Common Variation at 1q24.1 (ALDH9A1) Is a Potential Risk Factor for Renal Cancer

Marc Y. R. Henrion1, Mark P. Purdue2‡, Ghislaine Scelo10‡, Peter Broderick1, Matthew Frampton1, Alastair Ritchie4, Angela Meade6, Peng Li10, James McKay10, Mattias Johansson10, Mark Lathrop11, James Larkin5, Nathaniel Rothman2, Zhaoming Wang2,6, Wong-Ho Chow2,3, Victoria L. Stevens7, W. Ryan Diver7, Demetrius Albanes3, Jarmo Virtamo8, Paul Brennan10, Timothy Eisen9‡, Stephen Chanock2‡, Richard S. Houlston1‡*

1 Division of Genetics and Epidemiology, The Institute of Cancer Research, London, United Kingdom, 2 Division of Cancer Epidemiology and Genetics, Department Health and Human Services, National Cancer Institute, National Institutes of Health, Bethesda, Maryland, United States of America, 3 Division of Cancer Prevention and Population Sciences, Department of Epidemiology, The University of Texas M.D. Anderson Cancer Center, Houston, Texas, United States of America, 4 MRC Clinical Trials Unit at University College London, Aviation House, London, United Kingdom, 5 Royal Marsden NHS Foundation Trust, London, United Kingdom, 6 Cancer Genomics Research Laboratory, Leidos Biomedical Research Inc., Gaithersburg, Maryland, United States of America, 7 Epidemiology Research Program, American Cancer Society, Atlanta, Georgia, United States of America, 8 Department of Chronic Disease Prevention, National Institute for Health and Welfare, Helsinki, Finland, 9 Cambridge University Health Partners, Cambridge, United Kingdom, 10 International Agency for Research on Cancer, Lyon, France, 11 McGill University and Genome Quebec Innovation Centre, Montreal, Quebec, Canada

‡ MYRH, MPP, and GS are co-first authors on this work. TE, SC, and RSH are co-last authors on this work.
* richard.houlston@icr.ac.uk

Abstract

So far six susceptibility loci for renal cell carcinoma (RCC) have been discovered by genome-wide association studies (GWAS). To identify additional RCC common risk loci, we performed a meta-analysis of published GWAS (totalling 2,215 cases and 8,566 controls of Western-European background) with imputation using 1000 Genomes Project and UK10K Project data as reference panels and followed up the most significant association signals [22 single nucleotide polymorphisms (SNPs) and 3 indels in eight genomic regions] in 383 cases and 2,189 controls from The Cancer Genome Atlas (TCGA). A combined analysis identified a promising susceptibility locus mapping to 1q24.1 marked by the imputed SNP rs3845536 ($P_{\text{combined}}=2.30\times10^{-8}$). Specifically, the signal maps to intron 4 of the ALDH9A1 gene (aldehyde dehydrogenase 9 family, member A1). We further evaluated this potential signal in 2,461 cases and 5,081 controls from The International Agency for Research on Cancer (IARC) GWAS of RCC cases and controls from multiple European regions. In contrast to earlier findings no association was shown in the IARC series ($P=0.94$; $P_{\text{combined}}=2.73\times10^{-5}$). While variation at 1q24.1 represents a potential risk locus for RCC, future replication analyses are required to substantiate our observation.
Leukaemia Lymphoma Research. JL is funded by the RMH/ICR Biomedical Research Centre for Cancer. National Health Service (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 NCI kidney GWAS was funded by the Intramural Research Program of the National Cancer Institute, National Institutes of Health (NIH). The funders, except MRC CTU, had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: Zhaoming Wang is employed by Leidos Biomedical Research Inc., a US government contractor. Timothy Eisen is Chief Investigator of SORCE and has received research support from Bayer, GlaxoSmithKline, Pfizer and AstraZeneca, has taken part and been compensated for advisory boards for GlaxoSmithKline, Aveo and Astellas and is currently on leave of absence from the University of Cambridge to work as Chief Clinician Scientist at AstraZeneca. SORCE, though coordinated by the Medical Research Council (MRC) Clinical Trials Unit (CTU) at UCL and funded principally by the MRC CTU at UCL and Cancer Research UK, has been funded partially by an educational grant from a commercial source, Bayer. This does not alter the authors’ adherence to PLOS ONE policies on sharing data and materials.

