



Whole-Exome Sequencing Identifies Somatic Mutations Associated With Mortality in Metastatic Clear Cell Kidney Carcinoma

Alejandro Mendoza-Alvarez¹, Beatriz Guillen-Guio¹, Adrian Baez-Ortega², Carolina Hernandez-Perez³, Sita Lakhwani-Lakhwani¹, Maria-del-Carmen Maeso⁴, Jose M. Lorenzo-Salazar⁵, Manuel Morales³ and Carlos Flores^{1,5,6,7*}

¹ Research Unit, Hospital Universitario Nuestra Señora de Candelaria, Universidad de La Laguna, Santa Cruz de Tenerife, Spain, ² Department of Veterinary Medicine, University of Cambridge, Cambridge, United Kingdom, ³ Service of Medical Oncology, Hospital Universitario Nuestra Señora de Candelaria, Santa Cruz de Tenerife, Spain, ⁴ Department of Pathology, Hospital Universitario Nuestra Señora de Candelaria, Santa Cruz de Tenerife, Spain, ⁵ Genomics Division, Instituto Tecnológico y de Energías Renovables, Santa Cruz de Tenerife, Spain, ⁶ Instituto de Tecnologías Biomédicas, Universidad de La Laguna, Santa Cruz de Tenerife, Spain, ⁷ Centro de Investigación Biomédica en Red de Enfermedades Respiratorias, Instituto de Salud Carlos III, Madrid, Spain

OPEN ACCESS

Edited by:

Heather Cunliffe,
University of Otago, New Zealand

Reviewed by:

Shicheng Guo,
Marshfield Clinic Research Institute,
United States
Cristin Gregor Print,
The University of Auckland,
New Zealand

*Correspondence:

Carlos Flores
cflores@ull.edu.es

Specialty section:

This article was submitted to
Cancer Genetics,
a section of the journal
Frontiers in Genetics

Received: 08 October 2018

Accepted: 29 April 2019

Published: 15 May 2019

Citation:

Mendoza-Alvarez A, Guillen-Guio B, Baez-Ortega A, Hernandez-Perez C, Lakhwani-Lakhwani S, Maeso M-d-C, Lorenzo-Salazar JM, Morales M and Flores C (2019) Whole-Exome Sequencing Identifies Somatic Mutations Associated With Mortality in Metastatic Clear Cell Kidney Carcinoma. *Front. Genet.* 10:439. doi: 10.3389/fgene.2019.00439

Clear cell renal cell carcinoma (ccRCC) is among the most aggressive histologic subtypes of kidney cancer, representing about 3% of all human cancers. Patients at stage IV have nearly 60% of mortality in 2–3 years after diagnosis. To date, most ccRCC studies have used DNA microarrays and targeted sequencing of a small set of well-established, commonly altered genes. An exception is the large multi-omics study of The Cancer Genome Atlas Kidney Renal Clear Cell Carcinoma (TCGA-KIRC), which identified new ccRCC genes based on whole exome-sequencing (WES) data, and molecular prognostic signatures based on transcriptomics, epigenetics and proteomics data. Applying WES to simultaneously interrogate virtually all exons in the human genome for somatic variation, here we analyzed the burden of coding somatic mutations in metastatic ccRCC primary tumors, and its association with patient mortality from cancer, in patients who received VEGF receptor-targeting drugs as the first-line therapy. To this end, we sequenced the exomes of ten tumor–normal pairs of ccRCC patient tissues from primary biopsies at >100× mean depth and called somatic coding variation. Mutation burden analysis prioritized 138 genes linked to patient mortality. A gene set enrichment analysis evidenced strong statistical support for the abundance of genes involved in the development of kidney cancer ($p = 2.31 \times 10^{-9}$) and carcinoma ($p = 1.22 \times 10^{-5}$), with 49 genes having direct links with kidney cancer according to the published records. Two of these genes, *SIPA1L2* and *EIF3A*, demonstrated independent associations with mortality in TCGA-KIRC project data. Besides, three mutational signatures were found to be operative in the tumor exomes, one of which (COSMIC signature 12) has not been previously reported in ccRCC. Selection analysis yielded no detectable evidence of overall positive or negative selection, with the exome-wide

number of nonsynonymous substitutions per synonymous site reflecting largely neutral tumor evolution. Despite the limited sample size, our results provide evidence for candidate genes where somatic mutation burden is tentatively associated with patient mortality in metastatic ccRCC, offering new potential pharmacological targets and a basis for further validation studies.

