Human Molecular Genetics, 2013, Vol. 22, No. 4 **825–831**
doi:10.1093/hmg/dds489
Advance Access published on November 25, 2012

Common variation at 2q22.3 (*ZEB2*) influences the risk of renal cancer

Marc Henrion[1,†], Matthew Frampton[1,†], Ghislaine Scelo[2,†], Mark Purdue[3,†], Yuanqing Ye[4,†], Peter Broderick[1], Alastair Ritchie[5], Richard Kaplan[5], Angela Meade[5], James McKay[2], Mattias Johansson[2], Mark Lathrop[6], James Larkin[7], Nathaniel Rothman[3], Zhaoming Wang[3,8], Wong-Ho Chow[3,4], Victoria L. Stevens[9], W. Ryan Diver[9], Susan M. Gapstur[9], Demetrius Albanes[3], Jarmo Virtamo[10], Xifeng Wu[4,‡], Paul Brennan[2,‡], Stephen Chanock[3,‡], Timothy Eisen[11,‡] and Richard S. Houlston[1,*,‡]

[1]Division of Genetics and Epidemiology, Section of Cancer Genetics, Institute of Cancer Research, Surrey SM2 5NG, UK, [2]International Agency for Research on Cancer, Lyon, France, [3]Division of Cancer Epidemiology and Genetics, Department Health and Human Services, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA, [4]Division of Cancer Prevention and Population Sciences, Department of Epidemiology, The University of Texas M.D. Anderson Cancer Center, Houston, TX, USA, [5]MRC Clinical Trials Unit, Aviation House, 125 Kingsway, London WC2B 6NH, UK, [6]Commissariat à l'Énergie Atomique, Institut Génomique, Centre National de Génotypage, Evry 91000, France, [7]Royal Marsden NHS Foundation Trust, London, UK, [8]Core Genotyping Facility, SAIC-Frederick Inc., National Cancer Institute-Frederick, Frederick, MD, USA, [9]Epidemiology Research Program, American Cancer Society, Atlanta, GA, USA, [10]Department of Chronic Disease Prevention, National Institute for Health and Welfare, Helsinki FIN-00300, Finland and [11]Cambridge University Health Partners, Cambridge, UK

Received August 7, 2012; Revised October 29, 2012; Accepted November 15, 2012

Genome-wide association studies (GWASs) of renal cell cancer (RCC) have identified four susceptibility loci thus far. To identify an additional RCC common susceptibility locus, we conducted a GWAS and performed a meta-analysis with published GWASs (totalling 2215 cases and 8566 controls of European background) and followed up the most significant association signals [nine single nucleotide polymorphisms (SNPs) in eight genomic regions] in 3739 cases and 8786 controls. A combined analysis identified a novel susceptibility locus mapping to 2q22.3 marked by rs12105918 ($P = 1.80 \times 10^{-8}$; odds ratio 1.29, 95% CI: 1.18–1.41). The signal localizes to intron 2 of the *ZEB2* gene (zinc finger E box-binding homeobox 2). Our findings suggest that genetic variation in *ZEB2* influences the risk of RCC. This finding provides further insights into the genetic and biological basis of inherited genetic susceptibility to RCC.

INTRODUCTION

Worldwide renal cancer accounts for around 2% of all cancer, affecting over 270 000 individuals and accounting for around 116 000 cancer-related deaths each year (1). Overall, renal cell carcinoma (RCC) represents 90% of cancers of the kidney in adults.

In addition to the established modifiable risk factors for RCC, which include cigarette smoking, obesity and hypertension, there is compelling evidence for inherited genetic predisposition (2). While germline inactivating mutations in *VHL* (von Hippel–Lindau syndrome), *MET* (hereditary papillary renal carcinoma), *BHD* (Birt–Hogg–Dube syndrome) and *FH* (hereditary leiomyomatosis and RCC), genes confer an increased risk of RCC (3), these are rare and collectively do not account for the 2-fold increased risk of RCC seen in relatives of RCC cases (4). Evidence that multiple low-risk variants contribute to the heritability of RCC has been provided by

*To whom correspondence should be addressed. Tel: +44 2087224175; Fax: +44 2087224365; Email: richard.houlston@icr.ac.uk
†Co-first authorship.
‡Co-last authorship.

recent genome-wide association studies (GWASs) which have identified common risk variants at 2p21, 11q13.3 and 12p11.33 (5–7).

To identify an additional novel RCC susceptibility locus, we conducted an independent primary scan of RCC and performed a genome-wide meta-analysis with one previously published GWAS followed by analysis of the top nine single nucleotide polymorphisms (SNPs) through *in silico* replication in two published GWASs (5,6).