Introduction

Worldwide kidney cancer accounts for around 2% of all malignancies the disease affecting 270,000 individuals and causing 116,000 cancer-related deaths each year [1]. In adults 90% of kidney cancers are renal cell carcinomas (RCC) [2].

Besides the well-recognised modifiable risk factors for RCC—smoking and obesity-related traits, as well as the inverse relationship between risk and alcohol consumption, there is strong evidence for an inherited genetic predisposition [3]. Rare germline mutations in VHL (von Hippel—Lindau syndrome), MET (hereditary papillary renal carcinoma), BHD (Birt—Hogg—Dube syndrome) and FH (hereditary leiomyomatosis and RCC syndrome) dramatically increase the risk of RCC [4], but contribute little to the overall two-fold familial risk [5]. Evidence for polygenic susceptibility to RCC has recently been vindicated by genome-wide association studies (GWAS) that have identified risk SNPs (single nucleotide polymorphisms) at 2p21, 2q22.3, 8q24.21, 11q13.3, 12p11.33 and 12q24.31 [2,6—9].

To identify additional RCC risk SNPs, we imputed over 10 million SNPs in two published GWAS datasets, using data from the 1000 Genomes Project [10] and UK10K projects as reference (see Materials & Methods for details). This allowed us to recover untyped genotypes, thereby maximising the prospects of identifying novel risk variants for RCC. We then conducted a genome-wide meta-analysis of the two imputed studies.

Results

For the meta-analysis we made use of data from two previously published GWAS of RCC: (i) UK-GWAS, 1,045 RCC cases genotyped on Illumina Omni Express BeadChips with 2,699 individuals from the Wellcome Trust Case Control Consortium 2 (WTCCC2) 1958 birth cohort and 2,501 UK Blood Service which had been genotyped on Hap1.2M-Duo arrays serving as controls [2]; (ii) The National Cancer Institute (NCI) GWAS (NCI-GWAS), consisting of four European case-control series, totalling 1,311 cases and 3,424 controls, genotyped on HumanHap HapMap 500, 610 or 660W BeadChips [7].

Post quality control these GWAS provided data on a total of 2,215 cases and 8,566 controls. To maximise identification of novel risk variants, we imputed over 10 million SNPs using 1000 Genomes Project and UK10K data as reference. Quantile-quantile (Q-Q) plots for all SNPs post-imation did not show substantive over-dispersion (λ = 1.02 and 1.01 for UK-GWAS and NCI-GWAS respectively; S1 Fig.).

We pooled the data from these two GWAS and used an inverse-variance weighted fixed-effects meta analysis model to compute odds ratios (OR), confidence intervals (CI) and P-values for each SNP. Results from this meta-analysis, annotated with known risk loci, are shown on Fig. 1. We excluded SNPs that (i) directly mapped to previously published risk loci (2p21, 2q22.3, 8q24.21, 11q13.3, 12p11.33 and 12q24.31; S1 Table), (ii) were in linkage disequilibrium (LD; at a threshold of r2 > 0.8) with SNPs from these loci or (iii) had P > 0.01 in either the UK or the NCI dataset. After applying these filters, we considered 22 SNPs and 3 indels in eight regions of LD that showed evidence for association with RCC risk at P < 1.0 × 10−6 (S2 Table). To validate these potential associations, we conducted replication in cases and controls obtained from combining The Cancer Genome Atlas (TCGA) Kidney Renal Clear Cell Carcinoma (KIRC) and Cancer Genetic Markers of Susceptibility (CGEMS) datasets (383 cases and 2,189 controls; S3 Table).

