) have identified 15 independent genomic regions associated with bladder cancer risk. In search for additional susceptibility variants, we followed up on four promising single-nucleotide polymorphisms (SNPs) that had not achieved genome-wide significance in 6911 cases and 11 814 controls (rs6104690, rs4510656, rs5003154 and rs4907479, $P < 1 \times 10^{-6}$), using additional data from existing GWAS datasets and targeted genotyping for studies that did not have GWAS data. In a combined analysis, which included data on up to 15 058 cases and 286 270 controls, two SNPs achieved genome-wide statistical significance: rs6104690 in a gene desert at 20p12.2 ($P = 2.19 \times 10^{-11}$) and rs4907479 within the *MCF2L* gene at 13q34 ($P = 3.3 \times 10^{-10}$). Imputation and fine-mapping analyses were performed in these two regions for a subset of

5551 bladder cancer cases and 10 242 controls. Analyses at the 13q34 region suggest a single signal marked by rs4907479. In contrast, we detected two signals in the 20p12.2 region—the first signal is marked by rs6104690, and the second signal is marked by two moderately correlated SNPs ($r^2 = 0.53$), rs6108803 and the previously reported rs62185668. The second 20p12.2 signal is more strongly associated with the risk of muscle-invasive (T2-T4 stage) compared with non-muscle-invasive (Ta, T1 stage) bladder cancer (case–case $P \leq 0.02$ for both rs62185668 and rs6108803). Functional analyses are needed to explore the biological mechanisms underlying these novel genetic associations with risk for bladder cancer.

Introduction

Each year ~380 000 bladder cancer cases are diagnosed worldwide (1,2). While smoking is estimated to explain ~50% of bladder cancer, genetic susceptibility has also been noted to contribute to its etiology (2–4). Family history of bladder cancer in a first degree relative is associated with a ~1.7-fold increased risk, comparable with many other common adult cancers (e.g. breast, prostate, colon) (5,6). To date, candidate gene and genome-wide association studies (GWAS) have identified 15 genomic regions that harbor bladder cancer genetic susceptibility variants. These include 1p13.3 (*GSTM1*), 2q37.1 (*UGT1A* cluster), 3q26.2 (*TERC*), 3q28 (*TP63*), 4p16.3 (*TMEM129* and *TACC3-FGFR3*), 5p15.33 (*TERT-CLPTM1L*), 8p22 (*NAT2*), 8q24.21, 8q24.3 (*PSCA*), 11p15.5 (LSP1), 15q24 (CYP1A2), 18q12.3 (*SLC14A1*), 19q12 (*CCNE1*), 20p12.2 and 22q13.1 (*CBX6*, *APOBEC3A*) (7–19). Based on analysis of the reported signals that reached a conclusive threshold of genome-wide significance (20), we estimate that many additional common genetic variants for bladder cancer are yet to be discovered (11).

To identify new bladder cancer susceptibility variants, we followed up on four promising SNPs ($P < 1 \times 10^{-6}$) that did not achieve genome-wide significance in our previously reported meta-analysis of three independently published GWAS performed in individuals of European ancestry [National Cancer Institute (NCI)-GWAS1, NCI-GWAS2 and the Texas Bladder Cancer Study (TXBCS)-GWAS] (11,12,19). In addition, we genotyped five promising SNPs identified in a genome-wide interaction study of smoking and bladder cancer risk (21).

Results

In our previous meta-analysis of three bladder cancer GWAS: NCI-GWAS1 (8,11), NCI-GWAS2 (19) and TXBCS-GWAS (12) totaling 6911 cases and 11 814 controls of European descent, we identified four SNPs with promising associations of $P < 1 \times 10^{-6}$ (19). These SNPs were genotyped in an independent set of samples (4427 cases and 5881 controls) with individual TaqMan assays. We also obtained genotype data from existing GWAS data for 1724 cases and 265 722 controls from Iceland and 1996 cases and 2853 controls from the Netherlands (9,22) (see the Material and Methods section). Details of the studies and the genotyping data are summarized in Supplementary Material, Table S1. In a combined meta-analysis, two of the four promising SNPs achieved the threshold of genome-wide significance: rs4907479 at 13q34 ($P = 3.3 \times 10^{-10}$) and rs6104690 at 20p12.2 ($P = 2.19 \times 10^{-11}$). Study-specific estimates are shown in Figure 1. Two other promising SNPs rs4510656 and rs5003154 did not achieve genome-wide significance with additional data (Supplementary Material, Fig. S1).

We also evaluated five SNPs identified as suggestive in a genome-wide scan for interaction of smoking and bladder cancer risk (21). The most promising initial signals were rs1711973 (*FOXF2*) at 6p25.3 in never-smokers ($P = 5.18 \times 10^{-7}$, OR = 1.34) and rs12216499 (*RSPH3-TAGAP-EZR*) at 6q25.3 in ever-smokers ($P = 6.35 \times 10^{-7}$, OR = 0.75) (21). However, the current analysis in an additional set of almost 1000 never-smokers and 3000 ever-smokers did not provide supportive evidence for association of these variants with bladder cancer risk (Supplementary Material, Table S2).

To further refine the association signals with bladder cancer risk, we imputed the 13q34 and 20p12.2 regions in a subset of 5551 bladder cancer cases and 10 242 controls from NCI-GWAS1 and NCI-GWAS2. Imputation was done within 1 Mb windows centered on the GWAS markers and using the 1000 Genomes reference panel (Phase 3 October 2014). For the 13q34 region, we analyzed 1370 imputed and 146 genotyped markers (Supplementary Material, Table S3). Among the 1516 markers evaluated, we identified 29 additional SNPs in high linkage disequilibrium (LD, $r^2 > 0.8$) with rs4907479 and associated with bladder cancer risk ($P < 2.0 \times 10^{-4}$). These variants are located within a 24 Kb genomic region, in the first two introns of the *MCF2L* gene (Fig. 2). Fine-mapping analysis showed that these variants are highly correlated and analyses adjusting for the GWAS SNP (rs4907479) did not reveal an independent signal, thus pointing toward a single susceptibility locus marked by rs4907479 ($P = 1.92 \times 10^{-5}$, OR = 1.13).

Since bladder cancer risk variants at 20p12.2 have previously been reported (18,19), we sought to confirm and clarify these previous associations and determine if additional signals were present. After analysis of 2344 imputed and 246 genotyped markers across the 20p12.2 region, we observed three markers with comparably strong signals: rs6104690 ($P = 3.97 \times 10^{-5}$, OR = 1.11), rs6108803 ($P = 1.82 \times 10^{-6}$, OR = 1.18) and rs62185668 ($P = 1.39 \times 10^{-5}$, OR = 1.14, Table 1, Fig. 3 and Supplementary Material, Table S4). Per-allele odds ratio (OR) estimates adjusting for the various 20p12.2 marker combinations are presented in Table 1. Regardless of the models, rs6104690 and rs6108803 showed significant associations with bladder cancer risk ($P \leq 0.03$), while the association for rs62185668 was no longer significant ($P = 0.25$) when adjusting for the newly identified 20p12.2 SNP rs6108803. Logistic models that adjusted for the effects of two other markers showed significant residual associations for rs6104690 ($P = 0.03$, OR = 1.07) and rs6108803 ($P = 0.03$, OR = 1.12), but not rs62185668 ($P = 0.70$, OR = 1.02) (Table 1). We observed only a weak association (OR = 1.07, 1.02–1.13; $P = 1.00 \times 10^{-2}$, Fig. 3) with bladder cancer risk for the previously reported rs4813953 (18). Haplotype analysis at 20p12.2 showed the strongest associations with bladder cancer risk when at least two risk alleles were present ($P \leq 4.10 \times 10^{-6}$). We observed the most significant association with bladder cancer risk for the combination of rs6104690 and rs6108803; a haplotype with risk alleles of both markers had an OR = 1.21, $P = 2.00 \times 10^{-7}$ (Supplementary Material, Table S5). Further analysis in the same set of samples showed that the presence of risk alleles of all three SNPs did not improve the association (OR = 1.20, $P = 5.00 \times 10^{-7}$) above what was seen in the two-SNP haplotype analysis (Supplementary Material, Table S6), implying that the bladder cancer association signal in the 20p12.2 region could be efficiently captured by genotyping two markers, rs6104690 and rs6108803.

