cancer risk ($P = 0.0518$) in the SBCGS-1 stage 1 samples. The function of the *RCCD1* gene is unknown.

SNP rs4951011 is located in intron 2 of the *ZC3H11A* gene (encoding zinc finger CCCH domain–containing protein 11A) (NM_014827) at 1q32.1 (**Fig. 2a**) and the 5′ UTR of the *ZBED6* gene (encoding zinc finger BED domain–containing protein 6) (NM_001174108) (not shown). The ZBED6 protein has recently been recognized as a novel transcription factor in placental mammals[25]. The function of the *ZC3H11A* gene is not clear. ChromHMM annotation using human mammary epithelial cell (HMEC) data from ENCODE suggests that rs4951011 might be located in a strong enhancer region marked by peaks of several active histone methylation modifications (monomethylation of histone H3 at lysine 4 (H3K4me1), trimethylation of histone H3 at lysine 4 (H3K4me3), acetylation of histone H3 at lysine 9 (H3K9ac) and acetylation of histone H3 at lysine 27 (H3K27ac)). A search of RegulomeDB and HaploReg indicated that rs4951011 might be located in a predicted HNF1 motif and map to a DNase I hypersensitivity site in the MCF-7 breast cancer cell line (**Supplementary Table 7**). Expression levels of the *ZC3H11A* gene were significantly higher in breast tumor tissue than in adjacent normal tissue ($P = 0.0049$) in TCGA data (**Supplementary Table 6**). Analyses using TCGA data showed no evidence that this SNP or other SNPs correlated with it are *cis*-eQTLs for any genes in this locus. Recently, SNP rs4245739 in the *MDM4* gene, ~752 kb downstream of rs4951011 ($r^2 = 0$ in both ASN and CEU data), was associated with ER-negative breast cancer risk[5]. In our study, rs4245739 had a low MAF (0.03–0.05) and was not associated with breast cancer risk ($P = 0.1861$ in stage 1).

In summary, our large GWAS conducted among East Asian women identified three new breast cancer susceptibility loci at 1q32.1, 5q14.3 and 15q26.1 and suggested a possible association with a fourth locus at 18q11.2. The associations of these loci with breast cancer risk might be mediated through the regulation of cell growth control, tumor cell migration and invasion, or metastasis. Further studies of possible mechanisms through which these loci and genes are involved in breast tumorigenesis are warranted. Results from this study provide additional insights into the genetics and biology of breast cancer.

URLs. 1000 Genomes Project, http://www.1000genomes.org/; The Cancer Genome Atlas (TCGA), http://cancergenome.nih.gov/; DRIVE GAME-ON Consortium, http://epi.grants.cancer.gov/gameon/; EIGENSTRAT, http://genepath.med.harvard.edu/~reich/EIGENSTRAT.htm; ENCODE Project, http://www.genome.gov/10005107; HaploReg, http://www.broadinstitute.org/mammals/haploreg/haploreg.php; HapMap Project, http://hapmap.ncbi.nlm.nih.gov/; HumanExome BeadChip, http://genome.sph.umich.edu/wiki/Exome_Chip_Design; LocusZoom, v1.1, http://csg.sph.umich.edu/locuszoom/; MACH1.0, http://www.sph.umich.edu/csg/abecasis/MaCH/; Mach2dat, http://genome.sph.umich.edu/wiki/Mach2dat:_Association_with_MACH_output; METAL, http://www.sph.umich.edu/csg/abecasis/metal; PLINK version 1.07, http://pngu.mgh.harvard.edu/~purcell/plink/; R version 3.0.0, http://www.r-project.org/; RegulomeDB, http://regulome.stanford.edu/; SAS version 9.3, http://www.sas.com/; UCSC Genome Browser, http://genome.ucsc.edu/.

METHODS

Methods and any associated references are available in the online version of the paper.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

The content of this paper is solely the responsibility of the authors and does not necessarily represent the official views of the funding agencies. The authors wish to thank the study participants and research staff for their contributions and commitment to this project. We thank R. Courtney, J. Wu, J. He, H. Cai, X. Guo, B. Rammer and K. Kreth for their help with sample preparation and genotyping, statistical and bioinformatics analyses for the project, and editing and preparation of the manuscript at Vanderbilt.

This research was supported in part by US National Institutes of Health grants R01CA124558, R01CA148667 and R37CA070867 (to W.Z.); R01CA118229, R01CA092585 and R01CA064277 (to X.-O.S.); R01CA122756 (to Q.C.); and R01CA137013 (to J. Long), US Department of Defense Idea Awards BC011118 (to X.-O.S.) and BC050791 (to Q.C.), and Ingram Professorship and Research Reward funds (to W.Z.). Sample preparation and genotyping assays at Vanderbilt were conducted at the Survey and Biospecimen Shared Resources and the Vanderbilt Microarray Shared Resource, which are supported in part by the Vanderbilt-Ingram Cancer Center (P30CA068485). SeBCS was supported by the BRL (Basic Research Laboratory) program through the National Research Foundation of Korea funded by the Ministry of Education, Science and Technology (2011-0001564). KOHBRA/KOGES was supported by a grant from the National R&D Program for Cancer Control, Ministry for Health, Welfare and Family Affairs, Republic of Korea (1020350).

Studies participating in the ABCC include (principal investigator, grant support) the Shanghai Breast Cancer Study (W.Z. and X.-O.S., R01CA064277), the Shanghai Women's Health Study (W.Z., R37CA070867), the Shanghai Breast Cancer Survival Study (X.-O.S., R01CA118229), the Shanghai Endometrial Cancer Study (X.-O.S., R01CA092585, controls only), the Seoul Breast Cancer Study (D.K., BRL program through the National Research Foundation of Korea funded by the Ministry of Education, Science and Technology (2012-0000347)), the BioBank Japan Project (S.-K.L., the Ministry of Education, Culture, Sports, Science and Technology of Japan), the Hwasun Cancer Epidemiology Study-Breast (S.-S.K., the National R&D Program for Cancer Control, Ministry of Health and Welfare, Republic of Korea, 1020010), the Taiwan Study (C.-Y.S., Institute of Biomedical Sciences, Academia Sinica, Taiwan and National Science Council, DOH97-01), the Hong Kong Study (U.S.K., Research Grant Council, Hong Kong, China, HKU 7520/05M and 76730M), the Korean-NCC Study (M.K.K., grant-in-aid from the Korea National Cancer Center, NCC-0910310), the Nagano Breast Cancer Study (S.T., Grants-in-Aid for the Third-Term Comprehensive Ten-Year Strategy for Cancer Control from the Ministry of Health, Labor and Welfare of Japan, for Scientific Research on Priority Areas, 17015049 and for Scientific Research on Innovative Areas, 221S0001, from the Ministry of Education, Culture, Sports, Science and Technology of Japan), the Hospital-Based Epidemiologic Research Program at the Aichi Cancer Center (K. Tajima, Grant-in-Aid for Scientific Research (C) (24590776) and Grant-in-Aid for Scientific Research on Priority Areas of Cancer (17015018) and Scientific Research on Innovative Areas (221S0001) from the Ministry of Education, Culture, Sports, Science and Technology of Japan), the Malaysian Breast Cancer Genetic Study (S.-H.T., the Malaysian Ministry of Science, Technology and Innovation, Malaysian Ministry of Higher Education (UM.C/HIR/MOHE/06) and the Cancer Research Initiatives Foundation; controls were recruited by the Malaysian More than a Mammo Programme, which was supported by a grant from Yayasan Sime Darby LPGA) and the Singapore Breast Cancer Study (M.H., National Medical Research Council Start-Up Grant and Centre Grant (NMRC/CG/NCIS/2010); additional controls were recruited by the Singapore Consortium of Cohort Studies–Multi-Ethnic Cohort (SCCS-MEC), which was funded by the Biomedical Research Council, grant 05/1/21/19/425). The DRIVE GAME-ON Consortium is funded by US NIH grant U19CA148065 (D. Hunter). We thank I. Soong, A. Chan and T.Y. Leung for their assistance in facilitating recruitment of the breast cancer cases included in the Hong Kong Study.

AUTHOR CONTRIBUTIONS

W.Z. conceived and directed the Asia Breast Cancer Consortium (ABCC) and the Shanghai Breast Cancer Genetics Study. Q.C. and W.Z. wrote the manuscript with significant contributions from B.Z., J.S., J. Long, R.J.D., B.L. and X.-O.S. Q.C., B.Z., J. Long and W.W. coordinated the project. Q.C. directed the laboratory operations. J.S. performed the genotyping experiments. B.Z., J. Long and W.W. managed the study data. B.Z., J. Long and W.W. performed the statistical analyses with significant contributions from B.L. and C.L. Q.C., J. Long, R.J.D., Y. Zhang and B.L. performed the bioinformatics analyses. H.S., S.-K.L., S.-S.K., W.L., J.-Y.C., D.-Y.N., C.-Y.S., K. Matsuo, S.-H.T., M.K.K., U.S.K., M.I., M.H., A.T., K.A., K. Matsuda, M.-H.S., M.H.P., Y. Zheng, Y.-B.X., B.-T.J., S.K.P., P.-E.W., C.-N.H., H. Ito, Y.K., P.K., S.M., S.H.A., H.S.K., K.Y.K.C., E.P.S.M., H. Iwata, S.T., H.M.,

J. Liao, Y.N., M.K., Y.-T.G., X.-O.S., D.K. and W.Z. contributed to the collection of the data and biological samples for the original studies. All authors have reviewed and approved the final manuscript.

COMPETING FINANCIAL INTERESTS
The authors declare no competing financial interests.

1. Mavaddat, N., Antoniou, A.C., Easton, D.F. & Garcia-Closas, M. Genetic susceptibility to breast cancer. *Mol. Oncol.* **4**, 174–191 (2010).
2. Zhang, B., Beeghly-Fadiel, A., Long, J. & Zheng, W. Genetic variants associated with breast-cancer risk: comprehensive research synopsis, meta-analysis, and epidemiological evidence. *Lancet Oncol.* **12**, 477–488 (2011).
3. Zheng, W. *et al.* Common genetic determinants of breast-cancer risk in East Asian women: a collaborative study of 23 637 breast cancer cases and 25 579 controls. *Hum. Mol. Genet.* **22**, 2539–2550 (2013).
4. Michailidou, K. *et al.* Large-scale genotyping identifies 41 new loci associated with breast cancer risk. *Nat. Genet.* **45**, 353–361 (2013).
5. Garcia-Closas, M. *et al.* Genome-wide association studies identify four ER negative–specific breast cancer risk loci. *Nat. Genet.* **45**, 392–398 (2013).
6. Zheng, W. *et al.* Genome-wide association study identifies a new breast cancer susceptibility locus at 6q25.1. *Nat. Genet.* **41**, 324–328 (2009).
7. Long, J. *et al.* Identification of a functional genetic variant at 16q12.1 for breast cancer risk: results from the Asia Breast Cancer Consortium. *PLoS Genet.* **6**, e1001002 (2010).
8. Cai, Q. *et al.* Genome-wide association study identifies breast cancer risk variant at 10q21.2: results from the Asia Breast Cancer Consortium. *Hum. Mol. Genet.* **20**, 4991–4999 (2011).
9. Long, J. *et al.* Genome-wide association study in east Asians identifies novel susceptibility loci for breast cancer. *PLoS Genet.* **8**, e1002532 (2012).
10. Haiman, C.A. *et al.* A common variant at the *TERT-CLPTM1L* locus is associated with estrogen receptor–negative breast cancer. *Nat. Genet.* **43**, 1210–1214 (2011).
11. Kim, H.C. *et al.* A genome-wide association study identifies a breast cancer risk variant in *ERBB4* at 2q34: results from the Seoul Breast Cancer Study. *Breast Cancer Res.* **14**, R56 (2012).
12. Li, Y., Willer, C.J., Ding, J., Scheet, P. & Abecasis, G.R. MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genet. Epidemiol.* **34**, 816–834 (2010).
13. Willer, C.J., Li, Y. & Abecasis, G.R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).
14. ENCODE Project Consoritum. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
15. Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. *Nature* **490**, 61–70 (2012).
16. Zeller, T. *et al.* Genetics and beyond—the transcriptome of human monocytes and disease susceptibility. *PLoS ONE* **5**, e10693 (2010).
17. Boyle, A.P. *et al.* Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res.* **22**, 1790–1797 (2012).
18. Ward, L.D. & Kellis, M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res.* **40**, D930–D934 (2012).
19. Adélaïde, J. *et al.* Integrated profiling of basal and luminal breast cancers. *Cancer Res.* **67**, 11565–11575 (2007).
20. Draheim, K.M. *et al.* ARRDC3 suppresses breast cancer progression by negatively regulating integrin β4. *Oncogene* **29**, 5032–5047 (2010).
21. Rafiq, S. *et al.* Identification of inherited genetic variations influencing prognosis in early-onset breast cancer. *Cancer Res.* **73**, 1883–1891 (2013).
22. Zhu, C. & Jiang, W. Cell cycle–dependent translocation of PRC1 on the spindle by Kif4 is essential for midzone formation and cytokinesis. *Proc. Natl. Acad. Sci. USA* **102**, 343–348 (2005).
23. Li, C., Lin, M. & Liu, J. Identification of *PRC1* as the p53 target gene uncovers a novel function of p53 in the regulation of cytokinesis. *Oncogene* **23**, 9336–9347 (2004).
24. Mustacchi, G. *et al.* Identification and validation of a new set of five genes for prediction of risk in early breast cancer. *Int. J. Mol. Sci.* **14**, 9686–9702 (2013).
25. Markljung, E. *et al.* ZBED6, a novel transcription factor derived from a domesticated DNA transposon regulates IGF2 expression and muscle growth. *PLoS Biol.* **7**, e1000256 (2009).

[1]Division of Epidemiology, Department of Medicine, Vanderbilt Epidemiology Center, Vanderbilt-Ingram Cancer Center, Vanderbilt University School of Medicine, Nashville, Tennessee, USA. [2]Department of Biomedical Sciences, Seoul National University College of Medicine, Seoul, South Korea. [3]Genetic Epidemiology Branch, Division of Epidemiology and Genetics, National Cancer Institute, Rockville, Maryland, USA. [4]Laboratory for Statistical Analysis, Center for Integrative Medical Sciences, RIKEN, Yokohama, Japan. [5]Department of Preventive Medicine, Chonnam National University Medical School, Gwangju, South Korea. [6]Jeonnam Regional Cancer Center, Chonnam National University Hwasun Hospital, Hwasun, South Korea. [7]Shanghai Municipal Center for Disease Control and Prevention, Shanghai, China. [8]Cancer Research Institute, Seoul National University College of Medicine, Seoul, South Korea. [9]Department of Surgery, Seoul National University College of Medicine, Seoul, South Korea. [10]Institute of Biomedical Sciences, Academia Sinica, Taipei, Taiwan. [11]Taiwan Biobank, Academia Sinica, Taipei, Taiwan. [12]Department of Public Health, College of Public Health, China Medical University, Taichung, Taiwan. [13]Department of Preventive Medicine, Kyushu University Faculty of Medical Sciences, Fukuoka, Japan. [14]Cancer Research Initiatives Foundation, Sime Darby Medical Centre, Subang Jaya, Malaysia. [15]University Malaya Cancer Research Institute, Faculty of Medicine, University of Malaya, Kuala Lumpur, Malaysia. [16]Division of Cancer Epidemiology and Management, National Cancer Center, Gyeonggi-do, South Korea. [17]Department of Pathology, Li Ka Shing Faculty of Medicine, University of Hong Kong, Hong Kong, China. [18]Epidemiology Division, Research Center for Cancer Prevention and Screening, National Cancer Center, Tokyo, Japan. [19]Saw Swee Hock School of Public Health, National University of Singapore, Singapore. [20]Department of Surgery, National University of Singapore, Singapore. [21]Department of Medical Epidemiology and Biostatistics, Karolinska Institute, Stockholm, Sweden. [22]Laboratory for Genotyping Development, Center for Integrative Medical Sciences, RIKEN, Yokohama, Japan. [23]Laboratory of Molecular Medicine, Human Genome Center, Institute of Medical Science, The University of Tokyo, Tokyo, Japan. [24]Department of Surgery, Chonnam National University Medical School, Gwangju, South Korea. [25]Department of Epidemiology, Shanghai Cancer Institute, Shanghai, China. [26]Division of Cancer Epidemiology and Genetics, National Cancer Institute, US National Institutes of Health, US Department of Health and Human Services, Bethesda, Maryland, USA. [27]Department of Preventive Medicine, Seoul National University College of Medicine, Seoul, South Korea. [28]Division of Epidemiology and Prevention, Aichi Cancer Center Research Institute, Nagoya, Japan. [29]Department of Surgery, Nagano Matsushiro General Hospital, Nagano, Japan. [30]Department of Surgery, College of Medicine, University of Ulsan, Asan Medical Center, Seoul, South Korea. [31]Center for Breast Cancer, National Cancer Center, Gyeonggi-do, South Korea. [32]Department of Obstetrics and Gynaecology, Queen Mary Hospital, Li Ka Shing Faculty of Medicine, University of Hong Kong, Hong Kong, China. [33]Department of Breast Oncology, Aichi Cancer Center Central Hospital, Nagoya, Japan. [34]Research Center for Cancer Prevention and Screening, National Cancer Center, Tokyo, Japan. [35]Singapore Eye Research Institute, Singapore. [36]Department of Ophthalmology, National University of Singapore and National University Health System, Singapore. [37]Department of Medicine, University of Chicago, Chicago, Illinois, USA. [38]A full list of members and affiliations appears in the **Supplementary Note**. [39]Department of Molecular Physiology and Biophysics, Vanderbilt University School of Medicine, Nashville, Tennessee, USA. [40]Department of Biostatistics, Vanderbilt University School of Medicine, Nashville, Tennessee, USA. Correspondence should be addressed to Q.C. (qiuyin.cai@vanderbilt.edu).

