nature genetics

Genome-wide association study of renal cell carcinoma identifies two susceptibility loci on 2p21 and 11q13.3

Mark P Purdue[1,72], Mattias Johansson[2,72], Diana Zelenika[3,72], Jorge R Toro[1,72], Ghislaine Scelo[2,72], Lee E Moore[1,72], Egor Prokhortchouk[4,5,72], Xifeng Wu[6], Lambertus A Kiemeney[7,8], Valerie Gaborieau[2], Kevin B Jacobs[1,9], Wong-Ho Chow[1], David Zaridze[10], Vsevolod Matveev[10], Jan Lubinski[11], Joanna Trubicka[11], Neonila Szeszenia-Dabrowska[12], Jolanta Lissowska[13], Péter Rudnai[14], Eleonora Fabianova[15], Alexandru Bucur[16], Vladimir Bencko[17], Lenka Foretova[18], Vladimir Janout[19], Paolo Boffetta[20], Joanne S Colt[1], Faith G Davis[21], Kendra L Schwartz[22], Rosamonde E Banks[23], Peter J Selby[23], Patricia Harnden[24], Christine D Berg[25], Ann W Hsing[1], Robert L Grubb III[26], Heiner Boeing[27], Paolo Vineis[28–30], Françoise Clavel-Chapelon[31,32], Domenico Palli[33], Rosario Tumino[34], Vittorio Krogh[35], Salvatore Panico[36], Eric J Duell[37], José Ramón Quirós[38], Maria-José Sanchez[39,40], Carmen Navarro[40,41], Eva Ardanaz[40,42], Miren Dorronsoro[40,43], Kay-Tee Khaw[44], Naomi E Allen[45], H Bas Bueno-de-Mesquita[46], Petra H M Peeters[28,47], Dimitrios Trichopoulos[48,49], Jakob Linseisen[50,51], Börje Ljungberg[52], Kim Overvad[53], Anne Tjønneland[54], Isabelle Romieu[2], Elio Riboli[28], Anush Mukeria[10], Oxana Shangina[10], Victoria L Stevens[53], Michael J Thun[55], W Ryan Diver[53], Susan M Gapstur[55], Paul D Pharoah[56,57], Douglas F Easton[56,57], Demetrius Albanes[1], Stephanie J Weinstein[1], Jarmo Virtamo[58], Lars Vatten[59], Kristian Hveem[59], Inger Njølstad[60], Grethe S Tell[61], Camilla Stoltenberg[62], Rajiv Kumar[63], Kvetoslava Koppova[64], Olivier Cussenot[65], Simone Benhamou[66,67], Egbert Oosterwijk[8], Sita H Vermeulen[7,68], Katja K H Aben[7,69], Saskia L van der Marel[68], Yuanqing Ye[6], Christopher G Wood[70], Xia Pu[6], Alexander M Mazur[4,5], Eugenia S Boulygina[5], Nikolai N Chekanov[4], Mario Foglio[3], Doris Lechner[3], Ivo Gut[3], Simon Heath[3], Hélène Blanche[71], Amy Hutchinson[1,9], Gilles Thomas[1,9], Zhaoming Wang[1,9], Meredith Yeager[1,9], Joseph F Fraumeni Jr[1], Konstantin G Skryabin[4,5,73], James D McKay[2,73], Nathaniel Rothman[1,73], Stephen J Chanock[1,73], Mark Lathrop[3,73] & Paul Brennan[2,73]

We conducted a two-stage genome-wide association study of renal cell carcinoma (RCC) in 3,772 affected individuals (cases) and 8,505 controls of European background from 11 studies and followed up 6 SNPs in 3 replication studies of 2,198 cases and 4,918 controls. Two loci on the regions of 2p21 and 11q13.3 were associated with RCC susceptibility below genome-wide significance. Two correlated variants ($r^2 = 0.99$ in controls), rs11894252 ($P = 1.8 \times 10^{-8}$) and rs7579899 ($P = 2.3 \times 10^{-9}$), map to *EPAS1* on 2p21, which encodes hypoxia-inducible-factor-2 alpha, a transcription factor previously implicated in RCC. The second locus, rs7105934, at 11q13.3, contains no characterized genes ($P = 7.8 \times 10^{-14}$). In addition, we observed a promising association on 12q24.31 for rs4765623, which maps to *SCARB1*, the scavenger receptor class B, member 1 gene ($P = 2.6 \times 10^{-8}$). Our study reports previously unidentified genomic regions associated with RCC risk that may lead to new etiological insights.

Kidney cancer accounts for approximately 2% of new cancer diagnoses worldwide[1] and is the deadliest urologic malignancy, with an estimated 5-year survival rate between 50% and 60% (ref. 2). Approximately 80–90% of kidney cancers develop in the renal parenchyma and are known as renal cell carcinoma (RCC). Epidemiological studies have conclusively identified three risk factors for RCC, all of which are modifiable: hypertension, obesity and smoking[2,3]. Furthermore, there is evidence that genetic factors influence susceptibility to RCC; for instance, the lifetime risk for disease increases approximately twofold for those with a first-degree relative with RCC[4–7]. The tumor is also commonly observed in pedigrees with von Hippel-Lindau (VHL) syndrome, as well as other genetic disorders, such as hereditary papillary renal cell carcinoma, Birt-Hogg-Dubé syndrome, and hereditary leiomyomatosis and renal cell cancer (HLRCC)[2,8]. However, familial RCC cases represent less than 5% of RCC cases overall[9]. To date, candidate gene studies have not yielded genetic variants that conclusively replicate. In search of common genetic variants with moderate effect sizes, we have therefore conducted a genome-wide association study (GWAS) of RCC.

A full list of author's affiliations appears at the end of this paper.