Analysis by tumor stage and grade (Table 2 and Supplementary Material, Table S7) did not show significant associations with tumor characteristics for the 13q34 signal rs4907479. For

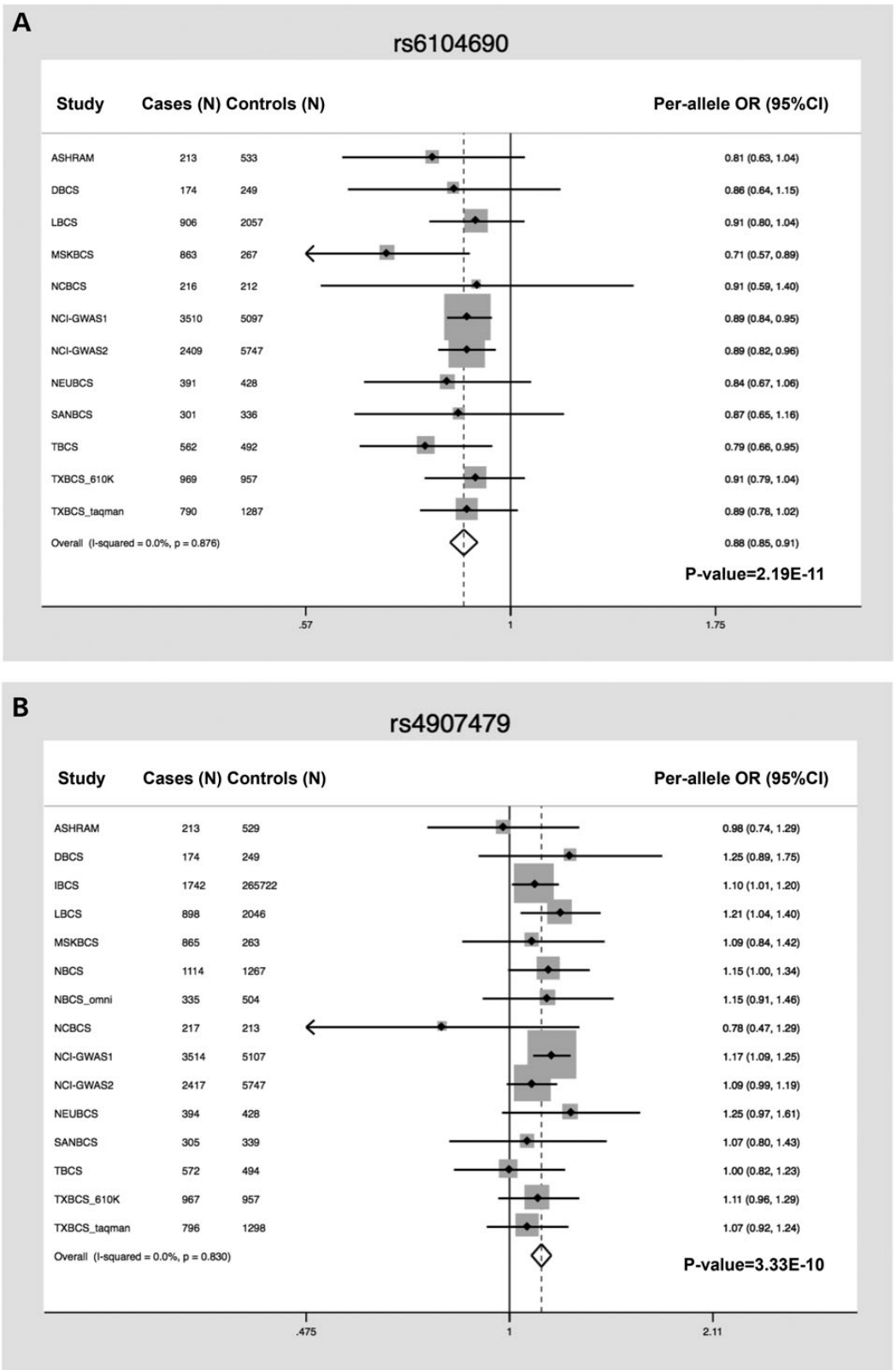

Figure 1. Forest plots of meta-analyses results with bladder cancer risk for SNPs rs6104690 at 20p12.2 and rs4907479 at 13q34. Metaplots for SNPs rs6104690 at 20p12.2 (**A**) and rs4907479 at 13q34 (**B**). Details of individual studies are presented in Supplementary Material, Table S1. The New England Bladder Cancer Study (NEBCS) represents a single study comprised of Maine (ME) and Vermont (VT) components genotyped in NCI-GWAS1, and the New Hampshire (NH) component genotyped in NCI-GWAS2. Fixed-effects meta-analysis by study was used to calculate the combined OR, 95% CI and P-trend for the variant allele.