ONLINE METHODS

Study populations. ABCC comprises 22,780 cases and 24,181 controls from 14 studies (**Supplementary Table 1**), including 15,483 Chinese women, 18,367 Korean women and 13,111 Japanese women. Chinese participants came from 6 studies based in Shanghai (n = 12,219; the Shanghai Breast Cancer Study (SBCS), the Shanghai Breast Cancer Survival Study (SBCSS), the Shanghai Endometrial Cancer Study (SECS; controls only) and the Shanghai Women's Health Study (SWHS))[6,26–28], Taiwan (n = 2,131)[29] and Hong Kong (n = 1,133)[30]. Korean participants came from 5 studies: the Seoul Breast Cancer Study (SeBCS; n = 6,179)[11], the Hwasun Cancer Epidemiology Study-Breast (HCES-Br; n = 6,573)[31–33], the Korea Genome Epidemiology Study (KoGES; n = 3,209)[34], the Korean Hereditary Breast Cancer study (KOHBRA; n = 1,397)[35] and the Korean National Cancer Center (n = 1,009). Japanese participants came from three studies: the Biobank Japan Project (BBJ; n = 11,021)[36], the Nagoya Study (n = 1,288)[37] and the Nagano Breast Cancer Study (n = 802)[38] (**Supplementary Table 1**). Detailed descriptions of these participating studies are presented in the **Supplementary Note**. The protocols for all participating studies were approved by their relevant institutional review boards, and all participants of the studies provided written informed consent. We estimated that our study had a statistical power of >80% to identify an association with an OR of 1.09 or greater at $P < 5 \times 10^{-8}$ for SNPs with a MAF as low as 0.25.

Genotyping methods. *Stage 1 genotyping.* Stage 1 included 2 GWAS in which 5,285 Chinese women and 4,777 Korean women were scanned primarily using Affymetrix Genome-Wide Human SNP Array 6.0. Genotyping protocols for stage I have been described elsewhere[6–9,11]. In the Chinese GWAS (SBCGS-1), the initial 300 samples were genotyped using the Affymetrix GeneChip Mapping 500K Array Set. The remaining 4,985 samples were genotyped using Affymetrix Genome-Wide Human SNP Array 6.0. After quality control exclusions, the final data set included 2,867 cases and 2,285 controls for 677,157 markers. For the Korean GWAS (SeBCS1), Affymetrix Genome-Wide Human SNP Array 6.0 Array was also used[11]. After quality control exclusions, the final data set included 2,246 cases and 2,052 controls for 555,117 markers. Detailed descriptions of quality control and criteria for sample and marker exclusion are presented in the **Supplementary Note**.

Stage 2 genotyping. Genotyping assays for 3,944 cases and 3,980 controls (SBCGS-2) in stage 2 were completed using Illumina Infinium assays as the add-on content to the Illumina HumanExome BeadChip (see URLs). Genotype calling was carried out using the Illumina GenTrain version 2.0 clustering algorithm in GenomeStudio version 2011.1. Cluster boundaries were determined using study samples. Further quality control procedures were conducted using PLINK (see URLs). Detailed descriptions of quality control and criteria for sample and marker exclusion are presented in the **Supplementary Note**. Of the 4,598 SNPs selected, assays for 4,071 SNPs were successfully designed using Illumina Infinium assays. A total of 3,850 SNPs were successfully genotyped, and 3,678 SNPs were included in the analyses of 3,472 breast cancer cases and 3,595 controls.

Stage 3 genotyping. Genotyping assays for the 50 SNPs in stage 3 were completed at the Vanderbilt Molecular Epidemiology Laboratory using the iPLEX Sequenom MassArray platform for 19,423 samples from the Taiwan, Hong Kong, HCES-Br, KOHBRA/KoGES, SeBCS2, Korea-NCC, Nagoya and Nagano studies. Detailed descriptions of quality control and criteria for sample and marker exclusion are presented in the **Supplementary Note**. For the BBJ1 study, the SNP data needed for the study were extracted from either genotyped (n = 8) or imputed (n = 14; mean RSQR = 0.96) data generated using the OmniExpress BeadChip. Breast cancer cases included in the BBJ2 study were genotyped using multiplex PCR Invader assays. SNP data for the BBJ2 controls were extracted from data generated using the OmniExpress BeadChip.

We also selected 16 SNPs that showed a promising trend in the other studies included in stage 3 for additional genotyping assays among 2,021 cases and 1,958 controls included in a case-control study conducted in Malaysia and Singapore that used the iPLEX Sequenom MassArray platform at the Cancer Research Initiatives Foundation, Sime Darby Medical Centre, Malaysia. However, because of a potential concern about genetic admixture shown in our previous study[3] and an unusual pattern of associations observed in these studies (**Supplementary Table 8**), we did not include these samples in the final analysis.

Statistical analyses. PLINK version 1.07 (see URLs)[39] was used to analyze the genome-wide data obtained in stage 1. To evaluate the population structure in the Chinese GWAS (SBCGS-1), we performed principal-component analyses using EIGENSTRAT software[40] in a set of approximately 6,000 independent SNPs that met the following criteria: (i) location at a distance of >200 kb from the SNP of interest, (ii) a MAF of >0.2, (iii) $r^2 < 0.1$ and (iv) a genotype call rate of >99%. The inflation factor (λ) was estimated to be 1.0426. Similar analyses were performed for the Korean GWAS (SeBCS1) and yielded a λ value of 1.0431 (ref. 11). We also rescaled the inflation statistic to an equivalent value for a study including 1,000 cases and 1,000 controls ($\lambda_{1,000}$) using the formula $\lambda_{1,000} = 1 + 500 \times (\lambda - 1) \times (1/n_{cases} + 1/n_{controls})$[41]. The $\lambda_{1,000}$ value was 1.02 for both studies included in stage 1 and 1.01 in the meta-analysis of both studies. These data suggest that any population substructure, if present, should not have any appreciable effect on the results. OR estimates associated with each SNP and 95% CIs were estimated using logistic regression implemented in PLINK with adjustment for age and the first two principal components.

We used the program MACH 1.0 (see URLs)[12] to impute genotypes for autosomal SNPs (n = 2,416,663) that were present in the CHB and JPT HapMap Phase 2 release 22 data for samples included in the Chinese and Korean GWAS. Only SNPs with a high imputation quality score (RSQR > 0.50) and MAF > 0.05 in these two GWAS were included in the analyses. Dosage data for imputed SNPs for samples in each GWAS were analyzed using the program Mach2dat (see URLs)[12]. Associations between genotype dosage (0, 1 or 2) of the effect allele and breast cancer risk were assessed using logistic regression models after adjusting for age and the first two principal components. ORs associated with each SNP and 95% CIs were estimated under a log-additive model. We also used SAS version 9.3 (see URLs) to analyze genotype data, which yielded results virtually identical to those generated with dosage data using Mach2dat. We obtained summary ORs and 95% CIs for SNPs from the two GWAS using METAL software (see URLs)[13] to run a fixed-effects inverse variance meta-analysis.

Individual data were obtained from all studies except for the two BioBank Japan studies (BBJ1 and BBJ2). Case-control differences in selected demographic characteristics and major risk factors were evaluated using t tests (for continuous variables) and χ^2 tests (for categorical variables). Summary associations between SNPs and breast cancer risk were generated on the basis of a fixed-effects inverse variance meta-analysis conducted using METAL software[13]. Analyses stratified by ancestry and ER status were also carried out. Heterogeneity across studies, among ancestry groups and according to ER status was assessed with a Cochran's Q test. P values of $<5 \times 10^{-8}$ in the combined analysis were considered to be statistically significant.

We assessed associations of breast cancer risk with the three newly identified risk variants among women of European ancestry in collaboration with the DRIVE GAME-ON Consortium (see URLs). Included in this analysis were data from 16,003 cases and 41,335 controls recruited in 12 studies. Genome-wide scan data from these studies were imputed and meta-analysis was performed; summary data are presented herein.

We generated forest plots using R version 3.0.0 (see URLs). Regional association plots were drawn using the web-based tool LocusZoom, version 1.1 (see URLs)[42]. LD matrices used in this study were reported on the basis of HapMap release 22 data. All genomic references are based on NCBI Build 36.

Functional annotation. Functional annotation was performed using data from the ENCODE Project (see URLs) accessed through the UCSC Genome Browser (see URLs). Enhancer and transcription elongation regions were predicted in HMECs using ChromHMM annotation. DNase I–hypersensitive areas, transcription factor binding sites and microRNA binding sites were evaluated in all cell types, including breast cancer cells, in ENCODE. RegulomeDB (see URLs)[17], a database that annotates SNPs with known and predicted regulatory elements in the intergenic regions of the human genome using data from the Gene Expression Omnibus (GEO), the ENCODE Project and published literature, was also used to predict the possibility of transcription factor binding sites and DNase I hypersensitivity. In addition, we performed functional

annotation using HaploReg v2 (see URLs)[18], a tool for exploring the annotations of the noncoding genome at variants on haplotype blocks.

eQTL analysis. We used TCGA breast cancer data (**Supplementary Note**) to perform an eQTL analysis for normal and tumor tissue samples separately. Detailed descriptions of the eQTL analysis are presented in the **Supplementary Note**. We focused only on SNPs and genes located within the 1-Mb regions flanking the three newly identified risk loci to identify *cis*-eQTLs. A significance threshold P value of <0.01 was used to identify candidate *cis*-eQTLs.

Differential gene expression analysis. To identify differentially expressed genes located in the three identified risk loci, we analyzed data from a total of 87 pairs of tumor-normal breast tissue samples included in TCGA (**Supplementary Note**). Detailed descriptions of the differential gene expression analysis are presented in the **Supplementary Note**.

26. Gao, Y.T. *et al.* Association of menstrual and reproductive factors with breast cancer risk: results from the Shanghai Breast Cancer Study. *Int. J. Cancer* **87**, 295–300 (2000).
27. Shu, X.O. *et al.* Soy food intake and breast cancer survival. *J. Am. Med. Assoc.* **302**, 2437–2443 (2009).
28. Zheng, W. *et al.* The Shanghai Women's Health Study: rationale, study design, and baseline characteristics. *Am. J. Epidemiol.* **162**, 1123–1131 (2005).
29. Ding, S.L. *et al.* Genetic variants of *BLM* interact with *RAD51* to increase breast cancer susceptibility. *Carcinogenesis* **30**, 43–49 (2009).
30. Chan, K.Y. *et al.* Functional polymorphisms in the *BRCA1* promoter influence transcription and are associated with decreased risk for breast cancer in Chinese women. *J. Med. Genet.* **46**, 32–39 (2009).
31. Song, H.R. *et al.* Sex-specific differences in the association between *ABO* genotype and gastric cancer risk in a Korean population. *Gastric Cancer* **16**, 254–260 (2013).
32. Cui, L.H. *et al.* Methylenetetrahydrofolate reductase C677T polymorphism in patients with gastric and colorectal cancer in a Korean population. *BMC Cancer* **10**, 236 (2010).
33. Kweon, S.S. *et al.* Cohort Profile: The Namwon Study and the Dong-gu Study. *Int. J. Epidemiol.* **43**, 558–567 (2014).
34. Cho, Y.S. *et al.* A large-scale genome-wide association study of Asian populations uncovers genetic factors influencing eight quantitative traits. *Nat. Genet.* **41**, 527–534 (2009).
35. Han, S.A. *et al.* The Korean Hereditary Breast Cancer (KOHBRA) Study: Protocols and Interim Report. *Clin. Oncol. (R Coll Radiol)* **23**, 434–441 (2011).
36. Elgazzar, S. *et al.* A genome-wide association study identifies a genetic variant in the *SIAH2* locus associated with hormonal receptor–positive breast cancer in Japanese. *J. Hum. Genet.* **57**, 766–771 (2012).
37. Hamajima, N. *et al.* Gene-environment Interactions and Polymorphism Studies of Cancer Risk in the Hospital-based Epidemiologic Research Program at Aichi Cancer Center II (HERPACC-II). *Asian Pac. J. Cancer Prev.* **2**, 99–107 (2001).
38. Itoh, H. *et al.* Serum organochlorines and breast cancer risk in Japanese women: a case-control study. *Cancer Causes Control* **20**, 567–580 (2009).
39. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
40. Price, A.L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **38**, 904–909 (2006).
41. Freedman, M.L. *et al.* Assessing the impact of population stratification on genetic association studies. *Nat. Genet.* **36**, 388–393 (2004).
42. Pruim, R.J. *et al.* LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics* **26**, 2336–2337 (2010).

doi:10.1038/ng.3041

NEOPLASIA
www.neoplasia.com

Volume 16 Number 8 August 2014 pp. 656–664 656

Late Cornified Envelope Group I, a Novel Target of p53, Regulates PRMT5 Activity[1]

Zhenzhong Deng*,†, Koichi Matsuda†, Chizu Tanikawa†, Jiaying Lin*,†, Yoichi Furukawa‡, Ryuji Hamamoto* and Yusuke Nakamura*

*Section of Hematology/Oncology, Department of Medicine, The University of Chicago, Chicago, IL 60637 USA; †Laboratory of Molecular Medicine, Human Genome Center, Institute of Medical Science, The University of Tokyo, Tokyo, Japan; ‡Division of Clinical Genome Research, Advanced Clinical Research Center, Institute of Medical Science, The University of Tokyo, Tokyo, Japan

Abstract

p53 is one of the most important tumor suppressor genes involved in human carcinogenesis. Although downstream targets of p53 and their biologic functions in cancer cells have been extensively investigated, it is still far from the full understanding. Here, we demonstrate that *Late Cornified Envelope Group I* (*LCE1*) genes, which are located in the *LCE* gene clusters encoding multiple well-conserved stratum-corneum proteins, are novel downstream targets of p53. Exogenous p53 overexpression using an adenoviral vector system significantly enhanced the expression of *LCE1* cluster genes. We also observed induction of LCE1 expressions by DNA damage, which was caused by treatment with adriamycin or UV irradiation in a wild-type p53-dependent manner. Concordantly, the induction of *LCE1* by DNA damage was significantly attenuated by the knockdown of p53. Among predicted p53-binding sites within the *LCE1* gene cluster, we confirmed one site to be a p53-enhancer sequence by reporter assays. Furthermore, we identified LCE1 to interact with protein arginine methyltransferase 5 (PRMT5). Knockdown of LCE1 by specific small interfering RNAs significantly increased the symmetric dimethylation of histone H3 arginine 8, a substrate of PRMT5, and overexpression of LCE1F remarkably decreased its methylation level. Our data suggest that LCE1 is a novel p53 downstream target that can be directly transactivated by p53 and is likely to have tumor suppressor functions through modulation of the PRMT5 activity.

Neoplasia (2014) 16, 656–664

Introduction

p53 is the most frequently mutated tumor suppressor gene involved in human cancers [1,2]. Its tetramer protein product can activate the transcription of a number of target downstream genes and mediate a variety of biologic functions through the transcriptional regulation of those targets [3]. To elucidate the critical roles of p53 in human carcinogenesis, we and others have attempted to identify p53 target genes through multiple approaches. We have mainly applied the expression profile analysis after the exogenous introduction of wild-type p53 into cancer cells using the adenovirus vector system and identified more than 50 p53 downstream candidate genes [4]. Among them, we have performed the functional analysis of more than a dozen of target genes including *p53R2*, *p53AIP1*, and *p53RDL1* [5–8]. Here, we report the characterization of the Late Cornified Envelope Group I (LCE1) as a novel downstream target of p53.