RESULTS

In the primary scan, 1045 RCC cases were genotyped using the Illumina Omni Express BeadChip. The newly scanned cases comprised 856 cases ascertained through the TRANSORCE study of an ongoing collection of RCC cases ascertained as part of the Medical Research Council (MRC) SORCE trial through UK clinical oncology centres and 189 RCC cases collected through the Institute of Cancer Research and Royal Marsden NHS Hospitals Trust. For controls, we made use of publicly accessible Hap1.2M-Duo Custom array data generated on 2699 individuals from the Wellcome Trust Case Control Consortium 2 (WTCCC2) 1958 birth cohort (also known as the National Child Development Study) and 2501 individuals from the UK Blood Service Control Group. After applying strict quality control criteria (Materials and Methods), we restricted the analysis to the subset of genotyped SNPs common to Illumina Omni Express and Hap1.2M-Duo Custom arrays; accordingly, we analysed 451 487 SNPs for association with RCC risk for 944 cases and 5197 controls. A quantile–quantile (Q–Q) plot of observed versus expected χ^2-test statistics showed little evidence for an inflation of test statistics, thereby excluding the possibility of substantive hidden population substructure, cryptic relatedness among subjects or differential genotype calling (inflation factor $\lambda = 1.03$; Supplementary Material, Fig. S1).

We performed a meta-analysis of our primary scan data with that of the recently published GWAS of RCC, genotyped at the National Cancer Institute (NCI), which comprised four case–control series of European ancestry genotyped using Illumina HumanHap HapMap 500, 610 or 660 W BeadChips, totalling 1311 cases and 3424 control; the study design, population characteristics and genotyping platforms for the study have been previously described (6). To ensure consistency of genotyping, we restricted our analysis to genotyped SNPs that were common across the different BeadChips and did not make use of imputed data for the meta-analysis. After quality control procedures, 1271 cases and 3369 controls were used for the meta-analysis. Combining our primary scan and this GWAS provided data on 284 377 SNPs in 2215 RCC cases and 8566 controls for the meta-analysis.

Pooling data from these GWASs, we derived joint odds ratios (ORs) and confidence intervals (CIs) under a fixed-effects model for each SNP, and associated *P*-values. Excluding SNPs (including those correlated with $r^2 > 0.8$) mapping to the previously identified risk loci at 2p21, 11q13.3 and 12p11.33, we considered nine SNPs in eight regions of linkage disequilibrium (LD) that were significantly associated with RCC at $P < 5.0 \times 10^{-5}$ (Supplementary Material, Table S1). We evaluated these putative associations through *in silico* replication of nine SNPs at eight loci in independent series from MD Anderson Comprehensive Cancer Centre (MDACC; 894 cases and 1516 controls), the International Agency for Research on Cancer (IARC; 2461 cases and 5081 controls) and the Cancer Genome Atlas (TCGA) study combined with Cancer Genetic Markers of Susceptibility (CGEMS) controls (384 cases and 2189 controls). For the *in silico* replication effort, if the SNP had not been directly typed in a dataset we made use of imputed genotypes.

In the combined analysis of these datasets, rs12105918, which maps to chromosome 2q22.3 (145 208 193 bps; NCBI build 37), showed evidence for an association with RCC at genome-wide significance ($P = 1.80 \times 10^{-8}$; $P_{het} = 0.12$, $I^2 = 46\%$; Table 1). rs12105918 localizes to intron 2 of the *ZEB2* gene (zinc finger E box-binding homeobox 2; MIM:60580; Fig. 1), within a 103 kb block of LD. rs13389578, which is correlated with rs12105918 ($r^2 = 0.61$ in UK controls) provided additional support for the 2q22.3 association ($P = 2.14 \times 10^{-7}$; $P_{het} = 0.19$, $I^2 = 35\%$; Table 1).

The second strongest signal was provided by rs10054504 which maps to chromosome 5p13.3 (32 000 483 bps; NCBI build 37), within intron 4 of the *PDZD2* gene (PDZ domain-containing 2; MIM: 610697), but did not achieve genome-wide significance ($P = 7.68 \times 10^{-7}$; $P_{het} = 0.06$, $I^2 = 57\%$; Supplementary Material, Tables S2 and S3). At this time, this promising locus requires further study to confirm its association with RCC risk.

The risk of RCC associated with rs12105918 genotype showed a dose response, such that the estimated risks are compatible with a log-additive model; with relative risk to homozygotes with the high-risk alleles is increased 3.65-fold. We investigated the combined effect of 2q22.3 variation and the previously identified risk variants on chromosomes 2p.21 (two independent loci defined by rs7579899 and rs4953346), 11q13.3 (rs7105934), 12p11.23 (rs718314) on RCC risk using data from the UK-GWAS and US-GWAS datasets. There was no evidence of interactive effects between any of the loci ($P > 0.05$), compatible with the assumption that each locus has an independent role in defining RCC risk (Supplementary Material, Table S4).

