Title

Pathway Analysis of Renal Cell Carcinoma Genome-Wide Association Studies Identifies Novel Associations

Journal

Cancer Epidemiology Biomarkers & Prevention, 29(10)

ISSN

1055-9965

Authors

Purdue, Mark P
Song, Lei
Scélo, Ghislaine
et al.

Publication Date

2020-10-01

DOI

10.1158/1055-9965.epi-20-0472

Peer reviewed

HHS Public Access

Author manuscript

Cancer Epidemiol Biomarkers Prev. Author manuscript; available in PMC 2022 September 02.

Published in final edited form as:

Cancer Epidemiol Biomarkers Prev. 2020 October ; 29(10): 2065–2069.

doi:10.1158/1055-9965.EPI-20-0472.

Pathway analysis of renal cell carcinoma genome-wide association studies identifies novel associations

Mark P. Purdue[1], **Lei Song**[1], **Ghislaine Scélo**[2], **Richard S. Houlston**[3], **Xifeng Wu**[4], **Lori C. Sakoda**[5], **Khanh Thai**[5], **Rebecca E. Graff**[6], **Nathaniel Rothman**[1], **Paul Brennan**[7], **Stephen J. Chanock**[1], **Kai Yu**[1]

[1]Division of Cancer Epidemiology and Genetics, National Cancer Institute, Rockville, USA

[2]Department of Medical Sciences, University of Turin

[3]Division of Genetics and Epidemiology, The Institute of Cancer Research, Sutton, UK

[4]Department of Big Data in Health Science, Zhejiang University School of Public Health, Hangzhou, China

[5]Division of Research, Kaiser Permanente Northern California, Oakland, USA

[6]Department of Epidemiology and Biostatistics, University of California, San Francisco, San Francisco, USA

[7]International Agency for Research on Cancer, Lyon, France.

Abstract

Background: Much of the heritable risk of renal cell carcinoma (RCC) associated with common genetic variation is unexplained. New analytic approaches have been developed to increase the discovery of risk variants in genome-wide association studies (GWAS), including multi-locus testing through pathway analysis.

Methods: We conducted a pathway analysis using GWAS summary data from six previous scans (10,784 cases and 20,406 controls) and evaluated 3,678 pathways and gene sets drawn from the Molecular Signatures Database. To replicate findings, we analyzed GWAS summary data from the UK Biobank (903 cases and 451,361 controls) and the Genetic Epidemiology Research on Adult Health and Aging cohort (317 cases and 50,511 controls).

Results: We identified 14 pathways / gene sets associated with RCC in both the discovery ($P < 1.36 \times 10^{-5}$, the Bonferroni correction threshold) and replication ($P < 0.05$) sets, 10 of which include components of the PI3K/AKT pathway. In tests across 2,035 genes in these pathways, associations (Bonferroni-corrected $P < 2.46 \times 10^{-5}$ in discovery and replication sets combined) were observed for *CASP9*, *TIPIN* and *CDKN2C*. The strongest SNP signal was for rs12124078 ($P_{Discovery} = 2.6 \times 10^{-5}$, $P_{Replication} = 1.5 \times 10^{-4}$, $P_{Combined} = 6.9 \times 10^{-8}$), a *CASP9* expression quantitative trait locus.

Corresponding author: Mark Purdue, Occupational and Environmental Epidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Rockville, MD 20850; purduem@mail.nih.gov; tel: +1 (240) 276-7265.

Conflict of interest statement: the authors have no conflicts of interest to disclose.

Conclusions: Our pathway analysis implicates genetic variation within the PI3K/AKT pathway as a source of RCC heritability and identifies several promising novel susceptibility genes, including *CASP9*, which warrant further investigation.

Impact: Our findings illustrate the value of pathway analysis as a complementary approach to analyzing GWAS data.

Keywords

genome-wide association study; kidney cancer; renal cell carcinoma; pathway analysis; meta-analysis

INTRODUCTION

Kidney cancer is one of the ten most common cancers in the United States, with around 74,000 new cases and 14,000 related deaths in 2019 (1). Renal cell carcinoma (RCC) is the most common type of kidney cancer, accounting for over 90% of kidney cancer diagnoses. RCC has a heritable basis, with relatives of patients having a two-fold increased risk (2, 3). While a number of rare familial RCC syndromes caused by inheritance of high-impact mutations have been identified, even collectively, they only account for less than 5% of RCC (4). Evidence for the role of common low-impact genetic variants influencing RCC risk has been established in genome-wide association studies (GWAS), which have so far identified 13 susceptibility loci (5).