In an analysis combining these three datasets, rs3845536, mapping to chromosome 1q24.1 (165,650,787 bps; NCBI build 37), achieved genome-wide significance (P = 2.30 × 10−8; Phet = 0.24, I2 = 29%; Table 1) for association with RCC risk. This association was driven by the NCI (P = 9.40 × 10−7) and UK (P = 4.61 × 10−7) studies and was not nominally significant in the TCGA study (P = 0.16). However, in the latter, smaller, study the effect is of similar size and in the
same direction as in the UK and NCI studies, thereby boosting the association signal in the meta-analysis.

rs3845536 localizes to intron 4 of the \textit{ALDH9A1} gene (aldehyde dehydrogenase, family 9, subfamily a, member 1; MIM 602733; Fig. 2), within a 64kb block of LD. We confirmed the high fidelity of imputation by directly genotyping rs3845536 in a random subset of the UK-GWAS (516 cases, $r^2 = 0.99$ and 402 controls, $r^2 = 0.98$, Materials and Methods). The RCC risk associated with rs3845536 genotype is compatible with a log-additive model, the OR for risk allele homozygotes being 1.51 (95% CI: 1.29–1.77).

We did not find evidence for interactions between 1q24.1 and any of the previously published risk loci—specifically we evaluated the interaction effects on RCC risk of rs3845536 with SNPs on 2p21 (rs7579899 and rs4953346), 2q22.3 (rs12105918), 8q24.1 (rs6470588 and rs6470589), 11q13.3 (rs7105934), 12p11.23 (rs718314) and 12q24.31 (rs4765623). The assumption of independent RCC risk loci was supported by the lack of significant interaction terms between the risk loci ($i.e.$ $P > 0.05$; S4 Table).

Using publicly available mRNA expression data, we evaluated the potential for \textit{cis}-regulation of \textit{ALDH9A1} or other nearby gene by rs3845536 variation. There was no statistically significant relationship between the genotype of rs3845536 or a SNP in LD with rs3845536 (at $r^2>0.8$) and expression of \textit{ALDH9A1} and the nearby transcripts \textit{MGST3} and \textit{TMCO1} (expression data for transcripts LOC440700 and BC071770, also in the region, were not available). Further, a Haploreg and RegulomeDb search did not yield evidence for rs3845536 or a correlated SNP to locate within a transcription regulatory region (data not shown). We also made use of TCGA clear cell data to examine the frequency of mutation of \textit{ALDH9A1}, \textit{MGST3}, \textit{LOC440700} and \textit{TMCO1} in renal cancer [11]. None of these genes have mutational frequencies in RCC $>1\%$ (no data were available for transcript BC071770).

To further examine this association we made use of data from the International Agency for Research on Cancer (IARC) GWAS of RCC which was based on eight independent case-control series from different European countries with 41.4\% of cases from Western and Northern Europe, and 58.6\% from Central and Eastern Europe. In the IARC series there was no evidence for an association between rs3845536 and risk of RCC ($P = 0.94$; Table 1). Hence overall, the association strength was markedly reduced with concomitant significant heterogeneity with inclusion of the IARC dataset ($P = 2.73 \times 10^{-5}$, $P_{\text{het}} = 9.1 \times 10^{-4}$, $I^2 = 82\%$; Table 1).

**Discussion**

We report a newly identified common variant on chromosome 1q24.1 annotating a potential RCC susceptibility locus candidate. If confirmed by additional studies there is a high likelihood...
Table 1. Risk of RCC associated with rs3845536.