The *LCE* clusters contain multiple well-conserved genes encoding stratum-corneum proteins [9,10] and are located on chromosome 1q21 in a region called as the epidermal differentiation complex [11,12]. This region is enriched for genes, which are expressed during

Address all correspondence to: Yusuke Nakamura, MD, PhD, Section of Hematology/Oncology, Department of Medicine, The University of Chicago, 5841 S. Maryland Ave., MC2115 Chicago, IL 60637. E-mail: ynakamura@bsd.uchicago.edu

[1] This article refers to supplementary materials, which are designated by Tables W1 to W6 and Figures W1 to W6 and are available online at www.neoplasia.com.

Received 10 June 2014; Revised 26 July 2014; Accepted 29 July 2014

1476-5586/14
http://dx.doi.org/10.1016/j.neo.2014.07.008

epidermal differentiation, including *loricrin*, *involucrin*, *filagrin*, the small proline-rich protein genes, and the *LCE* genes [9,13]. In mice, members in the LCE1 group are expressed in the relatively late stage of epithelial development and incorporated into the cornified envelope through cross-linking by transglutaminases [10]. In addition, real-time quantitative polymerase chain reaction (qPCR) analysis demonstrated that human *LCE1* and *LCE2* genes were primarily expressed in skin, whereas LCE4 and LCE5 gene expressions were undetectable in any human tissues examined [9]. In general, physiological functions of LCE proteins, especially their involvement in human cancer are still largely unknown.

Protein arginine methyltransferases (PRMTs) constitute of a large family of enzymes having the arginine methyltransferases activity responsible for catalyzing the formation of monomethyl arginine, asymmetric dimethyl arginine, and symmetric dimethyl arginine (SDMA) [14]. PRMT5 is one of the most well-characterized family members with SDMA activity and catalyzes formation of SDMA in proteins with a glycine and arginine-rich motif [15]. PRMT5 was reported to regulate various cellular functions including apoptosis, Golgi structure, pluripotency, cell growth, and snRNP biosynthesis [16–18]. One important key marker of the PRMT5 activity is the symmetrical dimethylation of histone 3 arginine 8 (H3R8me2s) level. Through hypermethylation of histone H3R8 around the promoter regions, PRMT5 could cause the transcriptional silencing of cell cycle regulator genes [19,20]. Since overexpression of PRMT5 has been reported in various types of human cancer, including melanoma, leukemia, lymphoma, glioma, as well as ovarian, breast, prostate, and lung cancers [16,21–23], this enzyme is considered as a good molecular target for development of novel cancer therapy [16].

In the present study, we demonstrate that LCE is a novel direct target of p53, can interact with PRMT5, and might modulate histone H3 methylation by PRMT5. This mechanism may be important for the interplay of two important cancer-related genes, *p53* and *PRMT5*, and our findings could indicate a possible role of LCE1 in human carcinogenesis.

Materials and Methods

Microarray Analysis

Replication-deficient recombinant adenovirus designed to express wild-type p53 (Ad-p53) or LacZ (Ad-LacZ) was generated and purified, as previously described [5,24]. Microarray analysis was then carried out as previously described [1,4,24]. In brief, poly(A)+ RNAs were isolated from U373MG cells at different time points after infection with Ad-p53 or Ad-LacZ. Each RNA sample was labeled and hybridized to a microarray consisting of 36,864 genes or Expressed Sequence Tags (ESTs) (http://www.ncbi.nlm.nih.gov/geo/index.cgi, Accession No. GSE14953).

Cell Culture and Transfection

Human cell lines, U373MG (glioblastoma), H1299 (lung carcinoma), and Human Embryonic Kidney 293T (HEK293T), were purchased from American Type Culture Collection (ATCC, Manassas, VA). HCT116 $p53^{-/-}$ and HCT116 $p53^{+/+}$ cell lines were obtained from Dr Bert Vogelstein (Johns Hopkins University, Baltimore, MD). U373MG glioblastoma cells and H1299 lung cancer cells were infected with Ad-p53 or Ad-LacZ at various multiplicity of infection (MOI) conditions and incubated at 37°C until the time of harvest. HEK293T cells were transfected with HA-Mock (empty vector of pCAGGSnHC) or HA-LCE1F (pCAGGSnHC-LCE1F) using FuGENE6. For protein-protein interaction experiments, HEK293T cells were transfected with HA-Mock or HA-LCE1F and FLAG-Mock or FLAG-PRMT5 using FuGENE6. To examine the co-localization, HCT116 $p53^{+/+}$ cells were transfected by HA-Mock or HA-LCE1F using FuGENE6. For the gene reporter assay, U373MG (mutated p53) and H1299 (p53 null) cells were transfected with a reporter plasmid and either a Mock empty vector of pcDNA3.1+ or a wild-type p53 expression pcDNA3.1+ vector in combination with a pRL-CMV vector using FuGENE6 [25,26]. Small interfering RNAs (siRNAs) that were commercially synthesized by Sigma-Aldrich (St Louis, MO) were transfected with Lipofectamine RNAiMAX reagent (Life Technologies, Carlsbad, CA). Two siRNA oligonucleotides of *LCE1A-F* were designed to target the region commonly conserved among *LCE1A-F*. Sequences of oligonucleotides are shown in Table W1. Western blot or real-time qPCR was applied to validate the efficiency of overexpression or knockdown experiments.

DNA-Damaging Treatments

Cells were seeded 24 hours before treatment. When cells reached 60% to 70% confluency, cells were incubated with 1 μg/ml adriamycin for 2 hours followed by further incubation of the drug-free medium. Then, the cells were harvested at different time points as indicated in the figure legends. For UV irradiation (UVR) experiments, the cells were washed twice with phosphate-buffered saline (PBS) and exposed to UV rays at different doses using XL-1500 Spectrolinker (Spectronics, Westbury, NY; peak emission, 254 nm). Cells were harvested 36 hours after UVR treatment.

Real-Time qPCR

Total RNA was isolated from cultured cells using RNeasy mini-spin column kits (Qiagen, Hilden, Germany) according to the manufacturer's procedure. cDNAs were synthesized with the SuperScript Pre-amplification System (Life Technologies). Real-time qPCR was conducted using the SYBR Green I Master on the LightCycler 480 (Roche Applied Science, Mannheim, Germany) according to the manufacturer's instructions. The primer sequences used in this manuscript are shown in Tables W1 and W2. Primers for *LCE1A*, *LCE1C*, *LCE1E*, and *LCE1F* were described previously [9]. Primers for *LCE1B*, *LCE1D* *LCE3A*, *LCE3B*, *LCE3C*, *LCE4A*, *XP33*, and *C1orf45* genes were designed by us. Except *LCE1D*, the specificity of all primers was confirmed by DNA sequencing of amplicons. For graphic representation of transcript data, all expression of target genes was shown relative to the housekeeping gene β_2-microglobulin expression in the same sample.

Prediction of Putative p53-Binding Sites

DNA sequences of an entire genomic region of *LCE1* including 10 kb of the 5′ upstream sequence were downloaded from the University of California Santa Cruz (UCSC) website (http://genome.ucsc.edu/), and the putative p53-binding sites (p53BSs) were screened according to the following criteria; at least 80% matched with the 20 nucleotides of consensus sequence RRRCWWGYYY__RRRCWWGYYY (R, purine; W, A, or T; Y, pyrimidine); we started the screening of 11 consensus sequences without any spacers between the two halves of p53BSs (Figure W1) and confirmed one site that was likely to be a direct p53-binding sequence.

Gene Reporter Assay

DNA fragments including each of the potential p53BSs of the *LCE1* gene cluster were amplified by KOD-Plus-DNA polymerase

(Toyobo, Osaka, Japan) and subcloned into the pGL3-promoter (pGL3-pro) vector. The primers for amplification are indicated in Table W1. To make a series of mutant vectors, a point mutation "T" was inserted into the site of the fourth and the fourteenth nucleotide "C" and into the seventh and the seventeenth nucleotide "G" of the consensus p53-BS using the KOD-Plus-Mutagenesis Kit (Toyobo). Since a functional p53BS is known in the *Fas* promoter region, a wild-type Fas promoter construct, pGL3-Fas, was used as a positive control. U373MG (mutated p53) and H1299 (p53 null) cells were plated in 12-well culture plates (5×10^4 cells per well) 24 hours before co-transfection of 125 ng of a reporter plasmid and either 125 ng of a Mock vector or a wild-type p53 expression vector in combination with 25 ng of a pRL-CMV vector. Cells were rinsed with PBS 36 hours after transfection and lysed in 250 μl of lysis buffer. Twenty microliters and 5 μl of lysates from U373MG and H1299, respectively, were sequentially measured using the PGD-S Dual Luciferase assay system following the manufacturer's procedure (Toyo Ink, Tokyo, Japan). The firefly luciferase value was normalized by the Renilla luciferase activity.

Expression Plasmid Construction

An entire coding sequence of PRMT5 was amplified using cDNA generated from mRNA of HEK293T cells and cloned into pCAGGSn3FC (PRMT5) vector. The entire coding sequences of *LCE1A-F* were amplified by the use of cDNA generated from mRNA that was extracted from HCT116 $p53^{+/+}$ cells after 70 J/m^2 of UVR treatment. Due to unknown reasons, we have been unsuccessful in cloning the *LCE1D* cDNA. An HA-epitope tag was placed at the C-terminus of the pCAGGSnHC vector and three-tandem FLAG epitope tags were placed at the C-terminus of the pCAGGSn3FC vector. The DNA sequences of expression constructs for LCE1 (pCAGGSnHC-LCE1) and PRMT5 (pCAGGSn3FC-PRMT5) were confirmed by DNA sequencing using ABI PRISM 3730XL Genetic Analyzer (Life Technologies).

Antibodies

The following primary antibodies were deployed: rabbit anti-HA (Y-11; Santa Cruz Biotechnology, Santa Cruz, CA; dilution used in WB: 1:1000), rabbit anti-PRMT5 (07-415; Millipore, Billerica, MA; dilution used in WB: 1:1000, ICC: 1:400), rabbit anti-p53 (sc6243, Santa Cruz; dilution used in WB: 1:1000), mouse anti-p53 [Ab-1, Calbiochem, San Diego, CA, dilution used in chromatin immunoprecipitation (ChIP): 1:100], rabbit anti-H3R8me2s (ab130740, Abcam, Cambridge, United Kingdom; dilution used in Western blot (WB): 1:1000), mouse anti–α-tubulin (clone DM1A, Millipore; dilution used in WB: 1:1000), rat anti-HA (3F10, Roche; dilution used in Immunocytochemistry (ICC): 1:800), and anti-FLAG (F7425, Sigma-Aldrich; dilution used in ICC: 1:1000).

Immunocytochemistry

Forty-eight hours after transfection with HA-Mock (empty vector of pCAGGSnHC) or pCAGGSnHC-HA-LCE1F into HEK293T or HCT116 $p53^{+/+}$ cells in four-well chambers, the cells were fixed by 1.7% formaldehyde or 4% paraformaldehyde in PBS and permeabilized with 0.2% Triton X-100 in PBS. After covering with blocking solution (3% BSA in 0.2% Triton X-100) for 1 hour at room temperature, cells were incubated with a rat anti-HA antibody and a rabbit PRMT5 antibody overnight under humidified atmosphere at 4°C. Further, the cells were stained with fluorescence-conjugated secondary antibodies and then counterstained with 4′,6-diamidino-2-phenylindole (DAPI) in VECTASHIELD Mounting Media (HT1200; Vector Laboratories, Burlingame, CA).

Immunoprecipitation and Mass-Spectrometric Analysis

Forty-eight hours after transfection with HA-Mock or HA-LCE1F, HEK293T cells were lysed in lysis buffer [50 mM Tris-HCl (pH 8.0), 0.4% NP-40, and 150 mM NaCl] containing Protease Inhibitor Cocktail Set III (Calbiochem). Whole-cell lysates were pre-cleared by incubation with normal mouse IgG (sc2025, Santa Cruz Biotechnology) and recombinant protein G-sepharose 4B (Life Technologies) at 4°C for 1 hour, followed by incubation with anti-HA agarose (A2095, Sigma-Aldrich) overnight. The proteins were separated in Mini-PROTEAN 5% to 20% gradient sodium dodecyl sulfate–polyacrylamide gel electrophoresis precast gels (Bio-Rad Laboratories, Hercules, CA) and stained with the silver-staining kit (Life Technologies). Protein bands that were specifically observed in the cell extracts transfected with HA-LCE1F compared with Mock were excised and analyzed by liquid chromatography–tandem mass spectrometry as previously described [27]. For the co-immunoprecipitation experiments, HEK293T cells were co-transfected with HA-LCE1F and FLAG-PRMT5 or FLAG-Mock as a control. An HA pull-down experiment is exactly the same as mentioned above. As for FLAG pull-down experiment, whole-cell lysates were pre-cleared by incubation with normal mouse IgG (sc2025, Santa Cruz Biotechnology) and recombinant protein G-sepharose 4B (Life Technologies) at 4°C for 1 hour, followed by incubation with mouse anti-FLAG M2 agarose (F3165, Sigma-Aldrich) overnight. Rat anti-HA (3 F10, Roche) and rabbit anti-Flag (F7425, Sigma-Aldrich) antibodies were used for Western blot. For the histone methylation analysis, histones were extracted by the histone purification mini kit (40026, Active Motif, Carlsbad, CA) following the manufacturer's procedure strictly.

ChIP Assay

ChIP assays were performed using ChIP Assay kit (17-295; Millipore) according to the manufacturer's protocol [28]. Briefly, HCT116 $p53^{+/+}$ cells were treated with 1 μg/ml adriamycin for 2 hours, and cells were cultured in the drug free-medium for 24 hours. Then, cells were harvested and the fragment of p53 and chromatin complexes was immnoprecipitated with an anti-p53 antibody (Ab-1, Calbiochem). After DNA fragments bound to p53 were eluted out, an aliquot was subjected to real-time qPCRs. Protein A agarose/Salmon Sperm DNA (16-157; Millipore) was used as a negative control. Primers were designed to amplify the region containing the p53BS2-binding site, and their sequence information is shown in Table W1.

Results

Identification of a Novel p53 Downstream Target

We had previously performed expression profile analysis to compare U373MG glioblastoma cells infected with wild-type p53 (Ad-p53) with those infected with LacZ (Ad-LacZ) to screen possible downstream genes that are regulated by p53 and identified more than 50 genes that were likely to be induced by wild-type p53. Among them, we confirmed that the transcriptional levels of LCE1B and LCE1C were elevated more than seven-fold higher in the cells infected with wild-type p53 than those with LacZ (data not shown). *LCE1B* and *LCE1C* belong to the *LCE* gene cluster containing multiple well-conserved genes that encode stratum-corneum proteins.

