Carcinogenesis, 2018, Vol. 39, No. 6, 752–757

doi:10.1093/carcin/bgy021
Advance Access publication April 4, 2018
Original Article

ORIGINAL ARTICLE

Germline genetic variants in somatically significantly mutated genes in tumors are associated with renal cell carcinoma risk and outcome

Xiang Shu[1], Jianchun Gu[1,2], Maosheng Huang[1], Nizar M.Tannir[3], Surena F.Matin[4], Jose A.Karam[4], Christopher G.Wood[4], Xifeng Wu[1,*] and Yuanqing Ye[1]

[1]Department of Epidemiology, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA, [2]Department of Oncology, Xinhua Hospital Affiliated to Shanghai Jiaotong University School of Medicine, Shanghai 200092, China, [3]Department of Urology and [4]Department of Genitourinary Medical Oncology, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA

*To whom correspondence should be addressed. Tel: +1 713 745 2485; Fax: +1 713 792 4657; Email: xwu@mdanderson.org

Abstract

Genome-wide association studies (GWAS) have identified 13 susceptibility loci for renal cell carcinoma (RCC). Additional genetic loci of risk remain to be explored. Moreover, the role of germline genetic variants in predicting RCC recurrence and overall survival (OS) is less understood. In this study, we focused on 127 significantly mutated genes from The Cancer Genome Atlas (TCGA) Pan-Cancer Analysis across 12 major cancer sites to identify potential genetic variants predictive of RCC risk and clinical outcomes. In a three-phase design with a total of 2657 RCC cases and 5315 healthy controls, two single nucleotide polymorphisms (SNPs) that map to *PIK3CG* (rs6466135:A, OR_{meta} = 0.85, 95% CI = 0.77–0.94, $P_{meta} = 1.4 \times 10^{-3}$) and *ATM* (rs611646:T, OR_{meta} = 1.17, 95% CI = 1.05–1.31, $P_{meta} = 3.5 \times 10^{-3}$) were significantly associated with RCC risk. With respect to RCC recurrence and OS, two separate datasets with a total of 661 stages I–III RCC patients (discovery: 367; validation: 294) were analyzed. The most significant association was observed for rs10932384:C (*ERBB4*) with both outcomes (recurrence: HR_{meta} = 0.52, 95% CI = 0.39–0.68, $P_{meta} = 3.81 \times 10^{-6}$; OS: HR_{meta} = 0.50, 95% CI = 0.37–0.67, $P_{meta} = 6.00 \times 10^{-6}$). In addition, six SNPs were significantly associated with either RCC recurrence or OS but not both ($P_{meta} < 0.01$). Rs10932384:C was significantly correlated with mutation frequency of *ERBB4* in clear cell RCC (ccRCC) patients ($P = 0.003$, Fisher's exact test). *Cis*-eQTL was observed for several SNPs in blood/transformed fibroblasts but not in RCC tumor tissues. In summary, we identified promising genetic predictors of recurrence and OS among RCC patients with localized disease.

Introduction

Kidney cancer remains one of the top 10 most commonly diagnosed cancers in the United States with an estimate of 63 990 new cases and 14 400 deaths in 2017 (1). Renal cell carcinoma (RCC) accounts for more than 80% of kidney cancer. There are several established modifiable lifestyle risk factors for kidney cancer including cigarette smoking, obesity and hypertension (2). Although the disease is curable at its early form, approximately 20–40% of localized patients may experience recurrence which leads to worse prognosis (3). The 5-year survival rate is only 11.7% for patients with distant metastasis (4). Reported prognostic factors are currently limited to clinicopathological variables such as clinical stage (5–7), grade (6,7), tumor size (7) and microvascular invasion (8).

Accumulating evidence supports that genetic factors play an important role in RCC development. Approximately 2–4% of RCC cases are hereditary which might contain mutations in high-penetrance predisposing genes including von Hippel–Lindau

Received: October 17, 2017; **Revised:** January 23, 2018; **Accepted:** March 26, 2018

