nature genetics

Article https:/

Multi-ancestry genome-wide association study of kidney cancer identifies 63 susceptibility regions

Received: 8 August 2023

Accepted: 13 March 2024

Published online: 26 April 2024

Check for updates

A list of authors and their affiliations appears at the end of the paper

Here, in a multi-ancestry genome-wide association study meta-analysis of kidney cancer (29,020 cases and 835,670 controls), we identified 63 susceptibility regions (50 novel) containing 108 independent risk loci. In analyses stratified by subtype, 52 regions (78 loci) were associated with clear cell renal cell carcinoma (RCC) and 6 regions (7 loci) with papillary RCC. Notably, we report a variant common in African ancestry individuals (rs7629500) in the 3′ untranslated region of *VHL*, nearly tripling clear cell RCC risk (odds ratio 2.72, 95% confidence interval 2.23–3.30). In *cis*-expression quantitative trait locus analyses, 48 variants from 34 regions point toward 83 candidate genes. Enrichment of hypoxia-inducible factor-binding sites underscores the importance of hypoxia-related mechanisms in kidney cancer. Our results advance understanding of the genetic architecture of kidney cancer, provide clues for functional investigation and enable generation of a validated polygenic risk score with an estimated area under the curve of 0.65 (0.74 including risk factors) among European ancestry individuals.

Kidney cancer is a commonly diagnosed malignancy with an estimated 430,000 new cases diagnosed worldwide in 2020[1]. The incidence of kidney cancer varies internationally; the highest rates are observed mainly in Western populations, at a 2:1 male:female ratio[2]. The 5-year relative survival for kidney cancer in the United States is 78% according to data from the Surveillance, Epidemiology and End Results Program (2013–2019), ranging from 93% for localized tumors to 17% for cancer with metastatic spread to distant sites at diagnosis[3]. In the United States, incidence rates of disease are higher among American Indian/Alaskan Native and non-Hispanic Black populations than among non-Hispanic white Americans, while the Hispanic and Asian/Pacific Islander populations experience lower rates[4]. These patterns are probably explained at least in part by differences in the prevalence of established risk factors (for example, obesity, smoking and hypertension)[5] and, possibly, host genetic factors.

The most common form of kidney cancer is renal cell carcinoma (RCC), derived from epithelial cells. The most common histologic subtypes of RCC are clear cell (ccRCC; over 75% of RCC cases), papillary (pRCC; 10%) and chromophobe (chRCC; 5%) carcinomas; rarer subtypes together account for less than 5% of cases[6]. Highly penetrant, rare mutations have been described in 14 genes (*BAP1*, *FLCN*, *FH*, *MET*, *PTEN*, *SDHB*, *SDHC*, *SDHD*, *ELOC*, *MITF*, *PRDM10*, *TSC1*, *TSC2* and *VHL*), accounting for between 3% and 5% of RCC[7–10].

The contribution of common low-penetrance genetic variants to kidney cancer heritability has been investigated through progressively larger genome-wide association studies (GWAS). The largest previous GWAS, based on 10,784 cases and 20,406 controls of European ancestry, identified 13 susceptibility regions associated with kidney cancer risk at genome-wide significance ($P < 5 \times 10^{-8}$)[11]. Functional investigation of select GWAS regions has elucidated local underlying biological mechanisms, namely altered regulation of key genes in four regions directly (for example, *MYC* for 8q24, *CCND1* for 11q13.3, *BHLHE41* for 12p12.1 and *DPF3* for 14q24)[12–15] and, in conjunction with obesity, cholesterol auxotrophy has been linked to *SCARB1* at 12q24.31 (ref. 16). Further investigation has revealed that kidney cancer susceptibility is related to allele-specific binding affinities of key renal transcription factors (TFs) such as hypoxia-inducible factors (HIFs) or paired box 8 (PAX8), particularly in enhancers[17–19]. Here we used a multi-ancestry

e-mail: purduem@mail.nih.gov; chanocks@mail.nih.gov

https://doi.org/10.1038/s41588-024-01725-7

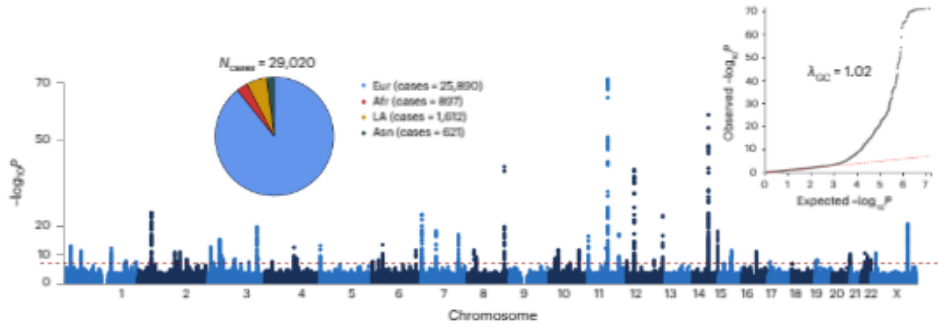


Fig. 1 | Manhattan plot for multi-ancestry GWAS of kidney cancer (29,020 cases and 835,670 controls) using fixed-effects meta-analysis model. Nominal statistical significance is shown as $-\log_{10}P$ of two-sided *z* statistics on the vertical axis. The red horizontal dashed line corresponds to genome-wide significance *P* value threshold of 5×10^{-8}. The left inset shows a pie chart for the number of cases across different population groups as defined by GrafPop analysis (Methods): Eur, Afr, LA and Asn. Most of the component substudies contained samples from a single population (Supplementary Table 1). For the NCI-3 scan, a multi-ancestry study, genetic ancestry was determined using GrafPop (Methods). The right inset shows the quantile–quantile (QQ) plot for the same GWAS with the estimated genomic inflation factor.

GWAS meta-analysis of kidney cancer to identify 63 susceptibility regions (50 novel) containing 108 independent risk loci.

Results

Study overview

A multi-ancestry kidney cancer meta-analysis combined summary statistics from seven published studies (six involving subjects of European ancestry, the seventh involving African Americans), three large biobanks (UK Biobank, FinnGen and Biobank Japan) and a newly genotyped study (National Cancer Institute 3 (NCI-3)). After quality control metrics were applied, the samples from the newly genotyped study cohort were stratified into six subcohorts: one based on samples collected specifically in Brazil (1,320 cases and 1,229 controls) and five based on genetic similarity clusters (European-like, 9,963 cases and 67,687 controls, denoted 'Eur' herein based on 1000 Genomes; African-like, 642 and 2,734, denoted 'Afr'; East Asian-like, 98 and 490, denoted 'Asn'; Latin American-like 1, 144 and 211, denoted 'LA1'; Latin American-like 2, 148 and 740, denoted 'LA2') from North America, South America and Europe (Methods and Supplementary Table 1). The newly genotyped and previously genotyped study samples were imputed to a total of 26,781,105 single-nucleotide polymorphisms (SNPs) across similarity clusters (imputation quality score >0.5 and minor allele frequency (MAF) >0.005) using the Trans-Omics for Precision Medicine (TOPMed) Imputation Server. With the addition of kidney cancer cases and controls from large biobanks, the final kidney cancer meta-analysis included 29,020 cases and 835,670 controls.

Multi-ancestry GWAS meta-analysis

We performed genome-wide association meta-analyses within four population strata (Eur, Afr, Asn and a stratum including Brazilian, LA1 and LA2 samples, denoted 'LA' herein) for kidney cancer and the two most common RCC subtypes, ccRCC ($n = 16,321$) and pRCC ($n = 2,193$), with adjustment for sex and principal components. Multi-ancestry meta-analyses of all constituent studies were conducted using inverse variance-weighted fixed-effects models, as well as random effects and Han–Eskin random effects models[20]. Little evidence of test statistic inflation was observed in the multi-ancestry meta-analyses (Supplementary Table 2).

The multi-ancestry fixed-effects meta-analysis of kidney cancer confirmed the 13 previously reported cytoband susceptibility regions (Supplementary Table 3) and identified 50 additional cytoband regions associated at a level of genome-wide significance ($P < 5 \times 10^{-8}$; Fig. 1). Performing linkage disequilibrium (LD) clumping based on the in-sample LD estimated from Eur participants within these 63 regions, we observed at least 108 loci independently associated with kidney cancer (2-Mb windows with $r^2 < 0.2$ in Supplementary Table 3 and regional plots in Supplementary Fig. 1). The significant associations are primarily driven by signals in Eur participants, which comprise the majority of samples in the meta-analysis (89% overall); 18 out of the 108 loci did not have a genome-wide significant *P* value in analysis restricted to Eur participants but had strong suggestive evidence of association (maximum *P* value of 4.1×10^{-7}).

At 3p25.3, a complex pattern of the underlying genetic architecture across ancestries was observed in which a variant mapping to the 3′ untranslated region (UTR) of *VHL* (rs7629500), and present only on African haplotypes among Afr and LA participants (effect allele frequency 10% and 2%, respectively, versus ~0% in Eur and Asn participants), was strongly associated with kidney cancer (odds ratio (OR) of 1.86 per effect allele; $P = 1.6 \times 10^{-13}$), while a second locus (rs139729777), located 106 kb upstream of *VHL*, was associated with risk across ancestries (OR of 1.09; $P = 8.9 \times 10^{-11}$).

Our genome-wide analyses of X chromosome variants, not investigated in past kidney cancer GWAS investigations, identified susceptibility loci at Xp22.2 and Xq25 (Supplementary Table 3). While the genetic mechanism driving the Xq25 signal is unclear, the Xp22.2 variant rs6629201 is an intron mapping to *REPS2* with a notable *cis*-expression quantitative trait locus (*cis*-eQTL) in Genotype-Tissue Expression Project (GTEx)[21] kidney cortex samples. *REPS2* encodes part of a protein complex that inhibits growth factor signaling. The risk T allele is associated with lower *REPS2* expression, which is also a predictor of poor ccRCC prognosis in The Cancer Genome Atlas Kidney Renal Clear Cell Carcinoma (TCGA–KIRC)[22] samples (proteinatlas.org)[23] and has also been implicated in the progression of other malignancies (for example, prostate, lung, esophageal and liver cancers)[24–28].

Conditional analyses were performed for the 108 individual loci distributed across 63 cytoband regions to identify secondary independent signals in the significant loci using in-sample LD estimated from Eur participants (Methods), identifying an additional two signals at 8q24.21 and 20q13.33 that became significant after adjusting for the GWAS index SNP (Supplementary Table 4). When we used random effects and Han–Eskin random effects models, genome-wide significant associations were observed for 51 cytoband regions ($n = 79$ loci) and 59 cytoband regions (99 loci), respectively (Supplementary Table 5). For each of the 108 loci associated with kidney cancer at genome-wide significance, we conducted additional meta-analyses stratified by sex, body mass index (BMI), smoking status, and history of diagnosed

https://doi.org/10.1038/s41588-024-01725-7

Table 1 | Regions containing loci associated with clear cell carcinoma at genome-wide significance ($P<5.0\times10^{-8}$) in the multi-ancestry GWAS meta-analysis that were not among the regions identified in genome-wide analyses for kidney cancer overall

Locus	Location[b]	rsID	Nearest gene	EA/ NEA	EAF	ccRCC: multi-ancestry meta-analysis (16,321 cases and 743,479 controls)			ccRCC: population-specific (Eur/ LA/Afr)[a]		Kidney cancer: multi-ancestry meta-analysis (29,020 cases and 835,670 controls)	
						OR (95% CI)	P	I^2 (%)	EAF	OR	OR (95% CI)	P
1p36.33	1161855	rs9442385	TTLL10	C/T	0.81	1.15 (1.10–1.21)	3.0×10^{-9}	0.0	0.83/0.88/0.81	1.15/1.20/1.10	1.10 (1.06–1.14)	1.8×10^{-6}
1q21.3	155007143	rs72885114	ZBTB76	A/G	0.10	0.87 (0.83–0.92)	3.1×10^{-6}	0.0	0.10/0.07/0.02	0.88/0.83/0.74	0.91 (0.87–0.94)	5.9×10^{-6}
7q21.2	92752382	rs147341073	CDK6	G/A	0.03	1.27 (1.17–1.38)	8.5×10^{-9}	0.0	0.03/0.03/0.01	1.30/0.95/0.62	1.19 (1.10–1.28)	1.2×10^{-5}
9q31.3	110026153	rs260189	PALM2AKAP2	C/T	0.62	1.09 (1.06–1.12)	3.2×10^{-11}	0.0	0.61/0.65/0.76	1.09/1.07/0.95	1.05 (1.03–1.08)	1.4×10^{-6}
10q22.1	72276671	rs4747241	DDIT4	T/C	0.39	1.08 (1.05–1.11)	4.5×10^{-10}	12.9	0.38/0.45/0.56	1.08/1.02/1.27	1.06 (1.04–1.08)	1.6×10^{-7}
11p11.2	47253513	rs10838681	NR1H3	A/G	0.28	0.91 (0.89–0.94)	6.3×10^{-9}	0.0	0.27/0.28/0.45	0.92/0.88/0.87	0.94 (0.92–0.97)	9.0×10^{-5}
22q13.31	46817636	rs714024	GRAMD4	T/C	0.56	1.09 (1.06–1.12)	1.6×10^{-8}	4.8	0.59/0.48/0.30	1.09/1.13/1.04	1.06 (1.04–1.08)	2.5×10^{-7}

EA, effect allele; NEA, non effect allele; EAF, effect allele frequency; P, P value (two-sided); I^2 (%), heterogeneity statistic for meta-analysis. [a]Asn samples were excluded from GWAS analyses of ccRCC and pRCC due to small numbers of cases with known histology. [b]Location extracted from human genome assembly GrCh38.

hypertension (Supplementary Table 6). Tests of heterogeneity across these strata did not identify differences in effects across strata for any loci ($108\times4=432$ tests; Bonferroni threshold 1.2×10^{-4}).