<table>
<thead>
<tr>
<th>Locus nearest genes</th>
<th>Variant</th>
<th>Position (hg19)</th>
<th>Allele</th>
<th>RAF cases</th>
<th>RAF controls</th>
<th>OR</th>
<th>CI (1.05–1.29)</th>
<th>P_{fixed}</th>
<th>IS</th>
<th>RAF cases</th>
<th>RAF controls</th>
<th>OR</th>
<th>CI (1.15–1.41)</th>
<th>P_{fixed}</th>
<th>IS</th>
<th>IARC meta-analysis^c</th>
<th>UK, US &amp; TCGA meta-analysis^c</th>
</tr>
</thead>
<tbody>
<tr>
<td>Iq24.1</td>
<td>MGST3, ALDH9A1, TMC01, LOC440700</td>
<td>rs3845536</td>
<td>165,650,787</td>
<td>C/T</td>
<td>0.68</td>
<td>0.64</td>
<td>1.16</td>
<td>0.64</td>
<td>0.64</td>
<td>0.69</td>
<td>0.66</td>
<td>1.30</td>
<td>1.17–1.44</td>
<td>9.40E–07</td>
<td>0.99</td>
<td>0.90 (0.85–0.94)</td>
<td>2.73E–05 82 9.12E–04</td>
</tr>
<tr>
<td></td>
<td></td>
<td>rs10918242</td>
<td>165,656,600</td>
<td>A/G</td>
<td>0.67</td>
<td>0.63</td>
<td>1.16</td>
<td>0.63</td>
<td>0.63</td>
<td>0.69</td>
<td>0.66</td>
<td>1.27</td>
<td>1.15–1.41</td>
<td>5.28E–06</td>
<td>0.99</td>
<td>0.99 (0.99–1.42)</td>
<td>7.05E–02 0.83</td>
</tr>
<tr>
<td></td>
<td></td>
<td>rs34072474</td>
<td>165,658,829</td>
<td>G/A</td>
<td>0.67</td>
<td>0.63</td>
<td>1.16</td>
<td>0.63</td>
<td>0.63</td>
<td>0.69</td>
<td>0.66</td>
<td>1.27</td>
<td>1.15–1.41</td>
<td>4.86E–06</td>
<td>0.99</td>
<td>0.98 (0.98–1.41)</td>
<td>7.74E–02 0.83</td>
</tr>
<tr>
<td></td>
<td></td>
<td>rs7541817</td>
<td>165,659,714</td>
<td>C/T</td>
<td>0.67</td>
<td>0.63</td>
<td>1.16</td>
<td>0.63</td>
<td>0.63</td>
<td>0.69</td>
<td>0.66</td>
<td>1.27</td>
<td>1.15–1.41</td>
<td>5.36E–06</td>
<td>0.98</td>
<td>0.98 (0.98–1.41)</td>
<td>8.01E–02 0.83</td>
</tr>
<tr>
<td></td>
<td></td>
<td>rs4307543</td>
<td>165,660,029</td>
<td>G/T</td>
<td>0.67</td>
<td>0.63</td>
<td>1.16</td>
<td>0.63</td>
<td>0.63</td>
<td>0.69</td>
<td>0.66</td>
<td>1.27</td>
<td>1.15–1.41</td>
<td>5.27E–06</td>
<td>0.98</td>
<td>0.98 (0.98–1.41)</td>
<td>8.04E–02 0.82</td>
</tr>
</tbody>
</table>

Shown are all variants in the locus achieving genome-wide significance (P_{fixed} < 5x10^-8) in the combined analysis of UK, NCI and TCGA data. Replication for rs3845536 is also shown.

RAF = risk allele frequency, OR = odds ratio, CI = confidence interval, IS = imputation accuracy score

^a nearest genes = genes within 50kb of rs3845536

^b alleles are given as risk & other allele

^c all meta-analysis results are for an inverse variance weighted, fixed effects model

^d the IARC results are for rs3845536 only and are the result of a meta-analysis of 8 studies from various European countries; the IS for each of the 8 studies was 0.99

doi:10.1371/journal.pone.0122589.t001
that the functional basis of the 1q24.1 risk locus is mediated through \textit{ALDH9A1} \textit{a priori} since the region of association is small and rs3845536 is intronic to \textit{ALDH9A1}. Although we did not observe an association between rs3845536 genotype and \textit{ALDH9A1} expression, a subtle relationship between the two, such as a cumulative, long term interaction, remains a possibility.