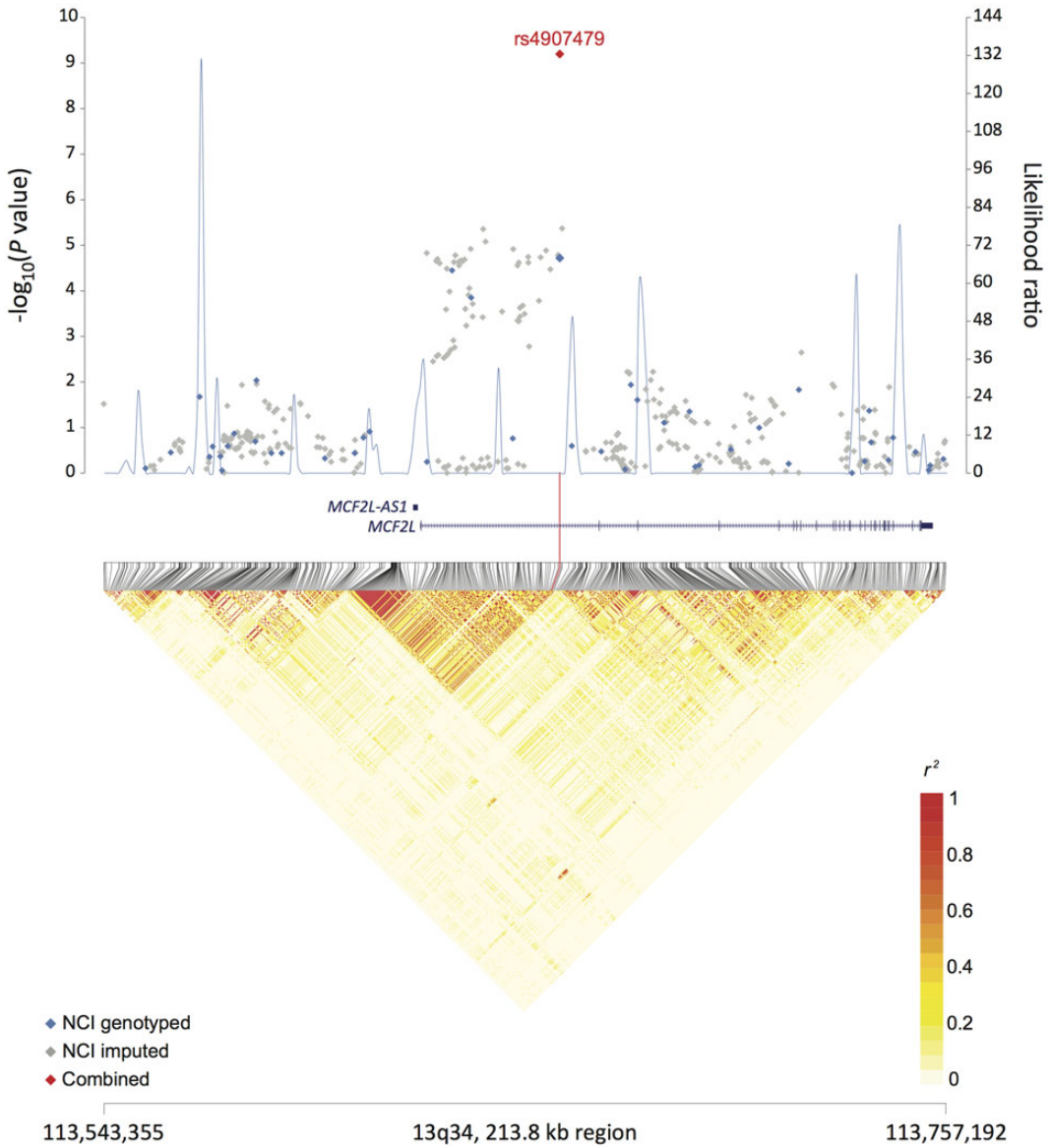

Figure 2. Association results and LD plot for the 13q34 region. The $-\log_{10}$(*P*-value) (left Y-axis) for NCI-GWAS1 and NCI-GWAS2 genotyped SNPs (blue) and imputed SNPs (gray) plotted on the genomic coordinates (X axis; NCBI genome build 37). The combined data for NCI-GWAS1, NCI-GWAS2, TXBCS-GWAS, IBCS-GWAS, NBCS-GWAS and TaqMan study data for the 13q34 locus marked by SNP rs4907479, are shown in red. Right Y-axis presents LR of putative recombination hotspots based on 5 sets of 100 randomly selected controls from NCI-GWAS1 and NCI-GWAS2 and shown as connected blue lines.

the 20p12.2 SNPs rs6108803 and rs62185668, the signal was stronger for muscle-invasive bladder cancer (MIBC, T2–T4 stages) compared with non-muscle-invasive bladder cancer (NMIBC, stages Ta and T1, case–case analysis, $P \leq 0.02$ for both markers), while there was no statistically significant difference for rs6104690. Case–case analysis adjusting for the 20p12.2 variants showed stronger association with MIBC for rs6108803 and rs62185668, which was not significantly affected by further adjustment for rs6104690 (Supplementary Material, Table S8). Association with tumor stage was consistent across studies ($I^2 = 0.0$, $P = 0.49$ for rs6108803, Supplementary Material, Fig. S2). SNPs in the 20p12.2 region were not significantly associated with tumor grade ($P > 0.56$) or high/low risk of progression tumor classification ($P > 0.52$, low risk defined as Ta stage with G1/G2 grade; high risk as T1–T4 or G3/G4 grade).

We analyzed the bladder cancer dataset containing data on 412 cases of The Cancer Genome Atlas (TCGA) (24), of which 391 had germline genetic data, to explore possible molecular phenotypes that might be related to the 13q34 and 20p12.2 signals. Since the SNPs of interest were not genotyped by TCGA, we used proxies for these variants based on European populations of the 1000 Genomes Project (Materials and Methods). We evaluated *MCF2L* mRNA expression in 375 bladder tumors in relation to rs2993291 (proxy for rs4907479, $r^2 = 0.96$) but observed no significant association (Supplementary Material, Fig. S3). The three SNPs at 20p12.2 (rs6104690, rs62185668 and rs6108803) are located within a 33 Kb region in a 1.2 Mb gene desert, at a distance of ~335 and 880 Kb from the closest genes, *JAG1* and *BTBD3*, respectively (Fig. 3). There are 6 proxy SNPs that are highly correlated ($r^2 > 0.8$) with rs6104690; 30 proxy SNPs for rs62185668; while there are no proxies for rs6108803 (only 4 SNPs are in $r^2 > 0.6$). For the TCGA analysis, we used rs6040291 as a proxy for rs6104690 ($r^2 = 1.0$), rs6074214 as a proxy for rs62185668 ($r^2 = 0.97$), while we could not perform analysis specifically for rs6108803

Table 1. Per-allele ORs and 95% CIs for SNPs at 20p12.2 locus with significant associations with bladder cancer risk in the combined NCI-GWAS1 and NCI-GWAS2 dataset of 5551 bladder cancer cases and 10 242 controls

SNP[a]	Alleles, risk-underlined	RAF cases/controls, %	OR (95% CI)	OR (95% CI), adjusted for rs6108803	OR (95% CI), adjusted for rs62185668	OR (95% CI), adjusted for rs6104690	OR (95% CI), adjusted for rs6104690, rs62185668, and rs6108803
rs6108803(G)	G/A	18.2/16.1	1.18 (1.10–1.26) $P = 1.82E-06$	–	1.12 (1.02–1.24) $P = 0.02$	1.14 (1.06–1.22)/ $P = 6.41E-04$	1.12 (1.01–1.24) $P = 0.03$
rs62185668 (I, G)	A/C	27.1/24.5	1.14 (1.07–1.20) $P = 1.39E-05$	1.05 (0.97–1.15) $P = 0.25$	–	1.09 (1.02–1.17) $P = 9.26E-03$	1.02 (0.93–1.12) $P = 0.70$
rs6104690 (G)	A/G	59.1/55.9	1.11 (1.06–1.17) $P = 3.97E-05$	1.07 (1.01–1.13) $P = 0.02$	1.07 (1.01–1.14) $P = 0.03$	–	1.07 (1.01–1.13) $P = 0.03$

[a]G indicates SNPs genotyped by GWAS and validated by TaqMan genotyping in a subset of samples; (I, G) indicates a SNP imputed with high confidence and then validated by TaqMan genotyping in a subset of samples. All ORs and 95% CIs were adjusted for age, gender, study groups, significant eigenvectors and smoking status; analysis is based on samples with genotype data for all three SNPs.
RAF, risk allele frequency.

(the best proxy for rs6108803 in TCGA was rs6074214 with $r^2 = 0.45$, this variant was already analyzed as a proxy for rs62185668).