Subtype analysis

In a multi-ancestry meta-analysis of ccRCC (16,321 confirmed cases and 743,479 controls), genome-wide significant associations were observed in 52 regions, which included 78 loci (Supplementary Table 7). Since the majority of kidney cancer cases are ccRCC, it is not surprising that 46 of the 63 regions associated with kidney cancer also achieved genome-wide significance for ccRCC in the subset with confirmed subtype pathology, while the 17 remaining kidney cancer regions were associated with ccRCC below this significance threshold (Supplementary Table 8). Conversely, the analysis of ccRCC also identified seven additional susceptibility regions (1p36.33, 1q21.3, 7q21.2, 9q31.3, 10q22.1, 11p11.2 and 22q13.31) that had not achieved genome-wide significance in analyses of overall kidney cancer (Table 1 and Supplementary Table 7). Notably, in a meta-analysis of the less-common pRCC subtype (2,193 cases and 740,816 controls), we report, to our knowledge, the first genome-wide significant associations to be identified for this subtype, in six regions (seven independent loci; Table 2), one of which is shared with ccRCC and kidney cancer (7q32.1, mapping to *IRF5*).

In tests for heterogeneity in effect sizes between ccRCC and pRCC subtypes for the 108 kidney cancer loci, we observed significant subtype differences for 7 loci in seven regions (1q22, 1q32.1, 3p25.3, 11q13.3, 14q24.2, 14q32.33 and 19p13.3; 108 tests, Bonferroni-corrected threshold of 4.6×10^{-4}), all of which involved stronger effect sizes for ccRCC (Supplementary Table 8). Notable among these findings was the *VHL* 3′ UTR variant rs7629500, which demonstrated a particularly strong association with ccRCC (OR 2.72, 95% confidence interval (CI) 2.23–3.30; $P=1.3\times10^{-23}$; Fig. 2a–c) and a null finding for pRCC (OR 0.96, 95% CI 0.70–1.33; $P=0.82$). Similarly, the nearby 3p25.3 locus rs139729777 demonstrated an effect specific to ccRCC (Supplementary Table 8). In analyses of cases subgrouped by age at diagnosis, we observed stronger effect sizes with earlier-onset kidney cancer for a few loci (for example, at 2p21, 3q11.2 and 4q25), although tests of heterogeneity were not significant at the Bonferroni-corrected threshold (Supplementary Table 8).

The overall genetic correlation between ccRCC and pRCC was estimated to be 0.51 (Methods), indicating that the subtypes have substantial overlap in the underlying genetic architecture of susceptibility. As noted earlier, one particular locus at 7q32.1 achieved genome-wide significance for both subtypes (index variants: rs3807306 and rs3778754 for ccRCC and pRCC, respectively; Fig. 3a–c). The consensus credible sets for ccRCC (8 variants) and pRCC (11 variants) at this locus overlapped substantially (7 variants: rs3778754, rs3807307, rs3823536, rs3778753, rs3778752, rs3778751 and rs3757387), indicating that the association results for the subtypes are driven by one or more common causal variants. This overlap was further confirmed using colocalization (probability of a shared causal variant across ccRCC and pRCC is 98.3%). We also found that the locus overlapped with potentially active enhancers (H3K27ac peaks) across both subtypes (Supplementary Table 9) and harbored significant *cis*-eQTLs for the expression of *IRF5* in both TCGA–KIRC (ccRCC) and TCGA pRCC (TCGA–KIRP, for pRCC) as well as GTEx whole blood (Fig. 3d–f and Supplementary Table 10). *IRF5* is a TF implicated in various auto-immune diseases including rheumatoid arthritis, inflammatory bowel disease and systemic lupus erythematosus, and has been reported to influence development of cancers of the stomach, breast and thyroid[29–31]. Notably, hyperactivation of *IRF5* in systemic lupus erythematosus has been shown to contribute to lupus nephritis severity and impair kidney function[31]. This suggests the potential function of this locus might be related to immune function, perhaps common to both subtypes. In fact, several variants present in the credible sets for both subtypes disrupted the TF-binding motifs for genes, including *NFKB2* and *RUNX2* (Fig. 3g,h), which have been implicated both in immune response and in different cancers.

Pleiotropy of kidney cancer loci

A search of the GWAS catalog for cancer pleiotropic effects for the identified kidney cancer loci and other correlated SNPs (Methods) yielded 20 cytoband regions (26 loci) with genome-wide significant associations for a total of 16 cancers (Extended Data Fig. 1 and Supplementary Table 11). Some of the shared loci map to regions with genes associated with multiple cancer types, thus raising the possibility of shared pathways in carcinogenesis. For instance, kidney cancer susceptibility loci at three of the six known loci within the *TERT-CLPTM1L* region on 5p15.33 (rs7734992, rs2853677 and rs33987166) are shared with ten other cancers (bladder, breast, colon/rectum, brain, lung, melanoma, non-melanoma cancers (NMSC), ovaries, prostate and testis). Kidney cancer also shares three independent loci at 11q22.3 within the large *ATM* gene, important in susceptibility to cancers of the breast, melanoma and prostate, as well as rare myeloproliferative neoplasms. Our findings also highlight the 10q24.33 region, which harbors the *OBFC1* gene associated with cancers of the breast, lung, ovaries, thyroid, melanoma and NMSC. Furthermore, the locus marked by rs78378222 on 17p13.1 maps to the 3′ end of *TP53*, which is also a locus for seven other cancers (breast, glioma, chronic lymphocytic leukemia, lung, melanoma, NMSC and prostate). In a separate search for pleiotropy with major kidney cancer risk factors within the GWAS catalog

https://doi.org/10.1038/s41588-024-01725-7

Table 2 | Summary of pRCC susceptibility loci in multi-ancestry GWAS meta-analysis at genome-wide significance ($P<5.0\times10^{-8}$). Associations with ccRCC and overall kidney cancer for these variants are also shown

Locus	Location[a]	rsID	Nearest gene	EA/NEA	EAF	pRCC (2,193 cases and 740,816 controls)			ccRCC (16,321 cases and 743,479 controls)		Kidney cancer (29,020 cases and 835,670 controls)	
						OR (95% CI)	P	I^2 (%)	OR	P	OR	P
1q32.1	202039789	rs9427472	*ELF3*	C/T	0.48	0.82 (0.77–0.88)	8.3×10^{-9}	2.5	0.95	8.5×10^{-3}	0.95	1.9×10^{-7}
1q32.1	204285634	rs4548504	*PLEKHA6*	T/C	0.56	0.76 (0.71–0.81)	8.7×10^{-17}	0.0	0.97	0.035	0.95	8.0×10^{-9}
4q31.21	143274817	rs1396196	*USP38*	G/A	0.55	0.79 (0.74–0.84)	2.5×10^{-13}	21.5	0.98	0.24	0.97	0.00067
7q32.1	128935488	rs3778754	*IRF5*	G/C	0.44	0.82 (0.77–0.87)	8.8×10^{-10}	0.0	0.91	1.6×10^{-6}	0.92	7.6×10^{-12}
12q13.12	49546847	rs17197593	*KCNH3*	T/C	0.02	1.76 (1.44–2.15)	3.8×10^{-8}	26.1	1.01	0.87	1.04	0.26
17p11.2	22643570	rs79138295	*KCNJ18*	T/C	0.05	1.57 (1.34–1.83)	1.1×10^{-8}	12.6	1.01	0.71	1.07	0.023
21q21.1	16389613	rs240471	*USP25*	A/C	0.38	1.29 (1.21–1.38)	8.9×10^{-14}	21.7	1.05	0.0013	1.05	1.1×10^{-5}

EA, effect allele; NEA, non-effect allele; EAF, effect allele frequency; P, P value (two-sided). [a]Location extracted from human genome assembly GrCH38.

and among publicly available summary statistics from genome-wide analyses of UK Biobank phenotypes (Methods), we identified regions associated with hypertension and/or blood pressure ($n=10$), BMI ($n=8$) and smoking ($n=3$; Extended Data Fig. 1 and Supplementary Table 12).

Fine-mapping and eQTL analysis

Fine-mapping of the kidney cancer susceptibility regions was performed to identify candidate causal variants in the independent loci significantly associated with kidney cancer, as well as in ccRCC and pRCC subtypes. For kidney cancer, a credible set of consensus between The SUm of SIngle Effects (SuSiE)[33] and Probabilistic Annotation INtegraTOR (PAINTOR)[34] packages (Methods) was generated and encompassed 1,321 unique variants (maximum P value of 5.3×10^{-6}), with a median of 12 variants per credible set (Supplementary Table 13); this set encompasses the high-interest variants identified by at least one of the programs. A substantial proportion (25.9%) of the credible sets contained five or fewer variants. Analysis in ccRCC and pRCC produced consensus credible sets encompassing 1,098 (median of 15) and 58 (median of 4) variants, respectively.

Analyses of kidney cancer risk variants using existing transcriptomics datasets revealed significant (false discovery rate (FDR) <5%) *cis*-eQTL effects on several genes within a ±500 kb window. Overall, among the significant loci, 48 variants from 34 cytoband regions displayed *cis*-eQTL for 83 genes across normal and tumor transcriptomes. Specifically, out of 108 sentinel kidney cancer variants, we identified 5 and 40 that displayed significant *cis*-eQTL effects on 5 and 70 genes in normal kidney cortex ($N=73$) and whole blood ($N=670$), respectively, from GTEx v8. The *cis*-eQTL effects of the associated risk variants were further pursued in TCGA–KIRC, and 17 variants having significant (FDR <5%) *cis*-eQTL effects on 20 genes were identified (Supplementary Table 10). A broader analysis using the 1,321 variants identified across credible sets from fine-mapping analysis revealed significant *cis*-eQTL effects on an additional 1 (total, 6) and 26 (total, 96) genes in kidney cortex and whole blood, respectively. Similar analysis in ccRCC and pRCC yielded 65 and 6 genes with significant *cis*-eQTL effects from 36 and 4 variants, respectively.

Functional enrichment

Previous work has focused on the role of HIFs in kidney carcinogenesis, including evidence that kidney cancer susceptibility loci are enriched for HIF-binding sites. We sought to investigate potential excess overlap among the HIF-binding sites as defined by chromatin immunoprecipitation sequencing (ChIP-seq) profiling of HIF-1α, HIF-2α and HIF-1β binding sites in 5 RCC4 and 786-O RCC cell lines with the 108 identified kidney cancer risk loci (Supplementary Table 14). Using a stringent definition of overlap, we found 11 kidney cancer risk variants (or variants in high LD) residing in HIF-binding sites defined by at least two of the five ChIP-seq profiles and a resampling/bootstrapping approach determined the overlap was higher than expected ($P=1.5\times10^{-4}$). The search was extended for ±25 kb from the HIF-binding peak identified by ChIP-seq, and a further six loci were detected overlapping with these sequences and demonstrating a similar enrichment ($P=1.1\times10^{-4}$).

In silico analysis using GWAS Analysis of Regulatory or Functional Information Enrichment with LD correction (GARFIELD)[35] showed that risk variants associated with kidney cancer were significantly enriched for putative regulatory annotations in several tissues and cell lines, including active promoter/enhancer regions and DNaseI hypersensitive hotspots (Extended Data Fig. 2 and Supplementary Table 15). To further characterize the functional underpinnings of the kidney cancer associated loci, we investigated the distribution of putative TF-binding sites using motif data and position weight matrices from publicly available databases[36]. For variants identified within the consensus credible sets, we identified 34.8% of the variants as mapping to known TF-binding sites. It has been previously established that kidney cancer susceptibility variants are enriched for allelic alterations in several known TF-binding motifs. Among such known TFs, we identified 12 variants across 8 loci mapping to *PAX8* binding sites, 28 variants across 19 loci mapping to *FOS* and 39 variants spanning 22 loci mapping to *JUND* binding motifs. Using a bootstrap approach, as above, we found that such overlaps were significant ($P=0.008$, 0.007 and 0.001 for *PAX8*, *FOS* and *JUND*, respectively).

Recently, Nassar et al.[37] conducted an epigenomic charting of the subtypes of RCC using assay for transposase-accessible chromatin with sequencing profiling[37]. Among the 78 loci (index variant and SNPs in high LD, $r^2>0.8$) identified for ccRCC, we found 19 loci overlapped with potential active promoters (H3K27ac peaks) of at least two of the ccRCC samples reported by Nassar et al.[37] (Supplementary Table 9). For pRCC, an overlap of four loci out of seven was identified. Both findings were determined to be significant ($P=2.3\times10^{-3}$ and 4.4×10^{-5}, respectively) using the previous resampling approach (Methods). Further, 16 of the 19 ccRCC loci that overlapped with H3K27ac peaks in ccRCC samples also overlapped with H3K27ac peaks in pRCC samples. Similarly, three of the seven pRCC loci that overlapped with H3K27ac peaks in pRCC samples were also observed under H3K27ac peaks in ccRCC samples, suggesting genetic similarities between subtypes for select loci.