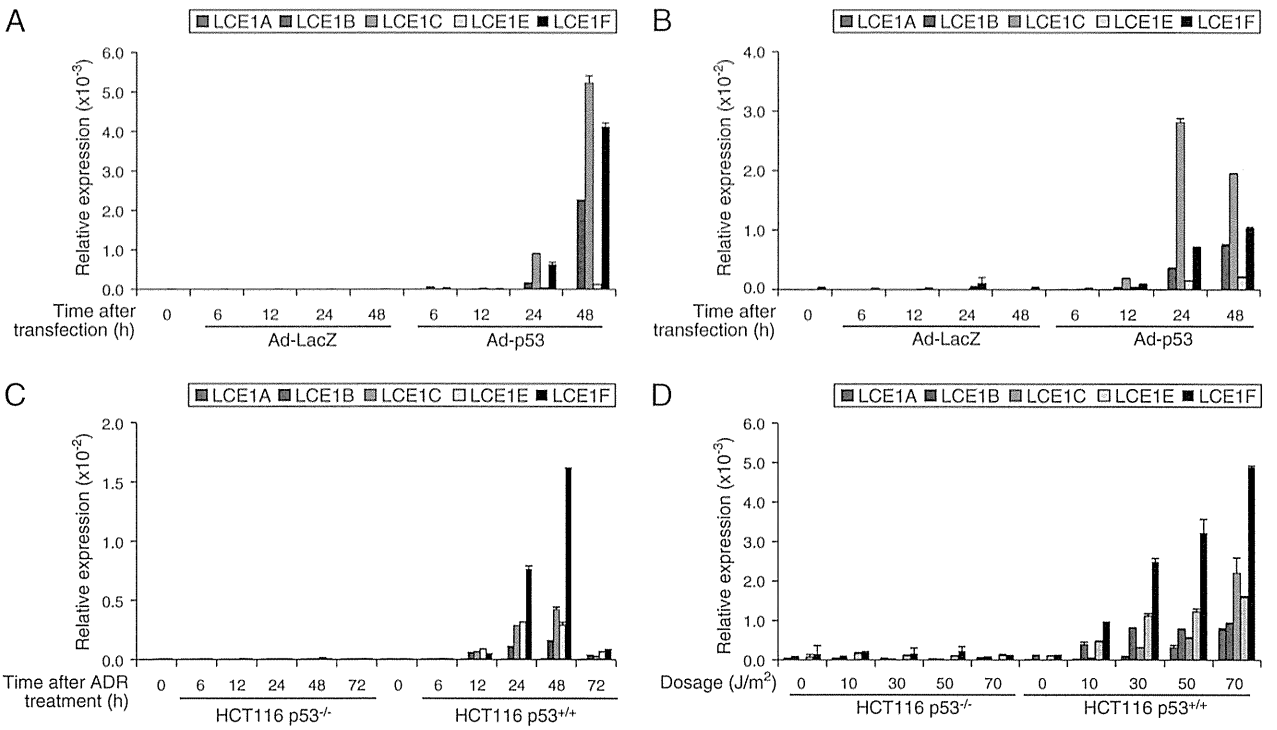

Figure 1. *LCE1* is upregulated by *p53*. Real-time qPCR analysis of *LCE1* mRNA expression in U373MG (*p53 mutant*) (A) and H1299 (*p53 null*) (B) cells at indicated time points after infection with Ad-p53 or Ad-LacZ at 8 MOIs. (C) Endogenous p53-dependent induction of *LCE1* mRNA transcription in $p53^{-/-}$ and $p53^{+/+}$ HCT116 cells after treatment with adriamycin (1 μg/ml) at various time points. (D) Cells were treated with UVR at different doses and harvested 36 hours after UVR. β_2-Microglobulin was used for the normalization of expression levels.

Within the *LCE* cluster, multiple genes form "groups" at chromosome 1q21 and are known to respond "group-wise" to various environmental stimuli like calcium and UV light [9], suggesting that other *LCE* members that were not included in our microarray might also be regulated by p53. We first compared the sequence similarity of the *LCE1B* and *LCE1C* transcripts with other members in the *LCE* cluster (Table W3) and selected genes showing high similarity (>80%) for further validation by real-time qPCR because the microarray results might reflect the cross-hybridization of other LCE members that were possibly induced by p53. Interestingly, expression levels of *LCE1* group genes were significantly increased in both U373MG (Figure 1*A*) and H1299 (Figure 1*B*) cells after 8 MOIs of Ad-p53 infection (Table W4), although the induction levels are very different. However, other genes in the LCE cluster were not induced by wild-type p53 introduction. In addition, endogenous p53, which was activated by adriamycin and UVR, could also induce expressions of the *LCE1* group genes in HCT116 $p53^{+/+}$ cancer cells but not in HCT116 $p53^{-/-}$ cancer cells (Figure 1, *C* and *D*), indicating that the LCE1 group genes can respond group-wise to the genotoxic stress condition in a p53-dependent manner [9].

LCE1 Is a Direct Target of p53

We then attempted to clarify whether *LCE1* is directly or indirectly regulated by p53. First, 11 putative p53BSs of 20 nucleotides with at least 80% match to the consensus p53-binding sequence (see Materials and Methods section) were selected from the *LCE1* cluster region (Figure W1*A*). We then constructed serial reporter vectors containing each of the 11 predicted p53BSs. Luciferase activity of p53BS2, p53BS6, and p53BS7 was likely to be enhanced by co-transfection of each luciferase vector and a p53 wild-type expression vector into U373MG (Figure W1*B*) and H1299 cells (Figure W1*C*) in comparison with those with the mock vector. Hence, these three p53BSs were further investigated by replacement of critical nucleotides in the core consensus sequence with other nucleotides (Figure 2*A*). As a result, we observed that substitutions in either of the p53BS2, p53BS6, and p53BS7 sequences significantly diminished the enhancement of the luciferase activity in both U373MG and H1299 cells (Figure 2, *B* and *C*).These results imply that p53 directly binds to these p53BSs in the *LCE1* cluster and regulates their transcriptions.

We subsequently transfected siRNA targeting p53 or that commonly targeting LCE1A-F into HCT116 $p53^{+/+}$ cells and then the cells were treated with adriamycin. As shown in Figure 2*D*, induction of *LCE1* genes by adriamycin treatment was significantly attenuated in the cells transfected with si-p53, also supporting that *LCE1* family genes are p53-direct targets. Additionally, ChIP analysis showed the direct binding of p53 to the *LCE1* cluster region in HCT116 $p53^{+/+}$ cells after adriamycin treatment (Figure W2). We then examined the expression of *LCE1F* in 82 cancer cell lines by real-time qPCR (Table W5) and confirmed that *LCE1F* was significantly downregulated in p53 mutant cells compared with p53 wild-type cells in colon and lung cancer cell lines (Figure W3). This is consistent with the induction of *LCE1F* by p53 in HCT116 colon and H1299 lung cancer cells as shown in Figure 1.

LCE1F Interacts with PRMT5

To further analyze possible biologic functions of LCE1F in cancer cells, we constructed an expression vector designed to express full-

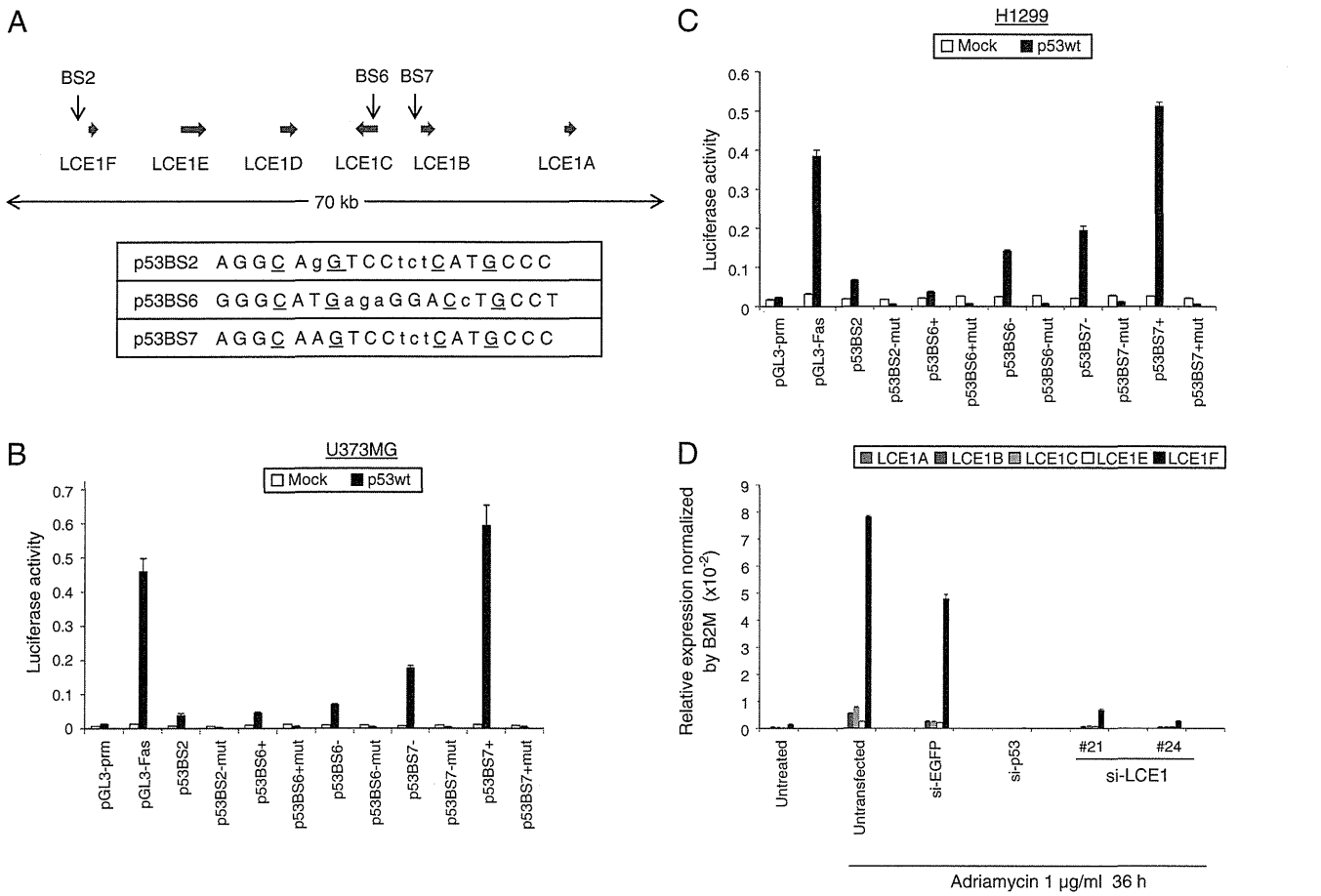

Figure 2. Identification of p53BSs in the *LCE1* cluster genes. (A) In the *LCE1* cluster genes, we identified 11 potential p53BSs that have at least 80% match to the consensus p53-binding sequence without any spacer nucleotides between the two halves of p53BSs. Among them, three possible p53 enhancer sequences are shown in the figure. Bold horizontal arrows indicate the locations and relative sizes of each *LCE1* gene. Vertical arrows show the potential p53BS locations. Identical nucleotides to the p53-binding sequence are written in capital letters. The underlined cytosine and guanine were substituted with thymine to introduce a mutant-type reporter vector at each p53BS. (B, C) Reporter assays of wild-type p53 and mutant-type p53 in U373MG (B) and H1299 (C) cells. Cells were co-transfected with a Mock (empty pcDNA3.1 +) or a wild-type p53 vector and each reporter plasmid, and luciferase assays were conducted 36 hours after transfection. Results are shown as the firefly luciferase activity normalized by the Renilla luciferase activity with 1 SD. (D) Real-time qPCR of *LCE1* family genes. HCT116 $p53^{+/+}$ cells were transfected with siRNAs targeting EGFP, p53, or LCE1 for 24 hours and then treated with 1 μg/ml adriamycin. Total RNA was extracted from the cells 36 hours after adriamycin treatment.

length LCE1F protein (pCAGGS-nHC-LCE1F) and transfected it into HEK293T cells. After confirmation of the LCE1F protein expression by Western blot analysis (Figure W4*A*), we performed immunocytochemical analysis to examine the subcellular localization of LCE1F. As shown in Figure W4*B*, LCE1F was strongly stained in both nucleus and cytoplasm.

In addition, we attempted to identify the interacting protein(s) of LCE1F. An HA-tagged LCE1F expression vector or an empty mock vector was transfected into HEK293T cells, and cell lysates were immunoprecipitated with anti–HA-conjugated agarose beads. Protein samples were separated by sodium dodecyl sulfate–polyacrylamide gel electrophoresis, and the gel was silver stained. We found a 70-kDa protein, which was strongly stained only in HA-LCE1F–overexpressing cells, conducted liquid chromatography–tandem mass spectrometry analysis of this protein, and identified it to be an arginine methyltransferase, PRMT5. This result was confirmed using an anti–PRMT5-specific antibody (Figures 3*A* and W5). Reversibly, cell lysates of HA-LCE1F–overexpressed cells were immunoprecipitated with an anti-PRMT5 antibody, and the immunoprecipitates were blotted with an anti-HA antibody, which also validated the interaction between LCE1F and PRMT5 (Figure 3*B*). Moreover, we co-transfected either FLAG-PRMT5 or FLAG-Mock with HA-LCE1F or HA-Mock, and cell lysates were immunoprecipitated with an anti-HA antibody (Figure 3*C*) or an anti-FLAG antibody (Figure 3*D*). Subsequent Western blot analysis clearly indicated the interaction between PRMT5 and LCE1F. These results imply that LCE1F interacts with PRMT5 in the cells. Interestingly, we confirmed that all LCE1 family proteins examined could interact with PRMT5 (Figure W6), suggesting that the interaction with PRMT5 is likely to be a common characteristic among LCE1 family members.

LCE1F Suppresses PRMT5 Activity

Since PRMT5 was reported to regulate the transcription of various genes through the methylation of arginine 8 on histone H3 (H3R8), which is considered to play an important role in human carcinogenesis [16,20,21], we further investigated the biologic

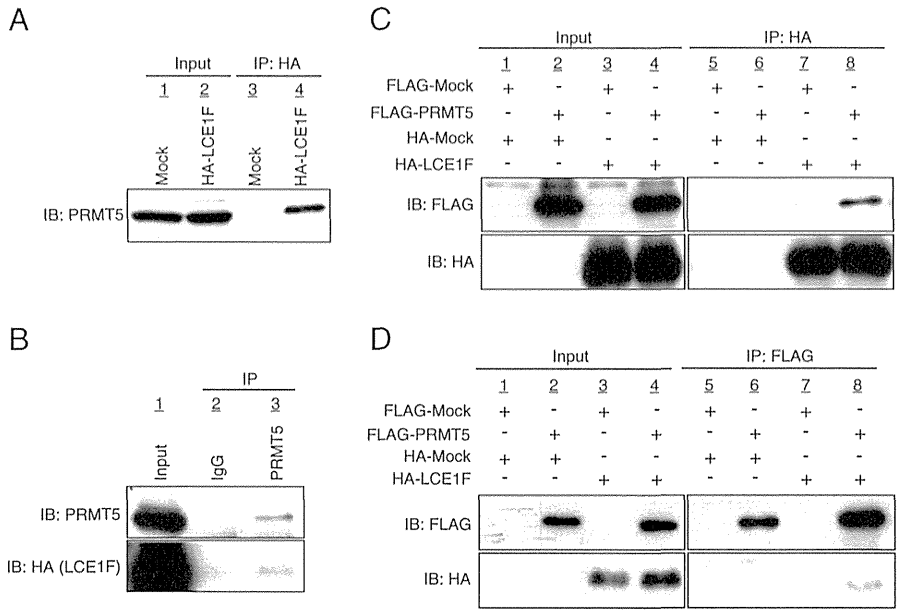

Figure 3. Interaction between LCE1F and PRMT5. (A) HEK293T cells were transfected with either a Mock (empty) or an HA-LCE1F expression vector and immunoprecipitated with anti-HA agarose. Samples were immunoblotted with an anti-PRMT5 antibody (Millipore). (B) HA-LCF1F–transfected HEK293T cells were immunoprecipitated with an anti-PRMT5 antibody or normal rabbit IgG (negative control). Samples were immunoblotted with an anti-PRMT5 or anti-HA antibody. (C, D) HEK293T cells were co-transfected with either FLAG-Mock or FLAG-PRMT5 and either HA-Mock or HA-LCE1F. Cells were immunoprecipitated with anti-HA agarose (C) or anti-FLAG antibody (D), and samples were immunoblotted with anti-FLAG or anti-HA antibodies.

significance of the interaction of PRMT5 and LCE1 members. We first performed immunocytochemical analysis and detected co-localization of LCE1F and PRMT5 in both the nucleus and the cytoplasm of the HCT116 $p53^{+/+}$ cells as shown in Figure 4. We then examined the methylation status of H3R8 after standardizing the quantity of PRMT5 using the anti-PRMT5 antibody. Using an antibody for symmetric dimethylation of H3R8 (H3R8me2s), we conducted Western blot analysis after transfection with siRNAs

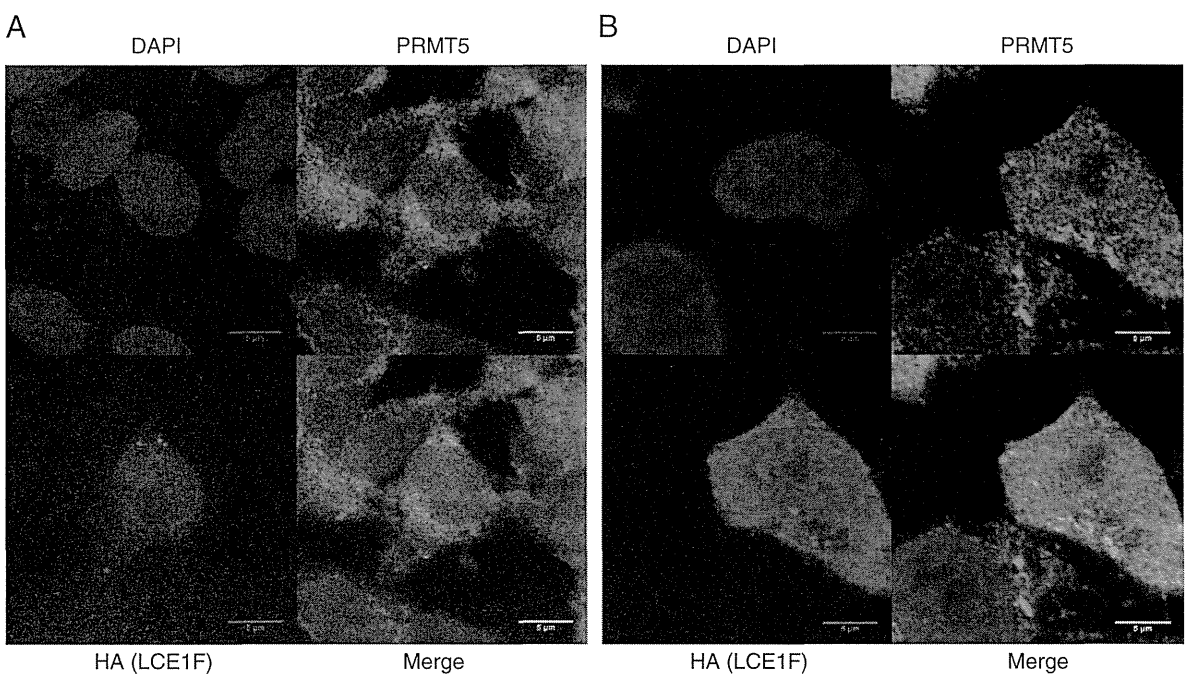

Figure 4. Co-localization of LCE1F and PRMT5. HCT116 $p53^{+/+}$ cells were transfected with HA-LCE1F. Cells were fixed by 1.7% formaldehyde and stained with an anti-HA antibody (red) and an anti-PRMT5 antibody (green).