Keywords: ccRCC, whole-exome sequencing, kidney cancer, somatic mutation, mortality

INTRODUCTION

Clear cell renal cell carcinoma (ccRCC) represents only 2–3% of all human cancers (Manley et al., 2017). Notwithstanding, over 30% of ccRCC patients have metastases at the time of diagnosis, and 60% die in the first 2–3 years after diagnosis (Casuscelli et al., 2017). ccRCC is characterized by the resistance to radiation, cytotoxic and hormone therapies. Current treatments for ccRCC include diverse chemotherapeutic agents targeting the vascular endothelial growth factor (VEGF) pathway (Sternberg et al., 2010).

Roughly a decade ago, genetic approaches to disease diagnosis were postulated as a costly new way to progress toward the paradigm shift aimed by precision medicine. In ccRCC, the vast majority of studies have been directed at assessing genes that are known to be directly involved in pathogenesis, most of them using DNA arrays for genetic screening. The drawbacks and advantages of holistic vs. targeted gene studies have been extensively discussed in the literature (Iglesias et al., 2014; Kong et al., 2018). Nowadays, high-throughput next generation sequencing (NGS) technologies have made genetic testing affordable and cost-effective, hence consolidating as a central instrument for the progress toward the implementation of precision medicine. Furthermore, the reduction in per-base sequencing cost has popularized the use of whole-exome sequencing (WES) for the investigation of the pathogenic impact of genetic variation in coding regions (Damiati et al., 2016; Fay et al., 2016; Lata et al., 2018). This is reflected by the sheer number of WES studies being published, including a multitude of analyses of cancer exomes (Samarakoon et al., 2014; Lata et al., 2018).

To our knowledge, research in ccRCC using WES has previously focused on the treatment response or toxicity variables in relation to chemotherapeutic treatment. Moreover, kidney cancer studies were often limited to genes which are frequently altered in this condition, most commonly focusing on a gene panel conformed by *VHL*, *PBRM1*, *BAP1*, *SETD2*, *TP53*, *PTEN*, *KDM5C*, and *TERT* genes (Casuscelli et al., 2017; Manley et al., 2017; Tennenbaum et al., 2017). One notable exception to this is the study of more than 400 ccRCC patients with different omics approaches (The Cancer Genome Atlas Research Network, 2013). While this large study revealed more than 19 commonly mutated genes in ccRCC, molecular prognostic signatures were only pursued with transcriptomics, epigenetics, and proteomics data. Here, for the first time, we apply high-depth WES to assess the association between somatic mutation burden in metastatic ccRCC primary tumors, which refers to the total

number of somatic mutations identified per gene per patient, and patient survival.

MATERIALS AND METHODS

Patient Population and Setting

A total of 13 metastatic ccRCC patients (stage IV) from the two tertiary hospitals of Tenerife (Spain), Hospital Universitario Nuestra Señora de Candelaria (HUNSC) and Hospital Universitario de Canarias (HUC), were included in the study. The patients were all of European ancestry (self-declared), aged 31–80 years old (mean age of 56 years), with a male percentage of 61.5%. Seven (54%) of these patients died of cancer-related causes during the course of the study. The study was approved by the HUNSC Ethics Committee and written informed consent was obtained from all patients.

Nephrectomies were performed with curative intentions in six patients. For the rest of individuals, surgery was performed with cytoreductive purposes (Flanigan et al., 2001; Mickisch et al., 2001). Patients were classified into prognosis groups according to the Heng scoring system (Heng et al., 2013). At the moment of the diagnosis of metastasis, five patients showed good prognosis, while 6 had an intermediate prognosis and 2 a bad prognosis. All patients received tyrosine kinase inhibitors of the VEGF pathway, namely pazopanib (Sternberg et al., 2010) or sunitinib (Motzer et al., 2013), as the first-line treatment, except for one patient with bad prognosis who received tamsulosin (Hudes et al., 2007) as first-line treatment and pazopanib as second-line treatment.

Formalin-fixed paraffin-embedded (FFPE) biopsies from the primary tumors were obtained in blocks for subsequent DNA extraction. After evaluation by a pathologist, hematoxylin-eosin stained tissues were used to determine the limits of tumoral tissues. Whenever possible, nontumoral (thereafter referred to as normal) and tumoral tissues for DNA isolation were obtained from independent tissue slices. The GeneRead DNA FFPE Kit (QIAGEN, Hilden, Germany) was used for DNA isolation according to manufacturer's instructions. The integrity and concentration of DNA was evaluated with the Qubit® 3.0 Fluorometer, using the dsDNA BR Assay Kit (Thermo Fisher Scientific, Waltham, MA, United States), and the TaqMan™ RNase P Detection Reagents Kit (Thermo Fisher Scientific, Waltham, MA, United States).