Abbreviations

CI	confidence interval
GWAS	Genome-wide association study
HR	hazard ratio
OR	odds ratio
OS	overall survival
RCC	renal cell carcinoma
SNP	single nucleotide polymorphism
SMG	significantly mutated genes
TCGA	The Cancer Genome Atlas
VHL	von Hippel–Lindau tumor suppressor

tumor suppressor (*VHL*), folliculin (*FLCN*), MET proto-oncogene, receptor tyrosine kinase (*MET*), fumarate hydratase (*FH*) and BRCA1 associated protein 1 (*BAP1*) (9,10). For the risk of RCC, large scale genome-wide association studies (GWAS) have identified several genetic susceptibility loci that map to 1p32.3, 2p21, 2q22.3, 3p22.1, 3q26.2, 8p21.3, 8q24.21, 10q24.33-q25.1, 11q13.3, 11q22.3, 12p11.23, 12q24.31 and 14q24.2 (11–15). However, there have been very limited studies performed to identify genetic predictors with respect to RCC clinical outcomes. Previous studies taking a candidate gene approach reported that common single nucleotide polymorphisms (SNPs) in *VHL*, hypoxia inducible factor 1 alpha subunit (*HIF1A*), vascular endothelial growth factor A (*VEGF*) and miRNA-related genes were significantly associated with RCC recurrence and overall survival (OS) (16–18). The limitations of these studies included relatively small sample size and lack of validation.

The molecular data generated by The Cancer Genome Atlas (TCGA) has greatly enhanced current knowledge of genomic, epigenomic, transcriptomic and proteomic alterations in a variety of cancers, including RCC. The Pan-Cancer analyses across 12 major cancer sites have identified 127 significantly mutated genes (SMG) (19). However, whether common germline genetic variants in these SMGs also contribute to RCC development and prognosis is unknown. Since these SMGs involved in a variety of known and emerging cellular process in tumorigenesis and cancer progression, the interplay between germline susceptibility and somatic mutations warrants further investigations.

In the present study, we hypothesized that common genetic variants located in 127 SMGs may play important roles in RCC development and prognosis. Through a multi-phase design, we sought to identify and validate new potential susceptibility loci for RCC and genetic predictors of recurrence and OS among localized patients.

Materials and methods

Study population

Cases and controls in the discovery analysis were ascertained from an ongoing RCC case–control study at the University of Texas MD Anderson Cancer Center since 2002 (20). All cases were newly diagnosed, histologically confirmed RCC patients. The majority of our cases had clear cell RCC (ccRCC, N = 594, 75%). The rest of cases had papillary RCC (N = 62, 8%), or other histology types (N = 133, 17%). There were no age, sex, ethnicity or cancer-stage restrictions on recruitment. Healthy control subjects without a history of cancer (except non-melanoma skin cancer) were recruited using random digit dialing method. Basic demographics, tobacco and alcohol use history, family history of cancer, weight and height to calculate body mass index and medical history were obtained by in-person interview. After the interview, a 40-ml blood sample was collected from each participant and delivered to the laboratory for molecular analysis. For smoking history, individuals who had never smoked or had smoked <100 cigarettes in a lifetime were classified as never smokers. Those subjects who had quit smoking >12 months prior to diagnosis/recruitment were considered former smokers. Clinical information including clinical stage, grade, comorbidities, pathological stage, histology, treatments, recurrence and vital status was extracted from medical records by trained staffs. Local recurrence was defined as the reappearance of renal tumor locally after previous tumor resection or partial nephrectomy. The study protocol was approved by the MD Anderson Cancer Center Institutional Review Board. All participants provided the written informed consent. A total of 789 newly diagnosed RCC cases and 797 controls were included in the discovery set. To assess the association with recurrence and OS, we restricted the analysis to 367 patients with localized disease (stages I–III patients). Cases and controls in the MD Anderson validation analysis were derived from a previously published RCC GWAS (11). After excluding subjects overlapped with the discovery set, this internal validation set composed of 557 cases and 1094 controls. For the association with recurrence and survival, 294 patients with localized disease (stages I–III patients) were included. A publicly available GWAS data was downloaded from National Cancer Institute The database of Genotypes and Phenotypes (dbGAP study Accession: phs000351.v1.p1), which consists of 1311 RCC cases and 3424 controls. Detailed information regarding the NCI GWAS populations were described previously (12). Since clinical and follow-up information was not available from dbGAP, the genotype data was used for validating SNPs identified for RCC risk only. In the current analysis, all subjects included in the discovery and validation sets were non-Hispanic whites.

SNP selection and genotyping

The 127 SMGs identified by TCGA Pan-Cancer effort across 12 major cancers (http://cancergenome.nih.gov/) were reported previously (19). For each gene, we identified tagging SNPs in a region 5 kb upstream of transcription start site and 5 kb downstream of 3′ transcription stop site with R^2 threshold of 0.8 in CEU population and MAF≥0.05. A total of 2159 tagging SNPs were initially selected by Tagger using HapMap database (NCBI B36 assembly, dbSNP b126) and sent to Illumina (San Diego, CA) for custom Infinium iSelect beadchip design. A total of 402 SNPs failed due to low design score (cutoff 0.42) or failure codes and therefore 1757 SNPs remained in the customized beadchip.