Since GWAS risk variants typically have a small effect size, they are difficult to identify in individual SNP-based GWAS after accounting for multiple testing, even with large sample numbers. Pathway-based analyses, involving joint testing of SNPs within gene sets defined by biological pathways, have the potential to empower the identification of new associations not captured by testing individual genetic variants (6, 7).

To gain further insight into the heritability of RCC, we evaluated 3,678 canonical pathways and gene sets using summary-level data from a GWAS meta-analysis. To validate our findings, we analyzed GWAS summary data from the UK Biobank and Kaiser Permanente Genetic Epidemiology Research on Adult Health and Aging (GERA) cohorts.

MATERIALS AND METHODS

Study Populations

The RCC GWAS meta-analysis, described previously (5), combined summary results from six independent GWAS totaling 10,784 RCC cases and 20,406 controls of European ancestry. Briefly, genotypes had been assayed across the scans using a combination of Illumina SNP arrays (Illumina Inc, San Diego, CA, USA). After performing imputation on all scans using 1,094 subjects from the 1000 Genomes Project (phase 1 release 3) as the reference panel, 7,437,091 SNPs were included in the meta-analysis. To facilitate the identification of novel genetic signals, we excluded from our pathway analysis 36,616 SNPs within 500kb of genetic variants previously reported to be associated with RCC at

genome-wide significance ($P < 5 \times 10^{-8}$). All tests of statistical significance used in this analysis were two sided.

To replicate study findings we made use of summary-level association statistics for 30,798,054 SNPs from a GWAS of RCC conducted among UK Biobank participants (903 cases, 451,361 controls) downloaded from GeneATLAS (http://geneatlas.roslin.ed.ac.uk/) (8). The SNP beta coefficients and standard errors in GeneATLAS were computed using mixed linear models; we transformed these summary statistics to odds ratios (ORs) using LMOR (9). Standard errors were calculated from the reported *P*-value and estimated OR.

For replication of SNP-level analyses, we used summary results from UK Biobank and, for selected SNPs (n=10), a GWAS of kidney cancer conducted among persons of European ancestry in the GERA cohort (317 cases and 50,511 controls). Details of the GERA GWAS have been previously described (10).

Pathway analysis

We downloaded definitions for 3,762 human-derived pathways and gene sets (C2 gene set collection) from the Broad Institute Molecular Signatures Database (MSigDB) v6.1 (http://software.broadinstitute.org/gsea/msigdb/collections.jsp) for the pathway-level analysis. Genomic definitions for genes were downloaded from human genes NCBI36 and reference genome GRCh37.p13 using the Ensemble BioMart tool.

We conducted gene- and pathway-level meta-analyses using the summary statistics-based adaptive rank truncated product (sARTP) method (https://www.rdocumentation.org/packages/ARTP2/versions/0.9.45/topics/sARTP). sARTP combines SNP associations across variants 20kb upstream and downstream of a given gene with adjustment for the size of genes and pathways through a resampling procedure to evaluate the global testing *P*-value, with proper adjustment of multiple comparisons (11). A web-based sARTP application tool is available allowing users to submit pathway analysis jobs online and receive results computed using NCI computing resources (https://analysistools.nci.nih.gov/pathway/). Significance of gene- and pathway-level associations were estimated from the null distribution generated from 10 million resampling steps. A panel of 503 European subjects (population codes: CEU, TSI, FIN, GBR, IBS) in the 1000 Genomes Project (phase 3, v5) was used in sARTP to estimate the linkage disequilibrium between SNPs. To mitigate the impact of population stratification, we applied genomic control inflation factors to rescale the standard errors of the log odds ratios for SNPs in each GWAS (lambda values 1.009 – 1.058) and in the meta-analysis (lambda = 1.037).

We successfully analyzed 3,678 of the 3,762 pathways and gene sets downloaded from MSigDB; tests of 84 pathways failed because of a lack of SNP coverage. To control the family-wise error rate in our discovery pathway analysis, we considered a Bonferroni-corrected P-value of 1.36×10^{-5} as being statistically significant (i.e., 0.05/3,678). For our analysis of promising pathway-level results in replication datasets, we considered an alpha of 0.05 as being significant. A schematic summarizing our pathway analysis approach is provided in Supplementary Figure 1.

Gene-level and SNP-level analyses within selected pathways

We evaluated gene-level and SNP-level test results among all constituent genes of pathways found to be associated with RCC in the discovery and replication sets, and combined association statistics through meta-analysis using fixed-effects models.