Expression of the closest genes, *JAG1* and *BTBD3*, was not associated with genotypes of these SNPs in 381 bladder tumors (Supplementary Material, Fig. S3A and B). Further, using TCGA bladder cancer data, we observed no evidence of associations for the proxy 13q34 and 20p12.2 markers with overall survival for 363 patients and bladder cancer recurrence for 250 patients (data not shown). Analysis of TCGA data through the CBio Cancer Genomics Portal (25,26) showed that *MCF2L* was not commonly mutated in bladder cancer. Based on tumor data from 412 patients, somatic alterations in *MCF2L* were detected only in 16 (4%) of all tumors, and gene amplifications represented most of these alterations (14 of 16 events).

We performed *in silico* annotation using ENCODE (27) and HaploReg (28) databases, compiling information on histone modification marks in cell lines, transcription factor (TF) binding sites and DNase hypersensitivity sites (DHS). We noted from these data that the 13q34 region contains two regions with enrichment of multiple functional marks suggestive of regulatory functions close to SNPs in high LD with the GWAS SNP rs4907479. Importantly, these two functional regions also showed DHS in an urothelial cell line (Supplementary Material, Fig. S3). Future work will explore these regions for their possible role in regulating *MCF2L* or other genes. Similar analysis for the 20p12.2 region showed enrichment of multiple functional signals close to rs6104690 and several other areas but without strong specific patterns (Supplementary Material, Fig. S4). Since the SNP markers at 20p12.2 region associated with bladder cancer risk map to a gene desert area, more work is needed to explore possible functional effects of genetic variants associated with bladder cancer.

The 20p12.2 and 13q34 regions also harbor markers associated with other phenotypes identified by published GWAS. SNP rs11842874 in the *MCF2L* gene at 13q34 was previously identified in a GWAS for osteoarthritis (29), but this variant was not associated with bladder cancer risk ($P = 0.67$, $r^2 = 0.01$ and $D' = 0.24$ with rs4907479, Supplementary Material, Table S3). A 20p12.2 region SNP rs1327235, which was previously associated with blood pressure (23), showed a nominal association with bladder cancer risk in our set (OR = 1.09, $P = 3.8 \times 10^{-4}$, $r^2 = 0.25$–0.41 with our best markers) (Fig. 3); there was no association for the bone density-associated SNPs rs3790160 and rs2273061 (30,31) ($P > 0.10$, $r^2 < 0.01$ with our best markers).

We have previously shown evidence for significant additive interactions with smoking for many bladder cancer susceptibility loci (15,19,32). There was a suggestion of an additive but not multiplicative interaction for the rs6108803 20p12.2 SNP and smoking (P-additive interaction = 0.04, P-multiplicative interaction = 0.66). All other SNPs did not show any evidence of interaction ($P \geq 0.28$) (data not shown).

We also estimated the proportion of familial risk explained, based on all genetic variants identified to date that show association with bladder cancer at a genome-wide significant level. We estimate that all significantly associated SNPs identified so far explain ~12% of familial risk for bladder cancer (33,34).

Discussion

Herein, we report a new bladder cancer susceptibility locus at 13q34 marked by rs4907479 ($P = 6.4 \times 10^{-10}$) and refine the previously reported 20p12.2 region as a multi-signal locus, with two associations, one marked by rs6104690 and a second marked by rs6108803 and rs62185668. Interestingly, the signal captured by rs6108803 and rs62185668 showed significantly stronger association with risk of MIBC compared with NMIBC, making this the first bladder GWAS signal to show a significantly stronger association with MIBC.

Fine-mapping analysis of the 13q24 region showed that the signal detected for rs4907479 can be represented by at least 29 correlated variants ($r^2 \geq 0.8$), all located within the first two introns of the *MCF2L* gene. The association signal for rs4907479 was similar in groups stratified by tumor stage and grade. MCF2L is a guanine nucleotide exchange factor (GEF) for members of the RHO subfamily of the RAS superfamily (35). The N-terminally truncated protein isoform of MCF2L was initially identified as an osteosarcoma oncogene (36). In support of the possible role of this gene in bone disease, an *MCF2L* genetic variant rs11842874 has been strongly associated in GWAS for osteoarthritis in Europeans, but this variant was not associated with bladder cancer in our study. *MCF2L* was not found to be commonly mutated in bladder tumors studied in TCGA (25). The functional role of this GWAS signal is unclear since it was not associated with mRNA expression of *MCF2L*, overall survival or bladder cancer recurrence in TCGA bladder cancer dataset. ENCODE and HaploReg (27,28) *in silico* analysis suggest that this region may be important for regulation of genes in the region given enrichment of DHS; hence, future work is needed to explore the molecular phenotype of this genetic association and its role in bladder cancer risk.