Elucidation of the basis of the 2q22.3 association will require fine mapping and functional analyses. However, to explore these regions further, we imputed unobserved genotypes in cases and controls using data from the 1000 Genomes Project (Phase 1 integrated variant set release). This analysis provided no statistical evidence for a stronger signal at 2q22.3 compared with that provided by rs12105918 (Fig. 1 and Supplementary Material, Table S5). To examine whether any directly genotyped or imputed SNPs were located within a putative transcription factor-binding site or enhancer element, we conducted a bioinformatics search of the region of association using the TRANSFAC Matrix Database and PReMod software. These analyses did not provide evidence for rs12105918 or any closely correlated SNPs mapping within a known or predicted transcription regulatory region.

To explore whether the rs12105918 association (or rs13389578) could possess *cis*-acting regulatory effects on *ZEB2*, we analysed publicly available mRNA expression

Table 1. Risk of RCC associated with rs12105918 and rs13389578

Study	Genotype counts cases (AA/AB/BB)	Genotype counts controls (AA/AB/BB)	RAF[a] cases	RAF[a] controls	OR[b]	CI[c]	*P*-value	Info score
rs12105918								
UK-GWAS	802/132/8	4622/561/10	0.079	0.056	1.45	1.20–1.75	1.34×10^{-4}	
US-GWAS	1079/173/12	3007/346/11	0.078	0.055	1.38	1.14–1.67	1.13×10^{-3}	
MDACC	780/106/8	1319/189/6	0.068	0.066	1.03	0.82–1.30	8.06×10^{-1}	
IARC[d]	2099.90/350.64/10.45	4443.80/620.08/17.10	0.075	0.064	1.23	1.05–1.43	8.68×10^{-3}	0.85
TCGA[d]	321.77/61.32/0.90	1933.99/248.95/6.03	0.082	0.060	1.60	1.13–2.28	8.32×10^{-3}	0.88
rs13389578								
UK-GWAS	763/169/12	4380/786/28	0.102	0.081	1.29	1.10–1.53	2.30×10^{-3}	
US-GWAS	1022/229/19	2861/480/21	0.105	0.078	1.35	1.14–1.59	4.53×10^{-4}	
MDACC	737/143/14	1248/254/14	0.096	0.093	1.03	0.85–1.25	7.66×10^{-1}	
IARC[d]	1976.40/463.29/21.28	4194.90/851/35.04	0.103	0.091	1.16	1.02–1.32	2.57×10^{-2}	0.84
TCGA	297/86/1	1828/344/17	0.115	0.086	1.38	1.07–1.77	1.16×10^{-2}	
Meta-analysis summary		rs12105918			rs13389578			
P-value[e]		1.80×10^{-8}			2.14×10^{-7}			
I^2		46.01%			34.69%			
Heterogeneity *P*-value		0.12			0.19			

[a]Risk allele frequencies (RAFs).
[b]ORs for an additive trend model fitted using logistic regression (for the US-GWAS, we have adjusted for the study centre; for the IARC data we have adjusted for country, sex and two eigenvectors).
[c]95% CIs for the ORs.
[d]Datasets in which the SNP under consideration has been imputed.
[e]*P*-value for an inverse variance weighted, fixed effects model.

data on lymphoblastoid cell lines (LCLs), adipose tissue, fibroblast, T cell, skin and RCC tumor tissue. There was no statistically significant relationship between the SNP genotype and *ZEB2* expression in any of these tissues after adjustment for multiple testing (Supplementary Material, Table S6).

DISCUSSION

In a new GWAS of RCC, we have identified a common variant on chromosome 2q22.3 that points to a novel susceptibility locus. Since rs12105918 is intronic to *ZEB2* and the region of LD does not encompass any other genes or transcripts, there is a high likelihood that the functional basis of the 2q22.3 association is mediated through *ZEB2 a priori*. Although we failed to demonstrate any association between the genotype and ZEB2 expression, this does not preclude the possibility of a subtle cumulative long-term relationship. Additional studies are needed to investigate the effect of genetic variation at 2q22.3 on kidney tissue sets (normal and cancerous).

ZEB2 is a member of the ZEB1/Drosophila Zfh1 family of two-handed zinc finger/homeodomain proteins. It functions as a DNA-binding transcriptional repressor interacting with activated SMADs, the transducers of TGF-beta signalling, and the nucleosome remodeling and histone deacetylation complex (8). Although germline inactivating mutations in *ZEB2* cause Mowat–Wilson syndrome (MIM: 235730; Hirschsprung disease, distinct facial appearance, mental retardation and variable multiple congenital anomalies, including renal anomalies, but not RCC) there is strong biological plausibility for directly implicating *ZEB2* in RCC susceptibility. Epithelial–mesenchymal transition (EMT), which allows for cellular dissociation from epithelial tissues, is a key embryonic process that is reactivated during tumourigenesis (9). Enforced expression of ZEB factors in epithelial cells results in a rapid EMT associated with a breakdown of cell polarity, loss of cell–cell adhesion and induction of cell motility. ZEB2 along with SNAI1, SNAI2, ZEB1, TWIST1 and TWIST2 are key EMT regulators (9). ZEB2 has also been reported to repress transcription of CDH1, CLDN4, CCND1, TERT, SFRP1, ALPL and miR-200b-200a-429 primary miRNA and upregulates transcription of mesenchymal markers (9). A role for hypoxia-inducible factor in the development of RCC is well established and TGFbeta, TNFalpha, IL1 and hypoxia signals directly upregulate ZEB2 to induce EMT, growth arrest, and senescence, whereas Hedgehog signals indirectly upregulate ZEB2 via TGFbeta (9,10). While the findings from our GWAS provide additional evidence for *ZEB2* being a key gene in RCC development, additional studies are needed to identify the functionally relevant common variants associated with increased RCC risk.