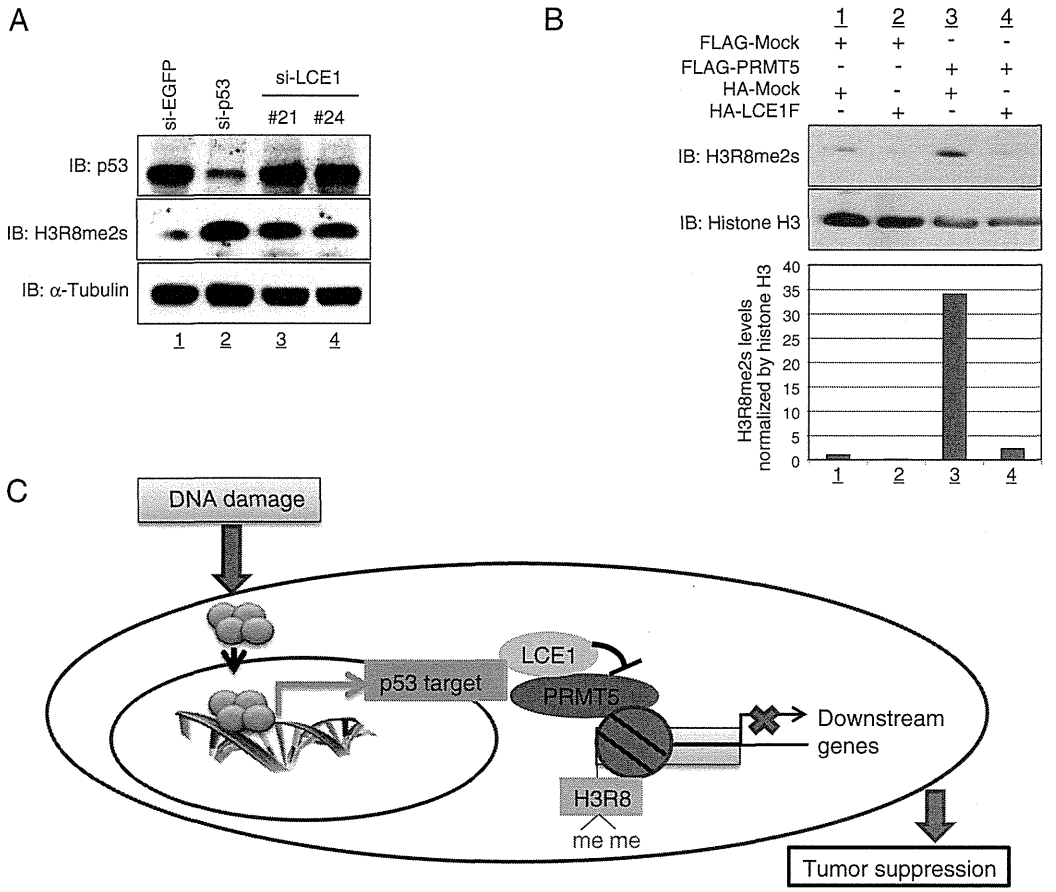

Figure 5. Possible mechanism of LCE1F on regulation of PRMT5 activity. (A) HCT116 $p53^{+/+}$ cells were transfected with siRNAs targeting EGFP, p53, or LCE1 for 24 hours and then treated with 1 μg/ml adriamycin. Proteins were extracted from the cells 36 hours after adriamycin treatment and immunoblotted with anti-p53, anti-H3R8 SDMA (H3R8me2s), or anti–α-tubulin (control) antibodies. (B) HEK293T cells were co-transfected with either FLAG-Mock or FLAG-PRMT5 and either HA-Mock or HA-LCE1F. Cells were harvested 48 hours after transfection and immunoblotted with anti-H3R8me2s or histone H3 (control) antibodies after histone purification. (C) A possible model of the network among p53, LCE1F, and PRMT5. p53 induced by DNA damage activates transcription of LCE1F, which suppresses the PRMT5 methyltransferase activity through their interaction.

targeting Enhanced green fluorescent protein (EGFP), p53, or LCE1 under treatment with adriamycin (Figure 5*A*) and found that knockdown of p53 and LCE1 significantly enhanced the methylation levels of H3R8 in HCT116 $p53^{+/+}$ cells, compared with the control cells, suggesting that LCE1 may modulate the PRMT5-dependent H3R8 methylation. To further validate this result, we co-transfected either FLAG-PRMT5 or FLAG-Mock and HA-LCE1F or HA-Mock into HEK293T cells and purified histone proteins from cell lysates. As shown in Figure 5*B*, in the presence of LCE1F, H3R8 methylation by PRMT5 protein was significantly suppressed. Taken together, our data indicate that LCE1F may negatively regulate PRMT5-dependent H3R8 methylation through the direct association to PRMT5.

Discussion

In this study, we showed that *LCE1* is a novel p53 target gene and that the LCE1F protein interacts with the arginine methyltransferase PRMT5. Through the interaction with PRMT5, LCE1F may suppress the PRMT5 methylation activity on arginine 8 of histone H3 (Figure 5*C*). Since PRMT5 is overexpressed in a wide range of human cancer and plays a critical role in tumorigenesis through the regulation of histone methylation, we unveiled a novel mechanism of tumor suppression mediated by p53.

The *LCE* gene cluster contains multiple conserved genes encoding stratum-corneum proteins [9,10,12,29]. Our study demonstrated that most of the members in the LCE1 group are transcriptionally regulated by the tumor suppressor p53 although the induction levels varied. Concordantly, *LCE1* genes were reported to be significantly upregulated in response to UVB irradiation of the skin cells [9]. UVR causes DNA damage, photoperoxidation of lipids, protein cross-linking, and isomerization of urocanic acid that lead to immunosuppression, photo-induced aging and cancer. p53 protein acts as a molecular sensor for the damages generated by UVR through mediating cell cycle arrest and apoptosis in damaged keratinocytes [30–34]. Taken together, cells may possess the function to express *LCE1* family genes through p53 activation to eliminate dangerous cells with DNA damages. Importantly, LCE1 family proteins show a high level of sequence similarity and LCE1F has more than 90% homology with other members (LCE1A-E; Table W6). In fact, as mentioned above, we confirmed the interaction between all of LCE1 proteins examined and PRMT5 (Figure W6). Since the expression of *LCE1* cluster genes is regulated by p53 as a whole (may not be all of the members), this protein family

members might play important roles to complementally or redundantly function as a tumor suppressor.

We identified PRMT5, a histone methyltransferase, as a key binding partner of the LCE1 proteins. Current progress of molecular medicine revealed that the enzymes relevant to histone methylation play critical roles in human carcinogenesis [28,35–45]. PRMT5 is one of the type II arginine methyltransferases, which catalyze the formation of symmetric dimethylation of arginine residues (SDMA) and regulates various cellular pathways [16,17,21]. A number of reports described the importance of this arginine methyltransferase in tumorigenesis [16,18,21,46–49]. Intriguingly, PRMT5 was reported to interact with BRG1- and hBRM-based hSWI/SNF chromatin remodelers and methylate arginine 8 on histone H3 [20]. The H3R8 methylation mediated by PRMT5-containing BRG1 and hBRM complexes directly repressed the expression of *suppressor of tumorigenicity 7* and *nonmetastatic 23*, tumor suppressor genes [20]. This line of pathway seems to be a key mechanism in the PRMT5-dependent tumorigenesis. Our data presented here have implied that LCE1, which is regulated by the tumor suppressor p53, negatively regulates H3R8 methylation mediated by PRMT5 (Figure 5, *A* and *B*), suggesting a new role of p53 in the regulation of histone modification. Since PRMT5 was reported to methylate p53 through the direct interaction and this methylation prevents p53-dependent apoptosis in cancer cells [50], we suggest an interesting feedback mechanism among p53, LCE1, and PRMT5.

Overall, our study elucidates a novel function of p53 as a tumor suppressor through the transcriptional regulation of the *LCE1* cluster genes. Further functional analysis may explore the importance of the LCE1 group proteins as tumor suppressors and the physiological relevance among p53 downstream genes.

Acknowledgment

We thank Koji Ueda and Yuki Funauchi for helpful discussion and their technical supports. This work was supported by Project for Development of Innovation Research on Cancer Therapeutics. Disclosure of potential conflicts of interest: Y.N. is a stockholder and a scientific advisor of Oncotherapy Science and also has research grants from Oncotherapy Science. No potential conflicts of interest were disclosed by the other authors.

Appendix A. Supplementary Materials

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.neo.2014.07.008.

References

[1] Soussi T (2011). TP53 mutations in human cancer: database reassessment and prospects for the next decade. *Adv Cancer Res* **110**, 107–139.

[2] Leroy B, Fournier JL, Ishioka C, Monti P, Inga A, Fronza G, and Soussi T (2013). The TP53 website: an integrative resource centre for the TP53 mutation database and TP53 mutant analysis. *Nucleic Acids Res* **41**, D962–D969.

[3] Beckerman R and Prives C (2010). Transcriptional regulation by p53. *Cold Spring Harb Perspect Biol* **2**, a000935.

[4] Tanikawa C, Furukawa Y, Yoshida N, Arakawa H, Nakamura Y, and Matsuda K (2009). XEDAR as a putative colorectal tumor suppressor that mediates p53-regulated anoikis pathway. *Oncogene* **28**, 3081–3092.

[5] Oda K, Arakawa H, Tanaka T, Matsuda K, Tanikawa C, Mori T, Nishimori H, Tamai K, Tokino T, and Nakamura Y, et al (2000). p53AIP1, a potential mediator of p53-dependent apoptosis, and its regulation by Ser-46-phosphorylated p53. *Cell* **102**, 849–862.

[6] Tanaka H, Arakawa H, Yamaguchi T, Shiraishi K, Fukuda S, Matsui K, Takei Y, and Nakamura Y (2000). A ribonucleotide reductase gene involved in a p53-dependent cell-cycle checkpoint for DNA damage. *Nature* **404**, 42–49.

[7] Tanikawa C, Matsuda K, Fukuda S, Nakamura Y, and Arakawa H (2003). p53RDL1 regulates p53-dependent apoptosis. *Nat Cell Biol* **5**, 216–223.

[8] Nakamura Y (2004). Isolation of p53-target genes and their functional analysis. *Cancer Sci* **95**, 7–11.

[9] Jackson B, Tilli CM, Hardman MJ, Avilion AA, MacLeod MC, Ashcroft GS, and Byrne C (2005). Late cornified envelope family in differentiating epithelia—response to calcium and ultraviolet irradiation. *J Invest Dermatol* **124**, 1062–1070.

[10] Marshall D, Hardman MJ, Nield KM, and Byrne C (2001). Differentially expressed late constituents of the epidermal cornified envelope. *Proc Natl Acad Sci U S A* **98**, 13031–13036.

[11] Mischke D, Korge BP, Marenholz I, Volz A, and Ziegler A (1996). Genes encoding structural proteins of epidermal cornification and S100 calcium-binding proteins form a gene complex ("epidermal differentiation complex") on human chromosome 1q21. *J Invest Dermatol* **106**, 989–992.

[12] Bergboer JG, Tjabringa GS, Kamsteeg M, van Vlijmen-Willems IM, Rodijk-Olthuis D, Jansen PA, Thuret JY, Narita M, Ishida-Yamamoto A, and Zeeuwen PL, et al (2011). Psoriasis risk genes of the late cornified envelope-3 group are distinctly expressed compared with genes of other LCE groups. *Am J Pathol* **178**, 1470–1477.

[13] Zhao XP and Elder JT (1997). Positional cloning of novel skin-specific genes from the human epidermal differentiation complex. *Genomics* **45**, 250–258.

[14] Bedford MT and Clarke SG (2009). Protein arginine methylation in mammals: who, what, and why. *Mol Cell* **33**, 1–13.

[15] Gkountela S, Li Z, Chin CJ, Lee SA, and Clark AT (2014). PRMT5 is required for human embryonic stem cell proliferation but not pluripotency. *Stem Cell Rev* **10**, 230–239.

[16] Nicholas C, Yang J, Peters SB, Bill MA, Baiocchi RA, Yan F, Sif S, Tae S, Gaudio E, and Wu X, et al (2013). PRMT5 is upregulated in malignant and metastatic melanoma and regulates expression of MITF and p27(Kip1.). *PLoS One* **8**, e74710.

[17] Powers MA, Fay MM, Factor RE, Welm AL, and Ullman KS (2011). Protein arginine methyltransferase 5 accelerates tumor growth by arginine methylation of the tumor suppressor programmed cell death 4. *Cancer Res* **71**, 5579–5587.

[18] Bao X, Zhao S, Liu T, Liu Y, Liu Y, and Yang X (2013). Overexpression of PRMT5 promotes tumor cell growth and is associated with poor disease prognosis in epithelial ovarian cancer. *J Histochem Cytochem* **61**, 206–217.

[19] Majumder S, Alinari L, Roy S, Miller T, Datta J, Sif S, Baiocchi R, and Jacob ST (2010). Methylation of histone H3 and H4 by PRMT5 regulates ribosomal RNA gene transcription. *J Cell Biochem* **109**, 553–563.

[20] Pal S, Vishwanath SN, Erdjument-Bromage H, Tempst P, and Sif S (2004). Human SWI/SNF-associated PRMT5 methylates histone H3 arginine 8 and negatively regulates expression of *ST7* and *NM23* tumor suppressor genes. *Mol Cell Biol* **24**, 9630–9645.

[21] Pal S, Baiocchi RA, Byrd JC, Grever MR, Jacob ST, and Sif S (2007). Low levels of miR-92b/96 induce PRMT5 translation and H3R8/H4R3 methylation in mantle cell lymphoma. *EMBO J* **26**, 3558–3569.

[22] Gu Z, Gao S, Zhang F, Wang Z, Ma W, Davis RE, and Wang Z (2012). Protein arginine methyltransferase 5 is essential for growth of lung cancer cells. *Biochem J* **446**, 235–241.

[23] Gu Z, Li Y, Lee P, Liu T, Wan C, and Wang Z (2012). Protein arginine methyltransferase 5 functions in opposite ways in the cytoplasm and nucleus of prostate cancer cells. *PLoS One* **7**, e44033.

[24] Tanikawa C, Nakagawa H, Furukawa Y, Nakamura Y, and Matsuda K (2012). CLCA2 as a p53-inducible senescence mediator. *Neoplasia* **14**, 141–149.

[25] Hamamoto R, Furukawa Y, Morita M, Iimura Y, Silva FP, Li M, Yagyu R, and Nakamura Y (2004). SMYD3 encodes a histone methyltransferase involved in the proliferation of cancer cells. *Nat Cell Biol* **6**, 731–740.