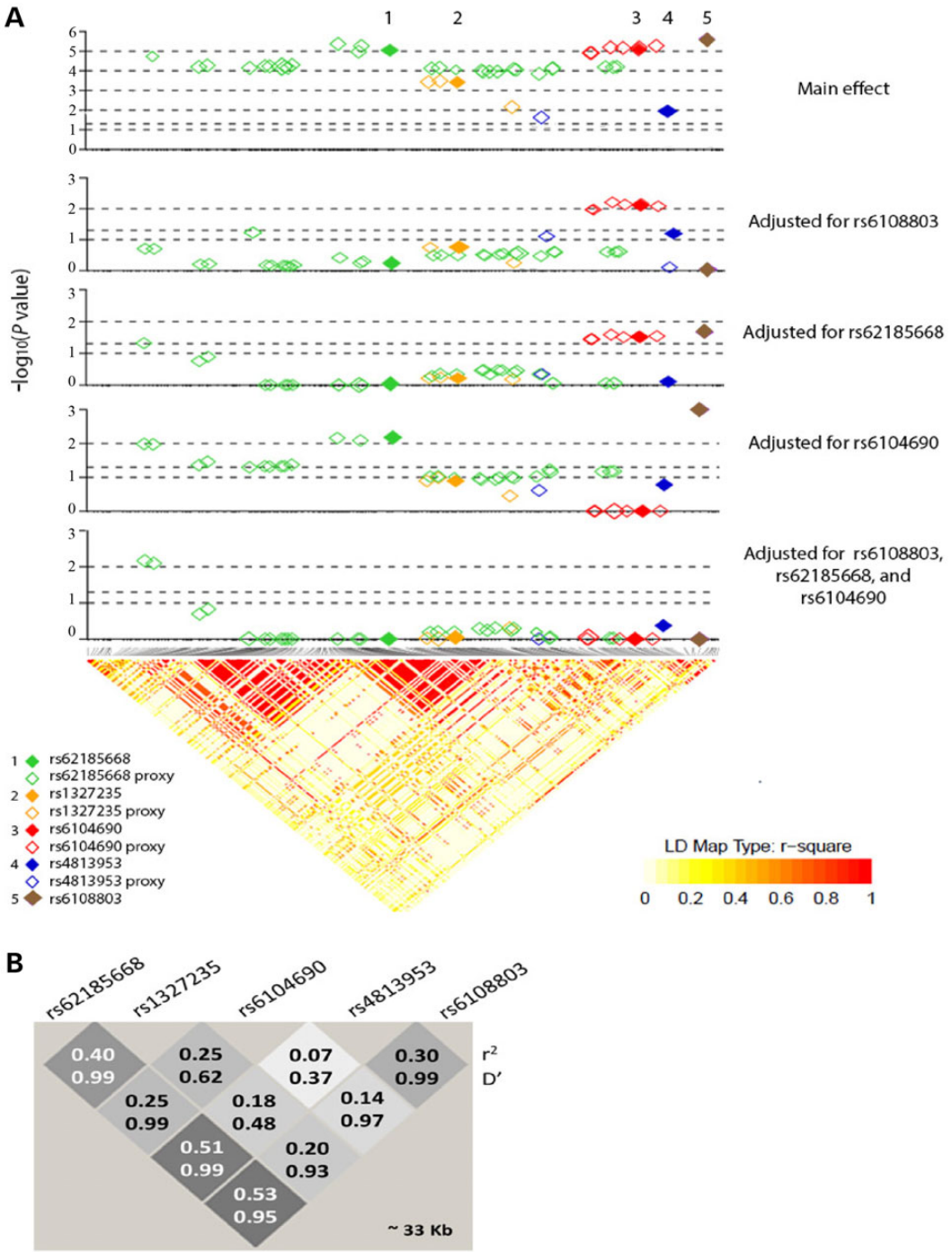

Figure 3. Fine-mapping association analysis of the 20p12.2 region. The results are shown for five SNPs of interest: validated GWAS candidate rs6104690 (19), novel finding rs6108803, previously reported rs62185668 and rs4813953 associated with bladder cancer (18), and rs1327235 associated with systolic blood pressure (23). The plots are based on the combined NCI-GWAS1 and NCI-GWAS2 dataset, which includes 5551 bladder cancer cases and 10 242 controls of European origin. (**A**) Association results for bladder cancer risk (Y-axis) are presented as –log10(*P*-value) for logistic regression models, assuming additive genetic effect and adjusting for age, gender, 11 study groups, significant eigenvectors, smoking (ever/never) and specified SNPs. SNPs of interest are marked as filled diamonds: rs6104690 (red), rs6108803 (brown), rs62185668 (green), rs4813953 (blue) and rs1327235 (orange); corresponding proxy SNPs ($r^2 \geq 0.8$) are presented as color-matched, un-filled diamonds. (**B**) Pairwise LD (r^2 and D′) of SNPs of interest across the 33 Kb in the 20p12.2 region.

In follow-up analyses, we defined the 20p12.2 region as a multi-signal locus, which includes at least two signals, the first signal represented by our initial SNP rs6104690 (19) and now a second signal marked by a novel SNP rs6108803 and a previously reported rs62185668 (18). A combination of two markers, rs6104690 and rs6108803 representing each of these 20p12.2 signals, most efficiently captured the bladder cancer association in this region in our combined dataset of NCI-GWAS1 and NCI-GWAS2. However, only the second signal, rs6108803/rs62185668 but not rs6104690, was associated with advanced tumor stage (MIBC).

A previously identified SNP rs7257330 upstream of *CCNE1* gene in 19q12 region showed an association with aggressive

Table 2. Per-allele ORs and 95% CIs for SNPs in the 20p12.2 region (rs6104690, rs6108803 and rs62185668) in the combined NCI-GWAS1 and NCI-GWAS2 dataset stratified by tumor stage as non-muscle-invasive (NMIBC, Ta-T1) and muscle-invasive (MIBC, T2-T4) bladder cancers

SNP		Controls	Cases	OR	Case–control			Case–Case
					95% CI		P	P
rs6104690	Stage at diagnosis							
	NMIBC (Ta-T1)	8403	3008	1.10	1.03	1.18	5.27E–03	Ref
	MIBC (T2-T4)		539	1.22	1.07	1.39	3.88E–03	0.20
rs6108803	Stage at diagnosis							
	NMIBC (Ta-T1)	7585	3010	1.10	1.01	1.21	3.33E–02	Ref
	MIBC (T2-T4)		539	1.36	1.15	1.60	2.62E–04	0.02
rs62185668	Stage at diagnosis							
	NMIBC (Ta-T1)	7966	2860	1.13	1.04	1.22	2.93E–03	Ref
	MIBC (T2-T4)		503	1.39	1.20	1.60	1.28E–05	0.01

Risk alleles are A (rs6104690), G (rs6108803) and A (rs62185668). Polytomous logistic regression was used to obtain OR and 95% CI for tumor subtypes adjusted for age, gender, study groups, significant eigenvectors and smoking status. Case–case *P* values were calculated with tumor type as an outcome and were used to test for differences in effect size between NMIBC and MIBC.

disease, which is based on a combination of tumor stage and grade information and corresponds to high risk of progression definition used here, but this was mostly driven by high grade (37). The 20p12.2 markers capture another important clinical difference, by tumor stage, an association with MIBC. However, our study and a previous study that analyzed rs62185668 (18) showed no difference in association for these markers by tumor grade or by low/high risk of cancer progression. MIBC represents up to 20% of all bladder cancer cases; this is a clinically severe cancer subtype that requires radical cystectomy and systemic chemotherapy. The risk of developing life-threatening metastatic disease remains high even after this treatment, resulting in relative 5-year survival rate of 15–63%, compared with 88–98% for NMIBC (http://www.cancer.org/cancer/bladdercancer/detailedguide/bladder-cancer-survival-rates). If validated in additional samples, associations at the 20p12.2 region might lead to a better understanding of genetic predisposition to MIBC. The associated 20p12.2 SNPs are located in a 1.2 Mb gene desert; expression of the closest genes *JAG1* and *BTBD3* was not associated with these variants in TCGA bladder cancer dataset, and alternative functional mechanisms will be explored.

We were unable to confirm associations for five loci previously identified as suggestive in a genome-wide interaction study of smoking and bladder cancer risk (21). These results indicate that additional large sample sets will be needed to explore loci with differential effects by smoking status.

In conclusion, we have identified a new susceptibility locus at 13q34 and refined our understanding of 20p12.2 as a multi-signal locus associated with bladder cancer risk in Europeans. Based on fine-mapping, we have identified optimal variants associated with bladder cancer risk that can be pursued in future studies. Comprehensive identification of the full range of bladder cancer susceptibility variants will provide a basis to further our understanding of the underlying biologic mechanisms and to explore the complex interplay of genes and environmental and occupational exposures (38) that contribute to bladder cancer risk.

Materials and Methods

Study participants

The samples and studies used are listed in Supplementary Material, Table S1. Cases and controls were non-Hispanic Caucasians of European ancestry. 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 study obtained informed consent from study participants and approval from the corresponding Institutional Review Boards (IRB). 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

Genotyping of cases and controls for NCI-GWAS1 and NCI-GWAS2 has previously been described (11,19) (Supplementary Material, Table S1). Genome-wide single-nucleotide variants (SNV) data for the first set of NBCS cases and controls were generated using the Illumina HumanHapCNV370-Duo (v1) or Illumina HumanHapCNV370-Quad (v3) BeadChip. A total of 1819 controls and 1601 bladder cancer patients passed pre-imputation QC (European ancestry, sample yield ≥96%, no gender mismatch, no duplicates). A second series of 1034 controls and 395 patients were successfully genotyped using the Illumina HumanOmniExpress-12 v1.1 BeadChip.