It is increasingly being recognized that some genetic variants can influence the risk of more than one cancer type. To explore the possibility that rs12105918 affects the risk of other malignancies, we examined the association with colorectal (11), breast (12), prostate (13) and lung cancers (14), acute lymphoblastic leukaemia (15), multiple myeloma (16), Hodgkin's lymphoma (17), glioma (18) and meningioma (19) using data from previously reported GWASs. However, for these cancers, there was no evidence of rs12105918 (or the correlated SNP rs1389578) having pleiotropic effects on tumour risk (i.e. $P > 0.05$).

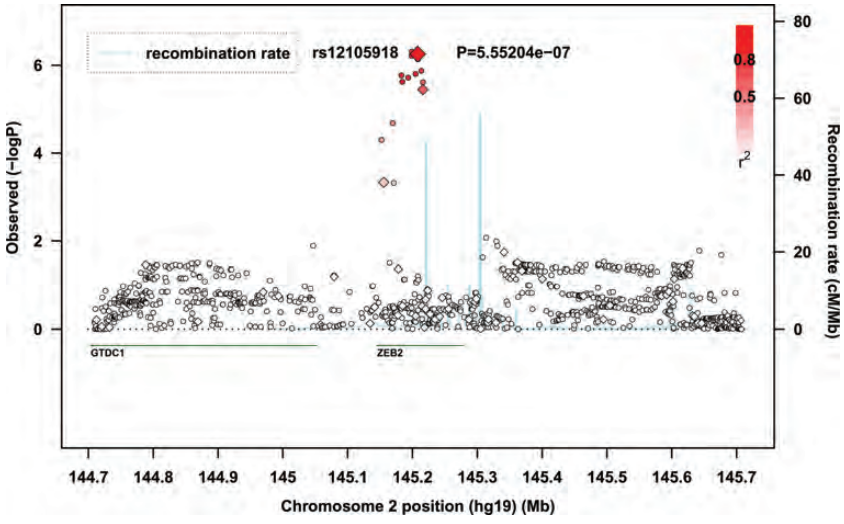

Figure 1. Regional plot of association results and recombination rates for the 2q22.3 risk locus. Association results of genotyped (triangles) and imputed (circles) SNPs in the GWAS samples and recombination rates. $-\log_{10}P$ values (y-axis) of the SNPs are shown according to their chromosomal positions (x-axis). rs12105918 in the combined analysis is denoted by a large triangle. The colour intensity of each symbol reflects the extent of LD with rs12105918: white ($r^2 = 0$) through to dark red ($r^2 = 1.0$), with r^2 estimated from the UK control samples. Genetic recombination rates (cM/Mb), estimated using 1000 Genomes Pilot 1 CEU samples, are shown with a light blue line. Physical positions are based on NCBI build 37 of the human genome. Also shown are the relative positions of genes and transcripts mapping to each region of association. Genes have been redrawn to show the relative positions; therefore, the maps are not to physical scale.

In summary, we have identified a novel RCC susceptibility locus, implicating genetic variation in *ZEB2* in the development of RCC. Given that the modest size of our new scan and that the established identified susceptibility loci at 2p21, 2q22.3, 11q13.3 and 12p11.33 collectively only account for ~4% of the familial RCC risk, it is likely that further risk variants can be identified through the meta-analysis of additional GWAS-based analyses.

MATERIALS AND METHODS

Ethics statement

Collection of blood samples and clinico-pathological information from all subjects was undertaken with informed consent and ethical review board approval from each site in accordance with the tenets of the Declaration of Helsinki.

Subjects and datasets

UK-GWAS

The UK-GWAS cases comprised adult patients with histologically proven RCC collected through two sources within the UK. First, 856 cases from SORCE, a MRC collection of surgically treated RCC cases ascertained through UK clinical oncology centres. Second, 189 RCC cases collected through the ICR and Royal Marsden NHS Hospitals Trust. Cases included 590 clear cell carcinomas (CCCs), 42 papillary carcinomas (PCs), 33 chromophobe carcinomas (CCs) and 19 mixed or other histological subtypes. DNA was extracted from EDTA-venous blood samples using the conventional methods and quantified using PicoGreen (Invitrogen).