Whole-Exome Sequencing

Enrichment, sequencing and read-mapping was accomplished by MacroGen Inc. Briefly, genomic DNA was enriched for

exome regions using the Ion AmpliSeq™ Exome RDY Kit (Thermo Fisher Scientific, Waltham, MA, United States) and Ion PI™ Chip Kit v3 (Thermo Fisher Scientific, Waltham, MA, United States). Exome-enriched DNA was sequenced on the Ion Proton™ platform (Thermo Fisher Scientific, Waltham, MA, United States), with two exomes per run to attain a theoretical depth of 100× per sample. Sequence data were aligned to the hg19/GRCh37 human reference genome using the Torrent Mapping Alignment Program v.5.0.13 included in the Torrent Suite Software for Sequencing Data Analysis v.5.0.4 (Thermo Fisher Scientific, Waltham, MA, United States).

Variant Calling and Annotation

Aligned sequence data were analyzed to identify somatic and germline single-nucleotide variants (SNVs) and small insertions and deletions (indels). We called genetic variation using a computational pipeline built on the variant caller Platypus v0.8.1 (Rimmer et al., 2014) (**Supplementary Figure S1**). As part of the pipeline and based on our previous experience, Platypus was run twice on each BAM file with two different settings: (i) default mode with additional options $-\text{minReads} = 3$ and $-\text{minPosterior} = 0$, (ii) default mode with options $-\text{minReads} = 3$, $-\text{minPosterior} = 0$, $-\text{minFlank} = 0$ and $-\text{trimReadFlank} = 10$. Variants (SNVs and indels) flagged with Platypus quality flags “badReads,” “MQ,” “strandBias,” “SC,” and “QD” were subsequently discarded, and the remaining variants were merged into a single file and genotyped across each sample. Variants that continued to be flagged with “badReads,” “MQ,” “strandBias,” “SC,” and “QD” after this genotyping were discarded. These procedures (which include more stringent filters for SNVs called near indels) and the low read support threshold allowed achieving very high sensitivity in the initial calling, while the stages of filtering, genotyping and re-filtering of the variant calls ensured high specificity. We then filtered germline variation and retained somatic variants for subsequent analyses. To that aim, we filtered out the variants that were present in any of the normal tissues, as well as the variants that were supported by less than 3 sequence reads. Remaining variants were considered putative somatic variation and annotated using the Ensembl Variant Effect Predictor (VEP) v91.0 (McLaren et al., 2016), particularly for gene elements, coding consequences and deleteriousness according to SIFT scoring. In order to introduce a germline variation reference for comparisons in annotations, we accessed and annotated all common (>5%) variation described in the Exome Aggregation Consortium (ExAC) database, which incorporates 60,706 sequencing data of unrelated individuals. Finally, we evaluated the overlap of the putative somatic dataset with the data deposited in COSMIC (v87, released 13-Nov-18)¹.

Mutation Burden and Selection Analysis

We analyzed the annotated somatic variants in each gene using bespoke analysis routines coded in the R programming language (R Development Core Team, 2008). To rank the associations between the mutation burden and patient mortality, a Fisher’s

exact test on the mutation count data was performed in R. A formal survival analysis adjusting for clinical and demographic confounders was not pursued because the minimal number of events that are needed per variable (Vittinghoff and McCulloch, 2007; Riley et al., 2018) was out of reach because of the limited sample size. However, with the exception of the treatment response, patients differing by mortality were not different for the major demographic and clinical variables (**Table 1**). Results were evaluated for inflation with a quantile-quantile (QQ) plot, using the qqplot v3.4.2 R package (Becker et al., 1988), and by estimating lambda with GenABEL v1.8-0 (Aulchenko et al., 2007). To assess evidence of positive or negative selection on somatic substitutions and detect any potential germline contamination in the somatic variant set, the dNdScv v0.0.0.9 R package (Martincorena et al., 2017) was employed to estimate exome-wide and per-gene number ratios of nonsynonymous substitutions per synonymous site (dN/dS).

