

ORIGINAL ARTICLE

A polymorphism of the *POLG2* gene is genetically associated with the invasiveness of urinary bladder cancer in Japanese males

Chanavee Ratanajaraya¹, Hiroyuki Nishiyama², Meiko Takahashi^{1,3}, Takahisa Kawaguchi¹, Ryoichi Saito², Yoshiki Mikami⁴, Mikita Suyama¹, Mark Lathrop^{5,6}, Ryo Yamada¹, Osamu Ogawa² and Fumihiko Matsuda^{1,3,5,7}

Urinary bladder cancer (UBC) is a common cancer with male predominance. Pathologically it is classified into two distinct tumor entities related to the risk of patients. The low-grade tumors with relatively well-differentiated tumor histology (G1 and G2) at stage Ta are non-invasive and pose a minimal risk, whereas high-grade tumors (G2 and G3) with stages T1 to T4 are aggressive with invasion, and therefore, pose a serious risk for the patients. DNA repair and metabolic process genes may have major roles in cancer progression and development. To identify genes associated with invasiveness of UBC, we have extensively genotyped 802 single nucleotide polymorphisms in 114 genes related to DNA repair mechanisms and metabolic processes. A genetic association study was performed between non-invasive (G1 and G2 with Ta) and invasive (G2 and G3 with T1 to T4) groups of Japanese UBC patients. We found that rs17650301 in *POLG2* showed marked difference in genotype distribution between the two groups in males ($P=6.93 \times 10^{-4}$), which was further confirmed in an independent sample set (overall $P=1.67 \times 10^{-4}$). We also found by an *in silico* analysis that the risk allele of rs17650301 increased the transcription of *POLG2*. In conclusion, rs17650301 is a good candidate marker for UBC invasiveness in Japanese males.

Journal of Human Genetics (2011) 56, 572–576; doi:10.1038/jhg.2011.60; published online 7 July 2011

Keywords: genotyping; invasiveness; *POLG2*; single nucleotide polymorphism; urinary bladder cancer

INTRODUCTION

Urinary bladder cancer (UBC) is one of the most common cancers of the urinary system and is much more common in males worldwide, including in Japan. UBC is classified by the stage and grade of the tumor, which are highly correlated with recurrence, progression and patient survival rates. Tumor stage is determined by the degree of invasiveness and metastasis, whereas tumor grade classification is based on the degree of differentiation. Non-invasive UBCs are designated as stage Ta, whereas stages T1, T2, T3 and T4 refer to invasion into the subepithelial connective tissue, muscle, perivesical tissue and adjacent organs, respectively. With regard to tumor grades, the degrees of differentiation are based on tumor histology: well differentiated (G1), moderately differentiated (G2) and poorly differentiated (G3) cancer.^{1,2} Most of the UBC cases clinically fall into one of the two distinct tumor entities at diagnosis related to the risk of patients.^{3,4} The first comprises non-invasive low-grade carcinoma of relatively well-differentiated histology (G1 or G2) with stage Ta, which rarely progresses to a higher stage and pose a minimal risk for the patients. On the other hand, the second refers to aggressive tumors

corresponding to stages T1 to T4 with higher grades (G2 to G3), showing high recurrence and progression rates, and therefore pose a serious risk for the patients.

Carcinogen exposure is one of the major risk factors contributing to UBC incidence.⁵ Genes involved in DNA repair mechanisms, DNA replication, transcription, DNA damage signaling, cell cycle and metabolic processes influence the development and progression of UBC. A number of genetic polymorphisms in these types of genes were reported to be associated with genetic susceptibility to UBC.^{6–8} However, genetic association analyses focusing on the invasiveness of UBC are yet to be performed.

In this study, we focused on 121 genes acting on the four pathways of DNA repair (base excision repair, nucleotide excision repair, double strand break repair and mismatch repair), as well as those related to DNA synthesis, cell cycle and metabolism. We designed a panel of 1536 single nucleotide polymorphism (SNP) markers from these genes for extensive genotyping study, of which 802 SNPs corresponding to 114 genes that passed quality control process were statistically analyzed for their association with UBC invasiveness in Japanese patients.