[26] Hamamoto R, Silva FP, Tsuge M, Nishidate T, Katagiri T, Nakamura Y, and Furukawa Y (2006). Enhanced SMYD3 expression is essential for the growth of breast cancer cells. *Cancer Sci* **97**, 113–118.

[27] Piao L, Nakagawa H, Ueda K, Chung S, Kashiwaya K, Eguchi H, Ohigashi H, Ishikawa O, Daigo Y, and Matsuda K, et al (2011). C12orf48, termed PARP-1 binding protein, enhances poly(ADP-ribose) polymerase-1 (PARP-1) activity and protects pancreatic cancer cells from DNA damage. *Genes Chromosomes Cancer* **50**, 13–24.

[28] Kogure M, Takawa M, Saloura V, Sone K, Piao L, Ueda K, Ibrahim R, Tsunoda T, Sugiyama M, and Atomi Y, et al (2013). The oncogenic polycomb histone methyltransferase EZH2 methylates lysine 120 on histone H2B and competes ubiquitination. *Neoplasia* **15**, 1251–1261.

[29] Tong L, Corrales RM, Chen Z, Villarreal AL, De Paiva CS, Beuerman R, Li DQ, and Pflugfelder SC (2006). Expression and regulation of cornified envelope proteins in human corneal epithelium. *Invest Ophthalmol Vis Sci* **47**, 1938–1946.

[30] Benjamin CL, Ullrich SE, Kripke ML, and Ananthaswamy HN (2008). p53 tumor suppressor gene: a critical molecular target for UV induction and prevention of skin cancer. *Photochem Photobiol* **84**, 55–62.

[31] Ziegler A, Jonason AS, Leffell DJ, Simon JA, Sharma HW, Kimmelman J, Remington L, Jacks T, and Brash DE (1994). Sunburn and p53 in the onset of skin cancer. *Nature* **372**, 773–776.

[32] Yamaizumi M and Sugano T (1994). U.v.-induced nuclear accumulation of p53 is evoked through DNA damage of actively transcribed genes independent of the cell cycle. *Oncogene* **9**, 2775–2784.

[33] Gujuluva CN, Baek JH, Shin KH, Cherrick HM, and Park NH (1994). Effect of UV-irradiation on cell cycle, viability and the expression of p53, gadd153 and gadd45 genes in normal and HPV-immortalized human oral keratinocytes. *Oncogene* **9**, 1819–1827.

[34] Maltzman W and Czyzyk L (1984). UV irradiation stimulates levels of p53 cellular tumor antigen in nontransformed mouse cells. *Mol Cell Biol* **4**, 1689–1694.

[35] Cho HS, Shimazu T, Toyokawa G, Daigo Y, Maehara Y, Hayami S, Ito A, Masuda K, Ikawa N, and Field HI, et al (2012). Enhanced HSP70 lysine methylation promotes proliferation of cancer cells through activation of Aurora kinase B. *Nat Commun* **3**, 1072.

[36] Hayami S, Kelly JD, Cho HS, Yoshimatsu M, Unoki M, Tsunoda T, Field HI, Neal DE, Yamaue H, and Ponder BA, et al (2011). Overexpression of LSD1 contributes to human carcinogenesis through chromatin regulation in various cancers. *Int J Cancer* **128**, 574–586.

[37] Hayami S, Yoshimatsu M, Veerakumarasivam A, Unoki M, Iwai Y, Tsunoda T, Field HI, Kelly JD, Neal DE, and Yamaue H, et al (2010). Overexpression of the JmjC histone demethylase KDM5B in human carcinogenesis: involvement in the proliferation of cancer cells through the E2F/RB pathway. *Mol Cancer* **9**, 59.

[38] Takawa M, Masuda K, Kunizaki M, Daigo Y, Takagi K, Iwai Y, Cho HS, Toyokawa G, Yamane Y, and Maejima K, et al (2011). Validation of the histone methyltransferase EZH2 as a therapeutic target for various types of human cancer and as a prognostic marker. *Cancer Sci* **102**, 1298–1305.

[39] Toyokawa G, Cho HS, Masuda K, Yamane Y, Yoshimatsu M, Hayami S, Takawa M, Iwai Y, Daigo Y, and Tsuchiya E, et al (2011). Histone lysine methyltransferase Wolf-Hirschhorn syndrome candidate 1 is involved in human carcinogenesis through regulation of the Wnt pathway. *Neoplasia* **13**, 887–898.

[40] Yoshimatsu M, Toyokawa G, Hayami S, Unoki M, Tsunoda T, Field HI, Kelly JD, Neal DE, Maehara Y, and Ponder BA, et al (2011). Dysregulation of PRMT1 and PRMT6, Type I arginine methyltransferases, is involved in various types of human cancers. *Int J Cancer* **128**, 562–573.

[41] Takawa M, Cho HS, Hayami S, Toyokawa G, Kogure M, Yamane Y, Iwai Y, Maejima K, Ueda K, and Masuda A, et al (2012). Histone lysine methyltransferase SETD8 promotes carcinogenesis by deregulating PCNA expression. *Cancer Res* **72**, 3217–3227.

[42] Cho HS, Kelly JD, Hayami S, Toyokawa G, Takawa M, Yoshimatsu M, Tsunoda T, Field HI, Neal DE, and Ponder BA, et al (2011). Enhanced expression of EHMT2 is involved in the proliferation of cancer cells through negative regulation of SIAH1. *Neoplasia* **13**, 676–684.

[43] Cho HS, Toyokawa G, Daigo Y, Hayami S, Masuda K, Ikawa N, Yamane Y, Maejima K, Tsunoda T, and Field HI, et al (2012). The JmjC domain-containing histone demethylase KDM3A is a positive regulator of the G_1/S transition in cancer cells via transcriptional regulation of the *HOXA1* gene. *Int J Cancer* **131**, E179-189.

[44] Toyokawa G, Cho HS, Iwai Y, Yoshimatsu M, Takawa M, Hayami S, Maejima K, Shimizu N, Tanaka H, and Tsunoda T, et al (2011). The histone demethylase JMJD2B plays an essential role in human carcinogenesis through positive regulation of cyclin-dependent kinase 6. *Cancer Prev Res (Phila)* **4**, 2051–2061.

[45] Kogure M, Takawa M, Cho HS, Toyokawa G, Hayashi K, Tsunoda T, Kobayashi T, Daigo Y, Sugiyama M, and Atomi Y, et al (2013). Deregulation of the histone demethylase JMJD2A is involved in human carcinogenesis through regulation of the G_1/S transition. *Cancer Lett* **336**, 76–84.

[46] Wang L, Pal S, and Sif S (2008). Protein arginine methyltransferase 5 suppresses the transcription of the RB family of tumor suppressors in leukemia and lymphoma cells. *Mol Cell Biol* **28**, 6262–6277.

[47] Cho EC, Zheng S, Munro S, Liu G, Carr SM, Moehlenbrink J, Lu YC, Stimson L, Khan O, and Konietzny R, et al (2012). Arginine methylation controls growth regulation by E2F-1. *EMBO J* **31**, 1785–1797.

[48] Yan F, Alinari L, Lustberg ME, Katherine Martin L, Cordero-Nieves HM, Banasavadi-Siddegowda Y, Virk S, Barnholtz-Sloan J, Bell EH, and Wojton J, et al (2014). Genetic validation of the protein arginine methyltransferase PRMT5 as a candidate therapeutic target in glioblastoma. *Cancer Res* **74**, 1752–1765.

[49] Shilo K, Wu X, Sharma S, Welliver M, Duan W, Villalona-Calero M, Fukuoka J, Sif S, Baiocchi R, and Hitchcock CL, et al (2013). Cellular localization of protein arginine methyltransferase-5 correlates with grade of lung tumors. *Diagn Pathol* **8**, 201.

[50] Jansson M, Durant ST, Cho EC, Sheahan S, Edelmann M, Kessler B, and La Thangue NB (2008). Arginine methylation regulates the p53 response. *Nat Cell Biol* **10**, 1431–1439.

Human Molecular Genetics, 2015, Vol. 24, No. 4 *1177–1184*
doi:10.1093/hmg/ddu512
Advance Access published on October 3, 2014

Genome-wide association study identified SNP on 15q24 associated with bladder cancer risk in Japanese population

Koichi Matsuda[1], Atsushi Takahashi[2], Candace D. Middlebrooks[3], Wataru Obara[4], Yasutomo Nasu[5], Keiji Inoue[6], Kenji Tamura[6], Ichiro Yamasaki[6], Yoshio Naya[7], Chizu Tanikawa[1], Ri Cui[1], Jonine D. Figueroa[3], Debra T. Silverman[3], Nathaniel Rothman[3], Mikio Namiki[8], Yoshihiko Tomita[9], Hiroyuki Nishiyama[10], Kenjiro Kohri[11], Takashi Deguchi[12], Masayuki Nakagawa[13], Masayoshi Yokoyama[14], Tsuneharu Miki[7], Hiromi Kumon[5], Tomoaki Fujioka[4], Ludmila Prokunina-Olsson[3], Michiaki Kubo[2], Yusuke Nakamura[15,16,*] and Taro Shuin[6]

[1]Laboratory of Molecular Medicine, Human Genome Center, Institute of Medical Science, The University of Tokyo, Tokyo, Japan, [2]Center for Integrative Medical Science, The Institute of Physical and Chemical Research (RIKEN), Kanagawa, Japan, [3]Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, USA, [4]Department of Urology, Iwate Medical University School of Medicine, Iwate, Japan, [5]Department of Urology, Okayama University Graduate School of Medicine, Dentistry and Pharmaceutical Sciences, Okayama, Japan, [6]Department of Urology, School of Medicine, Kochi University, Koichi, Japan, [7]Department of Urology, Graduate School of Medical Science, Kyoto Prefectural University of Medicine, Kyoto, Japan, [8]Department of Integrative Cancer Therapy and Urology, Kanazawa University Graduate School of Medical Science, Ishikawa, Japan, [9]Department of Urology, Yamagata University Faculty of Medicine, Yamagata, Japan, [10]Department of Urology, Faculty of Medicine, University of Tsukuba, Ibaragi, Japan, [11]Department of Nephro-urology, Nagoya City University Graduate School of Medical Sciences, Aichi, Japan, [12]Department of Urology, Graduate School of Medicine, Gifu University, Gifu, Japan, [13]Department of Urology, Graduate School of Medical and Dental Sciences, Kagoshima University, Kagoshima, Japan, [14]Department of Urology, Ehime University Graduate School of Medicine, Ehime, Japan, [15]Department of Medicine and [16]Department of Surgery, Center for Personalized Therapeutics, The University of Chicago, Chicago, IL, USA

Received June 30, 2014; Revised September 18, 2014; Accepted September 29, 2014

Through genome-wide association analysis and an independent replication study using a total of 1131 bladder cancer cases and 12 558 non-cancer controls of Japanese populations, we identified a susceptibility locus on chromosome 15q24. SNP rs11543198 was associated with bladder cancer risk with odds ratio (OR) of 1.41 and *P*-value of 4.03×10^{-9}. Subgroup analysis revealed rs11543198 to have a stronger effect in male smokers with OR of 1.66. SNP rs8041357, which is in complete linkage disequilibrium ($r^2 = 1$) with rs11543198, was also associated with bladder cancer risk in Europeans ($P = 0.045$ for an additive and $P = 0.025$ for a recessive model), despite much lower minor allele frequency in Europeans (3.7%) compared with the Japanese (22.2%). Imputational analysis in this region suggested CYP1A2, which metabolizes tobacco-derived carcinogen, as a causative candidate gene. We also confirmed the association of previously reported loci, namely *SLC14A1*, *APOBEC3A*, *PSCA* and *MYC*, with bladder cancer. Our finding implies the crucial roles of genetic variations on the chemically associated development of bladder cancer.

*To whom correspondence should be addressed at: Section of Hematology/Oncology, Department of Medicine, The University of Chicago, Knapp Center for Biomedical Discovery, Room 6130, MC2115, 900 E. 57th Street, Chicago, IL 60637, USA. Tel: +1 7738341405; Fax: +1 7737029268; Email: ynakamura@bsd.uchicago.edu

INTRODUCTION

Bladder cancer is one of the most frequent cancers, which causes ~150 000 death per year in the world (1). Both environmental and genetic factors are involved in the development of bladder cancer, and tobacco smoking is known to be the most important factor to increase the risk of bladder cancer; current or former smokers have a 2- to 6-fold higher risk than never-smokers (2,3). In addition, occupational exposures to industrial chemicals (4–6), arsenic contamination in drinking water (7) and infectious diseases (8) also increase the bladder cancer risk. Bladder cancer incidence in males is nearly 3-fold higher than that in females (3), probably due to the higher prevalence of tobacco smoking and occupational exposure in males. On the other hand, familial aggregation of bladder cancer has also been reported (9,10), suggesting the importance of genetic factors in bladder cancer development.

NAT2 and GSTM1 are involved in the detoxification of carcinogen (11), and an *NAT2* slow-metabolizer genotype and a *GSTM1* null genotype were indicated their association with an increased risk of bladder cancer (12,13). In addition, recent genome-wide association studies (GWASs) in European populations have identified multiple genetic factors associated with bladder cancer (14–18). However, no GWAS has yet been conducted in Asian populations. So, to identify genetic factors associated with the risk of bladder cancer, we conducted a genome-wide association study in the Japanese population.

RESULTS

In this study, we performed a two-stage GWAS approach. A study design and the information of subjects used in this analysis are shown in Supplementary Material, Fig. S1 and Table S1. Case samples in the screening and replication stages were recruited from a collaborative network consisting of 11 university-affiliated hospitals in Japan. The genotyping information for non-cancer controls was obtained from the BioBank Japan project. In the screening stage, a total of 539 bladder cancer cases were genotyped by using Human OmniExpress Exome chip. After an initial standard quality control procedure, we obtained genotyping results of 554 389 SNP loci in 531 cases and compared them with those of 5581 controls. Principle component analysis indicated that all subjects used in GWAS were of the Asian ancestry (Supplementary Material, Fig. S2). We performed statistical analysis using a Cochran–Armitage trend test and obtained the genomic inflation factor lambda to be 1.0493, indicating the low possibility of population stratification (Supplementary Material, Fig. S3). From the association analysis, 82 SNPs in 45 distinct genomic regions were indicated suggestive associations ($P < 5 \times 10^{-5}$). Among them, three SNPs on chromosome 6p21 passed the genome-wide significance threshold ($P = 1.76–3.39 \times 10^{-8}$, Fig. 1 and Supplementary Material, Table S2).

To validate these associations, we selected 64 SNPs through linkage disequilibrium analysis with the criteria of pair-wise r^2 of <0.8. In the replication analysis, we attempted genotyping of a new set of 592 cases and 6964 controls for these 64 SNPs and successfully obtained the genotype information at 59 SNP loci. Among these 59 SNPs, the only one SNP, rs11543198, on chromosome 15q24 revealed significant association with a P-value of 1.22×10^{-4} ($P < 8.47 \times 10^{-4} = 0.05/59$, Table 1 and Supplementary Material, Table S3). The combined analysis of the GWAS and replication data indicated SNP rs11543198 to have the genome-wide significance of the association ($P_{meta} = 4.03 \times 10^{-9}$, odds ratio (OR) = 1.41, Table 1).