Functional annotation of SNP associations

We queried public databases to explore the possible biologic effects of selected SNPs. We searched RegulomeDB (http://www.regulomedb.org) and HaploReg (https://pubs.broadinstitute.org/mammals/haploreg/haploreg.php) to assess the likelihood that SNPs map to regulatory elements, and the NHGRI-EBI GWAS Catalog (https://www.ebi.ac.uk/gwas/) to search for previously reported GWAS associations with other traits. We assessed potential SNP associations with gene expression in 527 TCGA (The Cancer Genome Atlas) renal cancer tumor cases (KIRC) using the PancanQTL database (http://bioinfo.life.hust.edu.cn/PancanQTL/) (12). We also explored eQTL kidney cortex expression data (n=73) using GTEx (https://gtexportal.org/home/) (13)

RESULTS

We identified 14 pathways and gene sets significantly associated with RCC in the discovery and replication sets, ranging in size from 12 to 732 genes and 402 to 31,986 SNPs (Table 1). Notably, 10 of the 14 pathways / gene sets include components of the Phosphatidylinositol-3-kinase (PI3K) /Akt signaling pathway, with *AKT1* and *CASP9* among the most significant gene-level signals for all 10 pathways. The four other gene sets were “Benporath cycling genes” (genes related to embryonic stem cell identity showing cell-cycle stage-specific expression), “Fortschegger PHF8 targets up” (genes upregulated in HeLa cells upon knockdown of PHF8 by RNAi), “West adrenocortical tumor dn” (down-regulated genes in pediatric adrenocortical tumors compared to the normal tissue) and “Pujana CHEK2 PCC network” (genes positively co-expressed with *CHEK2*).

We also explored gene- and SNP-level signals within the 14 significant pathways and gene sets (Supplementary Table 1). In testing across the 2,035 constituent genes, associations with RCC (Bonferroni-corrected $P < 2.46 \times 10^{-5}$ in discovery and replication sets combined) were observed for *CASP9* ($P = 3.7 \times 10^{-7}$), *TIPIN* ($P = 8.2 \times 10^{-6}$) and *CDKN2C* ($P = 1.7 \times 10^{-5}$). Promising gene signals in both the discovery and replication sets were also observed for *AKT1, ARID1A, EP300, FANCD2, HIST1H4, KCNK3, MAP2K1, RBPMS* and *RPL4*. We identified 4 highly promising SNPs with associations in both the discovery and replication sets at $P < 0.0001$ and $P < 0.05$, respectively (Table 2): rs12124078 (within the *CASP9* region; $P_{Combined} = 6.9 \times 10^{-8}$), rs41324853 (*CDKN2C*; $P_{Combined} = 2.4 \times 10^{-7}$), rs61758464 (*AKT1*; $P_{Combined} = 3.5 \times 10^{-7}$) and rs2979488 (*RBPMS*; $P_{Combined} = 4.4 \times 10^{-7}$).

We explored the potential functional impact of these SNPs by integration of publicly accessible resources (Table 3). The variant rs12124078 has a RegulomeDB score of 1f, being associated with *CASP9* expression across several non-kidney tissues. We confirmed that this eQTL relationship extends to kidney tissue, with the higher-risk A allele associated

with reduced *CASP9* expression in both TCGA ($P = 3.6 \times 10^{-7}$) and GTEx ($P = 0.0075$) datasets, as well as the majority of other tissue sets in GTEx (Supplementary Figure 2). While the three other SNPs had weaker predicted functional relevance, rs2979488 and rs61758464 were associated with expression of *RBPMS* and *ZBTB42* respectively in TCGA. In a search of the NHGRI-EBI GWAS Catalog for associations with other traits, rs12124078 and rs2979488 have previously been significantly associated with glomerular filtration rate and leukocyte count, respectively.

DISCUSSION

In this pathway-based meta-analysis of RCC GWAS summary results, we identified 14 pathways and gene sets associated with risk. In targeted SNP investigations across the 14 pathways in the discovery and replication sets, we observed an association approaching genome-wide significance overall for the variant rs12124078, which we found to be consistently associated with *CASP9* expression in kidney tissue. We also observed promising associations with genetic variation in close proximity to *AKT1*, *CDKN2C, TIPIN* and *RBPMS.*

The majority of the pathway findings appear to be driven by nucleotide variation in components of the PI3K/AKT signaling network, an important regulator of cell growth, proliferation, metabolism, survival, and apoptosis (14). This is one of the most frequently dysregulated signal transduction pathways in human cancers, including kidney cancer, with genetic alterations in constituent genes present in 15% of RCC (15). PI3K/AKT signaling is particularly important in the pathogenesis of clear cell RCC; aberrant pathway activation leads to upregulation of mammalian target of rapamycin (mTOR) signaling, which in turn upregulates hypoxia-inducible factor-mediated expression of angiogenic factors (16). PI3K-inhibiting therapeutic agents are used in treating metastatic RCC.