¹Center for Genomic Medicine, Kyoto University Graduate School of Medicine, Kyoto, Japan; ²Department of Urology, Kyoto University Graduate School of Medicine, Kyoto, Japan;

³Institut National de la Santé et de la Recherche Médicale (INSERM) Unit U852, Kyoto University Graduate School of Medicine, Kyoto, Japan; ⁴Department of Diagnostic Pathology, Kyoto University Hospital, Kyoto, Japan; ⁵Centre National de Génotypage, Institut Génomique, Commissariat à l'Énergie Atomique (CEA), Evry, France; ⁶Foundation Jean Dausset Centre d'Étude du Polymorphisme Humain (CEPH), Paris, France and ⁷CREST program, Japan Science and Technology Agency, Kawaguchi, Japan

Correspondence: Professor F Matsuda, Center for Genomic Medicine, Kyoto University Graduate School of Medicine, Yoshida, Kyoto 606-8501, Japan.

E-mail: fumi@genome.med.kyoto-u.ac.jp

Received 4 February 2011; revised 20 April 2011; accepted 7 May 2011; published online 7 July 2011

MATERIALS AND METHODS

Study population

DNA samples were collected at the Department of Urology, Kyoto University, Kyoto, Japan. The stage and the grade of UBC were judged by the histological examinations at diagnosis. Non-invasive group (group-N) corresponds to stage Ta patients with grades G1 or G2, whereas invasive group (group-I) refers to those of stages T1 to T4 with grades G2 or G3. The characteristics and the detailed clinical diagnosis of the study populations were summarized in Table 1 and Supplementary Table 1, respectively. All the patients provided their written informed consents according to the protocols approved by the ethical review board of Kyoto University Graduate School of Medicine.

Selection of candidate genes and construction of SNP genotyping panel

Identification of SNPs was performed by resequencing exons and flanking regions of 121 genes (summarized in Supplementary Table 2), using 32 each of DNA samples of Japanese, French, Congolese and Thai populations as reported.⁹ Alignment and genotyping were performed using the Genalys software (Centre National de Génotypage, Institut Génomique, Commissariat à l'Énergie Atomique (CEA), Evry, France, <http://software.cng.fr/>).¹⁰ As for the selection of SNP markers, SNPs located in linkage disequilibrium (LD) blocks encompassing the 121 genes and essentially covering all the major haplotypes with frequencies greater than 5% were initially chosen from the International HapMap Project data of Japanese, Caucasians and Africans.¹¹ Additional SNPs identified in at least one of the four populations by the SNP discovery were included in the selection. Finally, we generated a panel of 1536 SNPs, which were chosen as tag SNP markers for genetic analyses of multiple populations using GoldenGate technology (Illumina, San Diego, CA, USA).

SNP genotyping

Genomic DNAs were prepared from peripheral blood leukocytes using Qiagen DNA Extraction kits (Qiagen, Hilden, Germany). The first screening was performed using GoldenGate assay (Illumina). The validation analysis of marker rs17650301 was performed with additional samples by sequencing. The forward primer (5'-AGGCTGGTAGGGTCCAAAGT-3') and the reverse primer (5'-AGGGTTAGGTTGAGCATCCC-3') were used for the PCR and an internal primer (5'-GAAGTTTTCACCGTGTGCC-3') was used for the sequencing with BigDye Terminator v3.1 Cycle Sequencing kit (Life Technologies, Foster City, CA, USA).

Quality control and association analysis

Among the 1536 SNPs genotyped, 20 SNPs on X-chromosome and one triallelic SNP were initially removed before quality control process. During the quality

control, we first examined the presence of DNA samples showing high degrees of kinship by plink (Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA, USA, <http://pngu.mgh.harvard.edu/~purcell/plink/>)¹² and found there were no such samples. Three DNA samples with call rates smaller than 0.90 were removed. Regarding the SNP markers, 104 SNPs with calling rates smaller than 0.95 and 609 SNPs with minor allele frequencies smaller than 0.01 were also excluded. After these steps, the remaining 802 SNP markers corresponding to 114 genes were used for statistical analyses. The association was examined by trend exact test for genotype distribution. Odds ratios and 95% confidence intervals were estimated with additive model. The heterogeneity of the effect size between the first screening and the validation analysis was evaluated by Breslow–Day test.¹³ The LD structure of the *POLG2* locus was established by using Japanese data from the International HapMap Project.¹¹