Received 3 May; accepted 5 November; published online 5 December 2010; doi:10.1038/ng.723

Table 1 Summary results for six SNPs selected for replication in renal cell carcinoma genome-wide association study

Locus (gene region)	SNP ID (minor allele frequency)	IARC+NCI[a] 3,772/8,505[d]			Replication[b] 2,198/4,918[d]			All combined[c] 5,970/13,423[d]		
		OR[e]	95% CI[e]	P[e]	OR[e]	95% CI[e]	P[e]	OR[e]	95% CI[e]	P[e]
2p21 (*EPAS1*)	**rs11894252** (0.40)	1.18	(1.12–1.26)	1.9×10^{-8}	1.08	(1.00–1.16)	0.06	1.14	(1.09–1.20)	1.8×10^{-8}
2p21 (*EPAS1*)	**rs7579899** (0.40)	1.18	(1.11–1.25)	5.9×10^{-8}	1.11	(1.03–1.20)	0.008	1.15	(1.10–1.21)	2.3×10^{-9}
2p21 (*EPAS1*)	**rs6758592** (0.47)	1.13	(1.07–1.20)	2.5×10^{-5}	1.05	(0.97–1.14)	0.20	1.10	(1.05–1.15)	4.0×10^{-5}
3q26 (*PP13439*)	**rs9839909** (0.34)	0.82	(0.76–0.89)	4.3×10^{-6}	0.96	(0.90–1.03)	0.30	0.90	(0.86–0.95)	4.0×10^{-5}
11q13.31 (chr. 11)	**rs7105934** (0.07)	0.65	(0.55–0.76)	1.7×10^{-7}	0.71	(0.62–0.81)	6.8×10^{-7}	0.69	(0.62–0.76)	7.8×10^{-14}
12q24.31 (*SCARB1*)	**rs4765623** (0.34)	1.18	(1.11–1.25)	6.4×10^{-8}	1.07	(0.99–1.16)	0.09	1.15	(1.09–1.20)	2.6×10^{-8}

Chr., chromosome.

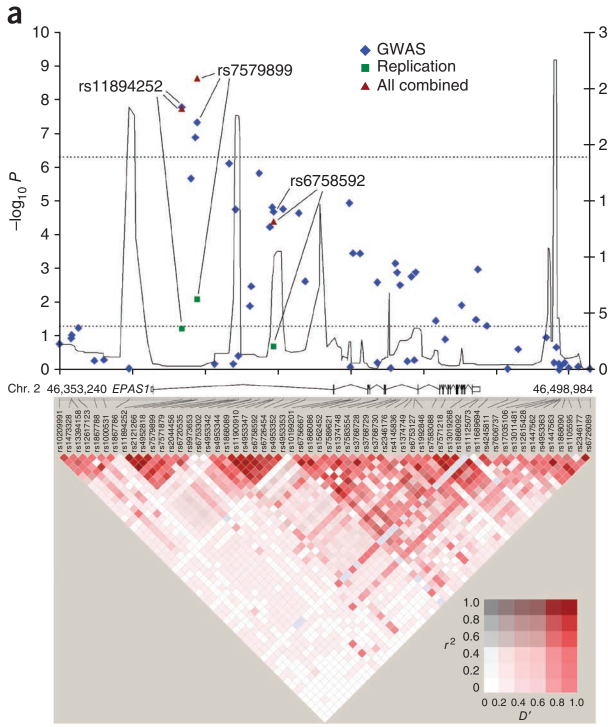
[a]All scanned samples from IARC-CNG and NCI combined by meta-analysis (Online Methods). [b]Samples include subjects from three replication studies: the MD Anderson Renal Cell Cancer Study, the Dutch Renal Cell Cancer Study and the IARC Replication Study (**Supplementary Note**). [c]Column shows combined results of the pooled GWAS data and the three replication studies by meta-analysis. [d]Number of cases/controls. [e]Odds ratios (OR) were estimated using the per-rare-allele log-additive model and unconditional logistic regression (Online Methods).

We report the findings of a two-stage GWAS of RCC based on two parallel scans followed by replication of six notable SNPs in three studies. The two scans were coordinated by (i) the International Agency for Research on Cancer (IARC) and the Centre National de Génotypage (CNG) based on 2,639 RCC cases and 5,392 controls of European background drawn from seven studies conducted in Europe with the Illumina Infinium HumanHap 300 and 610 BeadChips and (ii) the US National Cancer Institute (NCI) scan, based on 1,453 RCC cases and 3,531 controls of European background from four studies with the Illumina Infinium HumanHap 500, 610 and 660w BeadChips (**Supplementary Table 1**, Online Methods and **Supplementary Note**). All subjects from the IARC-CNG study were genotyped at the CNG with the exception of 305 cases and 323 controls from Russia that were genotyped at the Center 'Bioengineering' and at the Kurchatov Institute in Moscow. All subjects from the NCI study were scanned at the NCI Core Genotyping Facility. In addition, 1,438 controls from the Wellcome Trust Case Control Consortium were genotyped at the Sanger Institute, UK[10]. All RCC cases were defined on the basis of the International Classification of Diseases for Oncology, Second Edition (ICD-O-2), and included all cancers that were coded as C64.

Comparable quality control metrics were applied to the two scanned datasets, and following sample and SNP exclusions, genotype data for up to 577,547 SNPs were available for 2,461 cases and 5,081 controls in the IARC-CNG scan, and data for 585,576 SNPs were available for 1,311 cases and 3,424 controls in the NCI scan (Online Methods). We conducted the primary analyses using unconditional logistic regression models for genotype trend effects (1 degree of freedom) and adjusted for sex, country and eigenvectors, as well as for study in the data from the United States (Online Methods). In order to compute summary findings across both scans, we performed a meta-analysis using a fixed effects model with inverse-variance weighting followed by a pooled analysis with individual level data. Quantile-quantile plots of the combined results showed little evidence for inflation of the test statistics compared to the expected distribution ($\lambda = 1.018$ overall; **Supplementary Fig. 1**). We then applied genomic control, and we corrected all reported P values and CIs for the observed inflation. A Manhattan plot summarizing the combined results of 586,069 SNPs is shown in **Supplementary Figure 2**.

Based on the meta-analysis using SNPs genotyped in both centers, six SNPs were associated with RCC at a significance level approaching or surpassing genome-wide statistical significance ($P < 5 \times 10^{-7}$ in two-tailed tests)[10] and were selected for replication in three additional case-control series from Europe and the United States (2,198 RCC cases and 4,918 controls) (**Supplementary Table 1**). Performing genomic control on this data showed that hidden population substructures or differential genotype calling between cases and controls did not substantively influence these results (Online Methods). Three SNPs on 2p21 (rs11894252, rs7579899 and rs6758592) were selected, as well as single SNPs on 3q26.31 (rs9839909), 11q13.3 (rs7105934) and 12q24.31 (rs4765623). For the replication study, rs11894252 could not be optimized; thus we genotyped a highly correlated SNP, rs1867785 ($r^2 = 1.0$ in the HapMap European CEU population[11]) (Online Methods). For the other five SNPs, there was a high concordance between genotype calls on the Illumina BeadChip and the optimized TaqMan assays in both centers (concordance of 100% for IARC-CNG and 98.9%–100% for NCI)[12]. Because rs9839909 (3q26.31) and rs7105934 (11q13.3) were not included on the Illumina HumanHap 300 BeadChip, subjects genotyped with this chip in the GWAS (908 cases and 2,415 controls) were also genotyped by TaqMan and included in the replication phase. In a meta-analysis of the pooled GWAS and replication results, SNPs in three of the four regions achieved genome-wide significance and mapped to 2p21, 11q13.3 and 12q24.31 (**Table 1** and **Fig. 1**). Imputing SNPs in the implicated regions 2p21, 11q13.3 and 12q24.31 using the 1000 Genomes Project data[13] as a scaffold did not reveal additional SNPs with stronger, independent associations to those genotyped directly (**Supplementary Table 2**).