Heritability

Using individual-level data from Eur participants in NCI-1, NCI-2 and NCI-3 scans (13,692 kidney cancer cases), we estimated the liability-scale heritability of kidney cancer attributable to common variation to be 16.4% (95% CI 12.9–19.9%), which is approximately 42% of the total heritability of kidney cancer as estimated from twin studies (38%)[18]. Larger

https://doi.org/10.1038/s41588-024-01725-7

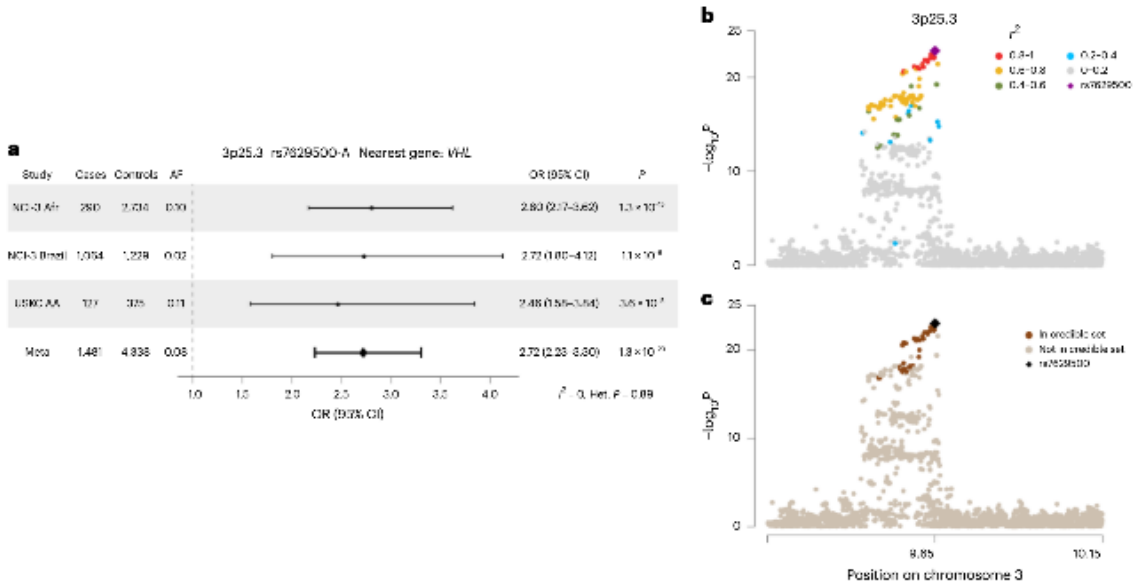


Fig. 2 | Locus 3p25.3. a, A forest plot showing the OR and 95% CI of index SNP rs7629500 (ref. allele G; effect allele A) in 3p25.3 on ccRCC across substudies containing the SNP and the overall meta-analysis. This variant was not present at detectable levels among Eur and Asn participants, as defined in the GrafPop analysis (Methods). The columns 'Study', 'Cases', 'Controls' and 'AF' show substudy, case and control sample sizes and allele frequency, respectively. USKCC AA, US Kidney Cancer Study—African Americans. Columns 'OR (95% CI)' and '*P*' show OR, the 95% CI and the corresponding study-specific *P* value. The final row shows the overall statistics from the multi-ancestry meta-analysis, denoted as 'Meta' (shown in bold), for the variant and the corresponding heterogeneity statistic (I^2) and one-sided heterogeneity test (Het.) *P* value. **b**, Regional plot of 3p25.3 for ccRCC with index SNP rs7629500 marked and other SNPs colored according to LD (r^2) with the index SNP in Afr samples. **c**, Regional plot of 3p25.3 showing the consensus credible set (Methods). For **b** and **c**, the vertical axis represents the $-\log_{10}P$ of the two-sided *z* statistics.

GWAS should continue to move us closer to a comprehensive map of the underlying genetic architecture of common variants for kidney cancer risk. Previous estimates of the heritability of kidney cancer suggest that more than 80% of the heritability of this cancer can be explained using the common variants alone in higher powered larger GWAS including up to 200,000 cases and an equal number of controls[39,40].

Polygenic prediction of kidney cancer risk

To further investigate the underlying architecture of kidney cancer susceptibility, we constructed a kidney cancer polygenic risk score (PRS) using the effect size estimates from the multi-ancestry meta-analysis (excluding UK Biobank) for 107 of the 108 kidney cancer loci (the *VHL* locus rs7629500, found only among Afr and LA participants, was not included). In a validation set of unrelated Eur UK Biobank participants (1,696 cases and 323,109 controls), a one standard deviation increase in the PRS was associated with increased risk for kidney cancer (OR of 1.50, 95% CI 1.43–1.58), with the association further strengthening when restricting to ccRCC as the outcome (OR of 1.69, 95% CI 1.57–1.81). Further comparison of PRS deciles (Supplementary Table 16) demonstrates the utility of risk stratification of the PRS, especially when comparing the lowest and highest deciles (OR of 4.24, 95% CI 3.30–5.51 for kidney cancer; OR of 6.90, 95% CI 4.59–10.86 for ccRCC). We evaluated the improvement in discriminative performance of this PRS in the validation set compared with a baseline model including sex. As shown in Fig. 4 and Supplementary Table 16, the PRS improved the predictive performance of the baseline model, increasing the estimated area under the curve (AUC) from 0.61 (95% CI 0.59–0.62) to 0.65 (95% CI 0.64–0.67). When we expanded the baseline model to include additional risk factors (age, smoking, BMI and hypertension), the addition of the PRS boosted the AUC from 0.71 (95% CI 0.70–0.72) to 0.74 (95% CI 0.72–0.75). The PRS also demonstrated better performance for predicting ccRCC cases in UK Biobank; its addition to the base model improved the AUC from 0.61 (95% CI 0.59–0.63) to 0.68 (95% CI 0.66–0.70), and its inclusion in the base plus risk factors model boosted the AUC from 0.70 (95% CI 0.68–0.72) to 0.74 (95% CI 0.72–0.76).

Discussion

This multi-ancestry GWAS meta-analysis of 29,020 kidney cancer cases and 835,670 controls expands the number of identified susceptibility regions from 13 to 63 and further advances knowledge of the genetic architecture of kidney cancer by identifying promising candidate functional variants localizing to eQTLs (for example, *IRF5*) and TF-binding sites (for example, HIF). In particular, our multi-ancestral approach identified associations in the *VHL* locus that highlight local differences in disease architecture due to population-specific allele frequencies.

Our meta-analysis discovered a ccRCC germline susceptibility locus with an unusually strong effect size that was common in some populations (for example, Afr participants) at the 3p25.3 locus tagged by rs7629500. We also observed a weaker nearby risk locus tagged by rs139729777 shared across populations. The rs7629500 variant maps to the 3′ UTR of *VHL* and the rs139729777 variant is approximately 100 kb telomeric of *VHL* (and an intron of *FANCD2*). *VHL* is a critical gene in the pathogenesis of ccRCC and mutations (and 3p deletion) are common in ccRCC tumors and the cause of von Hippel–Lindau syndrome, an inherited disorder characterized by a markedly elevated risk of ccRCC[7]. VHL inactivation is a critical early event in ccRCC development, inducing constitutive upregulation of HIF-mediated expression of oncogenic factors[41], and as such, these variants could promote VHL dysregulation. The paucity of Afr and LA samples in GTEx and TCGA data limits immediate investigations into functional effects, although

https://doi.org/10.1038/s41588-024-01725-7

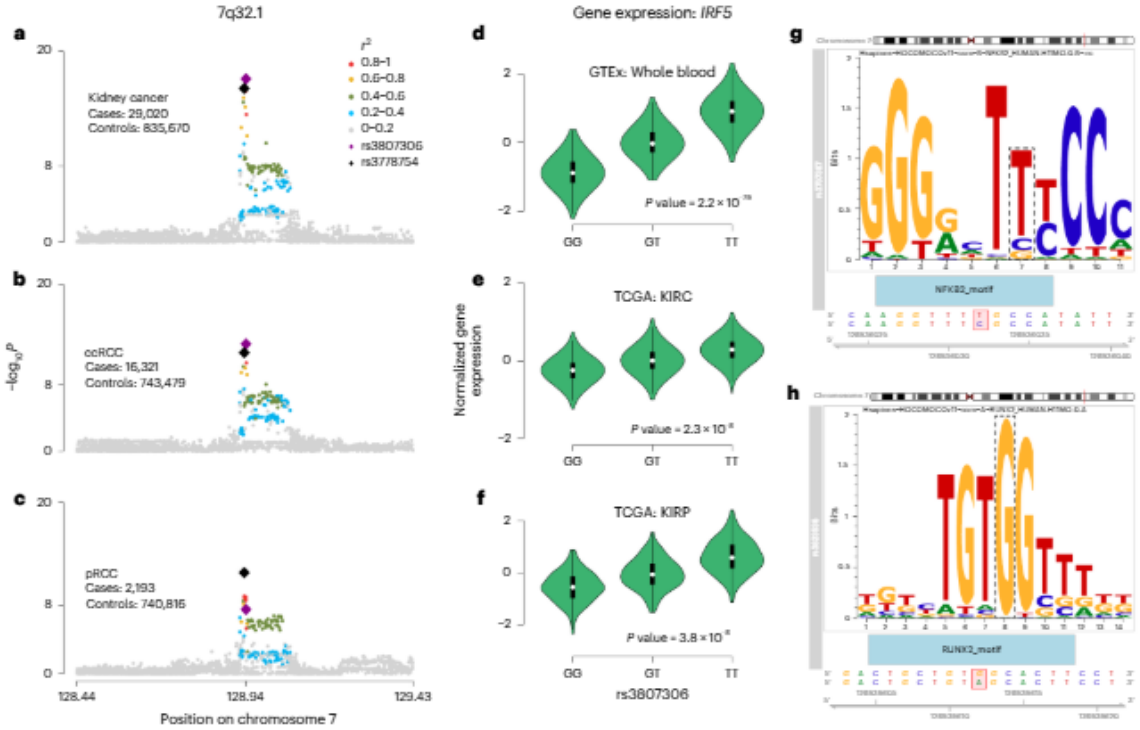


Fig. 3 | Locus 7q32.1. a–**c**, Regional plot of 7q32.1 for overall kidney cancer (**a**), ccRCC (**b**) and pRCC (**c**) with SNPs colored according to LD (r^2) with the index SNP rs3807306 in Eur samples. Index SNP rs3778754 for pRCC GWAS is marked in black. The vertical axis represents the $-\log_{10}P$ (two-sided) of the *z* statistics. **d**–**f**, eQTL effect of index SNP rs3807306 on the expression of *IRF5* in GTEx whole blood (**d**; $n = 670$), TCGA–KIRC (**e**; $n = 527$) and TCGA–KIRP (**f**; $n = 290$) is shown (Methods). Violin plots represent the distribution of gene expression between the 10th and 90th percentiles, with the median marked in white, across possible genotypes. **g**,**h**, Variants present in the fine-mapping credible sets (Methods) for ccRCC and pRCC disrupt the binding of immune-related and cancer-related TFs, such as rs3757387 for *NFKB2* (**g**) and rs3823536 for *RUNX2* (**h**), respectively.

3′ UTR variants have been shown to affect multiple post-transcriptional and translational processes[42].

The analysis of known ccRCC cases confirmed those for kidney cancer overall and identified a handful of novel loci, which is not surprising given that the majority of kidney cancers are this subtype. Some of the identified kidney cancer and ccRCC loci are located within or nearby genes involved in well-established pathways contributing to ccRCC pathogenesis (*VHL*, *AKT1* and *GRB10*), as well as genes more broadly recognized as tumor suppressors (*FANCD2* and *TP53*) and oncogenes (*MLLT10* and *CDK6*), while other novel loci are situated near genes involved in hypoxia response (*DDIT4*), cell cycle control (*CDKN1A*, *CDKN2C*, *CNEP1R1*, *INCENP*, *NCAPH2*, *STN1* and *TFDP2*), telomerase regulation (*POT1* and *TERT*) and insulin signaling (*INSR*). Follow-up functional investigations of these identified kidney cancer and ccRCC loci are warranted to identify the causal variants to elucidate the underlying biology.

Our multi-ancestry GWAS is, to the best of our knowledge, the first to report susceptibility loci for pRCC, identifying seven loci mapping to six regions, the majority of which are not associated with overall kidney cancer or ccRCC. As we described earlier, the 7q32.1 locus was significant in kidney cancer overall and in both subtypes, with several lines of evidence pointing to *IRF5*-related effects underlying this association. We also identified eQTLs localizing to other pRCC loci, offering new clues for the biological underpinning of pRCC susceptibility loci. For example, the rs1396196 locus at 4q31.21 is located near two plausible candidate genes. It lies 90 kb downstream of and is a GTEx whole blood *cis*-eQTL for *USP38*, which encodes a deubiquinase that promotes non-homologous end-joining repair by regulating the activity of HDAC1 (ref. 43). The locus is also 61 kb upstream of *GAB1* which encodes an adapter protein that binds directly with activated MET to mediate numerous oncogenic signals via the ERK and AKT/PKB pathways, including cell proliferation, survival and invasion. The proximity of this locus to *GAB1* is notable given that *MET* is a well-established pRCC proto-oncogene; autosomal dominant mutations in *MET* are responsible for hereditary papillary renal carcinoma[7] and *MET* alterations or chromosome 7 gain are found in the majority of sporadic type 1 pRCC tumors[44]. Expanded GWAS of pRCC in different populations could offer further insights into pRCC etiology, such as germline connections to *MET*, and could be informative in elucidating the basis for the approximate threefold excess of pRCC in the US Black population[45].