We also observed replicable pathway-level signals for four gene sets that are related to cancer; two involve cell cycle regulation (“Benporath cycling genes” and “Pujana CHEK2 PCC network”) (17, 18) , a third captures genes downregulated in pediatric aderenocortical tumors compared to normal tissue (“West adrenocortical tumor dn”) (19), and the fourth lists genes up-regulated upon knockdown of PHF8, a histone lysine demethylase and suspected transcription activator overexpressed in several types of cancer (“Fortschegger PHF8 targets up”) (20, 21). As all four gene sets are comparatively large, involving between 247 and 732 genes, it is possible that the observed RCC associations are reflective of signals from a subset of genes, such as *CDKN2C* and *RBPMS*.

When we conducted gene- and SNP-level investigations within the 14 significant pathways, the strongest evidence of an association with RCC was with *CASP9* and the nearby variant rs12124078, with the A allele associated with increased risk. *CASP9* encodes caspase-9, a critical initiator of cell apoptosis that is regulated by PI3K/AKT signaling; Akt phosphorylation at serine-196 inhibits caspase-9 protease activity, decreasing apoptosis (22, 23). Interestingly, the rs12124078 A allele has also been associated with lower glomerular filtration rate and decreased *CASP9* expression in peripheral blood monocytes (24). We have confirmed this eQTL, with the A allele being associated with reduced *CASP9* expression in

TCGA and GTEx kidney tissue and the majority of other GTEx tissue sets. Collectively, these findings are consistent with a reduction in caspase-9-mediated apoptotic activity potentially underlying the association between rs12124078 and RCC.

Genetic variation within the *AKT1*, *RBPMS*, *TIPIN* and *CDKN2C* gene regions also showed promising evidence of association with RCC. *AKT1* is a key member of the PI3K/AKT pathway, encoding a serine/threonine kinase regulating numerous mechanisms affecting cell growth, metabolism and angiogenesis (25). We found the nearby variant rs61758464 to be associated with RCC and while rs61758464 was not related to *AKT1* expression in renal tissue, it is notable that the risk allele was associated with reduced *AKT1* expression ($P = 2.2 \times 10^{-8}$; FDR = 2.5×10^{-4}) in blood eQTL data (26). We also observed in TCGA, but not GTEx, an association between this variant and expression of *ZBTB42*, which encodes a poorly characterized member of the C2H2 zinc finger protein family suspected to play a role in skeletal muscle development (27).

RBPMS encodes a member of the RNA recognition motif family of RNA-binding proteins. The function of RBPMS is poorly understood, although recent evidence suggests a role in mRNA transport and localization (28). The biologic basis for a role of *RBPMS* in RCC development remains to be established, although it has been shown to interact with VHL in cultured 786-O renal cancer cells (29). Intriguingly, *RBPMS* expression has been reported to be significantly elevated in tumor tissue of obese vs. non-obese clear cell RCC patients suggesting a possible link between BMI and renal cancer (30).

TIPIN, which was associated with RCC in gene-level testing, encodes a replisome-associated protein that contributes to genome maintenance by mediating Chk1 and Chk2 activation in response to DNA damage (31). *TIPIN* expression has been reported to be down-regulated in kidney tumor vs. matched normal tissue, possibly reflecting dysregulation of cell-cycle checkpoints (32).

Our findings for *CDKN2C*, involved in cell cycle regulation, and the nearby SNP rs41324853 likely reflects a previously reported GWAS risk locus. Although we filtered out GWAS results for SNPs within 500kb of previously identified GWAS hits prior to our analysis to prioritize the discovery of new loci, rs41324853 is 542kb from and moderately correlated with ($r^2 = 0.33$, D' = 0.93) the known RCC GWAS risk marker rs4381241 (5). When we ran a logistic model including both SNPs within the discovery set, rs41324853 was no longer associated with RCC risk ($P = 0.39$). Our eQTL analyses do not offer any further insight into the causal pathway underlying this locus.