For the discovery phase, genotyping was performed according to the standard Infinium II assay protocol for the iSelect HD BeadChips. Quality control measures were performed to exclude SNPs with a call rate <95% (26 SNPs) and those failed Hardy–Weinberg disequilibrium test ($P < 0.05$, 66 SNPs). After quality control, a total of 1665 SNPs remained for the final analysis. For validation using MD Anderson RCC GWAS data, genotyping was performed using HumanHap610/660W BeadChips (Illumina, San Diego, CA) and detailed description of quality control was provided elsewhere (11). For validation using NCI GWAS data, HumanHap 500, 610 or 660W BeadChips were used in the primary scan of the NCI population (12). If the candidate SNPs identified from the discovery phase were not directly genotyped in the MD Anderson validation population or NCI population, proxies located within the same LD block ($R^2 > 0.8$) and a range of ±500 kb were searched using SNiPA (http://snipa.helmholtz-muenchen.de/snipa/index.php).

Statistical analysis

Continuous host characteristics were analyzed using Student *t* test, whereas categorical variables were analyzed using Pearson chi-square test. Logistic regression was performed for testing the association between SNPs and RCC risk and estimate the odds ratios (ORs) and 95% confidence intervals (CIs) while adjusting for age, sex and smoking status. Recurrence was defined as a combination of local recurrence and progression to metastatic disease and patients who died or who were alive at the last follow-up were censored. Time to recurrence was calculated from date of diagnosis to date of first documented local recurrence and progression to metastatic disease or date of last follow-up or date of death, whichever came first. Overall survival was calculated from date of diagnosis to date of death or date of last follow-up, whichever came first. Cox regression was performed for testing the association between SNPs and risk of recurrence and death and estimating hazard ratios (HRs) and 95% CIs while adjusting for age, sex, clinical stage, grade and treatments in both discovery and validation sets. Assumption of proportional hazards were examined by the residual plots. All three genetic models (i.e. dominant, recessive and additive) were tested in the logistic

and Cox regression analyses. Nominal significant SNPs with $P < 0.05$ in the discovery set were considered candidates for validation. A fixed-effect model was selected if the test of heterogeneity was not significant ($P > 0.05$). Otherwise, a random-effect model was used to generate the estimates based on the meta-analysis. Interactions were evaluated by creating multiplicative interaction term between SNPs and sex or smoking status in the regression models. Associations between SNPs and somatic mutations (level 2 SNP genotyping data and level 3 Mutation Annotation Format files were downloaded from TCGA portal; dbGaP Study Accession: phs000178.v1.p1) were assessed by Fisher's exact test. SNPs that were not directly assayed were imputed by IMPUTE2 (21). *Cis*-expression quantitative trait loci (*cis*-eQTL) analysis was assessed for the SNPs within the same LD block of the validated SNPs ($R^2 > 0.8$) using SNiPA (http://snipa.helmholtz-muenchen.de/snipa/index.php) (22) and further examined using data downloaded from The Cancer Genome Atlas (TCGA, dbGaP Study Accession: phs000178.v9.p8). Genotype data for SNPs within the same LD block of the validated SNPs were imputed by IMPUTE2 (21) using the SNP genotyping data downloaded from TCGA. Detailed approach for *cis*-eQTL analysis using TCGA data was described elsewhere (23). In short, multivariate linear regression was first used on somatic copy number alterations and the CpG methylation levels in the promoter region to estimate residual expression level from the tissue gene expression level and the resulting residual expression was then regressed on the germline genotypes. Enhancer and DNase enrichment analyses were evaluated using Haploreg (http://www.broadinstitute.org/mammals/haploreg/haploreg_v2.php) (24). All tests were performed using STATA v13.0 (College Station, TX) and Plink (http://pngu.mgh.harvard.edu/purcell/plink/) (25) with a two-sided α level of 0.05 as threshold for significance.

Results

Patient characteristics

This study included a total of 2657 cases and 5315 controls for risk analysis with the discovery phase of 789 cases and 797 controls, followed by internal validation of 557 cases and 1094 controls and finally external validations of 1311 cases and 3424 controls (Supplementary Table 1, available at *Carcinogenesis* Online). We restricted the clinical outcome analysis in 661 RCC patients with localized disease (stages I–III) including 367 patients in the discovery phase and 294 patients in the validation phase (Supplementary Table 2, available at *Carcinogenesis* Online). Almost all of the stages I–III RCC patients (over 99%) received surgery in both the discovery and validation data and only a small percentage of patients received other treatments such as chemotherapy, radiotherapy and targeted therapies.