The increasing incidence of kidney cancer, frequent lack of symptoms at diagnosis (even at advanced stages) and wide variation in survival between early and advanced-stage disease support the importance of developing effective screening approaches[46]. In this regard, the kidney cancer PRS constructed from identified kidney cancer loci demonstrated good discriminatory performance in our validation dataset; its inclusion in a risk prediction model including major risk factors resulted in an AUC of 0.74 (versus 0.71 with risk factors only). The impact of this PRS was slightly greater when the model was used to

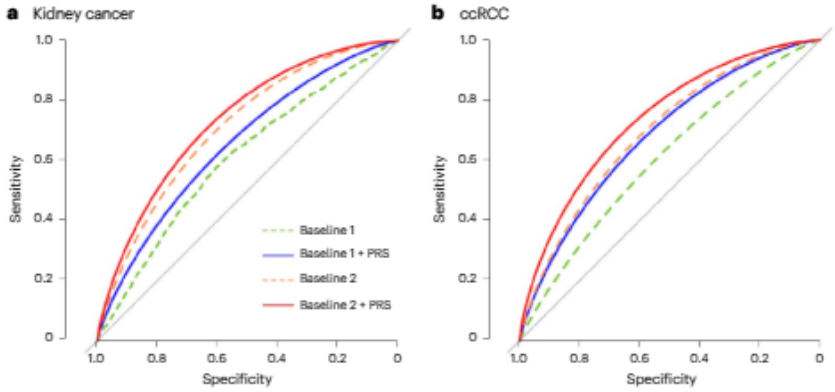


Fig. 4 | Performance of the PRS. a,**b**, Receiver operating characteristic (ROC) curve, demonstrating the area under the curve (AUC) for predicting kidney cancer (**a**) and ccRCC (**b**) in a validation set of unrelated Eur participants (1,696 cases and 323,109 controls) in UK Biobank. The PRS was constructed using the independent loci identified in the multi-ancestry meta-analysis GWAS results of overall kidney cancer. Two baseline models are considered for benchmarking, containing differing sets of predictors (Methods). Baseline 1 includes sex and genetic principal components, and baseline 2 includes age, sex, genetic principal components, body mass index, smoking status and hypertension. For full results, see Supplementary Table 16.

predict ccRCC (AUC 0.74 with PRS, 0.70 without), further underscoring the importance of this histologic subtype as a driver of the observed GWAS signals for kidney cancer overall. While the comparative rarity of kidney cancer in the general population is a barrier to broad-based screening, a targeted risk-stratified approach could be cost-effective. In this regard, and given the strength of its AUC, our PRS merits further investigation in well-defined clinical and preventive studies. However, as the PRS validation was limited to Eur participants in the UK Biobank, the generalizability of these findings to other populations should be the focus of ongoing investigation.

As more kidney cancer genetic susceptibility regions remain to be discovered with ever larger GWAS that expand into more diverse populations, the underlying genetic architecture will come into better focus, a key step towards clinical use of both rare variants and more precise PRS instruments. Deeper investigation of kidney cancer susceptibility regions promises to reveal clues into the etiology of kidney cancer and perhaps elucidate novel strategies for detection or treatment. Our study demonstrates genetic factors common to different histologic subtypes and at the same time, defines subtype-specific variants, which in turn could lead to new avenues for research.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41588-024-01725-7.

References

1. Sung, H. et al. Global Cancer Statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **71**, 209–249 (2021).
2. Znaor, A., Lortet-Tieulent, J., Laversanne, M., Jemal, A. & Bray, F. International variations and trends in renal cell carcinoma incidence and mortality. *Eur. Urol.* **67**, 519–530 (2015).
3. Cancer stat facts: kidney and renal pelvis cancer. *National Cancer Institute* https://seer.cancer.gov/statfacts/html/kidrp.html (2023).
4. Cancer Facts & Figures 2022. *American Cancer Society* https://www.cancer.org/research/cancer-facts-statistics/all-cancer-facts-figures/cancer-facts-figures-2022.html (2022).
5. Chow, W. H., Scelo, G. & Tarone, R. E. in *Schottenfeld and Fraumeni Cancer Epidemiology and Prevention* 4th edn (eds Thun, M. J. et al.) Ch. 51, 961–976 (Oxford Univ. Press, 2018).
6. Lopez-Beltran, A. et al. 2009 update on the classification of renal epithelial tumors in adults. *Int. J. Urol.* **16**, 432–443 (2009).
7. Haas, N. B. & Nathanson, K. L. Hereditary kidney cancer syndromes. *Adv. Chronic Kidney Dis.* **21**, 81–90 (2014).
8. Andreou, A. et al. Elongin C (ELOC/TCEB1)-associated von Hippel-Lindau disease. *Hum. Mol. Genet.* **31**, 2728–2737 (2022).
9. Lang, M. et al. Clinical and molecular characterization of microphthalmia-associated transcription factor (MITF)-related renal cell carcinoma. *Urology* **149**, 89–97 (2021).
10. Schmidt, L. S. et al. PRDM10 RCC: a Birt–Hogg–Dube-like syndrome associated with lipoma and a highly penetrant, aggressive renal tumors morphologically resembling type 2 papillary renal cell carcinoma. *Urology* **179**, 58–70 (2023).
11. Scelo, G. et al. Genome-wide association study identifies multiple risk loci for renal cell carcinoma. *Nat. Commun.* **8**, 15724 (2017).
12. Grampp, S. et al. Genetic variation at the 8q24.21 renal cancer susceptibility locus affects HIF binding to a MYC enhancer. *Nat. Commun.* **7**, 13183 (2016).
13. Schodel, J. et al. Common genetic variants at the 11q13.3 renal cancer susceptibility locus influence binding of HIF to an enhancer of cyclin D1 expression. *Nat. Genet.* **44**, 420–425 (2012).
14. Bigot, P. et al. Functional characterization of the 12p12.1 renal cancer-susceptibility locus implicates BHLHE41. *Nat. Commun.* **7**, 12098 (2016).
15. Colli, L. M. et al. Altered regulation of DPF3, a member of the SWI/SNF complexes, underlies the 14q24 renal cancer susceptibility locus. *Am. J. Hum. Genet.* **108**, 1590–1610 (2021).
16. Riscal, R. et al. Cholesterol auxotrophy as a targetable vulnerability in clear cell renal cell carcinoma. *Cancer Discov.* **11**, 3106–3125 (2021).
17. Grampp, S. et al. Multiple renal cancer susceptibility polymorphisms modulate the HIF pathway. *PLoS Genet.* **13**, e1006872 (2017).
18. Schmid, V. et al. Co-incidence of RCC-susceptibility polymorphisms with HIF *cis*-acting sequences supports a pathway tuning model of cancer. *Sci. Rep.* **9**, 18768 (2019).
19. Patel, S. A. et al. The renal lineage factor PAX8 controls oncogenic signalling in kidney cancer. *Nature* **606**, 999–1006 (2022).

20. Han, B. & Eskin, E. Random-effects model aimed at discovering associations in meta-analysis of genome-wide association studies. *Am. J. Hum. Genet.* **88**, 586–598 (2011).
21. GTEx Consortium. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* **369**, 1318–1330 (2020).
22. Cancer Genome Atlas Research Network et al. The Cancer Genome Atlas Pan-Cancer analysis project. *Nat. Genet.* **45**, 1113–1120 (2013).
23. Uhlen, M. et al. A pathology atlas of the human cancer transcriptome. *Science* **357**, eaan2507 (2017).
24. Singhal, S. S., Yadav, S., Drake, K., Singhal, J. & Awasthi, S. Hsf-1 and POB1 induce drug sensitivity and apoptosis by inhibiting Ralbp1. *J. Biol. Chem.* **283**, 19714–19729 (2008).
25. Oosterhoff, J. K., Kuhne, L. C., Grootegoed, J. A. & Blok, L. J. EGF signalling in prostate cancer cell lines is inhibited by a high expression level of the endocytosis protein REPS2. *Int. J. Cancer* **113**, 561–567 (2005).
26. Oosterhoff, J. K., Penninkhof, F., Brinkmann, A. O., Anton Grootegoed, J. & Blok, L. J. REPS2/POB1 is downregulated during human prostate cancer progression and inhibits growth factor signalling in prostate cancer cells. *Oncogene* **22**, 2920–2925 (2003).
27. Zhang, H., Duan, C. J., Zhang, H., Cheng, Y. D. & Zhang, C. F. Expression and clinical significance of REPS2 in human esophageal squamous cell carcinoma. *Asian Pac. J. Cancer Prev.* **14**, 2851–2857 (2013).
28. He, X. Y. et al. Liver X receptor agonists exert antitumor effects against hepatocellular carcinoma via inducing REPS2 expression. *Acta Pharmacol. Sin.* **44**, 635–646 (2023).
29. Du, J. et al. Cytoplasmic localization of IRF5 induces Wnt5a/E-cadherin degradation and promotes gastric cancer cells metastasis. *Cancer Gene Ther.* **30**, 866–877 (2023).
30. Bi, X. et al. Loss of interferon regulatory factor 5 (IRF5) expression in human ductal carcinoma correlates with disease stage and contributes to metastasis. *Breast Cancer Res* **13**, R111 (2011).
31. Massimino, M. et al. IRF5 promotes the proliferation of human thyroid cancer cells. *Mol. Cancer* **11**, 21 (2012).
32. Guiteras, J. et al. The gene silencing of IRF5 and BLYSS effectively modulates the outcome of experimental lupus nephritis. *Mol. Ther. Nucleic Acids* **24**, 807–821 (2021).
33. Zou, Y., Carbonetto, P., Wang, G. & Stephens, M. Fine-mapping from summary data with the 'Sum of Single Effects' model. *PLoS Genet.* **18**, e1010299 (2022).
34. Kichaev, G. et al. Integrating functional data to prioritize causal variants in statistical fine-mapping studies. *PLoS Genet.* **10**, e1004722 (2014).
35. Iotchkova, V. et al. GARFIELD classifies disease-relevant genomic features through integration of functional annotations with association signals. *Nat. Genet.* **51**, 343–353 (2019).
36. Kulakovskiy, I. V. et al. HOCOMOCO: towards a complete collection of transcription factor binding models for human and mouse via large-scale ChIP-Seq analysis. *Nucleic Acids Res.* **46**, D252–D259 (2018).
37. Nassar, A. H. et al. Epigenomic charting and functional annotation of risk loci in renal cell carcinoma. *Nat. Commun.* **14**, 346 (2023).
38. Mucci, L. A. et al. Familial risk and heritability of cancer among twins in nordic countries. *JAMA* **315**, 68–76 (2016).
39. Zhang, Y., Qi, G., Park, J. H. & Chatterjee, N. Estimation of complex effect-size distributions using summary-level statistics from genome-wide association studies across 32 complex traits. *Nat. Genet.* **50**, 1318–1326 (2018).
40. Zhang, Y. D. et al. Assessment of polygenic architecture and risk prediction based on common variants across fourteen cancers. *Nat. Commun.* **11**, 3353 (2020).
41. Linehan, W. M., Srinivasan, R. & Schmidt, L. S. The genetic basis of kidney cancer: a metabolic disease. *Nat. Rev. Urol.* **7**, 277–285 (2010).
42. Chan, J. J., Tabatabaeian, H. & Tay, Y. 3'UTR heterogeneity and cancer progression. *Trends Cell Biol.* **33**, 568–582 (2023).
43. Yang, Y. et al. The deubiquitinase USP38 promotes NHEJ repair through regulation of HDAC1 activity and regulates cancer cell response to genotoxic insults. *Cancer Res.* **80**, 719–731 (2020).
44. Cancer Genome Atlas Research Network et al. Comprehensive molecular characterization of papillary renal-cell carcinoma. *N. Engl. J. Med.* **374**, 135–145 (2016).
45. Olshan, A. F. et al. Racial difference in histologic subtype of renal cell carcinoma. *Cancer Med* **2**, 744–749 (2013).
46. Usher-Smith, J., Simmons, R. K., Rossi, S. H. & Stewart, G. D. Current evidence on screening for renal cancer. *Nat. Rev. Urol.* **17**, 637–642 (2020).