In the combined analysis[14], two SNPs on 2p21 achieved genome-wide significance, rs7579899 ($P = 2.3 \times 10^{-9}$, per allele odds ratio (OR) = 1.15, 95% CI 1.10–1.21) and rs11894252 ($P = 1.8 \times 10^{-8}$, OR = 1.14, 95% CI 1.09–1.20). Further, rs7579899 was significant in the independent replication analysis ($P = 0.008$, OR = 1.11, 95% CI 1.03–1.20), whereas rs1867785, a highly correlated surrogate for rs11894252, suggested a comparable effect that did not achieve independent significance ($P = 0.06$, OR = 1.08, 95% CI 1.00–1.16) (**Table 1**). When stratified by either SNP marker, the signal of the other was extinguished (data not shown). Together with the high correlation between the two markers ($r^2 = 0.99$ in controls), these results point toward a single common susceptibility locus for RCC. An additional SNP, rs4952818, achieved genome-wide significance in the combined scan ($P = 1 \times 10^{-7}$; **Fig. 1**), but its association was accounted for by rs11894252 and rs7579899 (adjusted $P = 0.45$ and adjusted $P = 0.36$, respectively) and was therefore not selected for replication. The third SNP selected for replication, rs6758592, was minimally correlated with the previous two SNPs ($r^2 = 0.12$ and $r^2 = 0.11$ with



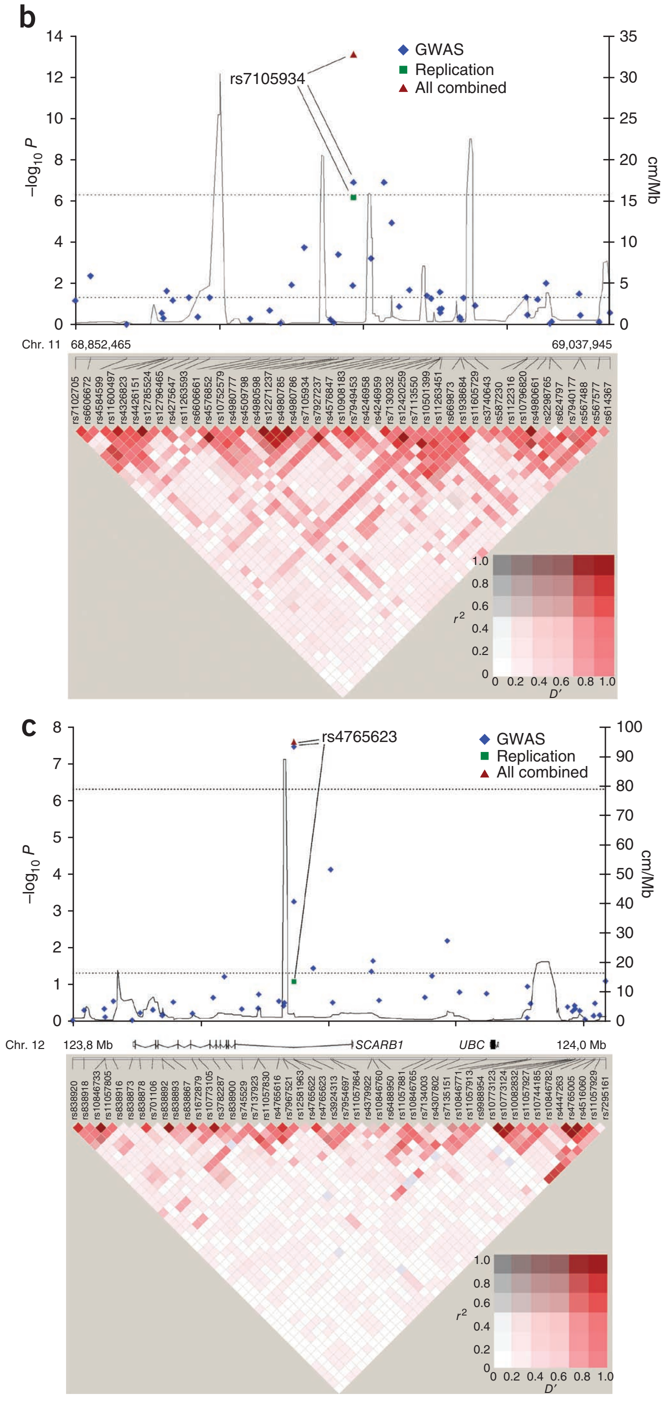


Figure 1 Association results, recombination and linkage disequilibrium plots for regions below genome-wide significance (2p21 and 11q13.3) and a region with a promising association (12q24.31) to RCC susceptibility. Results of pooled IARC-CNG and NCI GWAS data (GWAS), for SNPs selected for replication in replication studies combined by meta-analysis (replication), and of all studies combined by meta-analysis (all combined). P values for log-additive association results ($-\log_{10}$) are shown with recombination rates (cm/Mb) based on HapMap phase II data, and pairwise r^2 and superimposed D' values are displayed below for all SNPs included in the GWAS analysis. Coordinates refer to genome build 36.1. (**a**) A depiction of the region of 2p21 including the *EPAS1* gene region (46,353,240–46,498,984 bp). (**b**) A depiction of the region of 11q13.3 (68,852,465–69,037,945 bp). (**c**) A depiction of the region of 12q24.31 including the *SCARB1* gene region (123,800,267–124,008,657 bp).

rs11894252 and rs7579899, respectively) and only showed an association in the NCI data (NCI $P = 1.8 \times 10^{-7}$, IARC $P = 0.16$, heterogeneity $P = 0.0004$; **Supplementary Table 3**), which was not accounted for by rs11894252 and rs7579899 (adjusted $P = 1 \times 10^{-5}$ for both). Although rs6758592 did not replicate, the combined analysis yielded $P = 4.0 \times 10^{-5}$, suggesting that in the NCI scan data there could be evidence for a more complex genomic architecture underlying the association of this locus with RCC.