Strengths of our pathway-based GWAS meta-analysis, to our knowledge the first of its kind for RCC, include the large sample size and the use of an independent GWAS replication set for confirmation of pathway-, gene- and SNP-level findings. An additional strength is our use of the sARTP method for pathway analysis, which possesses many useful properties. Pathway analysis generally targets two types of null hypotheses: the competitive null hypothesis (33) (i.e., that genes in a candidate pathway are no more associated with the outcome than any other genes outside this pathway) and the self-contained null hypothesis (34) (i.e., that none of the genes in a pathway of interest is associated with the outcome).

The sARTP procedure focuses on the self-contained null hypothesis, as the main goal of this project is to identify outcome-associated genes or loci. As pointed out by Goeman et al. (35), tests for the competitive null hypothesis often assume that genotype measured at different genes are independent when evaluating the association significance level. This assumption, which is generally invalid in practice, is not required by sARTP when testing the self-contained null hypothesis. Other strengths of sARTP are its use of summary data and that it can handle large pathways, which might consist of thousands of genes and tens of thousands of SNPs. Other advantages of sARTP include its use of summary data and its capability handling large pathways, which might consist of thousands of genes and tens of thousands of SNPs (11).

A limitation of the sARTP approach involves the inherent assumption of mapping SNPs 20 kb upstream and downstream of each gene to identify candidate SNPs that may play a regulatory role in gene expression. This distance has been used for annotating SNPs/genes in previous pathway analyses (11, 36), as studies have shown functional variants are located approximately 16–20kb within transcription start sites (37–39). However, it does not capture *cis* regulatory effects outside this window, nor *trans* mechanisms.

In summary, our pathway-based analysis of the RCC GWAS meta-analysis has provided new promising genetic susceptibility regions that merit further investigation and functional follow-up.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Funding: This work was supported by the Intramural Research Program of the NIH and the National Cancer Institute (R01 CA201358). The Genetic Epidemiology Research on Adult Health and Aging (GERA) cohort was established through funding from the National Institute of Aging (RC2 AG036607), the Robert Wood Johnson Foundation, the Wayne and Gladys Valley Foundation, The Ellison Medical Foundation, and the Kaiser Permanente National and Regional Community Benefit Programs. 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.

Abbreviations

CI	confidence interval
eQTL	expression quantitative trait locus
GERA	Genetic Epidemiology Research on Adult Health and Aging cohort
GWAS	genome-wide association study
KIRC	kidney renal cell carcinoma
OR	odds ratio
MSigDB	Molecular Signatures Database

RCC	renal cell carcinoma
SNP	single-nucleotide polymorphism
sARTP	summary statistics-based adaptive rank truncated product
TCGA	the Cancer Genome Atlas