Mutational Signature Analysis

The sigfit v1.1.0 R package (Gori and Baez-Ortega, 2018) was used to identify mutational processes (Alexandrov and Stratton, 2014), by fitting the mutational signatures published in the COSMIC catalog² to the mutational profiles of the somatic SNVs in each tumor. The latter were obtained by classifying SNVs into 96 categories according to substitution type (interpreting the pyrimidine base in the Watson-Crick pair as the reference base) and the bases immediately 5′ and 3′ to the mutated base in the reference genome (Alexandrov and Stratton, 2014). Fitting of mutational signatures to somatic variants was initially performed using all 30 COSMIC signatures; subsequently, those signatures displaying significant activity and biological coherence were fitted again to obtain more precise estimates of signature activities.

Gene Set Enrichment Analysis

The mutational landscape of ccRCC was explored through gene set enrichment analysis (GSEA), which was performed on those genes with $p < 0.05$ in the Fisher’s exact test of mutation burden (described above). This was performed via the Enrichr tool (Chen et al., 2013; Kuleshov et al., 2016) focusing on disease links through the Jensen Diseases database, which compiles evidence of gene-disease associations through the analysis of existing

²<https://cancer.sanger.ac.uk/cosmic/signatures>

TABLE 1 | Demographic and clinical features of the study sample.

	Controls (n = 6)	Cases (n = 4)	p-value*
Gender (% male)	67	75	1.00
Mean age (years)	56	58	0.61
Hypertension (%)	33	25	1.00
Smokers (%)	50	50	1.00
Drug toxicity (%)	33	50	1.00
Treatment response (% refractory)	17	100	0.05

*The age comparison was conducted by the Wilcoxon signed-rank test; the other variables were compared by a chi-square test. P25, percentile 25; P75, percentile 75.

¹<https://cancer.sanger.ac.uk/cosmic/>

literature on genetic studies. A sensitivity analysis was performed at this stage by subsampling 71,000 random subsets of variable size (30–100 genes) from the prioritized gene set that were then evaluated in EnrichR, recording the proportion of times the terms were significant in the adjusted tests.

Validation of Results

In order to validate our results, we accessed The Cancer Genome Atlas Kidney Renal Clear Cell Carcinoma (TCGA-KIRC) project data. To avoid potential confounders due to differences in the biological factors underlying cancer, patient exposures, and survival among people of different racial and ethnic backgrounds, we focused on the 417 samples with WES data available from white-ethnicity patients. For each prioritized gene in the mutational burden analysis, we recorded the number of somatic variants in deceased and alive patients in TCGA-KIRC. For that, we selected “kidney” in “Cases by Major Primary Site” classification available in the main page of TCGA portal. In filtering sets, we select “TCGA” program and TCGA-KIRC project. Results were subsequently filtered by “white” ethnicity. Each of the prioritized genes was manually entered to annotate the total number of patients with data for each gene and how many patients were deceased or alive. These data were subsequently analyzed via Fisher’s exact test as described above.

Availability of Analysis Scripts and Data

A file with the putative somatic variation across all patients and the scripts used for the analyses in this study can be found in the GitHub repository³.

RESULTS

DNA Extraction and Sequencing

We extracted and quantified genetic material from the original 13 patient FFPE samples for further evaluation via qPCR amplification with TaqMan probes of the housekeeping gene RNAsp. Three of the samples were discarded from the study due to insufficient amount of extracted DNA and high fragmentation levels, caused by the formalin fixation process. We subsequently sequenced 20 paired DNA samples, extracted from normal and tumor tissues from the remaining 10 patients. The average age of the sequenced individuals was 55 years (range 31–80 years), where 70% were male and 40% died during the course of the study. Amplicon size ranged between 157 and 182 base pairs (bp), with a mean insert length of 172 bp. The Ion AmpliSeqTM Exome RDY Kit yielded a median of 91.17% reads covering the on-target region with at least 20× depth. Sequencing metric summaries are shown in **Supplementary Table S1**.

Variant Calling and Annotation

A total of 122,019 SNVs and 31,646 indels were initially identified by the variant calling pipeline (**Figure 1**). The elevated number of indels was likely due to characteristic sequencing errors at