SNPs associated with RCC risk

For RCC risk analysis, Table 1 showed the two SNPs that were significantly associated with RCC risk in the discovery and validation phase: rs6466135 (proxy: rs757903) and rs611646 (proxy: rs3824987). The A allele of rs6466135, which maps to *PIK3CG*, was significantly associated with reduced RCC risk ($P_{meta} = 1.4 \times 10^{-3}$, $P_{het} = 0.80$). ORs (95% CIs) were 0.80(0.65–0.98) in discovery phase, 0.87(0.77–0.97) in validation phase, and 0.85(0.77–0.94) in all three studies combined. rs611646 was located in the intron of *ATM* and the minor alleles of this SNP were significantly associated with increased risks of RCC ($OR_{meta} = 1.17$, 95% $CI_{meta} = 1.05–1.31$, $P_{meta} = 3.5 \times 10^{-3}$, $P_{het} = 0.78$).

SNPs associated with RCC outcomes

Among stages I–III patients, three SNPs in three different genes (*ERBB4*, *EGFR*, *RUNX1*) exhibited significant association with recurrence in both discovery and validation phases. The most significant association was observed for rs10932384 in the meta-analysis (effect allele: C, HR[95% CI] = 0.41[0.26–0.64], 0.60[0.42–0.86] and 0.52[0.39–0.68] in the discovery phase, validation phase and combined discovery and validation data, respectively; $P_{meta} = 3.8 \times 10^{-6}$, $P_{het} = 0.18$, Table 2). Significant associations with OS were observed for five SNPs in four genes (*ERBB4, TBL1XR1, TLR4, SETBP1*) in both discovery and validation phases. Again, rs10932384 was the most significant SNP (effect allele: C, HR[95% CI] = 0.51[0.27–0.96], 0.49[0.35–0.70], and 0.50[0.37–0.67] in the discovery phase, validation phase and combined discovery and validation data, respectively; $P_{meta} = 6.0 \times 10^{-6}$, $P_{het} = 0.92$, Table 2). All other identified SNPs were associated with either RCC recurrence or OS but not both. No significant interactions between SNPs and sex or smoking status were observed (data not shown).

Functional characterization of identified SNPs

To further investigate potential functions of identified SNPs, we conducted tests for the correlation between identified SNPs and mutation frequency of the gene that the SNP maps to using data from TCGA. A significant correlation was observed between rs10932384:C and *ERBB4* mutations in TCGA ccRCC patients ($P_{exact} = 0.003$, Table 3). For the *cis*-eQTL analysis using SNiPA, significant effect was found for four regions, i.e. rs6466135-*PIK3CG* in blood (rs6466134, $r^2 = 1.00$, $P = 5.48 \times 10^{-7}$), rs611646-*ATM* in transformed fibroblasts (rs227069, $r^2 = 0.94$, $P = 3.36 \times 10^{-7}$), rs984654-*EGFR* in blood (rs4947488, $r^2 = 0.80$, $P = 1.24 \times 10^{-4}$) and rs2770150-*TLR4* in blood (rs5030728, $r^2 = 0.88$, $P = 8.15 \times 10^{-35}$)

Table 1. SNPs associated with RCC risk by meta-analysis

SNP	Gene mapped	Model	OR[a]	95% CI	P	P_{het}
rs6466135, A[b]	*PIK3CG*					
MDA discovery		Dom	0.80	0.65–0.98	0.03	
MDA validation		Dom	0.87	0.70–1.08	0.205	
NCI validation		Dom	0.86	0.75–0.99	0.037	
Validation, meta		Dom	0.87	0.77–0.97	0.015	0.96
All stages, meta		Dom	0.85	0.77–0.94	0.0014	0.80
rs611646, T[c]	*ATM*					
MDA discovery		Dom	1.25	1.01–1.55	0.043	
MDA validation		Dom	1.18	0.94–1.48	0.164	
NCI validation		Dom	1.14	0.98–1.32	0.085	
Validation, meta		Dom	1.15	1.02–1.3	0.03	0.81
All stages, meta		Dom	1.17	1.05–1.31	0.0035	0.78

[a]Adjusted for age, sex and smoking status (Y/N) for MDA discovery and validation phases; sex and studies for NCI validation set.
[b]Proxy rs757903 ($r^2 = 1$) was analyzed in the MDA validation and NCI validation sets.
[c]Proxy rs3824987 ($r^2 = 0.99$) was analyzed in the MDA validation and NCI validation sets.