REFERENCES

1. American Cancer Society. Cancer Facts & Figures 2019. Atlanta: American Cancer Society;2019.
2. Hung RJ, Moore L, Boffetta P, Feng BJ, Toro JR, Rothman N, et al. Family history and the risk of kidney cancer: a multicenter case-control study in Central Europe. Cancer Epidemiol Biomarkers Prev 2007;16:1287–90. [PubMed: 17548699]
3. Karami S, Schwartz K, Purdue MP, Davis FG, Ruterbusch JJ, Munuo SS, et al. Family history of cancer and renal cell cancer risk in Caucasians and African Americans. Br J Cancer 2010;102:1676–80. [PubMed: 20442711]
4. Haas NB, Nathanson KL. Hereditary kidney cancer syndromes. Adv Chronic Kidney Dis 2014;21:81–90. [PubMed: 24359990]
5. Scelo G, Purdue MP, Brown KM, Johansson M, Wang Z, Eckel-Passow JE, et al. Genome-wide association study identifies multiple risk loci for renal cell carcinoma. Nat Commun 2017;8:15724. [PubMed: 28598434]
6. Moore JH, Asselbergs FW, Williams SM. Bioinformatics challenges for genome-wide association studies. Bioinformatics 2010;26:445–55. [PubMed: 20053841]
7. Wang K, Li M, Hakonarson H. Analysing biological pathways in genome-wide association studies. Nat Rev Genet 2010;11:843–54. [PubMed: 21085203]
8. Canela-Xandri O, Rawlik K, Tenesa A. An atlas of genetic associations in UK Biobank. Nat Genet 2018;50:1593–9. [PubMed: 30349118]
9. Lloyd-Jones LR, Robinson MR, Yang J, Visscher PM. Transformation of Summary Statistics from Linear Mixed Model Association on All-or-None Traits to Odds Ratio. Genetics 2018;208:1397–408. [PubMed: 29429966]
10. Rashkin SR, Graff RE, Kachuri L, Thai KK, Alexeff SE, Blatchins MA, et al. Pan-cancer study detects novel genetic risk variants and shared genetic basis in two large cohorts. bioRxiv 2019; 635367.
11. Zhang H, Wheeler W, Hyland PL, Yang Y, Shi J, Chatterjee N, et al. A Powerful Procedure for Pathway-Based Meta-analysis Using Summary Statistics Identifies 43 Pathways Associated with Type II Diabetes in European Populations. PLoS Genet 2016;12:e1006122.
12. Gong J, Mei S, Liu C, Xiang Y, Ye Y, Zhang Z, et al. PancanQTL: systematic identification of cis-eQTLs and trans-eQTLs in 33 cancer types. Nucleic Acids Res 2018;46:D971–D6. [PubMed: 29036324]
13. Consortium GT. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. Science 2015;348:648–60. [PubMed: 25954001]
14. Martini M, De Santis MC, Braccini L, Gulluni F, Hirsch E. PI3K/AKT signaling pathway and cancer: an updated review. Ann Med 2014;46:372–83. [PubMed: 24897931]
15. Ricketts CJ, De Cubas AA, Fan H, Smith CC, Lang M, Reznik E, et al. The Cancer Genome Atlas Comprehensive Molecular Characterization of Renal Cell Carcinoma. Cell Rep 2018;23:313–26 e5. [PubMed: 29617669]
16. Guo H, German P, Bai S, Barnes S, Guo W, Qi X, et al. The PI3K/AKT Pathway and Renal Cell Carcinoma. J Genet Genomics 2015;42:343–53. [PubMed: 26233890]
17. Ben-Porath I, Thomson MW, Carey VJ, Ge R, Bell GW, Regev A, et al. An embryonic stem cell-like gene expression signature in poorly differentiated aggressive human tumors. Nat Genet 2008;40:499–507. [PubMed: 18443585]

18. Pujana MA, Han JD, Starita LM, Stevens KN, Tewari M, Ahn JS, et al. Network modeling links breast cancer susceptibility and centrosome dysfunction. Nat Genet 2007;39:1338–49. [PubMed: 17922014]
19. West AN, Neale GA, Pounds S, Figueredo BC, Rodriguez Galindo C, Pianovski MA, et al. Gene expression profiling of childhood adrenocortical tumors. Cancer Res 2007;67:600–8. [PubMed: 17234769]
20. Fortschegger K, de Graaf P, Outchkourov NS, van Schaik FM, Timmers HT, Shiekhattar R. PHF8 targets histone methylation and RNA polymerase II to activate transcription. Mol Cell Biol 2010;30:3286–98. [PubMed: 20421419]
21. Shao P, Liu Q, Maina PK, Cui J, Bair TB, Li T, et al. Histone demethylase PHF8 promotes epithelial to mesenchymal transition and breast tumorigenesis. Nucleic Acids Res 2017;45:1687–702. [PubMed: 27899639]
22. Cardone MH, Roy N, Stennicke HR, Salvesen GS, Franke TF, Stanbridge E, et al. Regulation of cell death protease caspase-9 by phosphorylation. Science 1998;282:1318–21. [PubMed: 9812896]
23. Shultz JC, Goehe RW, Wijesinghe DS, Murudkar C, Hawkins AJ, Shay JW, et al. Alternative splicing of caspase 9 is modulated by the phosphoinositide 3-kinase/Akt pathway via phosphorylation of SRp30a. Cancer Res 2010;70:9185–96. [PubMed: 21045158]
24. Pattaro C, Kottgen A, Teumer A, Garnaas M, Boger CA, Fuchsberger C, et al. Genome-wide association and functional follow-up reveals new loci for kidney function. PLoS Genet 2012;8:e1002584.
25. Kumar A, Rajendran V, Sethumadhavan R, Purohit R. AKT kinase pathway: a leading target in cancer research. ScientificWorldJournal 2013;2013:756134.
26. Jansen R, Hottenga JJ, Nivard MG, Abdellaoui A, Laport B, de Geus EJ, et al. Conditional eQTL analysis reveals allelic heterogeneity of gene expression. Hum Mol Genet 2017;26:1444–51. [PubMed: 28165122]
27. Devaney SA, Mate SE, Devaney JM, Hoffman EP. Characterization of the ZBTB42 gene in humans and mice. Hum Genet 2011;129:433–41. [PubMed: 21193930]
28. Farazi TA, Leonhardt CS, Mukherjee N, Mihailovic A, Li S, Max KE, et al. Identification of the RNA recognition element of the RBPMS family of RNA-binding proteins and their transcriptome-wide mRNA targets. RNA 2014;20:1090–102. [PubMed: 24860013]
29. Lai Y, Song M, Hakala K, Weintraub ST, Shiio Y. The interaction of the von Hippel-Lindau tumor suppressor and heterochromatin protein 1. Arch Biochem Biophys 2012;518:103–10. [PubMed: 22234250]
30. Eckel-Passow JE, Serie DJ, Bot BM, Joseph RW, Hart SN, Cheville JC, et al. Somatic expression of ENRAGE is associated with obesity status among patients with clear cell renal cell carcinoma. Carcinogenesis 2014;35:822–7. [PubMed: 24374825]
31. Leman AR, Noguchi E. Local and global functions of Timeless and Tipin in replication fork protection. Cell Cycle 2012;11:3945–55. [PubMed: 22987152]
32. Mazzoccoli G, Piepoli A, Carella M, Panza A, Pazienza V, Benegiamo G, et al. Altered expression of the clock gene machinery in kidney cancer patients. Biomed Pharmacother 2012;66:175–9. [PubMed: 22436651]
33. Burren OS, Guo H, Wallace C. VSEAMS: a pipeline for variant set enrichment analysis using summary GWAS data identifies IKZF3, BATF and ESRRA as key transcription factors in type 1 diabetes. Bioinformatics 2014;30:3342–8. [PubMed: 25170024]
34. Swanson DM, Blacker D, Alchawa T, Ludwig KU, Mangold E, Lange C. Properties of permutation-based gene tests and controlling type 1 error using a summary statistic based gene test. BMC Genet 2013;14:108. [PubMed: 24199751]
35. Goeman JJ, Buhlmann P. Analyzing gene expression data in terms of gene sets: methodological issues. Bioinformatics 2007;23:980–7. [PubMed: 17303618]
36. Walsh N, Zhang H, Hyland PL, Yang Q, Mocci E, Zhang M, et al. Agnostic Pathway/Gene Set Analysis of Genome-Wide Association Data Identifies Associations for Pancreatic Cancer. J Natl Cancer Inst 2018;
37. Wang L, Jia P, Wolfinger RD, Chen X, Zhao Z. Gene set analysis of genome-wide association studies: methodological issues and perspectives. Genomics 2011;98:1–8. [PubMed: 21565265]