Our finding on 2p21 is notable because the candidate gene in this region, *EPAS1*, has previously been implicated in RCC[15–19]. The two SNPs on 2p21, rs11894252 and rs7579899, are distributed across a 4.2-kb region of intron 1 in *EPAS1*, which encodes the hypoxia-inducible factor 2α (HIF-2α) and is a key gene in the VHL-HIF pathway. The VHL complex targets HIF subunits for ubiquitin-mediated degradation[20]. Accumulation of HIF-2α leads to upregulation of vascular endothelial growth factor (VEGF) and epidermal growth factor receptor (EGFR). The inactivation of VHL in renal carcinoma cell lines leads to unchecked HIF-2α–mediated expression of HIF-responsive tumorigenic factors, most notably VEGF[16,17]. Further, tumor formation in VHL-deficient renal carcinoma cells has been found to be suppressed by inhibition of HIF-2α[18,19]. The findings from our GWAS provide further evidence that *EPAS1* is a key gene in RCC development, but additional studies are needed to identify the functionally relevant common variants associated with increased risk.

A variant, rs7105934, on 11q13.3 was associated with RCC in the combined analysis ($P = 7.8 \times 10^{-14}$, OR = 0.69, 95% CI 0.62–0.76). This SNP was independently replicated with a comparable risk estimate to the initial GWAS results ($P = 6.8 \times 10^{-7}$, OR = 0.71, 95% CI 0.62–0.81). Overall, the magnitude of the association with this relatively uncommon SNP (minor allele frequency = 0.08 in controls) is comparatively large compared to risk markers previously identified in the GWAS of other cancers[21]. This SNP maps to a 350-kb region

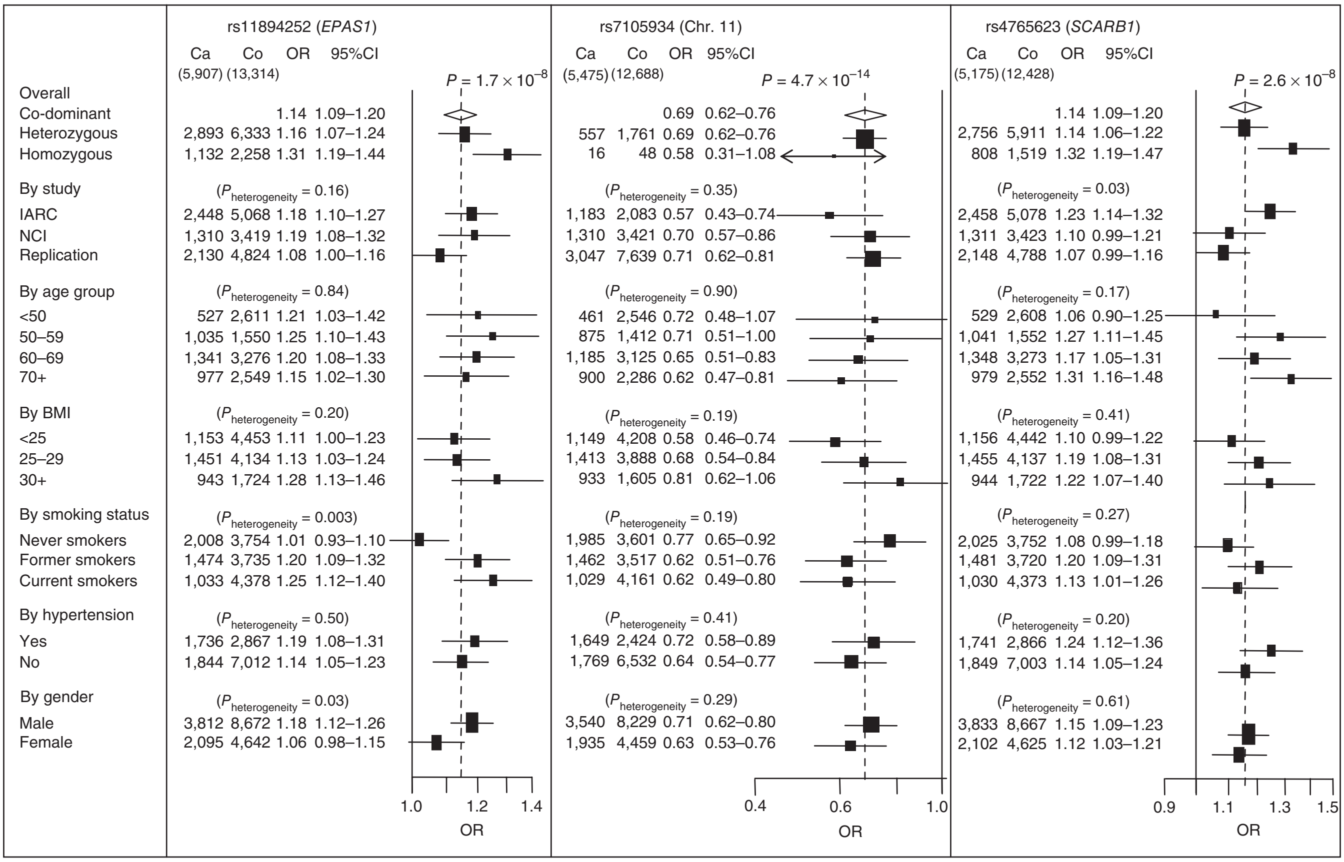


Figure 2 Forest plots for three SNPs showing significant or promising association to RCC susceptibility. Forest plots show stratified odds ratios (ORs) for SNPs selected for replication. The two highly correlated SNPs located at 2p21, rs7579899 and rs11894252, gave very similar results in the stratified analysis, and only the results from one of the SNPs (rs11894252) are shown in the figure. Apart from the ORs for heterozygous and homozygous individuals, ORs and 95% CIs were estimated by the per-rare-allele log-additive trend model. All models were adjusted for sex, study and country. The overall log-additive OR is shown by the broken vertical line. *P* values indicate heterogeneity for OR within each group.

of 11q13.3 containing no characterized genes; its flanking genes are *MYEOV* (encoding *Homo sapiens* myeloma overexpressed (in a subset of t(11;14)-positive multiple myelomas)) and *CCND1* (encoding cyclin D1), situated approximately 140 kb centromeric and 220 kb telomeric, respectively, from rs7105934. In the control samples, there was little evidence for linkage disequilibrium (LD) with markers in these genes ($r^2 < 0.01$ in scanned controls). Similarly, we did not observe LD with a complex susceptibility locus for prostate cancer also identified within 11q13 (refs. 22,23) nor with a SNP marker, rs614367, 89 kb telomeric to rs7105934 that was recently associated with breast cancer risk[24].