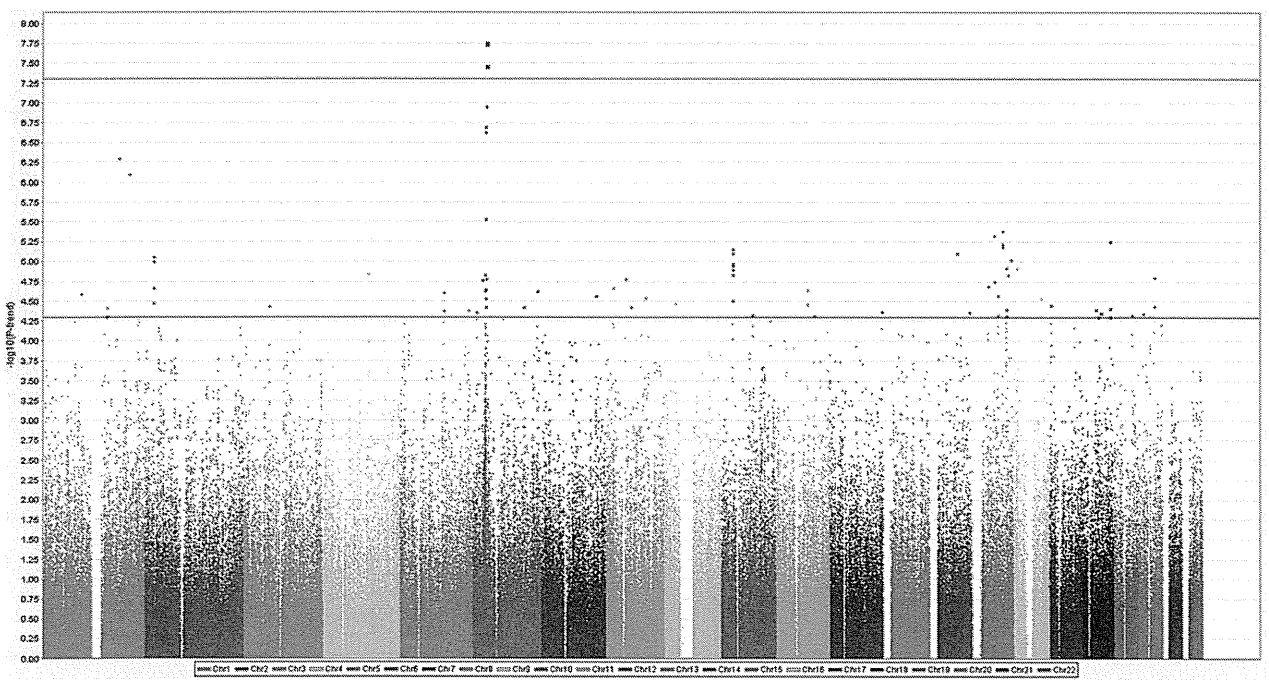

Figure 1. Manhattan plot showing the genome-wide P-values of association. The P-values were calculated by Cochran–Armitage trend test. The y-axis represents the $-\log_{10} P$-values of 554 389 SNPs, and x-axis shows their chromosomal positions. The horizontal blue line shows the threshold of $P \leq 5 \times 10^{-5}$ for selecting 84 SNPs for replication analysis.

For validation, we examined the association of this locus with bladder cancer risk in 3508 bladder cancer cases and 5101 controls of European ancestry (17). As rs11543198 was not genotyped in the European GWAS, we used SNP rs8041357, which is in complete linkage disequilibrium ($D' = 1, r^2 = 1$) with rs11543198 in Asians and Europeans based on the 1000 Genomes Project data (19). In agreement with the 1000 Genomes data, the frequency of the minor allele G was 22.2% in the Japanese controls, compared with 3.7% in Europeans (Supplementary Material, Table S4). Owing to much lower allele frequency, the association in Europeans was weaker but the direction and magnitude of association were similar in both populations.

As gender and smoking status are known to be important risk factors for bladder cancer, we conducted subgroup analysis. As a result, SNP rs11543198 showed a stronger effect among males and smokers (OR of 1.58 and 1.65, respectively) (Table 2). As males are more likely to use tobacco, we also conducted subgroup analysis stratified by both gender and smoking status. SNP rs11543198 exhibited stronger effect among males (OR of 1.66 for smoker and 1.49 for never-smoker) than females (OR of 1.29 for smoker and 1.12 for never-smoker), irrespective of smoking status (Table 2). Although we did not observe significant heterogeneity among each group, our findings suggested possible functional interaction of this genetic factor with tobacco smoking, occupational exposure to carcinogen or some hormonal factors.

We also conducted imputation analysis around this SNP region of chromosome 15q24 in our GWAS cohort and found that 29 SNPs within a 341-kb region including *SEMA7A-UBL7-ARID3B-CLK3-EDC3-CYP1A1-CYP1A2-CSK* genes showed similar association as rs11543198 ($P < 1 \times 10^{-5}$, Fig. 2 and Supplementary Material, Table S5). Among these 29 SNPs, rs1350194 in the 3′ flanking region of *CYP1A2* revealed the strongest association ($P = 5.80 \times 10^{-7}$ with OR of 1.55, Supplementary Material, Table S5). Moreover, an A allele of SNP rs2069514 (CYP1A2*1C), which is known to have the lower CYP1A2 activity (20), was indicated to reduce the disease risk ($P = 1.42 \times 10^{-6}$ with OR = 0.66, Supplementary Material, Table S5). These two SNPs (rs1350194 and rs2069514) were in modest linkage disequilibrium with rs11543198 ($r^2 =$ 0.215 and 0.468, respectively). As CYP1A2 metabolizes some polycyclic aromatic hydrocarbons (PAHs) to carcinogenic intermediates (21,22), the higher CYP1A2 activity generates more carcinogenic metabolites and then is likely to enhance the risk of tobacco-related cancers. Therefore, our result suggests that *CYP1A2* on chromosome 15q24 may be a causative gene candidate to increase the risk of bladder cancer.

In addition, we also examined previously reported susceptible loci in our GWAS cohort (14–18). We successfully genotyped 10 SNPs as shown in Table 3 and confirmed the significant association of the loci in the *SLC14A1*, *APOBEC3A*, *PSCA* and *MYC* genes with *P*-values of <0.05 (Table 3). Two SNPs on *TACC3* and *UGT1A* loci revealed weak trends, but the remaining four loci were not replicated. These results may imply the genetic heterogeneity between European and Asian populations.

Table 1. Summary of GWAS and replication analyses

SNP	Case			Control			P^{a}_{trend}	OR[b]	(95% CI)	P^{c}_{het}
	AA	AG	GG	AA	AG	GG				
rs11543198										
GWAS	16	140	375	283	1913	3385	6.22×10^{-6}	1.48	(1.25–1.75)	0.48
Replication	21	158	413	316	2382	4265	1.22×10^{-4}	1.36	(1.16–1.59)	
Combined[d]							4.03×10^{-9}	1.41	(1.26–1.59)	

A total of 1131 bladder cancer cases and 12 545 controls were analyzed.
[a]*P*-value obtained from Cochrane–Armitage trend test.
[b]Odds ratios (OR) and confidence interval (CI) are calculated using the non-susceptible allele (A) as reference.
[c]The *P*-values of heterogeneities across two stages were examined by using the Breslow–Day test.
[d]Meta-analysis of two stage was conducted by using a Cochran–Mantel–Haenszel test.

Table 2. Subgroup analysis stratified by gender and smoking status

SNP	Case			Control			Ptrend[a]	OR (95% CI) [b]	Phet[c]
	AA	AG	GG	AA	AG	GG			
rs11543198									
Male	16	148	427	309	2220	3876	3.01×10^{-8}	1.58 (1.34–1.87)	0.07
Female	6	45	97	229	1697	3049	0.32	1.16 (0.86–1.55)	
Smoker	11	95	294	295	2101	3714	8.63×10^{-7}	1.65 (1.34–1.87)	0.10
Nonsmoker	11	98	230	243	1816	3211	0.01	1.30 (0.86–1.55)	
Male/smoker	10	90	281	218	1579	2816	1.63×10^{-6}	1.66 (1.35–2.04)	0.15
Male/nonsmoker	6	58	146	91	641	1060	3.24×10^{-3}	1.49 (1.14–1.95)	
Female/smoker	1	5	13	77	522	898	0.54	1.29 (0.57–2.95)	
Female/nonsmoker	5	40	84	152	1175	2151	0.47	1.12 (0.82–1.54)	

A total of 739 bladder cancer cases and 11 380 controls were analyzed.
[a]*P*-value obtained from Cochrane–Armitage trend test.
[b]ORs and CI are calculated using the non-susceptible allele (A) as reference.
[c]The *P*-value of heterogeneities across four subgroups was examined by using the Breslow–Day test.

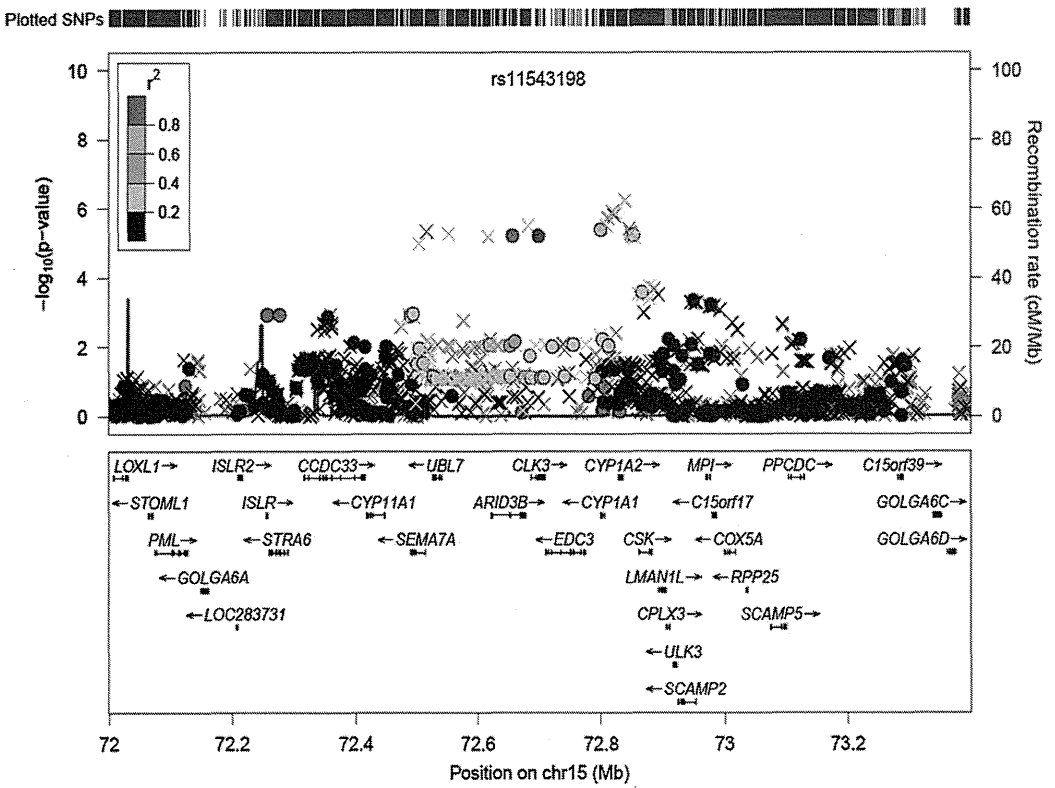

Figure 2. Regional association plots around rs11543198 on 15q24 (1.4 Mb). Upper panel: *P*-values of genotyped SNPs (circle) and imputed SNPs (cross) are plotted (as $-\log_{10}$ *P*-value) against their physical position on chromosome 15 (NCBI Build 36). SNP rs11543198 is represented by purple circle. The genetic recombination rates estimated from 1000 Genomes samples (JPT + CHB) are shown with a blue line. SNP's color indicates LD with rs11543198 according to a scale from $r^2 = 0$ to 1 based on pair-wise r^2-values from HapMap JPT. Lower panel; gene annotations from the University of California Santa Cruz genome browser.

DISCUSSION

Here, we reported the result of GWAS analysis using a total of 1131 Japanese bladder cancer cases and 12 558 controls. Our data indicated that SNP rs11543198 on chromosome 15q24 was significantly associated with bladder cancer risk. SNP rs11543198 is located within the *CLK3* gene, which encodes a serine/threonine protein kinase. CLK3 regulates localization of SR family of splicing factors; however, its association with human carcinogenesis was not reported so far. Although SNP rs11543198 alters amino acid sequence of minor isoforms of CLK3 (X1, X2, X5 and X6), this SNP does not affect amino acid sequence of catalytic domain. Therefore, SNP rs11543198 is not likely to be a causative variation.

Imputation analysis of 29 SNPs in the 341-kb region including this SNP locus suggested an SNP near the *CYP1A2* gene to be a causative candidate because of its stronger association than the marker SNP rs11543198 as well as its biological relevance. Interestingly, *CYP1A2* belongs to the cytochrome P450 superfamily, members of which encode monooxygenases involved in metabolism of various substrates including drugs and play essential roles in the synthesis of cholesterol, steroids and other lipids (23,24). CYP1A2, which is highly expressed in the liver, is activated by the exposure to PAHs (25,26) and metabolizes them. It is well know that CYP1A2 metabolizes heterocyclic aromatic amines contained in tobacco smoke (27,28) and generates carcinogenic intermediates (21,22). In addition, genetic variations in the *CYP1A2* locus are known to affect its enzymatic activity (29,30). CYP1A2 activity exhibits a significant degree of inter-individual diversity owing to both environmental and genetic factors. Hence, associations of *CYP1A2* variations with various cancers including bladder, breast, colorectal and lung cancers have been repeatedly investigated, but the results are controversial (31–34). Our first GWAS in the Asian population revealed that *CYP1A2* locus is significantly associated with bladder cancer. Moreover, SNP rs11543198 showed the stronger effect on smokers compared with never-smokers in both males and females. Considering the role of CYP1A2 in the metabolism of tobacco-derived carcinogen, our finding suggested the interesting gene-environmental interaction on the development of bladder cancer.

Among genes associated with bladder cancer in previously reported GWAS, our analyses revealed that SNPs in the *SLC14A1*, *APOBEC3A-CBX6*, *PSCA* and *MYC* loci were associated with bladder cancer risk in the Japanese population ($P < 0.05$). Association of these four SNPs with bladder cancer was also reported in the Chinese population (35,36). Thus, these variations are common bladder cancer loci among the European and Asian populations.

SLC14A1 functions as a urea transporter and regulates urine concentration and body-fluid balance. Although the function of SNP rs17674580 in SLC14A1 was not fully elucidated, our previous analysis revealed that *SLC14A1-SLC14A2* locus was associated with blood urea nitrogen level (37). Thus, this genetic

Table 3. Result of previously reported SNPs

SNP	Chr.	Chr. location	Gene	Relative Pos.	Case			Control			P^{a}_{trend}	OR[b]	RAF[b]		Allele[b]	
					11	12	22	11	12	22			Case	Control	1	2
rs17674580	18	43309911	*SLC14A1*	0	5	86	439	28	579	4619	1.8×10^{-4}	1.54	0.091	0.061	A	G
rs1014971	22	39332623	*APOBEC3A*	−20904	117	256	157	920	2545	1761	0.0074	1.19	0.462	0.420	T	C
rs2294008	8	143761931	*PSCA*	0	241	228	61	2079	2416	730	0.0092	1.20	0.670	0.629	T	C
rs9642880	8	128718068	*MYC*	−30247	63	226	241	491	2165	2569	0.039	1.15	0.332	0.301	T	G
rs798766	4	1734239	*TACC3*	0	15	174	341	154	1461	3611	0.056	1.17	0.192	0.169	A	G
rs11892031	2	234565283	*UGT1A8*	0	526	4	0	5143	83	0	0.13	2.11	0.996	0.992	T	G
rs1495741	8	18272881	*NAT2*	14158	55	232	244	559	2394	2628	0.62	1.03	0.322	0.315	A	G
rs401681	5	1322087	*CLPTM1L*	0	229	245	56	2359	2288	579	0.64	0.97	0.663	0.670	G	A
rs710521	3	189645933	*LEPREL1*	28584	298	203	29	2948	1996	282	0.92	0.99	0.754	0.755	T	C
rs8102137	19	30296853	*CCNE1*	−6048	2	89	439	37	834	4355	0.92	0.99	0.912	0.913	G	A

We analyzed 539 bladder cancer cases and 5581 controls at GWAS stage. Chr., chromosome; Pos., position in the NCBI Build 36.3.
[a]*P*-values were obtained from Cochrane-Armitage trend test.
[b]Allele 1 is a risk allele in the previous GWAS studies. OR was calculated using allele 2 as a reference.

variation would be associated with the kidney function and urine concentration and subsequently affect on the bladder cancer risk. SNP rs1014971 on chromosome 22q13.1 is located ~25 kb centromeric to *APOBEC3A*. As *APOBEC3A* was expressed only in peripheral blood leukocytes and spleen (38), the role of this variation in the carcinogenesis of bladder cancer is not yet clarified.

SNPs rs2294008 and rs9642880 are located on chromosome 8q24, and these loci are not in the same linkage disequilibrium (LD) block. SNP rs9642880, which is located in the LD block adjacent to *MYC*, a well-known oncogene, was shown to be associated with *MYC* mRNA and protein expression (39). SNP rs2294008 is associated with cell surface localization and higher expression level of PSCA (40,41). PSCA, which is up-regulated in various tumors including bladder cancer, was shown to be involved in cell renewal and proliferation (42). Therefore, the association of these SNPs with bladder cancer risk can be explained by the growth promoting effect of MYC and cell surface PSCA.

Taken together, we here demonstrated the important roles of genetic factors on the development of bladder cancer. Particularly, the stronger effect of rs11543198 on smokers is very important because it implies a possibility that a small change in the life style, quitting or avoiding smoking, may contribute to the improvement of individuals carrying the risk genotype. Although further prospective analysis is necessary, we hope that our finding would further emphasize the significance of tobacco control.