38. Fagny M, Paulson JN, Kuijjer ML, Sonawane AR, Chen CY, Lopes-Ramos CM, et al. Exploring regulation in tissues with eQTL networks. Proc Natl Acad Sci U S A 2017;114:E7841–E50. [PubMed: 28851834]

39. Veyrieras JB, Kudaravalli S, Kim SY, Dermitzakis ET, Gilad Y, Stephens M, et al. High-resolution mapping of expression-QTLs yields insight into human gene regulation. PLoS Genet 2008;4:e1000214.

Table 1.

Pathways and gene sets significantly associated with renal cell carcinoma in the GWAS meta-analysis ($P < 1.36 \times 10^{-5}$)[a] and UK Biobank ($P<0.05$)

			Discovery Set (GWAS Meta-Analysis; 10,784 cases and 20,406 controls		Replication Set (UK Biobank; 903 cases and 451,361 controls)	
MSigDB Pathway / Gene Set	**N_{Genes}**	**N_{SNPs}**	***P***	**Most Significant Genes[b]**	***P***	**Most Significant Genes[b]**
Benporath cycling genes	612	26858	2.00×10^{-7}	*TIPIN, CDKN2C, CENPQ*	0.016	*FANCD2,RAN,ZC3HC1*
Reactome AKT phosphorylates targets in the cytosol	12	402	6.50×19^{-7}	*AKT1, CASP9, CDKN1B*	0.00018	*CASP9,AKT1S1,AKT2*
BioCarta RAS pathway	21	957	7.50×10^{-7}	A*KT1, CASP9, MAP2K1*	0.0034	*CASP9*
Reactome PI3K/AKT activation	34	1557	8.50×10^{-7}	*AKT1, CASP9, IRS2*	0.0060	*CASP9*
Fortschegger PHF8 targets up	247	14870	1.30×10^{-6}	*TIPIN, INSR, GRB10*	0.013	*IGF1R,ATF4,CTSC*
Reactome PIP3 activates AKT signaling	25	1095	1.45×10^{-6}	*AKT1, CASP9, CDKN1B*	0.0045	*CASP9*
West adrenocortical tumor dn	491	31986	2.25×10^{-6}	*RBPMS, IRF5, ZFP36L2*	0.042	*CASP9,RPL38,KCNK3*
Reactome signaling by SCF-KIT	72	3976	5.10×10^{-6}	*AKT1, CASP9, MAP2K1*	0.0074	*CASP9,CDK1,SOCS1*
Reactome GAB1 signalosome	34	1832	6.15×10^{-6}	*AKT1, CASP9, CDKN1B*	0.0060	*CASP9*
KEGG prostate cancer	82	5107	6.25×10^{-6}	*AKT1, CASP9, MAP2K1*	0.00046	*CASP9,IGF1R,ATF4*
BioCarta HDAC pathway	29	2284	6.45×10^{-6}	*AKT1, INSR, MEF2D*	0.034	*IGF1R*
Reactome PI3K events in ERBB4 signaling	34	2757	7.05×10^{-6}	*AKT1, CASP9, CDKN1B*	0.0060	*CASP9*
PID PI3KCI AKT pathway	33	1451	9.00×10^{-6}	*AKT1, CASP9, CDKN1B*	0.00058	*CASP9*
Pujana CHEK2 PCC network	732	26725	9.50×10^{-6}	*POT1, CDKN2C, TFDP2*	0.016	*RAN,MIEF1,CENPA*