The ALDH gene superfamily is documented \cite{12} to include a variety of isozymes involved in the metabolism of aldehydes generated from chemically diverse endogenous and exogenous precursors. Aldehyde-mediated effects vary from homeostatic and therapeutic to cytotoxic, and genotoxic and several ALDHs have been implicated in human disease phenotypes or pathophysiologies \cite{12}. \textit{ALDH9A1} encodes \(\gamma\)-trimethylaminobutyraldehyde dehydrogenase that participates in the metabolism of \(\gamma\)-aminobutyraldehyde and aminoaldehydes derived from polyamines \cite{12}. High levels of \textit{ALDH9A1} expression are seen in the kidney \cite{13} with significant enrichment of dehydrogenases including \textit{ALDH9A1} in RCC \cite{14}. TNF signalling is well established to play a role in RCC development \cite{15} and it is notable that \textit{ALDH9A1} influences expression of TNF alpha induced protein 3 \cite{16}. Although speculative these data are consistent with the hypothesis of xenobiotic metabolism associated with apoptosis and tumorigenesis playing a role in RCC oncogenesis. While our finding adds evidence that \textit{ALDH9A1} is implicated in RCC development, further studies are required to determine the variants that are functionally relevant.

To interrogate whether rs3845536 has pleiotropic effects on the risks of other cancer types, we investigated the association with colorectal \cite{17} and lung cancers \cite{18}, acute lymphoblastic leukaemia \cite{19}, multiple myeloma \cite{20}, glioma \cite{21} and meningioma \cite{22} using data from previously reported GWAS. However, our data did not support this hypothesis and we did not
observe, for any of these cancers, a significant effect of rs3845536 genotype (or a correlated SNP at $r^2 \geq 0.8$) on tumor risk.

In summary, we report a potential RCC risk susceptibility locus candidate at rs3845536. This finding implicates genetic variation in ALDH9A1 in the development of RCC. Similar to other GWAS hits, rs3845536 is a common variant and confers moderate risk of RCC. However compelling our finding is from analysis of UK, NCI and TCGA data due to the failure to validate the association in the IARC series the observation has to be viewed with a degree of caution at this juncture and further replication is required. We note that due to both the modest size of our discovery dataset and the fact that published RCC susceptibility loci at 2p21, 2q22.3, 8q24.21, 11q13.3, 12p11.33 and 12q24.31 account for <5% of the familial risk additional risk variants are likely to be identifiable through expanded GWAS analyses.

Materials and Methods

Ethics statement

Collection of blood samples and clinico-pathological information from all subjects was undertaken with written informed consent with ethical board approvals from the Royal Marsden NHS Hospitals Trust (CCR 1552/1922) and the United Kingdom Multicentre Research Ethics Board (07/MRE01/10). Details about Ethics approval for the NCI, TCGA and IARC studies are detailed previously [7].

Subjects and datasets

GWAS datasets have been previously reported [2]. (i) UK-GWAS was based on 1,045 RCC cases (including 590 clear cell carcinomas (CCCs), 42 papillary carcinomas (PCs), 33 chromophobe carcinomas (CCs) and 19 mixed or other histological subtypes) genotyped using Human OmniExpress-12 BeadChips, with 856 cases from the MRC SORCE trial and 189 cases collected through The Institute of Cancer Research (ICR) and Royal Marsden NHS Hospitals Trust and 5,200 controls genotyped using Hap1.2M-Duo Custom array with 2,699 individuals from the Wellcome Trust Case Control Consortium 2 (WTCCC2) 1958 birth cohort and 2,501 from the UK Blood Service. (ii) NCI-GWAS was based on 1,453 RCC cases and 3,599 controls of European background genotyped using Illumina HumanHap HapMap 500, 610 or 660W BeadChips. Data were publicly available on 1,311 cases (including 534 CCCs, 93 PCs, 86 other histological subtypes) and 3,424 controls [7].