Cases were genotyped using the Human OmniExpress-12 BeadChip according to the manufacturer's recommendations (Illumina Inc, San Diego, CA, USA). Genotyping quality control was tested using duplicate DNA samples, together with direct sequencing significant SNPs in a subset of samples to confirm genotyping accuracy. For all SNPs, >99% concordant results were obtained. For controls, we made use of publicly accessible Hap1.2M-Duo Custom array data generated on 2699 individuals from the Wellcome Trust Case Control Consortium 2 (WTCCC2) 1958 birth cohort (also known as the National Child Development Study) and 2501 individuals from the UK Blood Service Control Group. We excluded individuals from analysis if they failed one or more of the following thresholds: overall successfully genotyped SNPs < 97% ($n = 3$), discordant sex information ($n = 1$), outliers in a plot of heterozygosity versus missingness ($n = 24$), duplication or cryptic relatedness to the estimated identity by descent (IBD) 0.185 ($n = 14$) and evidence of non-white European ancestry by PCA-based analysis in comparison with HapMap samples ($n = 62$; cut-off based on the minimum and maximum values of the top two principal components of the controls) (Supplementary Material, Fig. S3). We excluded SNPs from analysis if they failed one or more of the following thresholds: call rates <95% ($n = 0$); different missing genotype rates between cases and controls at $P < 10^{-5}$ ($n = 3426$); MAF < 0.01 ($n = 14$); departure from Hardy–Weinberg equilibrium in controls at $P < 10^{-5}$ ($n = 642$). The details of all sample exclusions are provided in Supplementary Material, Figure S4. The adequacy of the case–control matching and the possibility of differential genotyping of cases and controls were assessed using Q–Q plots of test statistics. The inflation factor λ_{GC} was calculated by dividing the median of the lower 90% of the test statistics by the

median of the lower 90% of the expected values from a χ^2 distribution with 1 d.f.

US-GWAS

The US NCI GWAS of RCC was based on 1453 RCC cases and 3599 controls of European background from three US and one European study genotyped using Illumina HumanHap HapMap 500, 610 or 660W BeadChips. The study design of each participating study and population characteristics have been previously described (6). Data were publicly available on 1311 cases (including 534 CCCs, 93 PCs, 86 other histological subtypes) and 3424 controls. After applying the same quality control as that performed for the UK-GWAS, 1271 cases and 3369 controls were available for the meta-analysis (Supplementary Material, Fig. S4). The inflation factor after adjustment for the study centre was 1.02 (Supplementary Material, Fig. S2).

Replication series

For *in silico* replication, we used data from three distinct studies: (1) The University of Texas MDACC GWAS which comprised 894 incident RCC cases (including 612 CCCs, 81 PCs, 39 CCs and 88 mixed or other histological subtypes) recruited from MDACC and 1516 healthy controls with no prior history of cancer (except non-melanoma skin cancer) of European descent (5). Genotyping of both cases and controls was performed using Illumina Infinium HumanHap660W arrays (2). The IARC GWAS comprising 2461 RCC cases (including 1340 CCCs, 95 PCs, 88 other histological subtypes) and 5081 controls of European background from seven European studies (6). Genotyping of cases and controls was performed using either Illumina HumanHap300, 550 or 610 Quad Beadchips (6). To harmonize data derived from the three arrays, we made use of imputation to recover untyped genotypes. The common set of SNPs did not include rs12105918 and rs13389578; however for 1365 cases and 2086 controls which were directly typed for these SNPs, the concordance between the typed and imputed SNP genotypes was high (97.8% for rs12105918 and 97.3% for rs13389578) (3). Data from TCGA on RCC cases were genotyped using the Affymetrix Genome-Wide Human SNP Array 6.0. For controls we made use of controls from the CGEMS studies of breast and prostate cancers which were genotyped using Illumina HumanHap550 and Phase 1A HumanHap300+Phase 1BHumanHap240 Beadchips respectively (12,13). After quality control including checks for relatedness, European ancestry and overlap with the US-GWAS data were available on 384 cases and 2189 controls (Supplementary Material, Fig. S4).

Statistical and bioinformatic analysis

Analyses were primarily undertaken using R (v2.14.2), STATA v.10 (State College) and PLINK (v1.07) software. The association between each SNP and the risk of RCC was assessed by the Cochran–Armitage trend test. ORs and associated 95% CIs were calculated by unconditional logistic regression. Prediction of the untyped SNPs was carried out using IMPUTEv2 (v2.2.2) based on the data from the 1000 Genomes Project (Phase 1 integrated variant set release). Imputed data were analyzed using SNPTEST v2.3.0 to account for uncertainties in SNP prediction. Association meta-analyses only included markers with info scores >0.8, imputed call rates/SNP >0.9 and MAFs >0.01. Meta-analyses were carried out with META v1.4 using the genotype probabilities from IMPUTEv2, where an SNP was not directly typed. We calculated Cochran's Q statistic to test for heterogeneity and the I^2 statistic to quantify the proportion of the total variation that was caused by heterogeneity. The I^2 values $\geq 75\%$ are generally considered to indicate substantial heterogeneity.