SNPs that had suggestive interaction with smoking (rs17621407, rs12216499, rs948798, rs846906 and rs1711973) (21) were genotyped with optimized TaqMan genotyping assays (ABI, Foster City, CA, USA) in eight additional studies from Europe and the United States (Supplementary Material, Table S1). Validation of imputed SNPs rs4813953 and rs62185668 and GWAS array genotyped SNPs rs61088036 and rs4907479 was done by TaqMan genotyping of 683 randomly selected DNA NCI-GWAS study samples representing cases and controls, with concordance rates of 99.4, 99.2, 100 and 99.7%, respectively. For the rs6104690 SNP at 20p12.2, comparison of the genotypes from the GWAS scan with TaqMan assays has been previously reported (19) and showed 100% concordance.

Imputation

IMPUTE version 2 (39) was used to infer additional genotypes in the 13q34 and 20p12.2 regions using genotype data for 5942 cases and 10 861 (whom we had individual-level genotype data on) from the combined dataset of bladder cancer NCI-GWAS1 and NCI-GWAS2 (11,19), and the 1000 Genomes Project Phase 3 integrated haplotypes (NCBI build 37 October 2014), which contains data for 2504 individuals from 21 populations (40). A 1 Mb window centered on SNPs rs4907479 at 13q34 or rs6104690 at 20p12.2 was used for imputation with a seed of 146 and 246

GWAS-genotyped SNPs, respectively. Imputation quality control included an assessment of overall concordance, which indicates how well the genotyped SNPs were imputed across samples (we used a threshold of 0.95), the average posterior probability and the IMPUTE2-info score of individual SNPs, which indicate how well individual SNPs were imputed across a dataset (we used a threshold of 0.9). This resulted in an overall genotype concordance score of 95% and a final SNP count of 1372 for the 13q34 region and an overall concordance of 97% with a final SNP count of 2344 for the 20p12.2 region. We calculated the Hardy–Weinberg equilibrium (HWE) and minor allele frequencies (MAF) in PLINK version 1.07 (10 August 2009), and variants with strong HWE deviations in controls ($P < 10^{-3}$) were reviewed and flagged. GTOOL software was used for all file conversions between pedigree and genotype file format.

Association testing on the combined NCI-GWAS1 and NCI-GWAS2 datasets that included both genotyped and imputed variants was performed using PLINK version 1.07 (10 August 2009) based on logistic regression models, considering an additive genetic effect and adjusting for age (in 5-year categories), gender, 11 study groups, significant eigenvectors (EV 1, 5 and 6) from the principal component analysis (PCA) as previously described (11,19) and smoking (ever/never). Additionally, models were adjusted for the specific SNPs to test for the presence of any additional independently or stronger associated SNPs. Calculation of LD metrics (D′ and r^2) and haplotype analysis for 20p12.2 locus SNPs (rs62185668, rs6104690 and rs6108803) were performed using PLINK.

Fixed-effects meta-analyses were used to determine associations for the SNPs in different sub-studies overall and in selected strata (tumor stage and grade) using STATA, Version 11.2. Heterogeneity in genetic effects across study groups was evaluated using the I^2 statistic. We evaluated SNP associations by stage, grade and high/low risk of progression tumors (low-risk tumors were defined as Ta stage with G1/G2 grade; high-risk tumors were T1–T4 stages or G3/G4 grade) using the combined set of NCI-GWAS1 and NCI-GWAS2 data. Polytomous logistic regression was used to obtain OR and 95% confidence interval (CI) for different tumor subtypes. Case–case *P*-values were calculated with tumor type as an outcome and were used to test for differences in effect size across subtypes. Polytomous logistic regression models for tumor grade and stage constraining the effect size to increase linearly across levels were also calculated and presented as case–case trend. Additive and multiplicative interactions were conducted using categorical variables (each SNP was coded as a dichotomous variable indicating the presence of any risk allele) to make the additive and multiplicative tests comparable as previously described (15,32).

Estimate of recombination hotspots

SequenceLDhot (41) that uses an approximate marginal likelihood method (42) 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 5 controls from each study). We used Phasev2.1 to infer the haplotypes as well as background recombination rates. The analysis was repeated with 5 non-overlapping sets of 100 pooled controls.

TCGA analysis

Expression (RNA-seq), genotypes (Affymetrix SNP6.0 arrays) and demographic and clinical data were obtained from TCGA (24) from 412 bladder cancer cases. Among the 412 cases, 391 had germline genotype data, 363 cases had data on overall survival and 250 cases had data on recurrence. Distributions of genotypes of SNPs rs2993291, rs6040291 and rs6074214, which were used as proxies for GWAS SNPs, were in HWE and comparable with the patterns in the 1000 Genomes populations. Total gene expression values for *JAG1*, *BTBD3* and *MCF2L* generated by TCGA as RSEM counts were log10-transformed (first adding 1 to all RSEM values). The log-transformed RSEM values were tested for association with SNPs based on an additive genetic model adjusted for age, sex and race using generalized linear models in SPSS v.21.

Survival analysis was performed using clinical and genotype bladder cancer TCGA data (24). Overall survival data defined as either months until patient death or last follow-up. Hazards ratios (HR) were estimated using Cox regression models with the number of risk alleles (0, 1 or 2) as the independent variable and overall survival as the outcome adjusting for age, gender, race and smoking status (ever/never).

Data access

Access to the NCI-GWAS1 and NCI-GWAS2 genotypes is available for investigators from certified scientific institutions after approval of a submitted Data Access Request through dbGAP identifier, phs000346.v2, at http://www.ncbi.nlm.nih.gov/gap.

URLs

CGEMS portal, http://cgems.cancer.gov/; Cancer Genomics Research Laboratory, CGR, http://cgf.nci.nih.gov/; GLU, http://code.google.com/p/glu-genetics/; EIGENSTRAT, http://genepath.med.harvard.edu/~reich/EIGENSTRAT.htm; STRUCTURE, http://pritch.bsd.uchicago.edu/structure.html; 1000 Genomes, http://www.1000genomes.org/; TCGA, http://cancergenome.nih.gov/; UCSC, https://genome.ucsc.edu/; STATA, http://www.stata.com; ; dbGAP, http://www.ncbi.nlm.nih.gov/gap; TCGA Research Network, http://cancergenome.nih.gov, ENCODE, https://genome.ucsc.edu/ENCODE/; HaploReg V.3, http://www.broadinstitute.org/mammals/haploreg/haploreg.php.

Supplementary Material

Supplementary material is available at *HMG* online.

Acknowledgements

The authors acknowledge the following: Marie Audouin, Cécile Gaffory, Valérie Ondet; Leslie Carroll (Information Management Services, Silver Spring, MD, USA); Gemma Castaño-Vinyals (Institut Municipal d'Investigació Mèdica, Barcelona, Spain); Francisco Fernández (Institut Municipal d'Investigació Mèdica, Barcelona, Spain);

Anna McIntosh (Westat, Inc., Rockville, MD, USA); Robert Saal (Westat, Rockville, MD, USA); Francisco Real [Spanish National Cancer Research Centre (CNIO), Madrid, Spain]; Maria Sala (Institut Municipal d'Investigació Mèdica, Barcelona, Spain); Kirk Snyder (Information Management Services, Inc., Silver Spring, MD); Anne Taylor (Information Management Services, Inc., Silver Spring, MD); Montserrat Torà (Institut Municipal d'Investigació Mèdica, Barcelona, Spain); Jane Wang (Information Management Services, Silver Spring, MD, USA).