Mark P. Purdue[1,96] ✉, Diptavo Dutta[2,96], Mitchell J. Machiela[2,96], Bryan R. Gorman[3,96], Timothy Winter[4], Dayne Okuhara[3], Sara Cleland[3], Aida Ferreiro-Iglesias[5], Paul Scheet[6], Aoxing Liu[7,8,9], Chao Wu[10], Samuel O. Antwi[11], James Larkin[12], Stênio C. Zequi[13,14,15,16], Maxine Sun[17], Keiko Hikino[18], Ali Hajiran[19], Keith A. Lawson[20], Flavio Cárcano[21], Odile Blanchet[22], Brian Shuch[23], Kenneth G. Nepple[24], Gaëlle Margue[25], Debasish Sundi[26], W. Ryan Diver[27], Maria A. A. K. Folgueira[28], Adrie van Bokhoven[29], Florencia Neffa[30], Kevin M. Brown[31], Jonathan N. Hofmann[1], Jongeun Rhee[1], Meredith Yeager[32], Nathan R. Cole[32], Belynda D. Hicks[32], Michelle R. Manning[32], Amy A. Hutchinson[32], Nathaniel Rothman[1], Wen-Yi Huang[33], W. Marston Linehan[34], Adriana Lori[27], Matthieu Ferragu[35], Merzouka Zidane-Marinnes[36], Sérgio V. Serrano[21], Wesley J. Magnabosco[37], BioBank Japan Project*, Ana Vilas[38], Ricardo Decia[39], Florencia Carusso[30], Laura S. Graham[40], Kyra Anderson[41], Mehmet A. Bilen[42], Cletus Arciero[43], Isabelle Pellegrin[44], Solène Ricard[25], FinnGen, Ghislaine Scelo[45], Rosamonde E. Banks[46], Naveen S. Vasudev[47], Naeem Soomro[48], Grant D. Stewart[49,50], Adebanji Adeyoju[51], Stephen Bromage[51], David Hrouda[52], Norma Gibbons[52], Poulam Patel[53], Mark Sullivan[54], Andrew Protheroe[55], Francesca I. Nugent[24], Michelle J. Fournier[56], Xiaoyu Zhang[20], Lisa J. Martin[20], Maria Komisarenko[20], Timothy Eisen[57], Sonia A. Cunningham[6], Denise C. Connolly[58], Robert G. Uzzo[59], David Zaridze[60], Anush Mukeria[60], Ivana Holcatova[61,62], Anna Hornakova[63], Lenka Foretova[64], Vladimir Janout[65], Dana Mates[66], Viorel Jinga[67], Stefan Rascu[68], Mirjana Mijuskovic[69], Slavisa Savic[70], Sasa Milosavljevic[71], Valérie Gaborieau[5], Behnoush Abedi-Ardekani[5], James McKay[5], Mattias Johansson[5], Larry Phouthavongsy[72],

Lindsay Hayman[73], Jason Li[73], Ilinca Lungu[72,73], Stephania M. Bezerra[74], Aline G. Souza[75], Claudia T. G. Sares[76], Rodolfo B. Reis[76], Fabio P. Gallucci[77], Mauricio D. Cordeiro[77], Mark Pomerantz[78], Gwo-Shu M. Lee[17], Matthew L. Freedman[17,78,79,80], Anhyo Jeong[23], Samantha E. Greenberg[81], Alejandro Sanchez[82], R. Houston Thompson[83], Vidit Sharma[83], David D. Thiel[84], Colleen T. Ball[11], Diego Abreu[39], Elaine T. Lam[40], William C. Nahas[77], Viraj A. Master[85], Alpa V. Patel[27], Jean-Christophe Bernhard[25], Neal D. Freedman[33], Pierre Bigot[35], Rui M. Reis[86,87], Leandro M. Colli[88], Antonio Finelli[20], Brandon J. Manley[89], Chikashi Terao[90], Toni K. Choueiri[17,91], Dirce M. Carraro[92], Richard Houlston[93], Jeanette E. Eckel-Passow[94], Philip H. Abbosh[95], Andrea Ganna[7,8,9], Paul Brennan[5,97], Jian Gu[6,97] & Stephen J. Chanock[4,97]

[1]Occupational and Environmental Epidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Rockville, MD, USA. [2]Integrative Tumor Epidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Rockville, MD, USA. [3]Booz Allen Hamilton Inc, McLean, VA, USA. [4]Laboratory of Genetic Susceptibility, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Rockville, MD, USA. [5]Genomic Epidemiology Branch, International Agency for Research on Cancer, Lyon, France. [6]Department of Epidemiology, Division of Cancer Prevention and Population Sciences, The University of Texas MD Anderson Cancer Center, Houston, TX, USA. [7]Institute for Molecular Medicine Finland, University of Helsinki, Helsinki, Finland. [8]Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA, USA. [9]Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA, USA. [10]Biosample Repository, Fox Chase Cancer Center—Temple Health, Philadelphia, PA, USA. [11]Department of Quantitative Health Sciences, Mayo Clinic, Jacksonville, FL, USA. [12]Department of Medical Oncology, Royal Marsden NHS Foundation Trust, London, UK. [13]Department of Urology, A.C. Camargo Cancer Center, São Paulo, Brazil. [14]National Institute for Science and Technology in Oncogenomics and Therapeutic Innovation INCIT-INOTE, São Paulo, Brazil. [15]Latin American Renal Cancer Group, São Paulo, Brazil. [16]Department of Surgery, Division of Urology, São Paulo Federal University, São Paulo, Brazil. [17]Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA, USA. [18]Laboratory for Pharmacogenomics, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan. [19]Department of Urology, Division of Urologic Oncology, West Virginia University Cancer Institute, Morgantown, WV, USA. [20]Department of Surgical Oncology, Division of Urology, Princess Margaret Cancer Centre, University Health Network, Toronto, Ontario, Canada. [21]Department of Medical Oncology, Barretos Cancer Hospital, Barretos, Brazil. [22]Biobank Resource Centre, CHU Angers, Angers, France. [23]Department of Urology, UCLA Jonsson Comprehensive Cancer Center, Los Angeles, CA, USA. [24]Department of Urology, Holden Comprehensive Cancer Center, University of Iowa, Iowa City, IA, USA. [25]Department of Urology, CHU Bordeaux, Bordeaux, France. [26]Department of Urology, The Ohio State University Wexner Medical Center, Columbus, OH, USA. [27]Department of Population Science, American Cancer Society, Atlanta, GA, USA. [28]Departments of Radiology and Oncology, Comprehensive Center for Precision Oncology—C2PO, Centro de Investigação Translacional em Oncologia, Instituto do Cancer do Estado de São Paulo, Hospital das Clinicas, Faculdade de Medicina Universidade de São Paulo, São Paulo, Brazil. [29]Department of Pathology, University of Colorado Anschutz Medical Campus, Aurora, CO, USA. [30]Tumor Bank, Hospital Militar, Montevideo, Uruguay. [31]Laboratory of Translational Genomics, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Rockville, MD, USA. [32]Cancer Genomics Research Laboratory, Frederick National Laboratory, Rockville, MD, USA. [33]Metabolic Epidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Rockville, MD, USA. [34]Urologic Oncology Branch, Center for Cancer Research, National Cancer Institute, Bethesda, MD, USA. [35]Department of Urology, CHU Angers, Angers, France. [36]Department of Pathology, CHU Angers, Angers, France. [37]Department of Urology, Barretos Cancer Hospital, Barretos, Brazil. [38]Department of Pathology, Hospital Pasteur, Montevideo, Uruguay. [39]Department of Urology, Hospital Pasteur, Montevideo, Uruguay. [40]Department of Medicine, Division of Medical Oncology, University of Colorado Anschutz Medical Campus, Aurora, CO, USA. [41]Oncology Clinical Research Support Team, University of Colorado Anschutz Medical Campus, Aurora, CO, USA. [42]Department of Hematology and Medical Oncology, Winship Cancer Institute, Emory University School of Medicine, Atlanta, GA, USA. [43]Department of Surgery, Winship Cancer Institute, Emory University School of Medicine, Atlanta, GA, USA. [44]Biobank, Bordeaux Biothèque Santé, CHU Bordeaux, Bordeaux, France. [45]Observational and Pragmatic Research Institute Pte Ltd, Singapore, Singapore. [46]Leeds Institute of Medical Research at St James's, University of Leeds, Leeds, UK. [47]Department of Oncology, Leeds Institute of Medical Research at St James's, University of Leeds, Leeds, UK. [48]Department of Urology, Newcastle Hospitals NHS Foundation Trust, Newcastle, UK. [49]Department of Urology, Western General Hospital, NHS Lothian, Edinburgh, UK. [50]Department of Surgery, University of Cambridge, Cambridge, UK. [51]Department of Urology, Stockport NHS Foundation Trust, Stockport, UK. [52]Department of Urology, Imperial College Healthcare NHS Trust, London, UK. [53]Division of Oncology, University of Nottingham, Nottingham, UK. [54]Department of Urology, Oxford University Hospitals NHS Foundation Trust, Oxford, UK. [55]Department of Oncology, Oxford University Hospitals NHS Foundation Trust, Oxford, UK. [56]Tissue Core, Moffitt Cancer Center, Tampa, FL, USA. [57]Department of Oncology, Cambridge University Hospitals NHS Foundation Trust, Cambridge, UK. [58]Cancer Signaling and Microenvironment, Biosample Repository Facility, Fox Chase Cancer Center—Temple Health, Philadelphia, PA, USA. [59]Department of Urology, Fox Chase Cancer Center—Temple Health, Philadelphia, PA, USA. [60]Department of Clinical Epidemiology, N.N. Blokhin National Medical Research Centre of Oncology, Moscow, Russia. [61]Institute of Public Health and Preventive Medicine, Second Faculty of Medicine, Charles University, Prague, Czech Republic. [62]Department of Oncology, Second Faculty of Medicine and University Hospital Motol, Charles University, Prague, Czech Republic. [63]Institute of Hygiene and Epidemiology, First Faculty of Medicine, Charles University, Prague, Czech Republic. [64]Department of Cancer Epidemiology and Genetics, Masaryk Memorial Cancer Institute, Brno, Czech Republic. [65]Faculty of Health Sciences, Palacky University, Olomouc, Czech Republic. [66]Department of Occupational Health and Toxicology, National Center for Environmental Risk Monitoring, National Institute of Public Health, Bucharest, Romania. [67]Urology Department, Academy of Romanian Scientists, Carol Davila University of Medicine and Pharmacy, Bucharest, Romania. [68]Urology Department, Carol Davila University of Medicine and Pharmacy, Bucharest, Romania. [69]Clinic of Nephrology, Faculty of Medicine, Military Medical Academy, Belgrade, Serbia. [70]Department of Urology, Clinical Hospital Center Dr Dragisa Misovic Dedinje, Belgrade, Serbia. [71]International Organisation for Cancer Prevention and Research, Belgrade, Serbia. [72]Ontario Tumour Bank, Ontario Institute for Cancer Research, Toronto, Ontario, Canada. [73]Diagnostic Development Program, Tissue Portal, Ontario Institute for Cancer Research, Toronto, Ontario, Canada. [74]Department of Pathology, A.C. Camargo Cancer Center, São Paulo, Brazil. [75]Departments of Medical Imaging, Hematology and Oncology, Division of Medical Oncology, Ribeirao Preto Medical School, University of São Paulo, Ribeirao Preto, Brazil. [76]Departments of Surgery and Anatomy, Division of Urology, Ribeirao Preto Medical School, University of São Paulo, Ribeirao Preto, Brazil. [77]Surgery Department, Urology Division, Instituto do Cancer do Estado de São Paulo, Hospital das Clinicas HCFMUSP, Faculdade de Medicina, Universidade de São Paulo, São Paulo, Brazil. [78]Dana-Farber Cancer Institute, Boston, MA, USA. [79]Center for Functional Cancer Epigenetics, Dana-Farber Cancer Institute, Boston, MA, USA. [80]Broad Institute of MIT and Harvard, Boston, MA, USA. [81]Department of Population Sciences, Genetic Counseling Shared Resource, Huntsman Cancer Institute,

Salt Lake City, UT, USA. [82]Department of Surgery, Division of Urology, Huntsman Cancer Institute and University of Utah, Salt Lake City, UT, USA. [83]Department of Urology, Mayo Clinic, Rochester, MN, USA. [84]Department of Urology, Mayo Clinic, Jacksonville, FL, USA. [85]Department of Urology, Winship Cancer Institute, Emory University School of Medicine, Atlanta, GA, USA. [86]Molecular Oncology Research Center, Barretos Cancer Hospital, Barretos, Brazil. [87]Life and Health Sciences Research Institute (ICVS), School of Medicine, University of Minho, Braga, Portugal. [88]Departament of Medical Image, Hematology and Oncology, Division of Medical Oncology, Ribeirao Preto Medical School, University of São Paulo, Ribeirao Preto, Brazil. [89]Genitourinary Oncology Program, Moffitt Cancer Center, Tampa, FL, USA. [90]Laboratory for Statistical and Translational Genetics, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan. [91]Department of Medicine, Brigham and Women's Hospital, Boston, MA, USA. [92]Clinical and Functional Genomics Group, CIPE (International Research Center), A.C. Camargo Cancer Center, São Paulo, Brazil. [93]Division of Genetics and Epidemiology, Institute of Cancer Research, Sutton, UK. [94]Department of Quantitative Health Sciences, Mayo Clinic, Rochester, MN, USA. [95]Department of Nuclear Dynamics and Cancer, Fox Chase Cancer Center—Temple Health, Philadelphia, PA, USA. [96]These authors contributed equally: Mark P. Purdue, Diptavo Dutta, Mitchell J. Machiela, Bryan R. Gorman. [97]These authors jointly supervised this work: Paul Brennan, Jian Gu, Stephen J. Chanock. *Lists of authors and their affiliations appear at the end of the paper. ✉e-mail: purduem@mail.nih.gov; chanocks@mail.nih.gov

BioBank Japan Project

Keiko Hikino[18] & Chikashi Terao[90]

FinnGen

Aoxing Liu[7,8,9] & Andrea Ganna[7,8,9]

Methods

Informed consent and study approval

The NCI-3 scan and meta-analysis were classified as exempt from ethics review by the National Institutes of Health Office of Human Subjects Research Protection (18-NCI-00636-2), given the absence of personally identifiable information. Each participating center and study obtained written informed consent from the study participants and approval from its institutional review board (Supplementary Table 17).