A third locus, marked by rs4765623 on 12q24.31, also achieved genome-wide significance overall ($P = 2.6 \times 10^{-8}$, OR = 1.15, 95% CI 1.09–1.20), although it did not independently replicate using a two-tailed significance test ($P = 0.09$, OR = 1.07, 95% CI 0.99–1.16). This SNP maps to intron 1 of *SCARB1*, the scavenger receptor class B, member 1 gene, which encodes a cell-surface receptor that binds to high-density lipoprotein cholesterol (HDL-C) and mediates HDL-C uptake[25–27]. Its role in cancer biology is not as well established, and the signal at this SNP was stronger in the European studies (scan and replication studies) than in the US studies (**Fig. 2** and **Supplementary Table 3**). Although this SNP marks a promising association, further confirmatory work is required to establish its association with RCC risk.

For each of the three regions associated with RCC risk, we conducted further pooled analyses stratified by study, age, gender and established modifiable risk factors: body mass index, smoking status and history of diagnosed hypertension. The associations with rs11894252 and rs7579899 were notable in former and current smokers but not in never smokers, suggesting an interaction with smoking (*P* heterogeneity = 0.003) (**Fig. 2**). This observation raises the possibility that the effect of *EPAS1* could be dependent on tobacco smoking, but further studies are needed to explore this promising finding. The associations with the two 2p21 (*EPAS1*) SNPs were stronger among men than women, possibly a result of the different risks by smoking status. The stratified analyses suggested no other evidence of interaction.

This study was well powered to detect common alleles with large effect sizes (greater than 90% power to detect a per-allele OR of 1.5 for a variant of allele frequency of 20% at an $\alpha = 5 \times 10^{-7}$), but the statistical power was limited for detecting effects of weaker size or those due to uncommon SNPs. Additional studies are needed to identify susceptibility markers of weaker effects or lower allele frequency.

Our study has identified previously unknown regions of the genome associated with risk of RCC. Two regions on 2p21 and 12q24.31 map to the candidate genes *EPAS1* and *SCARB1*, respectively, and one maps to a region of 11q13.3 with no characterized genes. Further fine mapping of these regions is required before investigating the optimal variants for studies into the biological underpinnings of the observed associations. Moreover, these loci should be pursued in follow-up studies in distinct populations, such as African Americans, who have an increased risk of RCC[2,3]. Similarly, it will be important to evaluate these regions in studies that address clinical endpoints, such as response to therapy and survival. The discovery of additional susceptibility loci should lead to further advances in understanding the etiology of RCC as well as its risk prediction and early detection.

URLs. CGEMS portal, http://cgems.cancer.gov/; CGF, http://cgf.nci.nih.gov/; GLU, http://code.google.com/p/glu-genetics/; EIGENSTRAT, http://genepath.med.harvard.edu/~reich/EIGENSTRAT.htm; STRUCTURE, http://pritch.bsd.uchicago.edu/structure.html; PLINK, http://pngu.mgh.harvard.edu/~purcell/plink/; SAS, http://www.sas.com/; MACH, http://www.sph.umich.edu/csg/abecasis/mach/index.html; ProbABEL, http://mga.bionet.nsc.ru/~yurii/ABEL/.

METHODS

Methods and any associated references are available in the online version of the paper at http://www.nature.com/naturegenetics/.

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS

The authors thank all of the participants who took part in this research and the funders and support staff who made this study possible. Funding for the genome-wide genotyping was provided by the French Institut National du Cancer (INCa) for those studies coordinated by IARC/CNG, and by the intramural research program of the National Cancer Institute (NCI), US National Institutes of Health (NIH) for those studies coordinated by the NCI. Additional acknowledgments can be found in the **Supplementary Note**.

AUTHOR CONTRIBUTIONS

M.P.P., M.J., J.R.T., G.S., L.E.M., V.G., W.-H.C., J.D.M., N.R., S.J.C. and P. Brennan contributed to the design and execution of the overall study. M.P.P., M.J., J.R.T., G.S., L.E.M., L.A.K., X.W., V.G., K.B.J., J.D.M., N.R., S.J.C. and P. Brennan contributed to the statistical analyses. M.P.P., M.J., S.J.C. and P. Brennan wrote the first draft of the manuscript. D. Zelenika, E.P., L.A.K., X.W., K.B.J., S.H.V., S.L.v.d.M., Y.Y., A.M.M., E.S.B., N.N.C., M.F., D.L., I.G., S.H., H. Blanche, A.H., G.S.T., Z.W., M.Y., K.G.S., S.J.C. and M.L. supervised or conducted the genotyping. The remaining authors conducted the epidemiologic studies and contributed samples to the GWAS and/or replication studies. All authors contributed to the writing of the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Ferlay, J., Bray, F., Pisani, P. & Parkin, D.M. *GLOBOCAN 2002: Cancer Incidence, Mortality and Prevalence Worldwide.* (IARC CancerBase No. 5. version 2.0, IARCPress, Lyon, France, 2004).
2. Scélo, G. & Brennan, P. The epidemiology of bladder and kidney cancer. *Nat. Clin. Pract. Urol.* **4**, 205–217 (2007).
3. Chow, W.H., Dong, L.M. & Devesa, S.S. Epidemiology and risk factors for kidney cancer. *Nat. Rev. Urol.* **7**, 245–257 (2010).
4. McLaughlin, J.K. *et al.* A population–based case–control study of renal cell carcinoma. *J. Natl. Cancer Inst.* **72**, 275–284 (1984).
5. Schlehofer, B. *et al.* International renal-cell-cancer study. VI. The role of medical and family history. *Int. J. Cancer* **66**, 723–726 (1996).
6. Gago-Dominguez, M., Yuan, J.M., Castelao, J.E., Ross, R.K. & Yu, M.C. Family history and risk of renal cell carcinoma. *Cancer Epidemiol. Biomarkers Prev.* **10**, 1001–1004 (2001).
7. Hung, R.J. *et al.* Family history and the risk of kidney cancer: a multicenter case-control study in Central Europe. *Cancer Epidemiol. Biomarkers Prev.* **16**, 1287–1290 (2007).
8. Linehan, W.M. *et al.* Hereditary kidney cancer: unique opportunity for disease-based therapy. *Cancer* **115**, 2252–2261 (2009).
9. Peto, J. & Houlston, R.S. Genetics and the common cancers. *Eur. J. Cancer* **37** Suppl 8, S88–S96 (2001).
10. Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* **447**, 661–678 (2007).
11. Frazer, K.A. *et al.* A second generation human haplotype map of over 3.1 million SNPs. *Nature* **449**, 851–861 (2007).
12. Packer, B.R. *et al.* SNP500Cancer: a public resource for sequence validation, assay development, and frequency analysis for genetic variation in candidate genes. *Nucleic Acids Res.* **34**, D617–D621 (2006).
13. 1000 Genomes Project Consortium. A map of human genome variation from population-scale sequencing. *Nature* **467**, 1061–1073 (2010).
14. Skol, A.D., Scott, L.J., Abecasis, G.R. & Boehnke, M. Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. *Nat. Genet.* **38**, 209–213 (2006).
15. Higgins, J.P. *et al.* Gene expression patterns in renal cell carcinoma assessed by complementary DNA microarray. *Am. J. Pathol.* **162**, 925–932 (2003).
16. Xia, G. *et al.* Regulation of vascular endothelial growth factor transcription by endothelial PAS domain protein 1 (EPAS1) and possible involvement of EPAS1 in the angiogenesis of renal cell carcinoma. *Cancer* **91**, 1429–1436 (2001).
17. Sowter, H.M., Raval, R.R., Moore, J.W., Ratcliffe, P.J. & Harris, A.L. Predominant role of hypoxia-inducible transcription factor (Hif)-1alpha versus Hif-2alpha in regulation of the transcriptional response to hypoxia. *Cancer Res.* **63**, 6130–6134 (2003).
18. Kondo, K., Kim, W.Y., Lechpammer, M. & Kaelin, W.G. Jr. Inhibition of HIF2alpha is sufficient to suppress pVHL-defective tumor growth. *PLoS Biol.* **1**, E83 (2003).
19. Zimmer, M., Doucette, D., Siddiqui, N. & Iliopoulos, O. Inhibition of hypoxia-inducible factor is sufficient for growth suppression of VHL$^{-/-}$ tumors. *Mol. Cancer Res.* **2**, 89–95 (2004).
20. Gunaratnam, L. & Bonventre, J.V. HIF in kidney disease and development. *J. Am. Soc. Nephrol.* **20**, 1877–1887 (2009).
21. Chanock, S. High marks for GWAS. *Nat. Genet.* **41**, 765–766 (2009).
22. Thomas, G. *et al.* Multiple loci identified in a genome-wide association study of prostate cancer. *Nat. Genet.* **40**, 310–315 (2008).
23. Eeles, R.A. *et al.* Multiple newly identified loci associated with prostate cancer susceptibility. *Nat. Genet.* **40**, 316–321 (2008).
24. Turnbull, C. *et al.* Genome-wide association study identifies five new breast cancer susceptibility loci. *Nat. Genet.* **42**, 504–507 (2010).
25. Kozarsky, K.F. *et al.* Overexpression of the HDL receptor SR-BI alters plasma HDL and bile cholesterol levels. *Nature* **387**, 414–417 (1997).
26. Rigotti, A. *et al.* A targeted mutation in the murine gene encoding the high density lipoprotein (HDL) receptor scavenger receptor class B type I reveals its key role in HDL metabolism. *Proc. Natl. Acad. Sci. USA* **94**, 12610–12615 (1997).
27. Ueda, Y. *et al.* Lower plasma levels and accelerated clearance of high density lipoprotein (HDL) and non-HDL cholesterol in scavenger receptor class B type I transgenic mice. *J. Biol. Chem.* **274**, 7165–7171 (1999).