WHI Program Office: (National Heart, Lung, and Blood Institute, Bethesda, Maryland) Jacques Rossouw, Shari Ludlam, Dale Burwen, Joan McGowan, Leslie Ford, and Nancy Geller.

Clinical Coordinating Center: Clinical Coordinating Center: (Fred Hutchinson Cancer Research Center, Seattle, WA) Garnet Anderson, Ross Prentice, Andrea LaCroix, and Charles Kooperberg.

Investigators and Academic Centers: (Brigham and Women's Hospital, Harvard Medical School, Boston, MA) JoAnn E. Manson; (MedStar Health Research Institute/Howard University, Washington, DC) Barbara V. Howard; (Stanford Prevention Research Center, Stanford, CA) Marcia L. Stefanick; (The Ohio State University, Columbus, OH) Rebecca Jackson; (University of Arizona, Tucson/Phoenix, AZ) Cynthia A. Thomson; (University at Buffalo, Buffalo, NY) Jean Wactawski-Wende; (University of Florida, Gainesville/Jacksonville, FL) Marian Limacher; (University of Iowa, Iowa City/Davenport, IA) Robert Wallace; (University of Pittsburgh, Pittsburgh, PA) Lewis Kuller; (Wake Forest University School of Medicine, Winston-Salem, NC) Sally Shumaker.

Women's Health Initiative Memory Study: (Wake Forest University School of Medicine, Winston-Salem, NC) Sally Shumaker.

For a list of all the investigators who have contributed to WHI science, please visit: https://cleo.whi.org/researchers/SitePages/Write%20a%20Paper.aspx. The Memorial Sloan Kettering investigators gratefully acknowledge the MSK Cancer Center Support Grant/Core Grant (P30 CA008748).

Conflict of Interest statement: None declared.

Funding

See Supplementary Material for funding details.

References

1. Ferlay, J., Shin, H.R., Bray, F., Forman, D., Mathers, C. and Parkin, D.M. (2010) Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. *Int. J. Cancer*, **127**, 2893–2917.
2. Silverman, D.T., Devessa, S.S., Moore, L. and Rothman, N. (2006) In: Fraumeni, J.F. Jr. and Schottenfeld, D. (eds), *Cancer Epidemiology and Prevention Third Edition*. Oxford University Press, New York, NY, pp. 1101–1127.
3. Kantor, A.F., Hartge, P., Hoover, R.N. and Fraumeni, J.F. Jr. (1985) Familial and environmental interactions in bladder cancer risk. *Int. J. Cancer*, **35**, 703–706.
4. Murta-Nascimento, C., Silverman, D.T., Kogevinas, M., Garcia-Closas, M., Rothman, N., Tardon, A., Garcia-Closas, R., Serra, C., Carrato, A., Villanueva, C. *et al.* (2007) Risk of bladder cancer associated with family history of cancer: do low-penetrance polymorphisms account for the increase in risk? *Cancer Epidemiol. Biomarkers Prev.*, **16**, 1595–1600.
5. Plna, K. and Hemminki, K. (2001) Familial bladder cancer in the National Swedish Family Cancer Database. *J. Urol.*, **166**, 2129–2133.
6. Dong, C. and Hemminki, K. (2001) Modification of cancer risks in offspring by sibling and parental cancers from 2,112,616 nuclear families. *Int. J. Cancer*, **92**, 144–150.
7. Garcia-Closas, M., Malats, N., Silverman, D., Dosemeci, M., Kogevinas, M., Hein, D.W., Tardon, A., Serra, C., Carrato, A., Garcia-Closas, R. *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. Garcia-Closas, M., Ye, Y., Rothman, N., Figueroa, J.D., Malats, N., Dinney, C.P., Chatterjee, N., Prokunina-Olsson, L., Wang, Z., Lin, J. *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.
9. Kiemeney, L.A., Thorlacius, S., Sulem, P., Geller, F., Aben, K.K., Stacey, S.N., Gudmundsson, J., Jakobsdottir, M., Bergthorsson, J.T., Sigurdsson, A. *et al.* (2008) Sequence variant on 8q24 confers susceptibility to urinary bladder cancer. *Nat. Genet.*, **40**, 1307–1312.
10. Moore, L.E., Baris, D.R., Figueroa, J.D., Garcia-Closas, M., Karagas, M.R., Schwenn, M.R., Johnson, A.T., Lubin, J.H., Hein, D. W., Dagnall, C.L. *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.
11. Rothman, N., Garcia-Closas, M., Chatterjee, N., Malats, N., Wu, X., Figueroa, J.D., Real, F.X., Van Den Berg, D., Matullo, G., Baris, D. *et al.* (2010) A multi-stage genome-wide association study of bladder cancer identifies multiple susceptibility loci. *Nat. Genet.*, **42**, 978–984.
12. Wu, X., Ye, Y., Kiemeney, L.A., Sulem, P., Rafnar, T., Matullo, G., Seminara, D., Yoshida, T., Saeki, N., Andrew, A.S. *et al.* (2009) Genetic variation in the prostate stem cell antigen gene PSCA confers susceptibility to urinary bladder cancer. *Nat. Genet.*, **41**, 991–995.
13. Kiemeney, L.A., Sulem, P., Besenbacher, S., Vermeulen, S.H., Sigurdsson, A., Thorleifsson, G., Gudbjartsson, D.F., Stacey, S.N., Gudmundsson, J., Zanon, C. *et al.* (2010) A sequence variant at 4p16.3 confers susceptibility to urinary bladder cancer. *Nat. Genet.*, **42**, 415–419.
14. Rafnar, T., Vermeulen, S.H., Sulem, P., Thorleifsson, G., Aben, K.K., Witjes, J.A., Grotenhuis, A.J., Verhaegh, G.W., Hulsbergen-van de Kaa, C.A., Besenbacher, S. *et al.* (2011) European genome-wide association study identifies SLC14A1 as a new urinary bladder cancer susceptibility gene. *Hum. Mol. Genet.*, **20**, 4268–4281.
15. Tang, W., Fu, Y.P., Figueroa, J.D., Malats, N., Garcia-Closas, M., Chatterjee, N., Kogevinas, M., Baris, D., Thun, M., Hall, J.L. *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.
16. Rafnar, T., Sulem, P., Stacey, S.N., Geller, F., Gudmundsson, J., Sigurdsson, A., Jakobsdottir, M., Helgadottir, H., Thorlacius, S., Aben, K.K. *et al.* (2009) Sequence variants at the TERT-CLPTM1L locus associate with many cancer types. *Nat. Genet.*, **41**, 221–227.
17. Matsuda, K., Takahashi, A., Middlebrooks, C.D., Obara, W., Nasu, Y., Inoue, K., Tamura, K., Yamasaki, I., Naya, Y., Tanikawa, C. *et al.* (2015) Genome-wide association study identified SNP on 15q24 associated with bladder cancer risk in Japanese population. *Hum. Mol. Genet.*, **24**, 1177–1184.
18. Rafnar, T., Sulem, P., Thorleifsson, G., Vermeulen, S.H., Helgason, H., Saemundsdottir, J., Gudjonsson, S.A., Sigurdsson, A., Stacey, S.N., Gudmundsson, J. *et al.* (2014) Genome-wide association study yields variants at 20p12.2 that associate with urinary bladder cancer. *Hum. Mol. Genet.*, **23**, 5545–5557.
19. Figueroa, J.D., Ye, Y., Siddiq, A., Garcia-Closas, M., Chatterjee, N., Prokunina-Olsson, L., Cortessis, V.K., Kooperberg, C., Cussenot, O., Benhamou, S. *et al.* (2014) Genome-wide association study identifies multiple loci associated with bladder cancer risk. *Hum. Mol. Genet.*, **23**, 1387–1398.
20. Park, J.H., Wacholder, S., Gail, M.H., Peters, U., Jacobs, K.B., Chanock, S.J. and Chatterjee, N. (2010) Estimation of effect size distribution from genome-wide association studies and implications for future discoveries. *Nat. Genet.*, **42**, 570–575.
21. Figueroa, J.D., Han, S.S., Garcia-Closas, M., Baris, D., Jacobs, E. J., Kogevinas, M., Schwenn, M., Malats, N., Johnson, A., Purdue, M.P. *et al.* (2014) Genome-wide interaction study of