LD blocks were defined on the basis of HapMap recombination rate (cM/Mb) as defined using the Oxford recombination hotspots and on the basis of the distribution of CIs defined by Gabriel *et al.* (20).

The familial relative risk of RCC attributable to a variant was calculated using the formula (21):

$$\lambda^* = \frac{p(pr_2 + qr_1)^2 + q(pr_1 + q)^2}{(p^2r_2 + 2pqr_1 + q^2)^2},$$

where P is the population frequency of the minor allele, $q = 1 - P$, and r_1 and r_2 are the relative risks (approximated by ODs) for heterozygotes and the rarer homozygotes relative to the more common homozygotes respectively. From λ^*, it is possible to quantify the influence of the locus on the overall familial risk of RCC in first-degree relatives of RCC patients. Assuming a multiplicative interaction between risk alleles, the proportion of the overall familial risk attributable to the locus is given by $\log(\lambda^*)/\log(\lambda_0)$, where λ_0, the overall familial risk of RCC, shown in epidemiological studies is 2.45 (4).

Relationship between SNP genotype and mRNA expression

The associations of SNP genotype with gene expression were investigated in three publicly available Sentrix Human-6 Expression BeadChips (Illumina, San Diego, USA) datasets. LCLs from HapMap3 CEU individuals (22), three cell types (fibroblast, lymphoblastoid cell line and T cell) derived from umbilical cords of 75 Geneva GenCord individuals (23), three tissue types (166 adipose, 156 LCLs and 160 skin samples) derived from a subset of healthy female twins of the MuTHER resource (24). We also examined the relationship between the genotype and expression in RCC using TCGA data generated using Agilent 244K Custom G4502A arrays.

URLs

The R suite can be found at http://www.r-project.org
Illumina: http://www.illumina.com
dbSNP: http://www.ncbi.nlm.nih.gov/projects/SNP
HapMap: http://www.hapmap.org
1000 Genomes: http://www.1000genomes.org
SNAP http://www.broadinstitute.org/mpg/snap
IMPUTE: https://mathgen.stats.ox.ac.uk/impute/impute
SNPTEST: http://www.stats.ox.ac.uk/~marchini/software/gwas/snptest

Wellcome Trust Case Control Consortium: www.wtccc.org.uk
Mendelian Inheritance In Man: http://www.ncbi.nlm.nih.gov/omim
Cancer Genome Atlas project: http://cancergenome.nih.gov
Genevar (GENe Expression VARiation): http://www.sanger.ac.uk/resources
SORCE: http://www.ctu.mrc.ac.uk
Cancer Genetic Markers of Susceptibility (CGEMS): cgems.cancer.gov

SUPPLEMENTARY MATERIAL

Supplementary Material is available at *HMG* online.

ACKNOWLEDGEMENTS

We thank the study participants and their families and the study investigators and coordinators for work in recruitment. This study made use of genotyping data from the 1958 Birth Cohort and National Blood Service samples, kindly made available by the Wellcome Trust Case Control Consortium 2. A full list of the investigators who contributed to the generation of the data is available at http://www.wtccc.org.uk/. The authors would like to acknowledge the participants and researchers from the following IARC/CNG studies, EPIC, HUNT2, NCI/IARC Central Europe study, ASHRAM, CeRePP, the Leeds cohort, the Search study and the Moscow case-control study. Further details of these studies may be found in the supplementary material of (6). The results published here are in whole or part based upon data generated by The Cancer Genome Atlas pilot project established by the NCI and NHGRI. Information about TCGA and the investigators and institutions who constitute the TCGA research network can be found at http://cancergenome.nih.gov/. Finally, we acknowledge the work of the following US individuals Lee E. Moore (Division of Cancer Epidemiology and Genetics, NCI, National Institutes of Health, Department Health and Human Services), Kevin B. Jacobs (Division of Cancer Epidemiology and Genetics, NCI, National Institutes of Health, Department Health and Human Services; Core Genotyping Facility, SAIC-Frederick Inc., National Cancer Institute-Frederick); Jorge R. Toro (Division of Cancer Epidemiology and Genetics, NCI, National Institutes of Health, Department Health and Human Services); Joanne S. Colt (Division of Cancer Epidemiology and Genetics, NCI, National Institutes of Health, Department Health and Human Services); Faith G. Davis (Division of Epidemiology/Biostatistics, School of Public Health, University of Illinois at Chicago); Kendra L. Schwartz (Karmanos Cancer Institute and Department of Family Medicine, Wayne State University); Christine D. Berg (Division of Cancer Prevention, NCI, National Institutes of Health, Department of Health and Human Services); Robert L. Grubb III (Division of Urologic Surgery, Washington University School of Medicine); Michelle A. Hildebrandt (Department of Epidemiology, Division of Cancer Prevention and Population Sciences, The University of Texas M.D. Anderson Cancer Center), Xia Pu (Department of Epidemiology, Division of Cancer Prevention and Population Sciences, The University of Texas M.D. Anderson Cancer Center); Amy Hutchinson (Division of Cancer Epidemiology and Genetics, NCI, National Institutes of Health, Department Health and Human Services; Core Genotyping Facility, SAIC-Frederick Inc., National Cancer Institute-Frederick); Joseph F. Fraumeni Jr (Division of Cancer Epidemiology and Genetics, NCI, National Institutes of Health, Department Health and Human Services) and Meredith Yeager (Division of Cancer Epidemiology and Genetics, NCI, National Institutes of Health, Department Health and Human Services; Core Genotyping Facility, SAIC-Frederick Inc., National Cancer Institute-Frederick).