Genotype quality control and imputation

NCI-3 scan. Samples obtained from collaborating centers in North America, South America and Europe (Supplementary Table 18) were genotyped by the NCI using the Illumina Global Screening Array with Multiple Disease content v2. We intersected and merged the newly generated genotype data with Global Screening Array with Multiple Disease content (v1) data from the Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial and Cancer Prevention Study II cohorts into a single file. For our sample quality control (QC), 512 replicate samples were removed, and 161 discordant genetic sex samples were removed. We checked for samples with a low call rate (<95%) or contamination, and none was found. Samples specifically collected in Brazil were set aside to form a Brazil-based cohort (consisting mostly of admixed American ancestry samples). GrafPop v1.0[47] was then used to classify the remaining samples by genetic similarity into the following groups: Eur, Afr, LA1, LA2 and Asn. GrafPop v1.0 groups study subjects into these classes based on their similarity to five large reference samples drawn from the National Institutes of Health database of Genotypes and Phenotypes (dbGaP[48]), defined using study reported population terms: (1) 'White, Caucasian, European, European American, and other equivalent terms', (2) 'Black, African, African American, Ghana, Yoruba, etc.', (3) 'Asian, East Asian, Chinese, Japanese, etc.', (4) 'Asian Indian, Pakistani', and (5) 'Mexican, Latino.' The LA1 cluster has greater similarity to the African reference (dbGaP sample 2) than LA2. A fraction of the cohort (753 samples) that did not classify into one of these ancestries was excluded from further analysis. A total of 97 samples with excess homozygosity (F_{ST} >0.1) or heterozygosity (F_{ST} <−0.1) calculated within ancestry, were removed. We further applied 1:5 case–control matching using PCAmatchR[49] v0.3.3 in the Asn and LA2 subsets only due to a high degree of population structure in our data for those ancestries.

We then performed marker QC separately within each group. We removed genotyped markers failing Hardy–Weinberg equilibrium (HWE) in controls ($P < 1 \times 10^{-15}$ in the large Eur sample; $P < 1 \times 10^{-6}$ otherwise; Eur: 4,310, Afr: 1,545, Asn: 1,912, LA1: 49, LA2: 385 and Brazil: 1,050 markers), differential case–control missingness ($P < 1 \times 10^{-15}$ in Eur; $P < 1 \times 10^{-6}$ otherwise; Eur: 32,941, Afr: 3,299, Asn: 260, LA1: 4, LA2: 56 and Brazil: 3), or with a large deviation from the Genome Aggregation Database v3.1.2 population-specific reference frequency in controls[50] (20% and greater; Eur: 1,822, Afr: 1,988, Asn: 2,011, LA1: 2,288, LA2: 1,921 and Brazil: 1,976), along with monomorphic markers (Eur: 64,215, Afr: 61,176, Asn: 52,964, LA1: 86,956, LA2: 79,976 and Brazil: 46,785). Markers with a low call rate (<90%) after subcohort stratification were also excluded (Eur: 0, Afr: 13, Asn: 32, LA1: 1, LA2: 15 and Brazil: 0).

Post-QC genotype sets from each ancestry were imputed using the TOPMed Imputation Server (reference panel version R2)[51,52]. Finally, post-imputation QC assessment led us to additionally filter 474 genotyped markers in the Eur samples where imputation of the masked genotype was poorly correlated with the actual genotype (leave-one-out R-squared–empirical R-squared >0.5), suggestive of genotyping errors.

Previously reported scans. One previously reported scan (International Agency for Research on Cancer (IARC); combines data from two IARC GWAS (IARC-1 and IARC-2))[11] did not have individual-level data readily available, so summary data were used. For the others with individual-level data (NCI-1, NCI-2, MD Anderson (MDA), MDA-OncoArray, and UK and US Kidney Cancer Study—African Americans)[5,53–57], we performed fresh rounds of genotype QC and imputation. No changes were made to the sample composition of the scans, with one exception: for the MDA-OncoArray scan, we used KING[58] to find and remove unexpected genetic duplicates across this and other genotyped MDA samples in the MDA GWAS and NCI-3 scan, and then used GrafPop v1.0 to classify MDA-OncoArray samples by genetic similarity; any samples that were not classified as Eur were excluded from further analysis due to small numbers. For all previously reported scans with individual-level data, genotyped markers with departures from HWE in controls ($P < 1 \times 10^{-6}$), differential case–control missingness ($P < 1 \times 10^{-6}$) or large deviations from population-specific Genome Aggregation Database v3.1.2 (ref. 59) reference frequency in controls (allele count Fisher's test $P < 1 \times 10^{-6}$) were removed. SNPs with low call rate (<95%), monomorphic SNPs, and ambiguous SNPs were also excluded. Remaining genotypes were imputed to the TOPMed panel[51] containing 97,256 samples and 308,107,085 variants using the TOPMed Imputation Server R2 (ref. 52).

UK Biobank. A detailed description of the design of the UK Biobank has been previously reported[60]. Briefly, 503,317 individuals visited an assessment center between 2006 and 2010 to complete a questionnaire collecting information on demographic, lifestyle and other health-related factors and provide biological samples and physical measures. Genotyping of DNA extracted from blood specimens was performed by Affymetrix (now part of Thermo Fisher Scientific) using either the UK BiLEVE Axiom (N = 49,950) or UK Biobank Axiom (N = 438,427) array. The released genotyping dataset, after application of quality control filtering, included 488,377 samples with 805,426 markers from both arrays. Genotype imputation (N = 97,059,328 variants) was performed using data from the Haplotype Reference Consortium and UK10K haplotype resource. We used GrafPop v1.0 to classify participants by genetic similarity; Eur participants were included in the analysis. This research has been conducted using the UK Biobank Resource under Application Number 86140.

FinnGen. FinnGen is a nationwide research project launched in 2017, combining genomic data and electronic health registers of hundreds of thousands of Finns to provide novel insights into the mechanism of human diseases[61]. Samples were genotyped with Thermo Fisher, Illumina and Affymetrix arrays and imputed with a Finnish population-specific reference panel (https://sisuproject.fi/).

Biobank Japan. The Biobank Japan Project (BBJ) is a national initiative that aims to create a biobank of genetic and clinical data from around 200,000 individuals with 47 target diseases, collected through a nationwide hospital-based genome cohort since 2013[62,63]. Genotyping of the collected DNA samples was conducted using either the Illumina HumanOmniExpressExome BeadChip or a combination of the Illumina HumanOmniExpress and HumanExome BeadChips[64]. Imputation of the genotypes was performed using a reference panel developed from 3,256 high-depth whole-genome sequences from the Japanese population combined with data from the 1000 Genomes Project[65] after standard quality control of array data.

Statistical analyses

PCA. Principal component analysis (PCA) was performed within each cohort for assessment of population substructure and for covariate adjustment in GWAS. Autosomal genotypes passing QC were LD pruned using PLINK 1.9 (ref. 66) ('–indep-pairwise 50 10 0.1'), excluding the major histocompatibility complex region. KING[58] v2.3 was used to construct a maximal unrelated set of samples with less than two degrees of relatedness. PCA was then performed on the unrelated set using PLINK v2.0 (ref. 67), and the full set including related samples were projected onto the eigenvectors.

GWAS. GWAS analyses of all scans except FinnGen and IARC (which was previously analyzed)[11] were performed in SAIGE[68] v1.1.6.2. For each GWAS, a full genetic relatedness matrix was constructed on the basis of the same LD-pruned set of cohort-specific SNPs used in the PCA. Additionally, we enabled leave-one-chromosome-out model fitting and approximate Firth effect size estimation was used for SNPs with $P < 0.05$. Sex and ancestry-specific principal components (PCs) were adjusted for in all scans, except for the UK scan, which had PCs only. A total of 20 PCs were used in the NCI-3 Eur GWAS due to observable structure in higher PCs; 10 PCs were used otherwise. For BBJ, PCA identified PC1, PC2 and PC5 as significantly associated with case–control status, after adjusting for sex; these PCs were included together with sex as model covariates for GWAS analysis conducted using SAIGE. For FinnGen, GWAS was performed using REGENIE, with sex, age (at genotyping), age squared, genotyping batch and the first ten PCs as covariates. GWAS analyses were also conducted for ccRCC and pRCC, as well as for kidney cancers categorized by age at diagnosis (<60 versus ≥60 years and <50 versus ≥65 years). We employed case–case comparisons to test for heterogeneity in SNP effects between the two histologic subtypes and the earlier- and later-onset cases[69]. In addition, we conducted GWAS analyses of kidney cancer separately for men and women, and analyses of identified kidney cancer loci stratified by BMI, smoking and hypertension. All statistical tests were two-sided. We performed a check for duplicates and relatives across studies for which individual-level data were available at NCI (all studies except for IARC, FinnGen and BBJ). In analyses of 458,752 samples, we identified 328 duplicates (0.07%), 466 first-degree relatives (0.10%) and 522 second-degree relatives (0.12%). As these samples represent an extremely small proportion of total samples, we did not re-analyze the data with them removed.

GWAS meta-analyses. We performed population-specific (Eur, Afr, Asn and a fourth stratum including Brazilian, LA1 and LA2 participants, denoted 'LA') and multi-ancestry GWAS meta-analyses for overall kidney cancer and the two major subtypes (ccRCC and pRCC). GWAS summary statistics were imported into the GWAS variant call format[70] v0.1.2 format using the bcftools[71] +munge plugin v1.16 (ref. 72). Where necessary, summary statistics were lifted over to genome build GRCh38 using the bcftools +liftover plugin (ibid.). Inverse variance-weighted fixed-effects meta-analyses were then performed using the bcftools + metal plugin (ibid.), a reimplementation of METAL[73] for GWAS variant call format. For multi-ancestry GWAS, we additionally performed random effects and Han–Eskin random effects[20] meta-analyses in METASOFT v2.0.1.

Clumping. We performed LD clumping in PLINK v1.9 (ref. 66) to identify independent loci at genome-wide significant regions. For clumping, we retained the variants that had at least 1% MAF in any one of the 1000 Genomes Project populations. Since the majority (~85%) of the dataset are Eur, we performed clumping with LD extracted from the Eur individuals in the sample having genotype-level data. This included 93,595 individuals from NCI-1, NCI-2, NCI-3 Eur, MDA and MDA-OncoArray cohorts. We used an $r^2 < 0.2$ threshold and a 2 Mb window, meaning that any two genome-wide significant SNPs having $r^2 > 0.2$ (as calculated from the in-sample LD among Eur participants in the study) and physically within 2 Mb of each other will be considered as belonging to the same locus, with the SNP with the lower P value being retained as the index SNP.

For the identified LD-clumped loci, we reviewed post-imputation QC statistics (Supplementary Table 19) from controls of studies that were imputed using the TOPMed Imputation Server (all but IARC, FinnGen, UK Biobank and BBJ). One marker (rs17789633) in one genotype set (NCI-3 Eur) showed significant deviation from HWE ($P < 3.0 \times 10^{-16}$). However, the difference in observed and expected percentage heterozygosity is small (0.494% versus 0.479%). Moreover, we observed no evidence of inter-study heterogeneity in our meta-analysis of this variant across Eur study samples ($I^2 = 0$, with associations at $P < 0.05$ in three other studies (IARC, MD and MDA-OncoArray; Supplementary Table 19). As a consequence, we retained the NCI-3 Eur data for this locus in our analysis.

Conditional analysis. We further performed conditional analysis on each of the LD-clumped loci to detect any additional SNPs within the region of 1 Mb upstream and downstream of the index variant for each of the initially identified LD-clumped locus. In-sample LD was extracted from the Eur participants in the study as described above. The analysis was performed using the standard parameter settings in the GCTA-COJO module v1.94.1 (ref. 74), using a forward stepwise selection approach, starting with the index SNP in each locus.

Allelic pleiotropy. We searched for evidence of allelic pleiotropy using two data sources. First, we used LDTrait[75] v5.6.4 to search the GWAS catalog (www.ebi.ac.uk; data as of 23 June 2023) for previously reported genome-wide significant associations with other cancers or kidney cancer risk factors (BMI/obesity, hypertension and smoking) for the identified kidney cancer loci and other correlated SNPs ($r^2 > 0.4$ in the 1000 Genomes EUR panel) within ±1 Mb using GRCh38. Database search terms are listed in the footnote of Supplementary Table 11. Second, we used the online database Open Targets Genetics (https://genetics.opentargets.org/; data as of 23 June 2023) to search UK Biobank GWAS summary statistics for associations between the kidney cancer loci and the aforementioned risk factors, using a Bonferroni-corrected significance threshold of 2.3×10^{-5} accounting for 108 loci and 20 search terms (see footnote of Supplementary Table 12).

Fine-mapping. We performed fine-mapping of kidney cancer in each of the loci identified through LD clumping. For each locus, we retained variants in the 1000 Genomes reference panel of MAF ≥5% within a region 1 Mb upstream and downstream of the sentinel (index) variant. Since the majority of individuals in our GWAS are from the Eur stratum, we extracted in-sample LD from the individuals for whom variant-level data were available. An exception is the locus on 3p25.3, indexed by the variant rs7629500; for this locus, we extracted in-sample LD from Afr participants ($n = 4{,}006$).

Subsequently, using the GWAS summary statistics and the extracted LD estimates, fine-mapping was performed by two different methods: SuSiE and PAINTOR v3 (refs. 33,34). SuSiE performs a Bayesian stepwise regression to identify for each locus the optimal number of causal signals (we specified a maximum of ten) and a 95% credible set corresponding to each causal signal. We used PAINTOR v3 to determine the posterior inclusion probability for each SNP for a given number of causal signals (maximum of three). The credible sets were obtained for each locus in the PAINTOR analysis, meaning the minimal set of variants that encompasses a pre-specified (95%) posterior probability. For each locus, the union of the unique SNPs in the credible sets identified by SuSiE and PAINTOR were considered for further testing as 'consensus' credible sets.

eQTL. eQTL analysis was performed for the associated loci identified through the GWAS by systematically querying different transcriptomic databases. This analysis was initially performed using the index SNPs of the loci associated with kidney cancer overall, ccRCC and pRCC, and then again repeated with the variants identified in the consensus credible sets (see above) for each of the susceptibility loci. We cataloged the genes that had at least one significant *cis*-eQTL (FDR <5%) in the queried set of SNPs using transcriptomic data from the following tissues in GTEx v8 (ref. 21) and TCGA[22]: GTEx kidney cortex ($N = 73$), GTEx whole blood ($N = 670$), TCGA–KIRC ($N = 527$) and TCGA–KIRP ($N = 290$).