Expression quantitative trait loci analysis

Allele-specific effect on the expression of rs17650301 was performed using Japanese data set GSE6536 of expression profile Gene Expression Omnibus (National Center for Biotechnology Information, National Institutes of Health, Bethesda, MD, USA, <http://www.ncbi.nlm.nih.gov/geo/>).¹⁴ The association *P*-values were obtained by Jonckheere–Terpstra method using R software (Department of Statistics, The University of Auckland, Auckland, New Zealand, <http://www.r-project.org/>) or SPSS (version 18, IBM, Armonk, NY, USA, <http://www-01.ibm.com/software/data/spss/products/statistics/statistics-desktop/>).

RESULTS

First screening

Among 1536 SNPs corresponding to 121 candidate genes in the genotyping panel, a total of 802 SNPs derived from 114 candidate genes were examined for their association with UBC invasiveness, using 106 group-N samples and 93 group-I samples (for details, see Materials and Methods). Although no SNPs showed significant association ($P=6.23 \times 10^{-5}$) after Bonferroni's correction for multiple testing, we identified 55 SNPs with nominal *P*-values smaller than 0.05, of which nine SNPs showed *P*-values smaller than 0.01 (Table 2). The strongest *P*-value was obtained for rs17650301 ($P=3.40 \times 10^{-3}$) located in intron 5 of DNA polymerase subunit $\gamma 2$ gene (*POLG2*) at 17q24.1. Five SNPs in the *POLL* gene, rs3730476, rs1055364, rs3730465, rs3730472 and rs3095795, were in complete LD with each other.

We then performed a subgroup analysis by dividing the UBC patients according to their gender. In the male study, six SNPs showed *P*-values smaller than 0.01, of which the strongest association was again obtained for rs17650301 ($P=6.93 \times 10^{-4}$) (Table 2). However, rs17650301 did not show association in the female study. In addition, rs2518968, which showed the second lowest *P*-value in the joint study ($P=4.50 \times 10^{-3}$), showed association in the male study ($P=4.81 \times 10^{-3}$), although it was much weaker than that of rs17650301.

Validation analysis

We focused on rs17650301 and performed a validation analysis by direct sequencing, using DNA samples of an independent set of UBC patients consisting of 65 group-N samples and 114 group-I samples. Although the association was not observed in the male–female joint study ($P=0.432$), we obtained a marginal *P*-value ($P=0.0396$) in the male study (Table 3). The combined analysis by pooling the genotypes of the two studies returned a *P*-value of 9.45×10^{-3} and 1.67×10^{-4} in the joint study and in the male study, respectively (Table 3). The *P*-value of heterogeneity did not show evidence of differentiation in effect size ($P=0.17$ for the joint study and $P=0.47$ for the male study). A logistic regression analysis adjusted for age and sex returned a *P*-value of 5.28×10^{-3} for the joint study.

Table 1 Characteristics of UBC patients used for statistical analysis

Study	Group-N	Group-I
First screening		
Number	106	93
Number of males (%)	75 (70.8)	70 (75.3)
Age at diagnosis (mean \pm s.d.)	68.5 \pm 11.9	68.8 \pm 10.5
Validation analysis		
Number	65	114
Number of males (%)	50 (76.9)	98 (86.0)
Age at diagnosis (mean \pm s.d.)	67.9 \pm 13.3	71.1 \pm 10.0
Combined analysis		
Number	171	207
Number of males (%)	125 (73.1)	168 (81.2)
Age at diagnosis (mean \pm s.d.)	68.3 \pm 12.4	70.1 \pm 10.3

Abbreviations: Group-I, invasive group; Group-N, non-invasive group; UBC, urinary bladder cancer.