Conflict of Interest statement. None declared.

FUNDING

SORCE is coordinated by the Medical Research Council (MRC) and funded principally by the MRC and Cancer Research UK with an educational grant from Bayer. Additional funding was provided by Cancer Research UK (C1298/A8362 supported by the Bobby Moore Fund). Marc Henrion was supported by Leukaemia Lymphoma Research. NHS funding for the Royal Marsden Biomedical Research Centre and Cambridge University Health Partners is acknowledged. Funding for the SEARCH team was provided by Cancer Research UK (C490/A10124). The US kidney GWAS was funded by the Intramural Research Program of the National Cancer Institute, NIH.

REFERENCES

1. Ferlay, J., Shin, H.R., Bray, F., Forman, D., Mathers, C. and Parkin, D.M. (2010) Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. *Int. J. Cancer*, **127**, 2893–2917.
2. Chow, W.H., Dong, L.M. and Devesa, S.S. (2010) Epidemiology and risk factors for kidney cancer. *Nat. Rev. Urol.*, **7**, 245–257.
3. Linehan, W.M., Pinto, P.A., Bratslavsky, G., Pfaffenroth, E., Merino, M., Vocke, C.D., Toro, J.R., Bottaro, D., Neckers, L., Schmidt, L.S. *et al.* (2009) Hereditary kidney cancer: unique opportunity for disease-based therapy. *Cancer*, **115**, 2252–2261.
4. Goldgar, D.E., Easton, D.F., Cannon-Albright, L.A. and Skolnick, M.H. (1994) Systematic population-based assessment of cancer risk in first-degree relatives of cancer probands. *J. Natl Cancer Inst.*, **86**, 1600–1608.
5. Wu, X., Scelo, G., Purdue, M.P., Rothman, N., Johansson, M., Ye, Y., Wang, Z., Zelenika, D., Moore, L.E., Wood, C.G. *et al.* (2012) A genome-wide association study identifies a novel susceptibility locus for renal cell carcinoma on 12p11.23. *Hum. Mol. Genet.*, **21**, 456–462.
6. Purdue, M.P., Johansson, M., Zelenika, D., Toro, J.R., Scelo, G., Moore, L.E., Prokhortchouk, E., Wu, X., Kiemeney, L.A., Gaborieau, V. *et al.* (2011) Genome-wide association study of renal cell carcinoma identifies two susceptibility loci on 2p21 and 11q13.3. *Nat. Genet.*, **43**, 60–65.
7. Han, S.S., Yeager, M., Moore, L.E., Wei, M.H., Pfeiffer, R., Toure, O., Purdue, M.P., Johansson, M., Scelo, G., Chung, C.C. *et al.* (2012) The chromosome 2p21 region harbors a complex genetic architecture for association with risk for renal cell carcinoma. *Hum. Mol. Genet.*, **21**, 1190–1200.
8. Verstappen, G., van Grunsven, L.A., Michiels, C., Van de Putte, T., Souopgui, J., Van Damme, J., Bellefroid, E., Vandekerckhove, J. and Huylebroeck, D. (2008) Atypical Mowat-Wilson patient confirms the importance of the novel association between ZFHX1B/SIP1 and NuRD corepressor complex. *Hum. Mol. Genet.*, **17**, 1175–1183.
9. Peinado, H., Olmeda, D. and Cano, A. (2007) Snail, Zeb and bHLH factors in tumour progression: an alliance against the epithelial phenotype? *Nat. Rev. Cancer*, **7**, 415–428.