Colocalization. Colocalization analysis was performed to see whether the signal at 7q32.1 for ccRCC and pRCC were driven by the same causal variant(s) using the coloc v5.2.3 R package.

Functional annotations. We evaluated the enrichment of SNPs significantly associated with kidney cancer and its subtypes (ccRCC and pRCC) in relation to several functional features as annotated in the Encyclopedia of DNA Elements[76] and Roadmap Epigenomics projects. Analysis of functional enrichment was performed using GARFIELD[35] v2 with default parameter settings.

Enrichment of HIF-binding sites. Excess overlap was investigated between significant SNPs and HIF-binding sites in specific RCC cell lines. First, we identified all SNPs in strong LD ($r^2 \geq 0.8$) with the index SNP at each LD-clumped locus using LD extracted from individual-level Eur data. We determined HIF-1α, HIF-2α and HIF-1β-binding sites using publicly available ChIP-seq data in RCC4 and 786-O RCC cell lines from ref. 18. We defined overlap as a locus coinciding with peaks in two or more of the five ChIP-seq datasets of HIF-binding sites. We then performed a bootstrapping analysis by randomly shuffling the GWAS significant loci 500,000 times across the genome and at least 5 Mb away from the originally identified loci, evaluating statistical significance against the null hypothesis that the susceptibility loci are not enriched for HIF-binding sites. We further extended our search for overlapping loci to ±25 kb around the ChIP-seq peaks and calculated the *P* value for the observed overlap similarly as before.

Identifying TF-binding motifs. To identify potential TF-binding sites for a given set of SNPs, motif analysis was performed using the motifbreakR[77] v2.14.2 package, which includes the HOCOMOCO[36] v11 positional weight matrices database, with default parameters (downloaded on 5 May 2023). We evaluated the enrichment of binding sites for several a priori selected TFs, by repeating the experiment 1,000 times with randomly selected SNPs having *P* value >0.01 and at least 1 Mb away from each of the SNPs in the initial analysis, allowing us to calculate the *P* value for overrepresentation of specific TF-binding sites.

Enrichment of RCC-related epigenomic annotations. We used data from a previous detailed epigenomic profiling of the major subtypes of RCC (ccRCC, pRCC and chRCC)[37] to investigate whether ccRCC and pRCC loci GWAS are enriched among the corresponding H3K27ac peaks across multiple samples. Any GWAS identified loci was termed to be overlapping only if it mapped to H3K27ac peaks for at least two samples. Given that the authors identified a striking similarity among epigenetic profiles of pRCC and chRCC, we also performed an overlap analysis of the GWAS loci associated with pRCC in the H3K27ac peaks of chRCC samples. Lastly, to identify the specificity of the overlapping loci, we investigated whether the loci identified for ccRCC overlapped with the H3K27ac peaks in at least one pRCC sample and vice versa.

Heritability and genetic correlation. GWAS heritability on the liability scale was estimated using the GCTA[78] v1.94.1 software and individual-level data from the NCI-1, NCI-2 and Eur samples in NCI-3 scans. Analyses assumed a disease prevalence of 1.66%, were restricted to SNPs of MAF >0.05, removed subjects with more than 5% of genotypes missing and adjusted for sex, study and the top ten PCs. Genetic correlation, in the observed scale, was computed through linkage disequilibrium score regression using the 1000 Genomes EUR reference panel[79].

PRS. We identified 324,805 unrelated Eur participants (1,696 cases and 323,109 controls) in UK Biobank for PRS analysis, with the PRS constructed from the identified kidney cancer susceptibility loci that are present in the Eur population stratum ($n = 107$). Two baseline risk prediction models were used for logistic regression: the first contained sex and genetic principal components as covariates, while the second model adjusted for these covariates as well as age at enrollment, BMI, smoking status and hypertension. UK Biobank bgen files were converted to a plink binary format using a hard-call-threshold 0.1. PLINK 2.0 (ref. 67) –score was then used to calculate the PRS. AUC values were calculated with the R pROC package.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

All GWAS summary statistics are available on dbGaP (phs003505.v1.p1) and GWAS Catalog (GCST90320043–GCST90320065). Individual-level data from the new NCI-3 scan are available on dbGaP (phs003505.v1.p1). Data from previously published scans are also available on dbGaP (NCI-1, phs000351.v1.p1; NCI-2, phs001736.v2.p1; USKC, phs000863.v1.p1; IARC-2, phs001271.v1.p1; MDA and MDA-OncoArray, phs003505.v1.p1). The individual-level data from the IARC-1 and UK scans have not been deposited in dbGaP or any other data archive site given decisions by the institutional ethics review boards for these projects. The data from these scans are available upon reasonable request through internal processes unique to each institution. Such requests can be made in writing to the principal investigators (IARC: P. Brennan, pbrennan@iarc.fr and UK: R.H., richard.houlston@icr.ac.uk); the time frame from request to receipt of data is approximately 4–6 weeks. The UK Biobank analysis was conducted via application number 86140 (https://www.ukbiobank.ac.uk/). The Finnish biobank data included in FinnGen can be accessed via Fingenious services at https://site.fingenious.fi/en/ (ref. 80) managed by FINBB. Finnish Health register data can be applied for via Findata at https://findata.fi/en/data/ (ref. 81). The full GWAS results of the BBJ are available via the website of the Japanese ENcyclopedia of GEnetic Associations by Riken (JENGER) at http://jenger.riken.jp/en/ (ref. 82; case–control GWAS no. 156). Function annotation enrichment was performed with the annotation data provided via the GARFIELD package at https://www.ebi.ac.uk/birney-srv/GARFIELD/ (ref. 83). Position weight matrices for transcription factor-binding sites as cataloged in HOCOMOCO were provided along with the motifbreakR R package, in the associated MotifDb database. ChIP-Seq data reported by Schmid et al.[18] are publicly available through the Gene Expression Omnibus (GEO) database under the accession codes: GSE120885 (HIF-1α, HIF-2α and HIF-1β ChIP-seq in RCC4 cells) and GSE67237 (HIF-2α and HIF-1β ChIP-seq in 786-O cells). Epigenomic charting data (H3K27ac peaks) generated by Nassar et al.[37] are publicly available through GEO database under accession code GSE188486; the sample attributes are mentioned in Supplementary Table 1 of the corresponding paper. GTEx v8 and TCGA data can be accessed via GTEx and Genomic Data Commons at https://gtexportal.org/home/ (ref. 84) and https://portal.gdc.cancer.gov/repository (ref. 85), respectively. Additionally, eQTLs for TCGA were queried via the PancanQTL database at http://gong_lab.hzau.edu.cn/PancanQTL/ (ref. 86).

Code availability

Code used in performing the liftover of summary statistics and fixed-effects GWAS meta-analyses (version 2022-12-23) is available via GitHub at https://github.com/freeseek/score (ref. 72). No previously unreported custom computer code or algorithm was used to generate results.

References

47. Jin, Y., Schaffer, A. A., Feolo, M., Holmes, J. B. & Kattman, B. L. GRAF-pop: a fast distance-based method to infer subject ancestry from multiple genotype datasets without principal components analysis. *G3* **9**, 2447–2461 (2019).
48. *Database of Genotypes and Phenotypes* (NCBI, 2014); https://www.ncbi.nlm.nih.gov/gap/
49. Brown, D. W., Myers, T. A. & Machiela, M. J. PCAmatchR: a flexible R package for optimal case-control matching using weighted principal components. *Bioinformatics* **37**, 1178–1181 (2021).

50. Chen, S. et al. A genomic mutational constraint map using variation in 76,156 human genomes. *Nature* **625**, 92–100 (2022).
51. Taliun, D. et al. Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *Nature* **590**, 290–299 (2021).
52. Das, S. et al. Next-generation genotype imputation service and methods. *Nat. Genet.* **48**, 1284–1287 (2016).
53. Purdue, M. P. et al. Genome-wide association study of renal cell carcinoma identifies two susceptibility loci on 2p21 and 11q13.3. *Nat. Genet.* **43**, 60–65 (2011).
54. Wu, X. et al. A genome-wide association study identifies a novel susceptibility locus for renal cell carcinoma on 12p11.23. *Hum. Mol. Genet* **21**, 456–462 (2012).
55. Shu, X. et al. Potential susceptibility loci identified for renal cell carcinoma by targeting obesity-related genes. *Cancer Epidemiol. Biomark. Prev.* **26**, 1436–1442 (2017).
56. Henrion, M. et al. Common variation at 2q22.3 (ZEB2) influences the risk of renal cancer. *Hum. Mol. Genet* **22**, 825–831 (2013).
57. Purdue, M. P. et al. A genome-wide association study of renal cell carcinoma among African Americans. *Cancer Epidemiol. Biomark. Prev.* **23**, 209–214 (2014).
58. Manichaikul, A. et al. Robust relationship inference in genome-wide association studies. *Bioinformatics* **26**, 2867–2873 (2010).
59. Karczewski, K. J. et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434–443 (2020).
60. Bycroft, C. et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).
61. Kurki, M. I. et al. FinnGen provides genetic insights from a well-phenotyped isolated population. *Nature* **613**, 508–518 (2023).
62. Nagai, A. et al. Overview of the BioBank Japan Project: study design and profile. *J. Epidemiol.* **27**, S2–S8 (2017).
63. *Index* (BioBank Japan, 2017); https://biobankjp.org/en/index.html
64. Terao, C. et al. Chromosomal alterations among age-related haematopoietic clones in Japan. *Nature* **584**, 130–135 (2020).
65. Tanaka, N. et al. Eight novel susceptibility loci and putative causal variants in atopic dermatitis. *J. Allergy Clin. Immunol.* **148**, 1293–1306 (2021).
66. Purcell, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
67. Chang, C. C. et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).
68. Zhou, W. et al. Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. *Nat. Genet.* **50**, 1335–1341 (2018).
69. Begg, C. B. & Zhang, Z. F. Statistical analysis of molecular epidemiology studies employing case-series. *Cancer Epidemiol. Biomark. Prev.* **3**, 173–175 (1994).
70. Lyon, M. S. et al. The variant call format provides efficient and robust storage of GWAS summary statistics. *Genome Biol.* **22**, 32 (2021).
71. Danecek, P. et al. Twelve years of SAMtools and BCFtools. *Gigascience* **10**, giab008 (2021).
72. *freeseek* (Github, 2022); https://github.com/freeseek/score
73. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).
74. Yang, J. et al. Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat. Genet.* **44**, 369–375 (2012).
75. Lin, S. H., Brown, D. W. & Machiela, M. J. LDtrait: an online tool for identifying published phenotype associations in linkage disequilibrium. *Cancer Res.* **80**, 3443–3446 (2020).
76. Consortium, E. P. et al. Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature* **583**, 699–710 (2020).
77. Coetzee, S. G., Coetzee, G. A. & Hazelett, D. J. motifbreakR: an R/Bioconductor package for predicting variant effects at transcription factor binding sites. *Bioinformatics* **31**, 3847–3849 (2015).
78. Yang, J. et al. Genetic variance estimation with imputed variants finds negligible missing heritability for human height and body mass index. *Nat. Genet.* **47**, 1114–1120 (2015).
79. Bulik-Sullivan, B. K. et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).
80. *Fingenius* (FINBB, 2024); https://site.fingenious.fi/en/
81. *Data* (FINDATA, 2017); https://findata.fi/en/data/
82. *Laboratory for Statistical and Translational Genetics* (Japanese ENcyclopedia of GEnetic associations by Riken, 2021); http://jenger.riken.jp/en/
83. *GARFIELD* (EMBL-EBI, 2015); https://www.ebi.ac.uk/birney-srv/GARFIELD/
84. *GTEx Portal* (GTEx, 2017); https://gtexportal.org/home/
85. *Repository* (Genomic Data Commons, 2024); https://portal.gdc.cancer.gov/repository
86. *Cis-eQTLs and Trans-eQTLs in 33 Cancer Types* (PancanQTL, 2018); http://gong_lab.hzau.edu.cn/PancanQTL/

Acknowledgements

We thank O. Jahagirdar and A. Klein for their efforts in implementing various computational pipelines used in this project. This research was supported by the Intramural Research Program of the National Cancer Institute, National Institutes of Health, US Department of Health and Human Services, as well as with Federal funds from the National Cancer Institute under contract no. 75N91019D00024. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the US Government. Where authors are identified as personnel of the International Agency for Research on Cancer/World Health Organization, the authors alone are responsible for the views expressed in this article and they do not necessarily represent the decisions, policy or views of the International Agency for Research on Cancer/World Health Organization. Acknowledgments and funding sources for participating centers are provided in Supplementary Table 17.