[1]Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Department Health and Human Services, Bethesda, Maryland, USA. [2]International Agency for Research on Cancer (IARC), Lyon, France. [3]Commissariat à l'énergie Atomique, Institut Genomique, Centre National de Genotypage, Evry, France. [4]Center 'Bioengineering' of Russian Academy of Sciences, Moscow, Russia. [5]Kurchatov Scientific Center, Moscow, Russia. [6]Department of Epidemiology, Division of Cancer Prevention and Population Sciences, The University of Texas MD Anderson Cancer Center, Houston, Texas, USA. [7]Department of Epidemiology, Biostatistics and Health Technology Assessment, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands. [8]Department of Urology, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands. [9]Core Genotyping Facility, SAIC-Frederick Inc., National Cancer Institute-Frederick, Frederick, Maryland, USA. [10]Russian N.N. Blokhin Cancer Research Centre, Moscow, Russia. [11]International Hereditary Cancer Center, Department of Genetics and Pathomorphology, Pomeranian Medical University, Szczecin, Poland. [12]Department of Epidemiology, Institute of Occupational Medicine, Lodz, Poland. [13]Maria Sklodowska-Curie Memorial Cancer Center and Institute of Oncology, Warsaw, Poland. [14]National Institute of Environmental Health, Department of Environmental Epidemiology, Budapest, Hungary. [15]Regional Authority of Public Health, Banská Bystrica, Slovakia. [16]Institute of Public Health, Bucharest, Romania. [17]Charles University in Prague, First Faculty of Medicine, Institute of Hygiene and Epidemiology, Prague, Czech Republic. [18]Department of Cancer Epidemiology and Genetics, Masaryk Memorial Cancer Institute, Brno, Czech Republic. [19]Palacky University, Olomouc, Czech Republic. [20]The Tisch Cancer Institute, Mount Sinai School of Medicine, New York, New York, USA. [21]Division of Epidemiology and Biostatistics, School of Public Health, University of Illinois at Chicago, Chicago, Illinois, USA. [22]Karmanos Cancer Institute and Department of Family Medicine, Wayne State University, Detroit, Michigan, USA. [23]Cancer Research UK Centre, Leeds Institute of Molecular Medicine, St James's University Hospital, Leeds, UK. [24]Department of Pathology, St James's University Hospital, Leeds, UK. [25]Division of Cancer Prevention, National Cancer Institute, National Institutes of Health, Department of Health and Human Services, Bethesda, Maryland, USA. [26]Division of Urologic Surgery, Washington University School of Medicine, St. Louis, Missouri, USA. [27]Department of Epidemiology, German Institute of Human Nutrition, Potsdam-Rehbruecke, Nuthetal, Germany. [28]School of Public Health, Imperial College London, London, UK. [29]Medical Research Council-Health Protection Agency (MRC-HPA)