As we previously described [2], we applied a number of pre-specified quality control metrics to the data. Specifically we used the following criteria to exclude individuals: overall successfully genotyped SNPs < 97%, discordant sex, outliers in a plot of heterozygosity versus missingness, duplication or relatedness to the estimated identity by descent (IBD) > 0.185 or evidence of non-European ancestry by PCA-based analysis using HapMap reference samples (S2 Fig.). SNP exclusion criteria included: call rate < 95%; different missing genotype rates between cases and controls at $P < 10^{-5}$; MAF < 0.01; departure from Hardy—Weinberg equilibrium in controls at $P < 10^{-5}$. An overview of all sample exclusions is given in S3 Fig. Adequacy of the case—control matching was assessed by inspection of Q—Q plots of test statistics and by means of the inflation factor $\lambda_{GC}$.

Replication series

For replication, we used, as detailed previously [2], data from TCGA and IARC. Briefly, the TCGA RCC clear cell cases (KIRC study, accession number phs000178.v7.p6) were genotyped using the Affymetrix Genome-Wide Human SNP Array 6.0. For controls we made use of data
on healthy individuals from the CGEMS breast and prostate cancer study, genotyped using Illumina HumanHap550 and Phase 1A HumanHap300+Phase 1B HumanHap240 Beadchips respectively. Both cases and controls were formally examined for an overlap with the NCI GWAS samples. Any TCGA or CGEMS sample found to be a duplicate of or related to a sample from the NCI GWAS was removed from the replication dataset. After further checking for relatedness and European ancestry 383 cases and 2,189 controls constituted the TCGA/CGEMS replication series. The International Agency for Research on Cancer (IARC) GWAS consisted of 2,461 RCC cases (including 1,340 CCCs, 95 PCs, 88 other histological subtypes) and 5,081 controls of European background from eight European studies) and has previously been described [7]. Genotyping of cases and controls was performed using either Illumina HumanHap300, 550 or 610 Quad Beadchips. Data derived from the three arrays were imputed to recover rs3845536 genotype.

Statistical and bioinformatic analyses

R (v3.02) and SNPTTEST (v2.4.1) software were used for analysis. Association between individual SNPs and RCC risk was evaluated by the Cochran—Armitage trend test. Unconditional logistic regression was used to calculate ORs and associated 95% CIs. The UK-GWAS did not require any covariates to adjust for, the NCI-GWAS required adjusting for study centre and the TCGA-GWAS required adjusting for the first principal component. Phasing of GWAS SNP genotypes was performed using SHAPEIT v2.644. Untyped SNPs were imputed using IMPUTEv2 (v2.3.0) with data from the 1000 Genomes Project (Phase 1 integrated variant set, v3.20101123, released on the IMPUTEv2 website on 9 December 2013) and UK10K (ALSPAC, EGAS00000100090 / EGAD000001000195, and TwinsUK, EGAS000001000108 / EGAD000001000194, studies only) used as reference panels. Analysis of imputed data was conducted using SNPTTEST v2.4.1 to account for uncertainties in SNP prediction. Association meta-analyses only included markers with info scores >0.4, imputed call rates/SNP >0.9 (UK & NCI studies) and MAFs >0.005. Meta-analyses were carried out with the R package meta v2.4–1, using the genotype probabilities from IMPUTEv2 for untyped SNPs. Heterogeneity was assessed using Cochran’s Q statistic and the proportion of the total variation due to heterogeneity was assessed using the I² statistic.

HapMap recombination rate (cM/Mb) was used to define LD blocks. The recombination rate defined using the Oxford recombination hotspots and on the basis of the distribution of CIs defined by Gabriel and co-workers [23].