Author contributions

M.P.P., D.D., M.J.M., B.R.G., P. Brennan, J.G. and S.J.C. contributed to the design and execution of the overall study. M.Y., N.R.C., B.D.H., M.R.M. and A.A.H. performed the experiments. M.P.P., D.D., M.J.M., B.R.G., T.W., D.O., S.C., A. Liu, K.H., K.M.B., J.R. and S.J.C. contributed to the design and execution of the statistical analysis. M.P.P., D.D., M.J.M., B.R.G. and S.J.C. contributed to the first draft of the paper. P. Brennan, J.G. and S.J.C. edited the paper. A.F.-I., P.S., A. Liu, C.W., S.O.A., J. Larkin, S.C.Z., M.S., K.H., A.H., K.A.L., F. Cárcano, O.B., B.S., K.G.N., G.M., D.S., W.R.D., M.A.A.K.F., A.v.B., F.N., J.N.H., N.R., W.Y.H., W.M.L., A. Lori, M.F., M.Z.-M., S.V.S., W.J.M., Biobank Japan Project, A.V., R.D., F. Carusso, L.S.G., K.A., M.A.B., C.A., I.P., S. Ricard, FinnGen, G.S., R.E.B., N.S.V., N.S., G.D.S., A.A., S.B., D.H., N.G., P.P., M.S., A.P., F.I.N., M.J.F., X.Z., L.J.M., M.K., T.E., S.A.C., D.C.C., R.G.U., D.Z., A.M., I.H., A.H., L.F., V. Janout, D.M., V. Jinga, S. Rascu, M.M., S.S., S.M., V.G., B.A.-A., J.M., M.J., L.P., L.H., J. Li, I.L., S.M.B., A.G.S., C.T.G.S., R.B.R., F.P.G., M.D.C., M.P., G.-S.M.L., M.L.F., A.J., S.E.G., A.S., R.H.T., V.S., D.D.T., C.T.B., D.A., E.T.L., W.C.N., V.A.M., A.V.P., J.-C.B., N.D.F., P. Bigot, R.M.R., L.M.C., A.F., B.J.M., C.T., T.K.C., D.M.C., R.H., J.E.E.-P., P.H.A., A.G., P. Brennan and J.G. contributed samples and/or data. All authors critically reviewed the paper. The following

authors contributed equally as co-first authors: M.P.P., D.D., M.J.M. and B.R.G. The following authors contributed equally to the work as co-second authors: T.W., D.O., S.C., A.F.I., P.A.S., A. Liu, C.W., S.O.A., J. Larkin, S.C.Z., M.S., K.H., A.H., K.A.L., F. Cárcano, O.B., B.S., K.G.N., G.M., D.S., W.R.D., M.A.A.K.F., A.v.B. and F.N. The following authors contributed equally as co-second-last authors: D.A., E.T.L., W.C.N., V.A.M., A.V.P., L.M.C., J.C.B., N.F., P. Bigot, R.M.R., A.F., B.J.M., C.T., T.K.C., D.M., R.H., J.E.E.-P., P.H.A. and A.G. The following authors jointly supervised this work: P. Brennan, J.G. and S.J.C.

Competing interests

D.S. has received funds from Janssen for consulting outside of the submitted work. N.S.V. has received grants, personal fees and non-financial support from Bristol Myers Squibb; personal fees and non-financial support from Ipsen and EUSA Pharma; and personal fees from Merck Serono, Pfizer, Eisai Ltd and 4D Pharma, all outside the submitted work. G.D.S. has received educational grants from Pfizer and AstraZeneca; consultancy fees from Pfizer, Merck, EUSA Pharma and MSD; travel expenses from Pfizer and speaker fees from Pfizer, all outside the submitted work. M.S. has received honoraria from Covidien/Medtronic for teaching on courses and speaker fees from Pfizer, all outside the submitted work. L.M.C. has received research funding from BMS, Novartis and GSK, all outside the submitted work. B.J.M. has received funding from the NCCN Hereditary Kidney Cancer Panel and Merck, all outside the submitted work. T.K.C. has received funding from Alkermes, AstraZeneca, Aravive, Aveo, Bayer, Bristol Myers Squibb, Calithera, Circle Pharma, Deciphera Pharmaceuticals, Eisai, EMD Serono, Exelixis, GlaxoSmithKline, Gilead, IQVIA, Infinity, Ipsen, Jansen, Kanaph, Lilly, Merck, Nikang, Nuscan, Novartis, Oncohost, Pfizer, Roche, Sanofi/Aventis, Scholar Rock, Surface Oncology, Takeda, Tempest, Up-To-Date, CME events (Peerview, OncLive, MJH, CCO and others), outside the submitted work. The other authors declare no competing interests.

Additional information

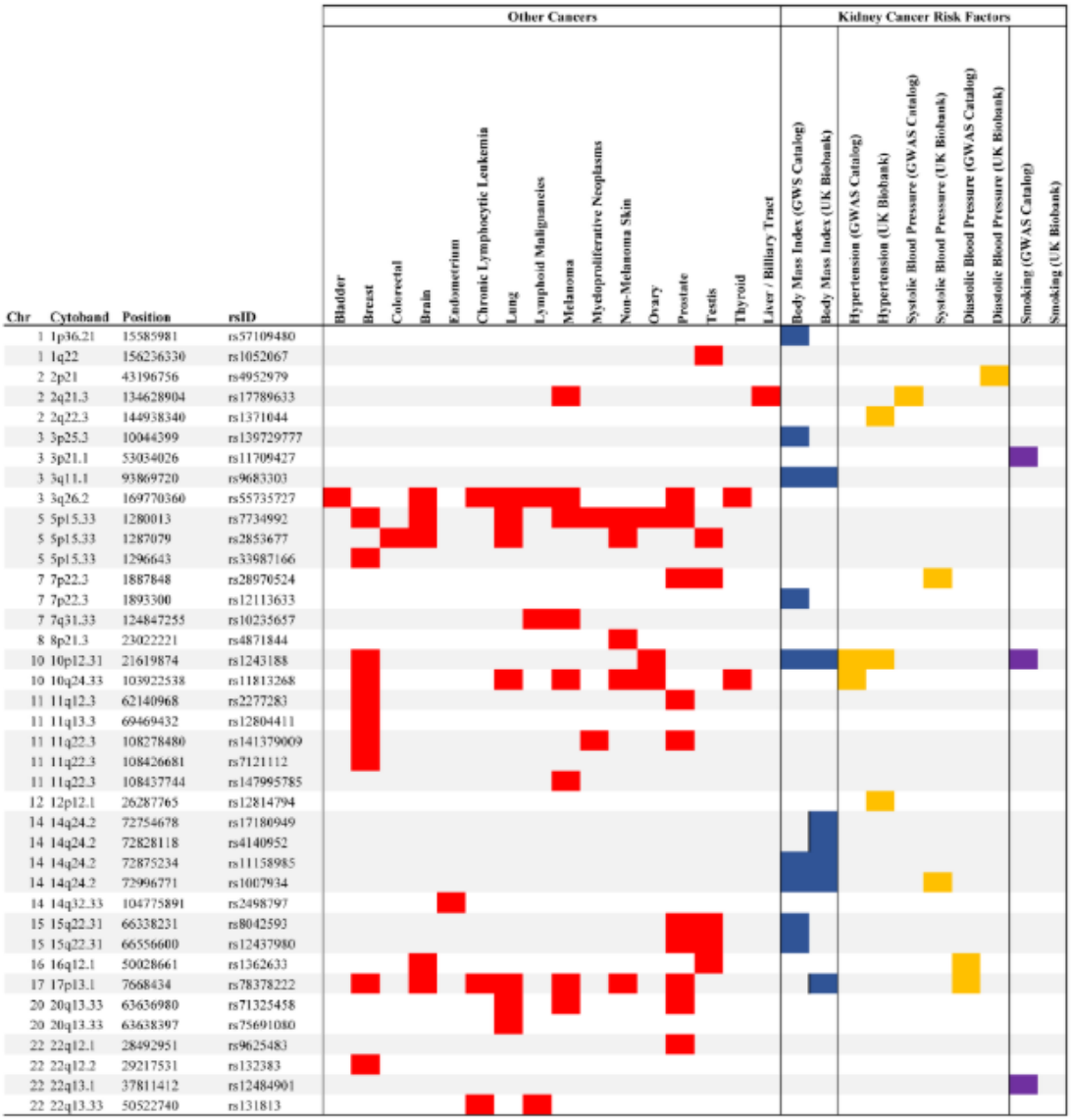
Extended data is available for this paper at https://doi.org/10.1038/s41588-024-01725-7.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41588-024-01725-7.

Correspondence and requests for materials should be addressed to Mark P. Purdue or Stephen J. Chanock.

Peer review information *Nature Genetics* thanks Christopher Amos and A. Ari Hakimi for their contribution to the peer review of this work. Peer reviewer reports are available.

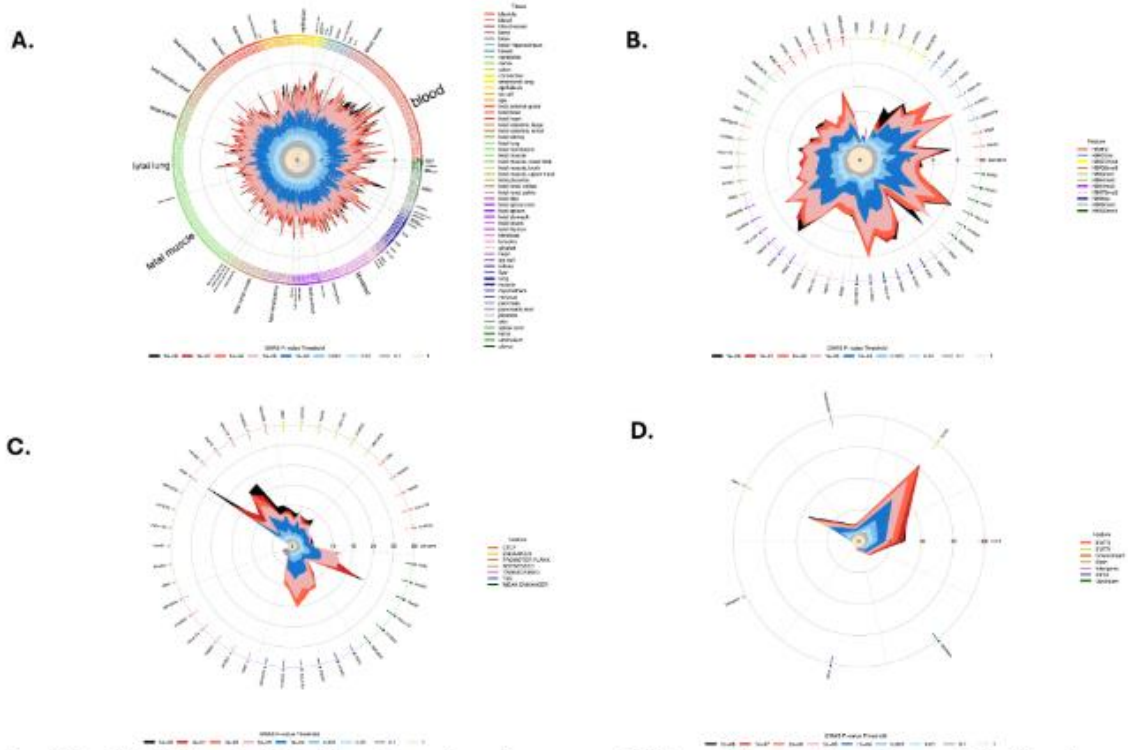
https://doi.org/10.1038/s41588-024-01725-7



Chr	Cytoband	Position	rsID
1	1p36.21	15585981	rs57109480
1	1q22	156236330	rs1052067
2	2p21	43196756	rs4952979
2	2q21.3	134628904	rs17789633
2	2q22.3	144938340	rs1371044
3	3p25.3	10044399	rs139729777
3	3p21.1	53034026	rs11709427
3	3q11.1	93869720	rs9683303
3	3q26.2	169770360	rs55735727
5	5p15.33	1280013	rs7734992
5	5p15.33	1287079	rs2853677
5	5p15.33	1296643	rs33987166
7	7p22.3	1887848	rs28970524
7	7p22.3	1893300	rs12113633
7	7q31.33	124847255	rs10235657
8	8p21.3	23022221	rs4871844
10	10p12.31	21619874	rs1243188
10	10q24.33	103922538	rs11813268
11	11q12.3	62140968	rs2277283
11	11q13.3	69469432	rs12804411
11	11q22.3	108278480	rs141379009
11	11q22.3	108426681	rs7121112
11	11q22.3	108437744	rs147995785
12	12p12.1	26287765	rs12814794
14	14q24.2	72754678	rs17180949
14	14q24.2	72828118	rs4140952
14	14q24.2	72875234	rs11158985
14	14q24.2	72996771	rs1007934
14	14q32.33	104775891	rs2498797
15	15q22.31	66338231	rs8042593
15	15q22.31	66556600	rs12437980
16	16q12.1	50028661	rs1362633
17	17p13.1	7668434	rs78378222
20	20q13.33	63636980	rs71325458
20	20q13.33	63638397	rs75691080
22	22q12.1	28492951	rs9625483
22	22q12.2	29217531	rs132383
22	22q13.1	37811412	rs12484901
22	22q13.33	50522740	rs131813

Extended Data Fig. 1 | Pleiotropic effects of kidney cancer loci for other cancers and risk factors. Pleiotropy matrix summarizing kidney cancer susceptibility loci with evidence of pleiotropic effects for other cancers and/or selected risk factors (body mass index, hypertension, blood pressure, smoking) from searches of GWAS Catalog and UK Biobank GWAS summary statistics. Cell colors indicating associations with specific traits: red, other cancers; blue, body mass index; orange, hypertension or blood pressure; violet, smoking. Locus-trait summary statistics listed in Supplementary Table 11 (GWAS Catalog) and 12 (UK Biobank).

https://doi.org/10.1038/s41588-024-01725-7



Extended Data Fig. 2 | *In silico* analysis of enrichment for putative regulatory annotations among kidney cancer loci. Enrichment of variants associated with overall RCC in (**a**) DNAse Hotspots (**b**) Histone modification sites (**c**) Chromatin states (**d**) different genic locations. The enrichments are depicted for variants at different p-value thresholds denoted by the colors and across different categories in each panel. The results were computed using GARFIELD v2.