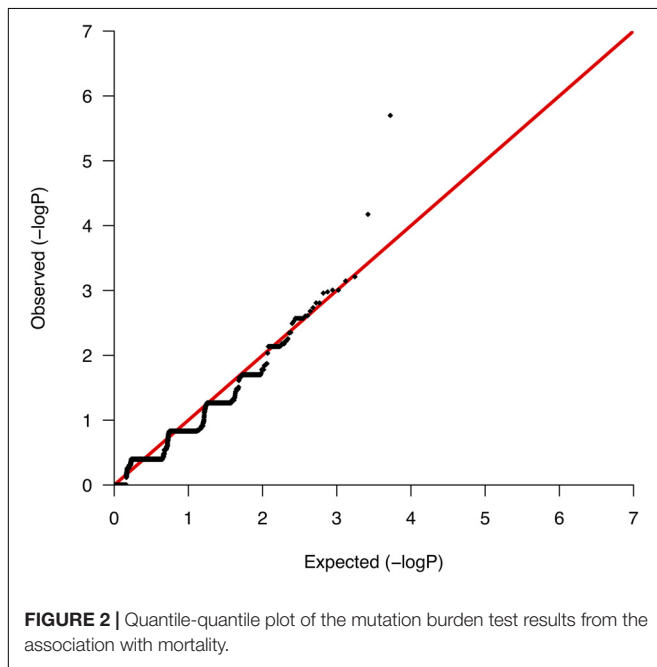
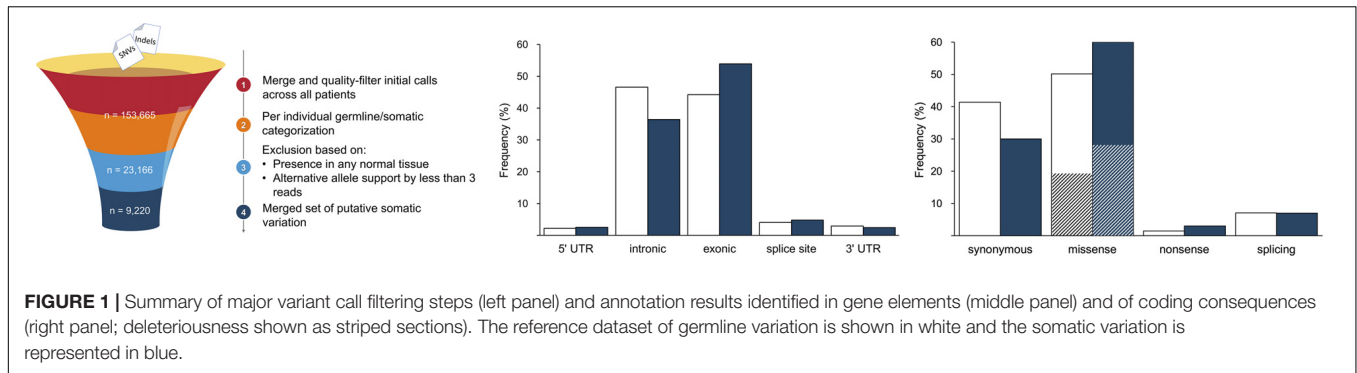
polynucleotide tracts, associated with the Ion Torrent sequencing chemistry (Fujita et al., 2017; Lata et al., 2018). The categorization of all these variants into germline and somatic sets per individual and a filtering of the flagged variants resulted in a total of 23,157 SNVs (18.98%) and 9 indels (0.28%) of somatic origin. A final filter based on their presence in normal tissues from other patients resulted in a refined set of 9,220 (40%) high-confidence somatic SNVs, which were considered for subsequent analyses; all indels were filtered out at this stage. This figure agrees with previous results (Miao et al., 2018), confirming that ccRCC is among the cancer types with lowest somatic mutation prevalence. To provide further support to the putative somatic calls, we aligned the 9,220 somatic variants to that of COSMIC. In agreement with other studies, we found that 1,012 (10.9%) variants were represented in COSMIC (Cai et al., 2016), supporting their credibility. Besides, to evidence other footprints of somatic mutations, we predicted their functional consequences. As expected, comparing the somatic variation to a suitable reference germline variation set obtained from ExAC (**Figure 1**), we found somatic variants to be more prevalent than germline variants in exonic regions (54.0% vs. 44.3% of germline). Furthermore, somatic variation involved a larger proportion of missense (60.0% vs. 50.0% of germline) and nonsense (3.0% vs. 1.0% of germline) substitutions. Among missense substitutions, the somatic set also displayed a larger proportion of predicted deleterious mutations, on the basis of SIFT score (54.0% vs. 40.0% of germline missense).