Table 2. SNPs associated with RCC clinical outcome by meta-analysis in stages I–III RCC patients

SNP[a]	Proxy[b]	r^2	Gene mapped	Model	Discovery		Validation		Combined		
					HR (95% CI)[c]	*P*	HR (95% CI)[†]	*P*	HR (95% CI)	P_{meta}	P_{het}
Recurrence											
rs10932384, C	NA	NA	*ERBB4*	Dom	0.41 (0.26–0.64)	8.6×10^{-5}	0.60 (0.42–0.86)	5.33×10^{-3}	0.52 (0.39–0.68)	3.81×10^{-6}	0.18
rs984654, G	NA	NA	*EGFR*	Rec	2.52 (1.18–5.38)	0.017	1.50 (1.11–2.02)	0.009	1.61 (1.21–2.12)	8.92×10^{-4}	0.21
rs11702779, A	NA	NA	*RUNX1*	Dom	0.62 (0.40–0.98)	0.042	0.69 (0.48–0.97)	0.033	0.66 (0.50–0.87)	0.003	0.76
Overall survival											
rs10932384, C	NA	NA	*ERBB4*	Dom	0.51 (0.27–0.96)	0.036	0.49 (0.35–0.70)	6.07×10^{-5}	0.50 (0.37–0.67)	6.00×10^{-6}	0.92
rs2862644, C	NA	NA	*TBL1XR1*	Dom	2.52 (1.27–4.99)	0.008	2.24 (1.39–3.60)	0.001	2.33 (1.58–3.44)	2.20×10^{-5}	0.78
rs2770150, G	rs2737191, G	1	*TLR4*	Add	1.78 (1.12–2.83)	0.015	1.51 (1.14–2.01)	0.004	1.58 (1.24–2.02)	2.09×10^{-4}	0.56
rs6507587, A	NA	NA	*SETBP1*	Dom	1.97 (1.02–3.81)	0.043	1.61 (1.09–2.37)	0.016	1.70 (1.22–2.37)	0.002	0.61
rs17325821, T	NA	NA	*ERBB4*	Dom	2.15 (1.15–4.01)	0.016	1.45 (1.02–2.06)	0.04	1.59 (1.17–2.17)	0.007	0.28

[a]Tagging SNPs genotyped in the MDA discovery set.
[b]Proxy SNPs analyzed in the MDA validation set.
[c]Adjusted for age, sex, smoking status (Y/N), clinical stage, Fuhrman grade; additionally adjusted for antiangiogenesis treatment and chemotherapy where was appropriate.

Table 3. Associations between identified SNPs and mutation frequency of significantly mutated genes in ccRCC

SNP	Gene	Mutation%	With mutation *N* (%)	Without mutation *N* (%)	P^
rs10932384[a]	*ERBB4*	1.8			
AA			0 (0)	131 (100)	
AC			2 (1.48)	133 (98.52)	
CC			3 (11.54)	23 (88.46)	0.003

[a]rs10932384 was imputed with a info score of 0.984.
^Fisher's exact test *P* value.

(Table 4). However, using the data downloaded from TCGA, none of the SNPs were significantly associated with tumor tissue gene expression levels (data not shown). We also performed Enhancer and DNase enrichment analyses and the results indicated that the enrichments were highly significant for multiple loci in various cell lines (Supplementary Tables 3 and 4, available at *Carcinogenesis* Online).

Discussion

To the best of our knowledge, this is the first study to investigate the predisposing and prognostic value of germline genetic variants in 127 SMGs reported by TCGA Pan-Cancer effort. We identified two loci associated with RCC development, three loci associated with recurrence and five loci associated with OS. The locus on *ERBB4* at 2q34, represented by rs10932384, was significantly associated with both recurrence and OS in localized RCC patients (stages I–III). We also observed that the mutation frequency of *ERBB4* was significantly correlated with C allele of rs10932384.