10. Krishnamachary, B., Zagzag, D., Nagasawa, H., Rainey, K., Okuyama, H., Baek, J.H. and Semenza, G.L. (2006) Hypoxia-inducible factor-1-dependent repression of E-cadherin in von Hippel–Lindau tumor suppressor-null renal cell carcinoma mediated by TCF3, ZFHX1A, and ZFHX1B. *Cancer Res.*, **66**, 2725–2731.
11. Dunlop, M.G., Dobbins, S.E., Farrington, S.M., Jones, A.M., Palles, C., Whiffin, N., Tenesa, A., Spain, S., Broderick, P., Ooi, L.Y. *et al.* (2012) Common variation near CDKN1A, POLD3 and SHROOM2 influences colorectal cancer risk. *Nat. Genet.*, **44**, 770–776.
12. Hunter, D.J., Kraft, P., Jacobs, K.B., Cox, D.G., Yeager, M., Hankinson, S.E., Wacholder, S., Wang, Z., Welch, R., Hutchinson, A. *et al.* (2007) A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer. *Nat. Genet.*, **39**, 870–874.
13. Yeager, M., Orr, N., Hayes, R.B., Jacobs, K.B., Kraft, P., Wacholder, S., Minichiello, M.J., Fearnhead, P., Yu, K., Chatterjee, N. *et al.* (2007) Genome-wide association study of prostate cancer identifies a second risk locus at 8q24. *Nat. Genet.*, **39**, 645–649.
14. Broderick, P., Wang, Y., Vijayakrishnan, J., Matakidou, A., Spitz, M.R., Eisen, T., Amos, C.I. and Houlston, R.S. (2009) Deciphering the impact of common genetic variation on lung cancer risk: a genome-wide association study. *Cancer Res.*, **69**, 6633–6641.
15. Papaemmanuil, E., Hosking, F.J., Vijayakrishnan, J., Price, A., Olver, B., Sheridan, E., Kinsey, S.E., Lightfoot, T., Roman, E., Irving, J.A. *et al.* (2009) Loci on 7p12.2, 10q21.2 and 14q11.2 are associated with risk of childhood acute lymphoblastic leukemia. *Nat. Genet.*, **41**, 1006–1010.
16. Broderick, P., Chubb, D., Johnson, D.C., Weinhold, N., Forsti, A., Lloyd, A., Olver, B., Ma, Y.P., Dobbins, S.E., Walker, B.A. *et al.* (2011) Common variation at 3p22.1 and 7p15.3 influences multiple myeloma risk. *Nat. Genet.*, **44**, 58–61.
17. Enciso-Mora, V., Broderick, P., Ma, Y., Jarrett, R.F., Hjalgrim, H., Hemminki, K., van den Berg, A., Olver, B., Lloyd, A., Dobbins, S.E. *et al.* (2010) A genome-wide association study of Hodgkin's lymphoma identifies new susceptibility loci at 2p16.1 (REL), 8q24.21 and 10p14 (GATA3). *Nat. Genet.*, **42**, 1126–1130.
18. Shete, S., Lau, C.C., Houlston, R.S., Claus, E.B., Barnholtz-Sloan, J., Lai, R., Il'yasova, D., Schildkraut, J., Sadetzki, S., Johansen, C. *et al.* (2010) Genome-wide high-density SNP linkage search for glioma susceptibility loci: results from the gliogene Consortium. *Cancer Res.*, **71**, 7568–7755.
19. Dobbins, S.E., Broderick, P., Melin, B., Feychting, M., Johansen, C., Andersson, U., Brannstrom, T., Schramm, J., Olver, B., Lloyd, A. *et al.* (2011) Common variation at 10p12.31 near MLLT10 influences meningioma risk. *Nat. Genet.*, **43**, 825–827.
20. Gabriel, S.B., Schaffner, S.F., Nguyen, H., Moore, J.M., Roy, J., Blumenstiel, B., Higgins, J., DeFelice, M., Lochner, A., Faggart, M. *et al.* (2002) The structure of haplotype blocks in the human genome. *Science*, **296**, 2225–2229.
21. Houlston, R.S. and Ford, D. (1996) Genetics of coeliac disease. *Q.J.M.*, **89**, 737–743.
22. Stranger, B.E., Montgomery, S.B., Dimas, A.S., Parts, L., Stegle, O., Ingle, C.E., Sekowska, M., Smith, G.D., Evans, D., Gutierrez-Arcelus, M. *et al.* (2012) Patterns of cis regulatory variation in diverse human populations. *PLoS Genet.*, **8**, e1002639.
23. Dimas, A.S., Deutsch, S., Stranger, B.E., Montgomery, S.B., Borel, C., Attar-Cohen, H., Ingle, C., Beazley, C., Gutierrez Arcelus, M., Sekowska, M. *et al.* (2009) Common regulatory variation impacts gene expression in a cell type-dependent manner. *Science*, **325**, 1246–1250.
24. Nica, A.C., Parts, L., Glass, D., Nisbet, J., Barrett, A., Sekowska, M., Travers, M., Potter, S., Grundberg, E., Small, K. *et al.* (2011) The architecture of gene regulatory variation across multiple human tissues: the MuTHER study. *PLoS Genet*, **7**, e1002003.