[a] Pathways/gene sets statistically significant at Bonferroni-adjusted α-level of 1.36×10^{-5} (0.05/3678).

[b] Top 3 gene-level test results with a *P*-value <0.001

Table 2.

Summary of SNPs within significant pathways with RCC associations observed in the discovery ($P < 1.0 \times 10^{-4}$) and replication ($P<0.05$) sets.

				Discovery		Replication[b]		Combined[c]	
SNP (Gene)	**Chr**	**Position[a]**	**A/a**	**OR (95% CI)**	***P***	**OR (95% CI)**	***P***	**OR (95% CI)**	***P***
rs12124078 (*CASP9*)	1	15869899	A/G	0.92 (0.88 to 0.96)	2.6×10^{-5}	0.84 (0.77 to 0.92)	1.5×10^{-4}	0.91 (0.87 to 0.94)	6.9×10^{-8}
rs41324853 (*CDKN2C*)	1	51449575	T/C	1.10 (1.05 to 1.14)	2.0×10^{-5}	1.15 (1.05 to 1.25)	2.3×10^{-3}	1.11 (1.06 to 1.15)	2.4×10^{-7}
rs2979488 (*RBPMS*)	8	30280630	A/G	1.10 (1.05 to 1.14)	1.3×10^{-5}	1.13 (1.03 to 1.25)	8.8×10^{-3}	1.11 (1.06 to 1.15)	4.4×10^{-7}
rs61758464 (*AKT1*)	14	10525780	G /A	0.87 (0.82 to 0.93)	5.2×10^{-6}	0.86 (0.76 to 0.98)	0.022	0.87 (0.83 to 0.92)	3.5×10^{-7}

[a] GRCh37.p13

[b] Replication includes results from UK Biobank (903 cases, 451 361 controls) and GERA Cohort (317 cases, 50 511 controls) combined by meta-analysis using fixed effects model.

[c] Summary results from meta-analysis using fixed effects model.

Table 3.

Exploration of SNP functional relevance: RegulomeDB score, expression quantitative trait locus (eQTL) analyses for selected SNP-gene pairs in tumor (TCGA-KIRC) and normal (GTEx) kidney tissue samples and previous genome-wide significant findings for other traits.

				eQTL Analyses				
				TCGA, KIRC (N=527)		GTEx, Kidney Cortex (N=73)		
SNP	**A/a**	**RegulomeDB Score**	**Gene**	**B** [a]	***P***	**β**	***P***	**In NHGRI-EBI GWAS Catalog (Trait)**
rs12124078	A/G	1f	*CASP9*	0.26	3.6×10^{-7}	0.39	0.0075	Glomerular filtration rate
rs41324853	T/C	4	- [b]					- [c]
rs2979488	A/G	4	*RBPMS*	−0.27	2.6×10^{-6}	0.02	0.89	Leukocyte count
rs61758464	G/A	5	*ZBTB42*	−0.28	1.2×10^{-4}	−0.12	0.51	-

[a] SNP-gene association at false discovery rate < 0.05 in database of TCGA eQTLs (12). β represents directional expression effect for rare allele.

[b] No eQTL identified.

[c] No entry in GWAS Catalog.