Centre for Environment and Health, Imperial College London, London, UK. [30]Human Genetics Foundation (HuGeF), Torino, Italy. [31]Inserm, Centre for Research in Epidemiology and Population Health, Institut Gustave Roussy, Villejuif, France. [32]Paris South University, UMRS 1018, Villejuif, France. [33]Molecular and Nutritional Epidemiology Unit Cancer Research and Prevention Institute-ISPO, Florence, Italy. [34]Cancer Registry, Azienda Ospedaliera 'Civile MP Arezzo', Ragusa, Italy. [35]Nutritional Epidemiology Unit, Fondazione Istituto Di Ricovero e Cura a Carattere Scientifico (IRCCS) Istituto Nazionale dei Tumori, Milano, Italy. [36]Department of Clinical and Experimental Medicine, Federico II University, Naples, Italy. [37]Unit of Nutrition, Environment and Cancer, Cancer Epidemiology Research Program, Catalan Institute of Oncology (ICO-IDIBELL), Barcelona, Spain. [38]Jefe Sección Información Sanitaria, Consejería de Servicios Sociales, Principado de Asturias, Oviedo, Spain. [39]Andalusian School of Public Health, Granada, Spain. [40]CIBER Epidemiología y Salud Pública (CIBERESP), Barcelona, Spain. [41]Department of Epidemiology, Regional Council of Health and Consumer Affairs, Murcia, Spain. [42]Public Health Institute of Navarra, Pamplona, Spain. [43]Public Health Division of Gipuzkoa, Basque Regional Health Department, San Sebastian, Spain. [44]Department of Gerontology, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK. [45]Cancer Epidemiology Unit, Nuffield Department of Clinical Medicine, University of Oxford, Oxford, UK. [46]National Institute for Public Health and the Environment (RIVM), Bilthoven, The Netherlands. [47]Julius Center for Health Sciences and Primary Care, University Medical Center, Utrecht, The Netherlands. [48]Department of Epidemiology, Harvard School of Public Health, Boston, Massachusetts, USA. [49]Bureau of Epidemiologic Research, Academy of Athens, Athens, Greece. [50]Division of Cancer Epidemiology, German Cancer Research Centre, Heidelberg, Germany. [51]Institute of Epidemiology, Helmholtz Centre Munich, Munich, Germany. [52]Department of Surgical and Perioperative Sciences, Urology and Andrology, Umeå University, Umeå, Sweden. [53]Department of Epidemiology, School of Public Health, Aarhus University, Aarhus, Denmark. [54]The Danish Cancer Society, Institute of Cancer Epidemiology, Copenhagen, Denmark. [55]Epidemiology Research Program, American Cancer Society, Atlanta, Georgia, USA. [56]Department of Oncology, University of Cambridge, Cambridge, UK. [57]Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK. [58]Department of Chronic Disease Prevention, National Institute for Health and Welfare, Helsinki, Finland. [59]Department of Public Health, Faculty of Medicine, Norwegian University of Science and Technology, Trondheim, Norway. [60]Department of Community Medicine, University of Tromsø, Tromsø, Norway. [61]Department of Public Health and Primary Health Care, University of Bergen, Bergen, Norway. [62]Division of Epidemiology, Norwegian Institute of Public Health, Oslo, Norway. [63]Division of Molecular Genetic Epidemiology, German Cancer Research Center, Im Neuenheimer Feld, Heidelberg, Germany. [64]Department of Environmental Hygiene, Regional Authority of Public Health, Banska Bystrica, Slovakia. [65]CeRePP, Tenon Hospital Assistance Publique Hôpitaux de Paris (APHP) (ER2-University Paris 6), Paris, France. [66]INSERM, U946, Fondation Jean Dausset-Centre d'Etude du Polymorphisme Humain (CEPH), Paris, France. [67]Centre National de la Recherche Scientifique (CNRS) UMR8200, Institute Gustave Roussy, Villejuif, France. [68]Department of Human Genetics, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands. [69]Department of Cancer Registry and Research, Comprehensive Cancer Centre East, Nijmegen, The Netherlands. [70]Department of Urology, The University of Texas MD Anderson Cancer Center, Houston, Texas, USA. [71]Fondation Jean Dausset-CEPH, Paris, France. [72]These authors contributed equally to this work. [73]The authors jointly directed this work. Correspondence should be addressed to S.J.C. (chanocks@mail.nih.gov) or P. Brennan (brennan@iarc.fr).

ONLINE METHODS

Genome-wide SNP genotyping. New genome-wide SNP genotyping was conducted in three laboratories (**Supplementary Table 1**) using Illumina Infinium BeadChips available at the time of genotyping. All US samples were genotyped at the NCI Core Genotyping Facility (CGF, Division of Cancer Epidemiology and Genetics (DCEG), National Cancer Institute, Bethesda, Maryland, USA), whereas the Centre National de Genotypage (CNG, Evry, France) genotyped all samples from Central Europe and the HUNT2/Tromsø studies, as well as cases from EPIC, the UK and France. All Moscow samples were genotyped at the Kurchatov Scientific Center (KSC, Moscow, Russian Federation). Controls for the UK cases were drawn from data generated from the 1958 British Birth Cohort by the Wellcome Trust Sanger Institute as part of the Wellcome Trust Case Control Consortium (WTCCC)[10]. Controls from PLCO, ATBC and CPS-II were drawn from previously scanned subjects[28–30]. Controls for the EPIC cases were drawn from data generated from EPIC controls by CGF as part of the Pancreatic Cancer Cohort Consortium (PanScan)[31,32].

Quality control assessment. Systematic quality control common to both centers was conducted separately for the European and US datasets before merging the two datasets, which included quality control steps specific for the performance of different arrays at distinct times in the two main laboratories. For SNP assays, exclusions included those with less than 90% completion rate and those with extreme deviation from fitness for Hardy-Weinberg equilibrium ($P < 1 \times 10^{-7}$). Monomorphic assays observed in either cases or controls only and SNPs with alleles ambiguously coded (AT- and CG-coding alleles) were excluded.

IARC-CNG scan. After excluding 46 expected duplicate samples, the number of attempted DNA samples was 8,031. We excluded 4 pairs (8 samples) of expected duplicates that were not identical, 23 unexpected duplicate pairs (46 samples) and 112 samples with low (<95%) success rate. Samples were excluded if heterozygosity rates for autosomal chromosomes were >6 standard deviations from the mean. We further excluded one self-reported male and one female with abnormal X-chromosome heterozygosity rates (>10% and < 20%, respectively). Using a set of 12,000 unlinked SNPs (pair-wise $r^2 < 0.004$) common to all GWAS arrays[33], 59 samples with less than 80% European ancestry were excluded based on STRUCTURE analysis[34]. Eleven samples were identified as first-degree relatives and excluded based on identity-by-descent. A principal component analysis (PCA) using the EIGENSTRAT software excluded 83 additional samples detected as outliers (6 standard deviations from the mean)[35].

After these quality control steps, of the 8,031 samples genotyped, 7,542 (2,461 cases and 5,081 controls) were retained. 577,547 SNPs were available for data pooling.

NCI scan. 2,109 samples (1,490 cases and 619 new controls) were genotyped on Illumina 610 or 660w BeadChips at the Core Genotyping Facility. 3,004 previously scanned (on 550 or 610 BeadChips) samples from PLCO, CPS-II and ATBC were included. Participants were excluded based on (i) unanticipated inter-study duplicates ($n = 5$), (ii) completion rates lower than 92–94% as per the quality control groups ($n = 38$ samples), (iii) abnormal heterozygosity values of <25% or >35% ($n = 4$; two overlap with low completion samples) (iv) expected duplicates ($n = 50$ pairs), (v) abnormal X-chromosome heterozygosity ($n = 10$) and (vi) phenotype exclusions (due to ineligibility or incomplete information) ($n = 57$). Using a set of 12,000 unlinked SNPs (pairwise $r^2 < 0.004$) common to the GWAS chips used herein[33], 80 subjects with less than 85% European ancestry were excluded based on STRUCTURE analysis[34] and PCA[35]. For the known 50 duplicate pairs, concordance was 99.95%.