TCGA Pan-Cancer analysis showed that ccRCC is distinguished from other common cancers by featuring higher percentage of somatic mutations in *VHL*, *PBRM1*, *SETD2* and *BAP1* (10–52% of ccRCC cases versus 0–7.9% of cases with other major cancers) (19). Previous studies have also related *PBRM1*, *SETD2* and *BAP1* somatic mutations to the prognosis of ccRCC (26,27). In the current study, we did not observe significant associations for common germline variants in these genes with RCC prognosis. Further analysis using the TCGA data showed no significant association of the *BAP1* SNPs with *BAP1* mutation status (Supplementary Table 5, available at *Carcinogenesis* Online). This discrepancy may reveal the fact that the somatic mutations in these genes but not the germline variants are potential prognostic factors for patients with RCC. In this study, we observed significant associations of SNPs in *PIK3CG* and *ATM* with RCC risk. A previous study has assessed a potential functional SNP (rs779805) in *VHL* which did not find a significant association with RCC risk (16). In our study and consistent with previous finding, rs779803, a proxy SNP to rs779805 ($r^2 = 1$), did not exhibit significant association with RCC risk. Several germline *BAP1* mutations were identified for RCC development (28,29). In addition, Wang *et al.* (30) showed the cooperative effect of VHL and BAP1 mutations in the development of RCC. Furthermore, several studies provided mechanistic insights into *BAP1* function on metabolic process and the ability to regulate gene-environment interactions (31–34). In our recent study, two *BAP1* SNPs, rs11708581 and rs390802, were significantly associated with risk of RCC and negatively correlated with *BAP1* gene expression (35). However, SNP rs11708581 failed the Illumina design and was not available, while SNP rs390802 showed consistent effect but not significant association which perhaps might be due to the smaller sample size in the current study. Therefore, additional studies are needed to further elucidate the role of common genetic variants in *BAP1* gene on RCC risk.

Limited studies have identified significant genetic predictors for RCC recurrence and OS. Lacking validation is the common limitation for most of the previous studies (16–18,36). With this two-phase study design including the discovery and validation data, we showed that rs10932384 in *ERBB4* at 2q34 was significantly and consistently associated with both recurrence and OS in stages I–III RCC patients. Furthermore, the effect sizes for recurrence and OS were similar (recurrence: $HR_{meta} = 0.52$; OS: $HR_{meta} = 0.50$). *ERBB4* plays an important role in regulating breast cancer growth and progression. However, previous

Table 4. cis-eQTL for identified SNPs based on SNiPA

SNP[a]	# of SNPs in LD[b]	Gene	cis-effect SNP[†]	R^2	Tissue	$P^{†}$
Risk						
rs6466135	4	*PIK3CG*	rs6466134	1.00	Blood	5.48×10^{-7}
rs611646	12	*ATM*	rs227069	0.94	Transformed fibroblasts	3.36×10^{-7}
Recurrence						
rs984654	15	*EGFR*	rs4947488	0.80	Blood	1.24×10^{-4}
Overall survival						
rs2770150	6	*TLR4*	rs5030728	0.88	Blood	8.15×10^{-35}

Cis-eQTL results obtained from SNiPA.
[a]Tagging SNPs genotyped in the discovery set.
[b]Number of SNPs located in the same LD block ($R^2 > 0.8$).
[†]SNP within the same LD block which has the strongest cis-effect on mapped gene.

studies have observed both oncogenic and tumor suppressive functions for *ERBB4* (37). Studies also showed that expression of *ERBB4* was downregulated in ccRCC compared with other subtypes of RCC, as well as normal tissues (38,39). In addition, reduced *ERBB4* expression was observed in immune cells among patients with relapsing remitting multiple sclerosis (40). Furthermore, *ERBB4* was also predicted to interact with many immunological networks at minimum relationship confidence of >0.75 for complement and coagulation cascades, DC Network, hematopoietic cell lineage, intestinal immune network for IgA production, RIG I like receptor signaling pathway, NOD-like receptor signaling pathway, leukocyte transendothelial migration and cytosolic DNA sensing pathway (http://immunet.princeton.edu) (41). Taken together, additional studies are needed to further elucidate the role of *ERBB4* in immune system together with cancer development and progression. We also found additional signals for either risk of recurrence or death but not both which were mapped to genes that may play crucial roles for cancer development and progression, such as *EGFR*. Therefore, additional studies are needed to further examine the role of these genes on RCC prognosis.

We further performed functional characterization of the identified variants. Interestingly, there is a significant correlation between the variant C allele of rs10932384 and *ERBB4* mutation frequency. The interplay between germline genetic variants and somatic mutations is intriguing and an intricate link between them may exist. One study demonstrated that germline background may alter the probability where a somatic event can occur and modify the likelihood of acquiring mutations in specific cancer genes (42). It has also been estimated that approximately 40% of mutations occurred in the cancer predisposition genes may be oncogenic (43). However, a recent study showed that the gene mutation frequencies in GWAS identified susceptibility regions were comparable to background mutation frequencies (44) which implied the minimal link between the GWAS identified cancer susceptibility regions and somatic mutations. Although somatic mutations in *ERBB4* are not a major feature for ccRCC (1.8% based on TCGA data), our results may provide new hypothesis regarding germline genetic predictors of RCC recurrence and prognosis and shed some light on the link between clinical outcomes associated germline variants and somatic mutations which warrants further investigations. In this study, we also found four loci significant in the *cis*-eQTL analysis using expression measured in the blood/transformed fibroblasts but not in the RCC tumor tissues. Therefore, additional studies are needed to understand the role of these SNPs on RCC prognosis.