smoking and bladder cancer risk. *Carcinogenesis*, **35**, 1737–1744.
22. Gudbjartsson, D.F., Helgason, H., Gudjonsson, S.A., Zink, F., Oddson, A., Gylfason, A., Besenbacher, S., Magnusson, G., Halldorsson, B.V., Hjartarson, E. *et al.* (2015) Large-scale whole-genome sequencing of the Icelandic population. *Nat. Genet.*, **47**, 435–444.
23. Ehret, G.B., Munroe, P.B., Rice, K.M., Bochud, M., Johnson, A. D., Chasman, D.I., Smith, A.V., Tobin, M.D., Verwoert, G.C., Hwang, S.J. *et al.* (2011) Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. *Nature*, **478**, 103–109.
24. The Cancer Genome Atlas Research Network. (2014) Comprehensive molecular characterization of urothelial bladder carcinoma. *Nature*, **507**, 315–322.
25. Gao, J., Aksoy, B.A., Dogrusoz, U., Dresdner, G., Gross, B., Sumer, S.O., Sun, Y., Jacobsen, A., Sinha, R., Larsson, E. *et al.* (2013) Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci. Signal*, **6**, pl1.
26. Cerami, E., Gao, J., Dogrusoz, U., Gross, B.E., Sumer, S.O., Aksoy, B.A., Jacobsen, A., Byrne, C.J., Heuer, M.L., Larsson, E. *et al.* (2012) The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov.*, **2**, 401–404.
27. Ram, O., Goren, A., Amit, I., Shoresh, N., Yosef, N., Ernst, J., Kellis, M., Gymrek, M., Issner, R., Coyne, M. *et al.* (2011) Combinatorial patterning of chromatin regulators uncovered by genome-wide location analysis in human cells. *Cell*, **147**, 1628–1639.
28. Ward, L.D. and Kellis, M. (2012) HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res.*, **40**, D930–D934.
29. Day-Williams, A.G., Southam, L., Panoutsopoulou, K., Rayner, N.W., Esko, T., Estrada, K., Helgadottir, H.T., Hofman, A., Ingvarsson, T., Jonsson, H. *et al.* (2011) A variant in MCF2L is associated with osteoarthritis. *Am. J. Hum. Genet.*, **89**, 446–450.
30. Estrada, K., Styrkarsdottir, U., Evangelou, E., Hsu, Y.H., Duncan, E.L., Ntzani, E.E., Oei, L., Albagha, O.M., Amin, N., Kemp, J.P. *et al.* (2012) Genome-wide meta-analysis identifies 56 bone mineral density loci and reveals 14 loci associated with risk of fracture. *Nat. Genet.*, **44**, 491–501.
31. Kung, A.W., Xiao, S.M., Cherny, S., Li, G.H., Gao, Y., Tso, G., Lau, K.S., Luk, K.D., Liu, J.M., Cui, B. *et al.* (2010) Association of JAG1 with bone mineral density and osteoporotic fractures: a genome-wide association study and follow-up replication studies. *Am. J. Hum. Genet.*, **86**, 229–239.
32. Garcia-Closas, M., Rothman, N., Figueroa, J.D., Prokunina-Olsson, L., Han, S.S., Baris, D., Jacobs, E.J., Malats, N., De Vivo, I., Albanes, D. *et al.* (2013) Common genetic polymorphisms modify the effect of smoking on absolute risk of bladder cancer. *Cancer Res.*, **73**, 2211–2220.
33. Lee, S.H., Wray, N.R., Goddard, M.E. and Visscher, P.M. (2011) Estimating missing heritability for disease from genome-wide association studies. *Am. J. Hum. Genet.*, **88**, 294–305.
34. Yang, J., Benyamin, B., McEvoy, B.P., Gordon, S., Henders, A.K., Nyholt, D.R., Madden, P.A., Heath, A.C., Martin, N.G., Montgomery, G.W. *et al.* (2010) Common SNPs explain a large proportion of the heritability for human height. *Nat. Genet.*, **42**, 565–569.
35. Liu, Z., Adams, H.C. III and Whitehead, I.P. (2009) The rho-specific guanine nucleotide exchange factor Dbs regulates breast cancer cell migration. *J. Biol. Chem.*, **284**, 15771–15780.
36. Horii, Y., Beeler, J.F., Sakaguchi, K., Tachibana, M. and Miki, T. (1994) A novel oncogene, ost, encodes a guanine nucleotide exchange factor that potentially links Rho and Rac signaling pathways. *EMBO J.*, **13**, 4776–4786.
37. Fu, Y.P., Kohaar, I., Moore, L.E., Lenz, P., Figueroa, J.D., Tang, W., Porter-Gill, P., Chatterjee, N., Scott-Johnson, A., Garcia-Closas, M. *et al.* (2014) The 19q12 bladder cancer GWAS signal: association with cyclin E function and aggressive disease. *Cancer Res.*, **74**, 5808–5818.
38. Figueroa, J.D., Koutros, S., Colt, J.S., Kogevinas, M., Garcia-Closas, M., Real, F.X., Friesen, M.C., Baris, D., Stewart, P., Schwenn, M. *et al.* (2015) Modification of Occupational Exposures on Bladder Cancer Risk by Common Genetic Polymorphisms. *J. Natl. Cancer Inst.*, **107**, doi: 10.1093/jnci/djv223.
39. Howie, B.N., Donnelly, P. and Marchini, J. (2009) A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet.*, **5**, e1000529.
40. Abecasis, G.R., Auton, A., Brooks, L.D., DePristo, M.A., Durbin, R.M., Handsaker, R.E., Kang, H.M., Marth, G.T. and McVean, G. A. (2012) An integrated map of genetic variation from 1,092 human genomes. *Nature*, **491**, 56–65.
41. Fearnhead, P. (2006) SequenceLDhot: detecting recombination hotspots. *Bioinformatics*, **22**, 3061–3066.
42. Fearnhead, P., Harding, R.M., Schneider, J.A., Myers, S. and Donnelly, P. (2004) Application of coalescent methods to reveal fine-scale rate variation and recombination hotspots. *Genetics*, **167**, 2067–2081.