The final participant count for the association analysis was 1,311 cases and 3,424 controls. 585,576 SNPs were available for analysis in one or more studies.

Each participating study obtained informed consent from the study participants and approval from its Institutional Review Board; each study also obtained Institutional Review Board certification permitting data sharing in accordance with the US NIH Policy for Sharing of Data Obtained in NIH Supported or Conducted Genome-Wide Association Studies (GWAS). The Cancer Genetic Markers of Susceptibility (CGEMS) data portal provides access to individual-level data from the NCI scan only to investigators from certified scientific institutions after approval of their submitted Data Access Request.

Merging datasets. The post–quality-control datasets were merged, normalizing strand differences when necessary. No incompatible encodings were detected, and the final dataset contained 586,069 SNPs (after excluding monomorphic and ambiguously coded AT and CG SNPs) for 3,772 cases and 8,505 controls.

Statistical analysis. Associations between the 586,069 SNPs and the risk of RCC were estimated using unconditional logistic regression by the OR and 95% CI using multivariate unconditional logistic regression assuming a co-dominant–trend genetic model (in which the effect of the variant is calculated by a log-additive model with 1 degree of freedom). PCA analysis revealed two significant ($P < 0.05$) eigenvectors when included in the null model (which comprised logistic regression with dummy variables for sex, country and study for the US data). The main effect model was adjusted by sex, country, the two eigenvectors showing significant effect ($P < 0.05$) in the null model and study for the US studies. For the replication studies, both an unadjusted and an adjusted analysis were conducted; adjustment included sex, country (study), smoking status (current, former or never), body mass index and diagnosis of hypertension.

The estimated inflation factors of the test statistic were 1.011 for IARC-CNG scan, 1.016 for the NCI scan and 1.018 for the pooled scan. All P values and CIs were corrected for the appropriate observed inflation factor (genomic control)[36].

Replication and TaqMan genotyping. In order to select a set of top-ranked SNPs for further follow-up, we initially combined the European and US datasets through a meta-analysis. Genomic control was applied to the IARC-CNG and NCI scans separately[36], and the results were subsequently combined using a fixed-effects meta-analysis model, and per-allele trend effect estimates and P values were computed using inverse variance weighting (first column of **Table 1**). The individual level genome-wide data were subsequently pooled, and association results of the six SNPs selected for replication were combined with results from the replication studies by meta-analysis (third column of **Table 1**). A separate analysis of the six SNPs selected for replication is shown in **Supplementary Table 3** using alternative genetic models, namely, the dominant and recessive models. The association results of the six SNPs selected for replication are also shown separately for each study participating in the GWAS in **Supplementary Table 4**.

TaqMan genotyping assays (ABI) for replication were optimized for five of six SNPs in the three notable regions to validate the Illumina results. rs11894252 could not be manufactured, but instead, rs1867785 ($r^2 = 1.0$ in CEU HapMap Phase II) was optimized[12]. TaqMan assays for replication were genotyped in three centers: MD Anderson Cancer Center (Houston, Texas, USA), Nijmegen, The Netherlands, and IARC. Concordance of known duplicates was greater than 99%. In an analysis of 1,126 samples from three studies scanned at NCI, the comparison of the Illumina calls with the TaqMan assays showed a concordance of 98.7–100%; no shifts from wildtype to homozygotes were observed. The Illumina Infinium genotype probe cluster plots for the four SNPs achieving genome-wide significance, rs11894252, rs7579899, rs7105934 and rs4765623, are shown in **Supplementary Figure 3.**

Imputation. In order to further interrogate the loci associated with RCC, we imputed additional SNPs within 1 Mb on either side of the implicated SNPs using the MACH software and data from the 1000 Genomes Project as a scaffold[13]. Unconditional logistic regression as implemented in the ProbABEL[37] software was used to analyze the posterior SNP dosages from MACH, adjusting for sex, country, the two eigenvectors showing significant effect ($P < 0.05$) in the null model and study for the US studies. Association results for all SNPs with r^2 (squared correlation between imputed and true genotypes) above 0.3 and minor allele frequency above 0.05 in the regions of 2p21 (*EPAS1*), 11q13.3, and 12q24.31 (*SCARB1*), are shown in **Supplementary Table 2**. Also shown in **Supplementary Table 2** are the association results for each imputed SNP after adjusting for one of the implicated SNPs in each region.

Data analysis. Data analysis and management were performed with GLU (Genotyping Library and Utilities version 1.0), PLINK, SAS version 9.2, Eigenstrat, MACH and ProbABEL.

28. Yeager, M. *et al.* Genome-wide association study of prostate cancer identifies a second risk locus at 8q24. *Nat. Genet.* **39**, 645–649 (2007).
29. Hunter, D.J. *et al.* A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer. *Nat. Genet.* **39**, 870–874 (2007).
30. Landi, M.T. *et al.* A genome-wide association study of lung cancer identifies a region of chromosome 5p15 associated with risk for adenocarcinoma. *Am. J. Hum. Genet.* **85**, 679–691 (2009).
31. Amundadottir, L. *et al.* Genome-wide association study identifies variants in the ABO locus associated with susceptibility to pancreatic cancer. *Nat. Genet.* **41**, 986–990 (2009).
32. Petersen, G.M. *et al.* A genome-wide association study identifies pancreatic cancer susceptibility loci on chromosomes 13q22.1, 1q32.1 and 5p15.33. *Nat. Genet.* **42**, 224–228 (2010).
33. Yu, K. *et al.* Population substructure and control selection in genome-wide association studies. *PLoS ONE* **3**, e2551 (2008).
34. Falush, D., Stephens, M. & Pritchard, J.K. Inference of population structure using multilocus genotype data: dominant markers and null alleles. *Mol. Ecol. Notes* **7**, 574–578 (2007).
35. Price, A.L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **38**, 904–909 (2006).
36. de Bakker, P.I. *et al.* Practical aspects of imputation-driven meta-analysis of genome-wide association studies. *Hum. Mol. Genet.* **17**, R122–R128 (2008).
37. Aulchenko, Y.S., Struchalin, M.V. & van Duijn, C.M. ProbABEL package for genome-wide association analysis of imputed data. *BMC Bioinformatics* **11**, 134 (2010).

doi:10.1038/ng.723