Our study has several strengths. First, the current study applied a multi-phase design. For the RCC risk analysis, we further have utilized a publicly available GWAS dataset to externally validate our findings. Second, the patients enrolled in our study were prospectively followed for several years after treatment. We also recognize a few limitations. Since no clinical data is available in NCI dataset, the findings with respect to RCC recurrence and OS can only be validated internally. Also, the observed correlation between rs10932384:C and somatic mutation frequency of *ERBB4* requires further validation.

In summary, we identified potential novel genetic susceptibility loci for RCC by targeting significant mutated genes. Of note, many loci, particularly *ERBB4* at chromosome 2q34 may harbor genetic variants with independent prognostic value in localized RCC patients. Future independent studies are warranted to validate the identified genetic predictors of RCC recurrence and OS.

Supplementary material

Supplementary material can be found at *Carcinogenesis* online.

Funding

This work was supported in part by the National Institutes of Health (grant R01 CA170298), and the Center for Translational and Public Health Genomics, Duncan Family Institute for Cancer Prevention, The University of Texas MD Anderson Cancer Center.
Conflict of Interest Statement: None declared.

References

1. Siegel, R.L. *et al.* (2017) Cancer statistics, 2017. CA. Cancer J. Clin., 67, 7–30.
2. Chow, W.H. *et al.* (2010) Epidemiology and risk factors for kidney cancer. Nat. Rev. Urol., 7, 245–257.
3. Chin, A.I. *et al.* (2006) Surveillance strategies for renal cell carcinoma patients following nephrectomy. Rev. Urol., 8, 1–7.
4. Howlader, N. *et al.* SEER Cancer Statistics Review, 1975–2010, National Cancer Institute. Bethesda, MD. http://seer.cancer.gov/csr/1975_2010/.
5. Fergany, A.F. *et al.* (2000) Long-term results of nephron sparing surgery for localized renal cell carcinoma: 10-year followup. J. Urol., 163, 442–445.
6. Russo, P. *et al.* (2008) Survival rates after resection for localized kidney cancer: 1989 to 2004. Cancer, 113, 84–96.
7. Leibovich, B.C. *et al.* (2003) Prediction of progression after radical nephrectomy for patients with clear cell renal cell carcinoma: a stratification tool for prospective clinical trials. Cancer, 97, 1663–1671.
8. VanPoppel, H. *et al.* (1997) Microscopic vascular invasion is the most relevant prognosticator after radical nephrectomy for clinically nonmetastatic renal cell carcinoma. J. Urol., 158, 45–49.