Table 2 Results of association analysis in the first screening

Study	SNP ID	A1/A2 ^a	Gene	Non-invasive (N)				Invasive (I)				P-value ^b
				A1/A1	A1/A2	A2/A2	Freq. A1	A1/A1	A1/A2	A2/A2	Freq. A1	
Male and female												
	rs17650301	A/C	POLG2	55	41	10	0.712	31	43	19	0.565	3.40×10 ⁻³
	rs2518968	C/G	BLM	30	55	21	0.542	46	36	11	0.688	4.50×10 ⁻³
	rs3730476	A/G	POLL	81	23	0	0.889	85	5	1	0.962	7.29×10 ⁻³
	rs1055364	C/A	POLL	83	23	0	0.892	87	5	1	0.962	7.31×10 ⁻³
	rs3730465	A/G	POLL	83	23	0	0.892	87	5	1	0.962	7.31×10 ⁻³
	rs3730472	T/G	POLL	83	23	0	0.892	87	5	1	0.962	7.31×10 ⁻³
	rs3095795	T/C	POLL	83	23	0	0.892	87	5	1	0.962	7.31×10 ⁻³
	rs1801127	G/A	OGG1	98	7	1	0.958	93	0	0	1.000	7.63×10 ⁻³
	rs2276332	A/C	ADH1A	84	22	0	0.896	86	7	0	0.962	8.95×10 ⁻³
Male												
	rs17650301	A/C	POLG2	43	26	6	0.747	22	32	16	0.543	6.93×10 ⁻⁴
	rs8192772	T/C	CYP2E1	38	30	5	0.726	53	15	1	0.877	1.79×10 ⁻³
	rs2518968	C/G	BLM	19	41	15	0.527	36	25	9	0.693	4.81×10 ⁻³
	rs2070676	G/C	CYP2E1	4	23	45	0.215	1	10	57	0.088	6.34×10 ⁻³
	rs9634161	T/C	RAD52	65	10	0	0.933	46	21	1	0.831	6.84×10 ⁻³
	rs4937	T/C	POLR2C	25	38	11	0.595	13	34	22	0.435	8.68×10 ⁻³
Female												
	rs12611088	G/A	XRCC1	12	18	1	0.677	18	5	0	0.891	4.40×10 ⁻³
	rs2023614	C/G	XRCC1	26	5	0	0.919	12	9	2	0.717	9.92×10 ⁻³
	rs3213266	G/A	XRCC1	26	5	0	0.919	12	9	2	0.717	9.92×10 ⁻³

Abbreviation: SNP, single nucleotide polymorphism.

SNP markers that showed statistical *P*-values smaller than 0.01 are shown.^aThe alleles A1 and A2 represent the reference and non-reference allele, respectively, in NCBI build 36.^bStatistical *P*-values were calculated by trend exact test.**Table 3 Summary of the association analysis of rs17650301 with UBC invasiveness**

Study	Non-invasive (N)				Invasive (I)				OR (95% CI)	P-value ^a
	AA	AC	CC	Freq. A	AA	AC	CC	Freq. A		
First screening										
Male and female	55	41	10	0.712	31	43	19	0.565	1.91 (1.26–2.89)	3.40×10^{−3}
Male	43	26	6	0.747	22	32	16	0.543	2.48 (1.51–4.08)	6.93×10^{−4}
Validation analysis										
Male and female	32	26	7	0.692	49	49	16	0.645	1.24 (0.78–1.97)	0.432
Male	28	19	3	0.750	39	45	14	0.628	1.78 (1.04–3.05)	0.0396
Combined analysis										
Male and female	87	67	17	0.705	80	92	35	0.609	1.53 (1.13–2.08)	9.45×10^{−3}
Male	71	45	9	0.748	61	77	30	0.592	2.04 (1.43–2.93)	1.67×10^{−4}

Abbreviations: CI, confidence intervals; OR, odds ratio; UBC, urinary bladder cancer.

Reference (A) and variant (C) alleles are based on NCBI build36.

P-values that are less than 0.05 are shown in bold characters.^aStatistical *P*-values were calculated by trend exact test.

Also, the *P*-value did not change ($P=1.67 \times 10^{-4}$) after adjusting for age in the male study.

Linkage disequilibrium block

The LD block of the DNA region encompassing rs17650301 in the Japanese population was established using genotype data of the International Hapmap Project.¹¹ In the LD block encompassing a 34-kb DNA between rs8068311 and rs6504238, there were no SNP markers showing strong LD ($r^2 \geq 0.8$) with rs17650301, except rs1991401

($r^2=1$). Rs1991401 is located in the 5' untranslated region of the *DDX5* (DEAD box polypeptide 5) gene adjacent to *POLG2* (Figure 1).