The fidelity of imputation, as assessed by the concordance between imputed and directly genotyped SNPs, was examined in a random subset of samples from the UK-GWAS. To quantify the fidelity of imputation we calculated Pearson’s correlation coefficient r² between directly genotyped values (counting the number of reference alleles, taking discrete values in {0, 1, 2}) and the imputed genotypes (taking real values in the interval [0,2]).

The familial relative risk of RCC attributable to a specific variant was calculated using the formula from [24]:

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

where the overall sibling relative risk λ₀ for RCC is 2.45 [5].

Fig. 2 has been produced using visPIG [25].

Analysis of TCGA data

The associations of SNP genotype with gene expression in RCC was investigated using TCGA data generated using Agilent 244K Custom G4502A arrays. The frequency of mutations was obtained using the CBioPortal for Cancer Genomics web server.
Supporting Information
Supporting information is available at PLOS ONE online.

URLs
   Illumina: http://www.illumina.com
   HapMap: http://www.hapmap.org
   1000Genomes: http://www.1000genomes.org
   visPIG: http://vispig.icr.ac.uk
   IMPUTE: https://mathgen.stats.ox.ac.uk/impute/impute
   SNPTEST: http://www.stats.ox.ac.uk/~marchini/software/gwas/snptest
   cBioPortal for Cancer Genomics: http://www.cbiportal.org
   Wellcome Trust Case Control Consortium: www.wtccc.org.uk
   The 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

Supporting Information
S1 Dataset. UK & NCI association test results with meta-analysis results. Tab-delimited ASCII text file with one header row.
   (TXT)
S1 Fig. Q-Q plots of Cochran-Armitage trend test statistics for association based on meta-analysis of UK-GWAS and NCI-GWAS pre-imputation (a-b); post-imputation (e-h) and rare SNPs post-imputation (i-l). The identity line is indicated as a blue dashed line.
   (TIF)
S2 Fig. first two principal components of the UK and NCI datasets, as used for removing samples based on ancestry during quality control. Case and control samples are indicated as grey and black crosses, with the HapMap reference populations shown as bold coloured discs.
   (TIF)
S3 Fig. GWAS data quality control. Details are provided of samples, SNPs and quality control (QC) used in each GWAS.
   (TIF)
S1 Table. Evidence for association at previously reported RCC susceptibility loci. At each locus values are given for the previously reported SNPs and the lead SNP in this study.
   (PDF)
S2 Table. UK & NCI meta-analysis for all variants taken through to the replication stage.
   (PDF)
S3 Table. UK, NCI & TCGA meta-analysis for all variants taken through to the replication stage. Shown in bold are the variants achieving $P_{\text{fixed}} < 5 \times 10^{-8}$.
   (PDF)
S4 Table. significance of the interaction terms of rs3845536 with previously published risk SNPs for RCC.

(PDF)

Acknowledgments
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.wtcc.org.uk/. 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/. This study makes use of data generated by the UK10K Consortium, derived from samples from the ALSPAC and TwinsUK studies. A full list of the investigators who contributed to the generation of the data is available from www.UK10K.org. Funding for UK10K was provided by the Wellcome Trust under award WT091310. 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; Cancer Genomics Research Laboratory, Leidos Biomedical Research Inc.); 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); 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; Cancer Genomics Research Laboratory, Leidos Biomedical Research Inc.); 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; Cancer Genomics Research Laboratory, Leidos Biomedical Research Inc.).

The authors would like to acknowledge the participants and researchers from the following IARC 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 Purdue et al, Nature Genetics, 2011.

Author Contributions
Conceived and designed the experiments: RSH SC P. Brennan MPP. Performed the experiments: AR AM PL JM MJ ML JL NR ZW WHC VLS WRD DA JV TE. Analyzed the data: MYRH MPP GS MF. Wrote the paper: MYRH RSH MPP GS P. Broderick SC.
References