9. Pavlovich, C.P. *et al.* (2004) Searching for the hereditary causes of renal-cell carcinoma. Nat. Rev. Cancer, 4, 381–393.
10. Popova, T. *et al.* (2013) Germline BAP1 mutations predispose to renal cell carcinomas. Am. J. Hum. Genet., 92, 974–980.
11. Wu, X. *et al.* (2012) A genome-wide association study identifies a novel susceptibility locus for renal cell carcinoma on 12p11.23. Hum. Mol. Genet., 21, 456–462.
12. Purdue, M.P. *et al.* (2011) Genome-wide association study of renal cell carcinoma identifies two susceptibility loci on 2p21 and 11q13.3. Nat. Genet., 43, 60–65.
13. Henrion, M. *et al.* (2013) Common variation at 2q22.3 (ZEB2) influences the risk of renal cancer. Hum. Mol. Genet., 22, 825–831.
14. Gudmundsson, J. *et al.* (2013) A common variant at 8q24.21 is associated with renal cell cancer. Nat. Commun., 4, 2776.
15. Scelo, G. *et al.* (2017) Genome-wide association study identifies multiple risk loci for renal cell carcinoma. Nat. Commun., 8, 15724.
16. Qin, C. *et al.* (2012) The polymorphisms in the VHL and HIF1A genes are associated with the prognosis but not the development of renal cell carcinoma. Ann. Oncol., 23, 981–989.
17. Kawai, Y. *et al.* (2007) Associations of single nucleotide polymorphisms in the vascular endothelial growth factor gene with the characteristics and prognosis of renal cell carcinomas. Eur. Urol., 52, 1147–1155.
18. Du, M. *et al.* (2014) Genetic variations in microRNAs and the risk and survival of renal cell cancer. Carcinogenesis, 35, 1629–1635.
19. Kandoth, C. *et al.* (2013) Mutational landscape and significance across 12 major cancer types. Nature, 502, 333–339.
20. Clague, J. *et al.* (2009) Family history and risk of renal cell carcinoma: results from a case-control study and systematic meta-analysis. Cancer Epidemiol. Biomarkers Prev., 18, 801–807.
21. Howie, B.N. *et al.* (2009) A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS Genet., 5, e1000529.
22. Arnold, M. *et al.* (2015) SNiPA: an interactive, genetic variant-centered annotation browser. Bioinformatics, 31, 1334–1336.
23. Li, Q. *et al.* (2013) Integrative eQTL-based analyses reveal the biology of breast cancer risk loci. Cell, 152, 633–641.
24. Ward, L.D. *et al.* (2012) HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. Nucleic Acids Res., 40(Database issue), D930–D934.
25. Purcell, S. *et al.* (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet., 81, 559–575.
26. Kapur, P. *et al.* (2013) Effects on survival of BAP1 and PBRM1 mutations in sporadic clear-cell renal-cell carcinoma: a retrospective analysis with independent validation. Lancet. Oncol., 14, 159–167.
27. Hakimi, A.A. *et al.*; ccRCC Cancer Genome Atlas (KIRC TCGA) Research Network investigators. (2013) Adverse outcomes in clear cell renal cell carcinoma with mutations of 3p21 epigenetic regulators BAP1 and SETD2: a report by MSKCC and the KIRC TCGA research network. Clin. Cancer Res., 19, 3259–3267.
28. Testa, J.R. *et al.* (2011) Germline BAP1 mutations predispose to malignant mesothelioma. Nat. Genet., 43, 1022–1025.
29. Peña-Llopis, S. *et al.* (2012) BAP1 loss defines a new class of renal cell carcinoma. Nat. Genet., 44, 751–759.
30. Wang, S.S. *et al.* (2014) Bap1 is essential for kidney function and cooperates with Vhl in renal tumorigenesis. Proc. Natl. Acad. Sci. USA, 111, 16538–16543.
31. Bononi, A. *et al.* (2017) Germline BAP1 mutations induce a Warburg effect. Cell Death Differ., 24, 1694–1704.
32. Bononi, A. *et al.* (2017) BAP1 regulates IP3R3-mediated Ca2+ flux to mitochondria suppressing cell transformation. Nature, 546, 549–553.
33. Amelio, I. (2017) Genes versus Environment: cytoplasmic BAP1 determines the toxic response to environmental stressors in mesothelioma. Cell Death Dis., 8, e2907.
34. Dai, F.Y. *et al.* (2017) BAP1 inhibits the ER stress gene regulatory network and modulates metabolic stress response. Proc. Natl. Acad. Sci. USA, 114, 3192–3197.
35. Lin, M. *et al.* (2017) Common, germline genetic variations in the novel tumor suppressor BAP1 and risk of developing different types of cancer. Oncotarget, 8, 74936–74946.
36. Tan, X. *et al.* (2013) Genetic variation in the GSTM3 promoter confer risk and prognosis of renal cell carcinoma by reducing gene expression. Br. J. Cancer, 109, 3105–3115.
37. Sundvall, M. *et al.* (2008) Role of ErbB4 in breast cancer. J. Mamm. Gland Biol. Neoplasia, 13, 259–268.
38. Behbahani, T.E. *et al.* (2012) Tyrosine kinase expression profile in clear cell renal cell carcinoma. World J. Urol., 30, 559–565.
39. Thomasson, M. *et al.* (2004) ErbB4 is downregulated in renal cell carcinoma–a quantitative RT-PCR and immunohistochemical analysis of the epidermal growth factor receptor family. Acta Oncol., 43, 453–459.
40. Tynyakov-Samra, E. *et al.* (2011) Reduced ErbB4 expression in immune cells of patients with relapsing remitting multiple sclerosis. Mult. Scler. Int., 2011, 561262.
41. Gorenshteyn, D. *et al.* (2015) Interactive big data resource to elucidate human immune pathways and diseases. Immunity, 43, 605–614.
42. Carter, H. *et al.* (2017) Interaction landscape of inherited polymorphisms with somatic events in cancer. Cancer Discov., 7, 410–423.
43. Rahman, N. (2014) Realizing the promise of cancer predisposition genes. Nature, 505, 302–308.
44. Machiela, M.J. *et al.* (2015) Limited evidence that cancer susceptibility regions are preferential targets for somatic mutation. Genome Biol., 16, 193.