Rs17650301 and expression of *POLG2*

Although rs17650301 is located in an intronic region, it may have functional impact in the transcription of *POLG2*. We performed an expression quantitative trait loci (eQTL) analysis by examining the expression profiles of 44 lymphoblastoid cells of Japanese origin in Gene Expression Omnibus database¹⁴ and found an increase of

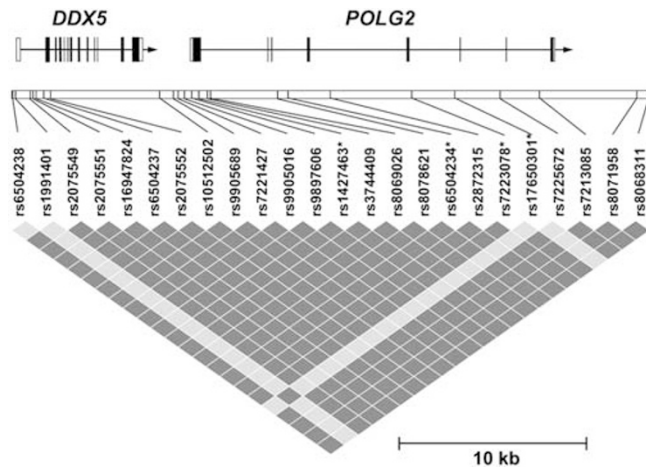


Figure 1 LD structure of the 34-kb region spanning the *POLG2* and *DDX5* genes. The structure and transcriptional polarity of *POLG2* and *DDX5* are shown according to NCBI Reference Sequence Build 36. Exons are shown by filled and open rectangles representing the coding and untranslated regions respectively. Pairwise LD estimation scores between SNPs within the region were converted into colors according to the color scheme of Haploview. The SNPs included in the genotyping panel for the first screening are indicated by an asterisk.

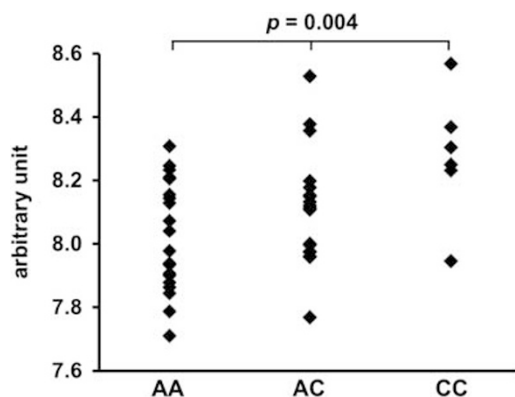


Figure 2 Allelic expression of *POLG2* with the genotypes of rs17650301. The levels of *POLG2* transcription in cell lines carrying the three genotypes are indicated by a square. 'C' and 'A' represent the risk and the alternative alleles, respectively.

POLG2 transcription by the risk allele (C) of rs17650301 ($P=0.004$, Figure 2). However, a search for *cis*-regulatory elements using Jaspardatabase (The Bioinformatics Centre, Department of Biology, University of Copenhagen, Copenhagen, Denmark, <http://jaspar.genereg-net/>)¹⁵ failed to identify any known motifs covering rs17650301.

DISCUSSION

In this study, we conducted a case-control association analysis of UBC invasiveness by taking a candidate gene approach of 1536 SNPs in 121 genes related to DNA repair mechanisms, DNA synthesis and metabolic processes. To our knowledge, this is the first genetic study focusing on the invasiveness of UBC. Statistical analysis was performed for 802 SNPs in 114 genes that passed quality control. The majority of the SNP markers (627 out of 734) excluded from the statistical test showed minor allele frequencies smaller than 0.01 in the Japanese population. This is due to the design of the genotyping panel for trans-ethnic study by covering SNPs identified in at least one of the four populations (Japanese, French, Thai and Congolese). Indeed,

there were as many as 934 SNPs whose minor allele frequencies in the SNP discovery were smaller than 0.01 in Japanese and greater than 0.01 in at least one of the three other populations.