MATERIALS AND METHODS

Study population

Characteristics of each cohort are shown in Supplementary Material, Table S1. In this study, we conducted GWAS and replication analyses using a total of 1131 bladder cancer cases and 12 558 controls. Case samples in GWAS and replication were obtained from a collaboration network consisting of Iwate Medical University, Okayama University, Kochi Medical School, Kyoto Prefectural University of Medicine, Kanazawa University, Yamagata University, University of Tsukuba, Nagoya City University, Gifu University, Kagoshima University and Ehime University. Control samples in GWAS and replication consisted of healthy volunteers ($n = 1919$) and subjects with other diseases ($n = 10\,639$, cerebral aneurysm, chronic obstructive pulmonary disease, glaucoma, nephrolithiasis, nephrotic syndrome, epilepsy, atopic dermatitis and Grave's disease) obtained from Biobank Japan Project supported by the Ministry of Education, Culture, Sports, Science and Technology of Japan. In the BioBank Japan Project, DNA and serum of patients were collected through a collaborating network of 66 hospitals throughout Japan. More than 200 000 individuals with 47 common diseases, irrespective of prior treatment, were enrolled in this project from 2003. A list of participating hospitals is provided at the BioBank Japan website (http://www.biobankjp.org/english/index.html). Subjects with a history of any cancers were excluded from controls. Smoking information of both cases and controls was obtained by oral interview. This project was approved by the ethical committees at each institute.

SNP genotyping

Genomic DNA was extracted from peripheral blood leukocytes or normal tissue using a standard method. In GWAS, we genotyped 539 bladder cancer cases and 5769 non-cancer controls (cerebral aneurysm, primary sclerosing cholangitis, chronic obstructive pulmonary disease, glaucoma and healthy volunteers) using Human OmniExpress and HumanExome (Supplementary Material, Fig. S1). A total of 951 117 SNPs including 925 436 autosomal SNPs were genotyped by both platforms. Among the 925 436 autosomal SNPs, 223 273 SNPs were monomorphic in our case–controls sample set. We excluded the following samples from analysis: closely related samples, gender mismatch and subjects whose ancestries were estimated to be distinct from East-Asian populations using principle component analysis. Then, we applied SNP quality control as follows: call rate ≥ 0.99 in case and control samples, minor allele frequency (MAF) of <0.01 and P-value of Hardy–Weinberg Equilibrium in control group $\geq 1 \times 10^{-6}$. Consequently, 554 389 SNPs on autosomal chromosomes passed the quality control filters. Among 82 SNPs showing $P < 5 \times 10^{-5}$, we selected 64 SNPs by linkage disequilibrium analysis with the criteria of pair-wise r^2 of >0.8. In the replication analysis, we genotyped 64 SNPs in 592 bladder cancer cases by using multiplex PCR-based Invader assay (Third Wave Technologies). A total of 6964 non-cancer controls (nephrolithiasis, epilepsy, atrophic dermatitis and Grave's disease) from Biobank Japan were genotyped by Human OmniExpress exome beadchip.

Statistical analysis

The association of SNPs with bladder cancer risk was tested by Cochran–Armitage trend test. To characterize population structure in the GWAS cohort, we performed principal component analysis (43). In the GWAS, the genetic inflation factor λ was derived by P-values obtained by Cochran–Armitage trend test for all the tested SNPs. The quantile–quantile plot was drawn using R program. The ORs were calculated using the non-susceptible allele as references, unless it was stated otherwise. The combined analysis of GWAS and replication stage was verified by conducting the Mantel–Haenszel method. Heterogeneity across two stages was examined by using the Breslow–Day test (44). We considered $P = 5 \times 10^{-8}$ (GWAS and meta-analysis) and 8.33×10^{-4} (0.05/59, replication analysis) as the significant threshold after Bonferroni correction for multiple testing.

Replication analysis in Europeans

Genotypes of rs8041357 in 3508 bladder cancer cases and 5101 controls of European ancestry were extracted from the previous GWAS data set (17). Each participating study obtained informed consent from study participants and approval from its respective Institutional Review Board for this study. Genome-wide genotyping was conducted using HumanHap 1M or HumanHap610-Quad BeadCHIP (Supplementary Material, Table S1). Genotypes of rs8041357 were in Hardy–Weinberg equilibrium in controls ($P > 0.05$). Association was evaluated for an additive model adjusting for age, sex, smoking status (ever/never), study sites and significant EVs as in the original GWAS (17). Owing to low allele frequency for allele G (3.7%), we also analyzed the results using a recessive genetic model, evaluating the effect in risk homozygotes (AA) versus a combined group of rare homozygotes (GG) and heterozygotes (AG), adjusting for the same covariates.

Imputation analysis

Imputation of ungenotyped SNPs was conducted by MACH (45) and minimac (46) using data of JPT/CHS/CHD subjects from 1000 genome project phase 1 (release 16 March 2012) as a reference. We excluded SNPs that met the following criteria: MAF < 0.01, Hardy–Weinberg Equilibrium P-value $< 1 \times 10^{-6}$ or large allele frequency difference between reference panel and GWAS (>0.16).

Software

For general statistical analysis, we employed R statistical environment version 2.9.1 (cran.r-project.org). The Haploview software version 4.2 (47) was used to draw Manhattan plot. Primer3-webv0.3.0 (http://frodo.wi.mit.edu) web tool was used to design primers. MACH (45) (http://www.sph.umich.edu/csg/abecasis/MACH/), minimac (46) (http://genome.sph.umich.edu/wiki/Minimac) and mach2dat (http://genome.sph.umich.edu/wiki/Mach2dat:_Association_with_MACH_output) were used for imputation analysis.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at *HMG* online.

ACKNOWLEDGEMENTS

We thank all the patients and the members of the Rotary Club of Osaka-Midosuji District 2660 Rotary International in Japan, who donated their DNA for this work. We also thank Satomi Takahashi (the University of Tokyo) and the technical staff of the Laboratory for Genotyping Development, Center for Integrative Medical Science, RIKEN for their technical support.

Conflict of Interest statement. None declared.

FUNDING

This work was conducted as a part of the BioBank Japan Project that was supported by the Ministry of Education, Culture, Sports, Science and Technology of the Japanese government. The work was supported in part by the Intramural Research Program (IRP) of the National Cancer Institute of the US National Institutes of Health.

REFERENCES

1. Jemal, A., Bray, F., Center, M.M., Ferlay, J., Ward, E. and Forman, D. (2011) Global cancer statistics. *CA Cancer J. Clin.*, **61**, 69–90.
2. Freedman, N.D., Silverman, D.T., Hollenbeck, A.R., Schatzkin, A. and Abnet, C.C. (2011) Association between smoking and risk of bladder cancer among men and women. *JAMA*, **306**, 737–745.
3. Castelao, J.E., Yuan, J.M., Skipper, P.L., Tannenbaum, S.R., Gago-Dominguez, M., Crowder, J.S., Ross, R.K. and Yu, M.C. (2001)

Gender- and smoking-related bladder cancer risk. *J. Natl Cancer Inst.*, **93**, 538–545.
4. Colt, J.S., Baris, D., Stewart, P., Schned, A.R., Heaney, J.A., Mott, L.A., Silverman, D. and Karagas, M. (2004) Occupation and bladder cancer risk in a population-based case-control study in New Hampshire. *Cancer Causes Control*, **15**, 759–769.
5. Kogevinas, M., t Mannetje, A., Cordier, S., Ranft, U., Gonzalez, C.A., Vineis, P., Chang-Claude, J., Lynge, E., Wahrendorf, J., Tzonou, A. *et al.* (2003) Occupation and bladder cancer among men in Western Europe. *Cancer Causes Control*, **14**, 907–914.
6. Goebell, P.J., Villanueva, C.M., Rettenmeier, A.W., Rubben, H. and Kogevinas, M. (2004) Environmental exposure, chlorinated drinking water, and bladder cancer. *World J. Urol.*, **21**, 424–432.
7. Chen, C.J., Kuo, T.L. and Wu, M.M. (1988) Arsenic and cancers. *Lancet*, **1**, 414–415.
8. Badawi, A.F., Mostafa, M.H., Probert, A. and O'Connor, P.J. (1995) Role of schistosomiasis in human bladder cancer: evidence of association, aetiological factors, and basic mechanisms of carcinogenesis. *Eur. J. Cancer Prev.*, **4**, 45–59.
9. Aben, K.K., Witjes, J.A., Schoenberg, M.P., Hulsbergen-van de Kaa, C., Verbeek, A.L. and Kiemeney, L.A. (2002) Familial aggregation of urothelial cell carcinoma. *Int. J. Cancer*, **98**, 274–278.
10. Murta-Nascimento, C., Silverman, D.T., Kogevinas, M., García-Closas, M., Rothman, N., Tardón, A., García-Closas, R., Serra, C., Carrato, A. and Villanueva, C. (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.
11. Hein, D.W. (2002) Molecular genetics and function of NAT1 and NAT2: role in aromatic amine metabolism and carcinogenesis. *Mutat. Res.*, **506**, 65–77.
12. Risch, A., Wallace, D.M., Bathers, S. and Sim, E. (1995) Slow N-acetylation genotype is a susceptibility factor in occupational and smoking related bladder cancer. *Hum. Mol. Genet.*, **4**, 231–236.
13. García-Closas, M., Malats, N., Silverman, D., Dosemeci, M., Kogevinas, M., Hein, D.W., Tardón, A., Serra, C., Carrato, A. and García-Closas, R. (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.
14. 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.
15. 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.
16. 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.
17. 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.
18. 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.
19. Consortium, G.P. (2010) A map of human genome variation from population-scale sequencing. *Nature*, **467**, 1061–1073.
20. Nakajima, M., Yokoi, T., Mizutani, M., Kinoshita, M., Funayama, M. and Kamataki, T. (1999) Genetic polymorphism in the 5′-flanking region of human CYP1A2 gene: effect on the CYP1A2 inducibility in humans. *J. Biochem.*, **125**, 803–808.
21. Lewtas, J., Walsh, D., Williams, R. and Dobiáš, L. (1997) Air pollution exposure–DNA adduct dosimetry in humans and rodents: evidence for non-linearity at high doses. *Mutat. Res.*, **378**, 51–63.
22. Boobis, A.R., Lynch, A.M., Murray, S., de la Torre, R., Solans, A., Farré, M., Segura, J., Gooderham, N.J. and Davies, D.S. (1994) CYP1A2-catalyzed conversion of dietary heterocyclic amines to their proximate carcinogens is their major route of metabolism in humans. *Cancer Res.*, **54**, 89–94.
23. Guengerich, F.P. (2008) Cytochrome p450 and chemical toxicology. *Chem. Res. Toxicol.*, **21**, 70–83.
24. Guengerich, F.P. (1994) Catalytic selectivity of human cytochrome P450 enzymes: relevance to drug metabolism and toxicity. *Toxicol. Lett.*, **70**, 133–138.
25. DeCaprio, A.P. (1997) Biomarkers: coming of age for environmental health and risk assessment. *Environ. Sci. Technol.*, **31**, 1837–1848.
26. Elovaara, E., Mikkola, J., Stockmann-Juvala, H., Luukkanen, L., Keski-Hynnilä, H., Kostiainen, R., Pasanen, M., Pelkonen, O. and Vainio, H. (2007) Polycyclic aromatic hydrocarbon (PAH) metabolizing enzyme activities in human lung, and their inducibility by exposure to naphthalene, phenanthrene, pyrene, chrysene, and benzo (a) pyrene as shown in the rat lung and liver. *Arch. Toxicol.*, **81**, 169–182.
27. Manabe, S., Tohyama, K., Wada, O. and Aramaki, T. (1991) Detection of a carcinogen, 2-amino-1-methyl-6-phenylimidazo [4, 5-b] pyridine (PhIP), in cigarette smoke condensate. *Carcinogenesis*, **12**, 1945–1947.
28. Wogan, G.N., Hecht, S.S., Felton, J.S., Conney, A.H. and Loeb, L.A. (2004) Environmental and chemical carcinogenesis. In: *Seminars in Cancer Biology*. Elsevier, Vol. 14, pp. 473–486.
29. Nakajima, M., Yokoi, T., Mizutani, M., Shin, S., Kadlubar, F.F. and Kamataki, T. (1994) Phenotyping of CYP1A2 in Japanese population by analysis of caffeine urinary metabolites: absence of mutation prescribing the phenotype in the CYP1A2 gene. *Cancer Epidemiol. Biomarkers Prev.*, **3**, 413–421.
30. Chida, M., Yokoi, T., Fukui, T., Kinoshita, M., Yokota, J. and Kamataki, T. (1999) Detection of three genetic polymorphisms in the 5′-flanking region and intron 1 of human CYP1A2 in the Japanese population. *Jpn. J. Cancer Res.*, **90**, 899–902.
31. Tian, Z., Li, Y.L., Zhao, L. and Zhang, C.L. (2013) Role of CYP1A2 1F polymorphism in cancer risk: evidence from a meta-analysis of 46 case-control studies. *Gene*, **524**, 168–174.
32. Zhenzhen, L., Xianghua, L., Ning, S., Zhan, G., Chuanchuan, R. and Jie, L. (2013) Current evidence on the relationship between three polymorphisms in the CYP1A2 gene and the risk of cancer. *Eur. J. Cancer Prev.*, **22**, 607–619.
33. Deng, S.Q., Zeng, X.T., Wang, Y., Ke, Q. and Xu, Q.L. (2013) Meta-analysis of the CYP1A2-163C>A polymorphism and lung cancer risk. *Asian Pac. J. Cancer Prev.*, **14**, 3155–3158.
34. Ma, Z., Guo, W., Gong, T., Niu, H.J., Wang, R.W. and Jiang, Y.G. (2014) CYP1A2 rs762551 polymorphism contributes to risk of lung cancer: a meta-analysis. *Tumour Biol.*, **35**, 2253–2257.
35. Wang, P., Ye, D., Guo, J., Liu, F., Jiang, H., Gong, J., Gu, C., Shao, Q., Sun, J., Zheng, S.L. *et al.* (2014) Genetic score of multiple risk-associated single nucleotide polymorphisms is a marker for genetic susceptibility to bladder cancer. *Genes Chromosomes Cancer*, **53**, 98–105.
36. Wang, M., Chu, H., Lv, Q., Wang, L., Yuan, L., Fu, G., Tong, N., Qin, C., Yin, C., Zhang, Z. *et al.* (2014) Cumulative effect of genome-wide association study-identified genetic variants for bladder cancer. *Int. J. Cancer*, **135**, 2653–2660.
37. Kamatani, Y., Matsuda, K., Okada, Y., Kubo, M., Hosono, N., Daigo, Y., Nakamura, Y. and Kamatani, N. (2010) Genome-wide association study of hematological and biochemical traits in a Japanese population. *Nat. Genet.*, **42**, 210–215.
38. Bogerd, H.P., Wiegand, H.L., Hulme, A.E., Garcia-Perez, J.L., O'Shea, K.S., Moran, J.V. and Cullen, B.R. (2006) Cellular inhibitors of long interspersed element 1 and Alu retrotransposition. *Proc. Natl Acad. Sci. USA*, **103**, 8780–8785.
39. Wang, M., Zhang, W., Yuan, L., Fu, G., Wei, Q. and Zhang, Z. (2009) Common genetic variants on 8q24 contribute to susceptibility to bladder cancer in a Chinese population. *Carcinogenesis*, **30**, 991–996.
40. Fu, Y.P., Kohaar, I., Rothman, N., Earl, J., Figueroa, J.D., Ye, Y., Malats, N., Tang, W., Liu, L., Garcia-Closas, M. *et al.* (2012) Common genetic variants in the PSCA gene influence gene expression and bladder cancer risk. *Proc. Natl Acad. Sci. USA*, **109**, 4974–4979.
41. Tanikawa, C., Urabe, Y., Matsuo, K., Kubo, M., Takahashi, A., Ito, H., Tajima, K., Kamatani, N., Nakamura, Y. and Matsuda, K. (2012) A genome-wide association study identifies two susceptibility loci for duodenal ulcer in the Japanese population. *Nat. Genet.*, **44**, 430–434, S431–432.
42. Gu, Z., Thomas, G., Yamashiro, J., Shintaku, I., Dorey, F., Raitano, A., Witte, O., Said, J., Loda, M. and Reiter, R. (2000) Prostate stem cell antigen (PSCA) expression increases with high Gleason score, advanced stage and bone metastasis in prostate cancer. *Oncogene*, **19**, 1288–1296.