There was no SNP showing a significant association after Bonferroni's correction for multiple testing ($P=6.23 \times 10^{-5}$). The Bonferroni method is a very stringent test; although the possibilities of false positives are decreased, it comes together with the risk of losing potential candidates with true associations. In order to overcome this drawback, we took a strategy to screen SNP markers for potential associations in two stages. The most important aspect of our results is that the association of rs17650301 with UBC invasiveness was reproducibly obtained in two independent sample collections, without evidence of differentiation in effect size. However, the detection of significant association will require replication analyses using other sample collections.

The genotyping of 802 SNP markers followed by the validation analysis identified rs17650301 located in intron 5 of the *POLG2* gene as the strongest candidate for the invasiveness of UBC in Japanese male patients. *POLG2* encodes the 55-kDa accessory subunit of mitochondrial DNA polymerase. This subunit, together with the 140-kDa catalytic subunit (POLG), stimulates the polymerase and exonuclease activities in the replication process of mitochondrial DNA.^{16,17} Although the association between *POLG2* polymorphisms and bladder cancer has not yet been established, its association with the risk of head and neck cancer was demonstrated in the French population.¹⁸ There is no strong biological evidence to support the functional importance of *POLG2* to the invasiveness of UBC. However, the increase of transcription level with the risk allele of rs17650301 may be a reason for the association. Because there are no other known SNPs in strong LD ($r^2 \geq 0.8$) with rs17650301 in the *POLG2* locus, rs17650301 is a good candidate marker for the invasiveness of UBC in Japanese males.

The strong association of rs17650301 with the UBC invasiveness was observed only in males. As such, the reasons for the observed higher UBC incidence in males than females remain uncertain. Different non-genetic risk factors including sex hormones, life style and environment may contribute to the disease onset, resulting in the sex-specific association. For instance, involvement of androgen and its receptor in bladder cancer is demonstrated in a mouse study.¹⁹ Epidemiologically, postmenopausal women have a higher risk of development and progression of UBC than premenopausal women.²⁰ Also, smoking is one of the major risks of urinary tract cancer and considered to increase the cancer risk by approximately threefold,^{21,22} and the percentage of smokers is much higher in males than in females in Japan (36.8 versus 9.1%, as of 2008 in the National Survey by the Ministry of Health, Labour and Welfare of Japan). If such environmental and life style-related factors contribute to the predisposition of bladder cancer together with *POLG2*, this may explain the observed effect of *POLG2* with UBC invasiveness in males only. However, further functional characterization of *POLG2* for its involvement in carcinogenesis of indolent and aggressive tumors are required for the elucidation of the molecular mechanism underlying the prognosis of UBC.

ACKNOWLEDGEMENTS

We would like to express our gratitude to all the study participants who made this research possible. We also thank M Kokubo and M Mizutani for technical assistance, and A Yoshizumi and H Uneme for informatics management. This study was supported, in part, by Core Research of Evolutional Science & Technology program (CREST) from Japan Science and Technology Agency (JST) and by Institut National de la Santé et de la Recherche Médicale.

- 1 Greene, F. L., Compton, C. C., Fritz, A. G., Shah, J. & Winchester, D. P. *AJCC Cancer Staging Atlas* (Springer, New York, 2006).
- 2 Meiden, A.v.d. in *Invasive Bladder Cancer* (eds Bassi, P. & Pagano, F.) 67–76 (Springer, London, 2007).
- 3 Gofrit, O. N., Pode, D., Lazar, A., Katz, R. & Shapiro, A. Watchful waiting policy in recurrent Ta G1 bladder tumors. *Eur. Urol.* **49**, 303–306 discussion 306–307 (2006).
- 4 McNally, M. L., Rashid, H. H. & Messing, E. M. in *Invasive Bladder Cancer* (eds Bassi, P. & Pagano, F.) 15–38 (Springer, London, 2007).
- 5 Taylor, J. A., Umbach, D. M., Stephens, E., Castranio, T., Paulson, D., Robertson, C. *et al.* The role of N-acetylation polymorphisms in smoking-associated bladder cancer: evidence of a gene-gene-exposure three-way interaction. *Cancer Res.* **58**, 3603–3610 (1998).
- 6 Sanyal, S., Festa, F., Sakano, S., Zhang, Z., Steineck, G., Norming, U. *et al.* Polymorphisms in DNA repair and metabolic genes in bladder cancer. *Carcinogenesis* **25**, 729–734 (2004).
- 7 Garcia-Closas, M., Malats, N., Silverman, D., Dosemeci, M., Kogevinas, M., Hein, D. W. *et al.* 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 (2005).
- 8 Figueroa, J. D., Malats, N., Rothman, N., Real, F. X., Silverman, D., Kogevinas, M. *et al.* Evaluation of genetic variation in the double-strand break repair pathway and bladder cancer risk. *Carcinogenesis* **28**, 1788–1793 (2007).
- 9 Danoy, P., Sonoda, E., Lathrop, M., Takeda, S. & Matsuda, F. A naturally occurring genetic variant of human XRCC2 (R188H) confers increased resistance to cisplatin-induced DNA damage. *Biochem. Biophys. Res. Commun.* **352**, 763–768 (2007).
- 10 Takahashi, M., Matsuda, F., Margetic, N. & Lathrop, M. Automated identification of single nucleotide polymorphisms from sequencing data. *J. Bioinform. Comput. Biol.* **1**, 253–265 (2003).
- 11 The International HapMap Consortium. A haplotype map of the human genome. *Nature* **437**, 1299–1320 (2005).
- 12 Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
- 13 Breslow, N. E. & Day, N. E. Statistical methods in cancer research. Volume I—The analysis of case-control studies. *IARC Sci. Publ.* 5–338 (1980).
- 14 Stranger, B. E., Forrest, M. S., Dunning, M., Ingle, C. E., Beazley, C., Thorne, N. *et al.* Relative impact of nucleotide and copy number variation on gene expression phenotypes. *Science* **315**, 848–853 (2007).
- 15 Portales-Casamar, E., Thongjuea, S., Kwon, A. T., Arenillas, D., Zhao, X., Valen, E. *et al.* JASPAR 2010: the greatly expanded open-access database of transcription factor binding profiles. *Nucleic Acids Res.* **38**, D105–D110 (2010).
- 16 Lim, S. E., Longley, M. J. & Copeland, W. C. The mitochondrial p53 accessory subunit of human DNA polymerase gamma enhances DNA binding, promotes processive DNA synthesis, and confers N-ethylmaleimide resistance. *J. Biol. Chem.* **274**, 38197–38203 (1999).
- 17 Yakubovskaya, E., Chen, Z., Carrodegua, J. A., Kisker, C. & Bogenhagen, D. F. Functional human mitochondrial DNA polymerase gamma forms a heterotrimer. *J. Biol. Chem.* **281**, 374–382 (2006).
- 18 Michiels, S., Danoy, P., Dessen, P., Bera, A., Boulet, T., Bouchardy, C. *et al.* Polymorphism discovery in 62 DNA repair genes and haplotype associations with risks for lung and head and neck cancers. *Carcinogenesis* **28**, 1731–1739 (2007).
- 19 Miyamoto, H., Yang, Z., Chen, Y. T., Ishiguro, H., Uemura, H., Kubota, Y. *et al.* Promotion of bladder cancer development and progression by androgen receptor signals. *J. Natl. Cancer Inst.* **99**, 558–568 (2007).
- 20 McGrath, M., Michaud, D. S. & De Vivo, I. Hormonal and reproductive factors and the risk of bladder cancer in women. *Am. J. Epidemiol.* **163**, 236–244 (2006).
- 21 Zeegers, M. P., Tan, F. E., Dorant, E. & van Den Brandt, P. A. The impact of characteristics of cigarette smoking on urinary tract cancer risk: a meta-analysis of epidemiologic studies. *Cancer* **89**, 630–639 (2000).
- 22 Puente, D., Hartge, P., Greiser, E., Cantor, K. P., King, W. D., Gonzalez, C. A. *et al.* A pooled analysis of bladder cancer case-control studies evaluating smoking in men and women. *Cancer Causes Control* **17**, 71–79 (2006).

Supplementary Information accompanies the paper on Journal of Human Genetics website (<http://www.nature.com/jhg>)