Finally, to evaluate the evidence for selection on somatic substitutions and identify any potential contamination from germline polymorphisms, the ratio of nonsynonymous substitutions per synonymous site (dN/dS) was measured for the set of somatic variants using a dN/dS model optimized for the analysis of selection in cancer (Martincorena et al., 2017). This algorithm accounts for variation in mutation rates, sequence context and trinucleotide mutability to reliably estimate dN/dS ratios, providing exome-wide and gene-wise estimates of dN/dS ratios and significant departures from neutrality (dN/dS = 1). Somatic variants identified in more than one tumor ($n = 464$) were excluded from the analysis in order to avoid spurious inflation of dN/dS estimates. The analysis yielded an exome-wide dN/dS ≈ 1, which is indicative of largely neutral evolution, in agreement with previous studies of selection in cancer (Martincorena et al., 2017). Similarly, none of the genes were found to display detectable evidence of selection on missense or truncating substitutions.

Gene-Based Mutation Burden and Mortality by ccRCC

We conducted comparative analyses between surviving and nonsurviving ccRCC patients, testing for differences per gene in the somatic mutation burden. We found a total of 5,267 genes with evidence of somatic variation among the 10 patients, where the most altered gene in terms of the number of mutations was *CDC27*, which harbored a total of 89 somatic variants. We then applied Fisher’s exact test to rank the differences in the somatic burden and prioritized 138 genes based on nominal significance

³<https://github.com/genomicsITER/mutational-burden-and-ccRCC-mortality>



(lowest nominal $p = 2.0 \times 10^{-6}$; **Supplementary Table S2**). A QQ-plot of the distribution of gene-based p -values nearly followed the null (**Figure 2**) suggesting a minimal lambda factor (1.071) and a minimal inflation of results. Interestingly, among the genes ranking higher, we found a number of genes expressed in kidney tissues and previously associated with a variety of human malignancies of neoplastic and nonneoplastic origin, such as *GPR155* (ranked 1st), *INPP5K* (ranked 3rd), and *KRT7* (ranked 4th) (**Supplementary Table S2**). Another notable result was the presence in the list of various mucin-encoding genes (*MUC5B*, *MUC12*, and *MUC16*), which are commonly observed mutated in many cancers and have been previously linked to colorectal, ovarian and hepatological cancers (Yin et al., 2013; Felder et al., 2014; Wang et al., 2018), as well as to severe fibrotic lung disorders (Seibold et al., 2011). In agreement with a previous targeted sequencing study (Tennenbaum et al., 2017), the mutation burden of *VHL*, which is the main hallmark of ccRCC, showed no differences between survivors and nonsurvivors, supporting its role only during early stages of tumorigenesis (Mandriota et al., 2002; Mitchell et al., 2018).

Gene Set Enrichment Analysis and Mutational Signatures

An enrichment analysis focused on the set of 138 prioritized genes based on the somatic mutation burden differences between survivors and nonsurvivors was performed to reveal disease links according to the Jensen Diseases database. Those genes most likely to be driving such relationships were prioritized. In agreement with a visual inspection of the prioritized gene list, this analysis showed a strong association between these genes and both kidney cancer development (adjusted $p = 2.32 \times 10^{-9}$) and carcinoma development (adjusted $p = 1.22 \times 10^{-5}$). A random subsampling of 71,000 gene sets of variable size out of the 138 prioritized genes also provided significant support to the enrichment of these two terms 90.9 and 74.5% of the times, respectively. This also evidenced that as few as 20% of randomly sampled genes from the prioritized gene set still detected a significant enrichment of kidney cancer 60% of the times, and that such statistical support stabilized when considering a random sample of 50% of the prioritized genes (**Supplementary Figure S2**). This sensitivity analysis reinforced that these observations are not significant by chance. A clustergram of the 49 genes that were directly associated with kidney cancer development is shown in **Figure 3**.

An analysis of mutational signatures revealed three signatures (COSMIC signatures 1, 5, and 12) with significant activity in the tumors (**Figure 4**). Signatures 1 and 5 correspond to endogenous mutational processes that are consistently operative in nearly all human cells (Alexandrov et al., 2015). On the other hand, signature 12, whose etiology is presently unknown, has been previously described only in liver cancer, and thus its presence in ccRCC tumors is unprecedented (Alexandrov et al., 2013). Strikingly, although with borderline significance, the mutational contribution of signature 12 in the tumors tends to associate with the age at diagnosis ($\rho = 0.71$, $p = 0.02$). Notwithstanding this result, the overall somatic mutation burden was not correlated with the age at diagnosis ($\rho = 0.32$, $p = 0.36$).

Validation of Results

In order to provide validation to our findings, we then retrieved the mutation count for all of the prioritized genes from the TCGA-KIRC dataset, stratified by survival, and again tested for

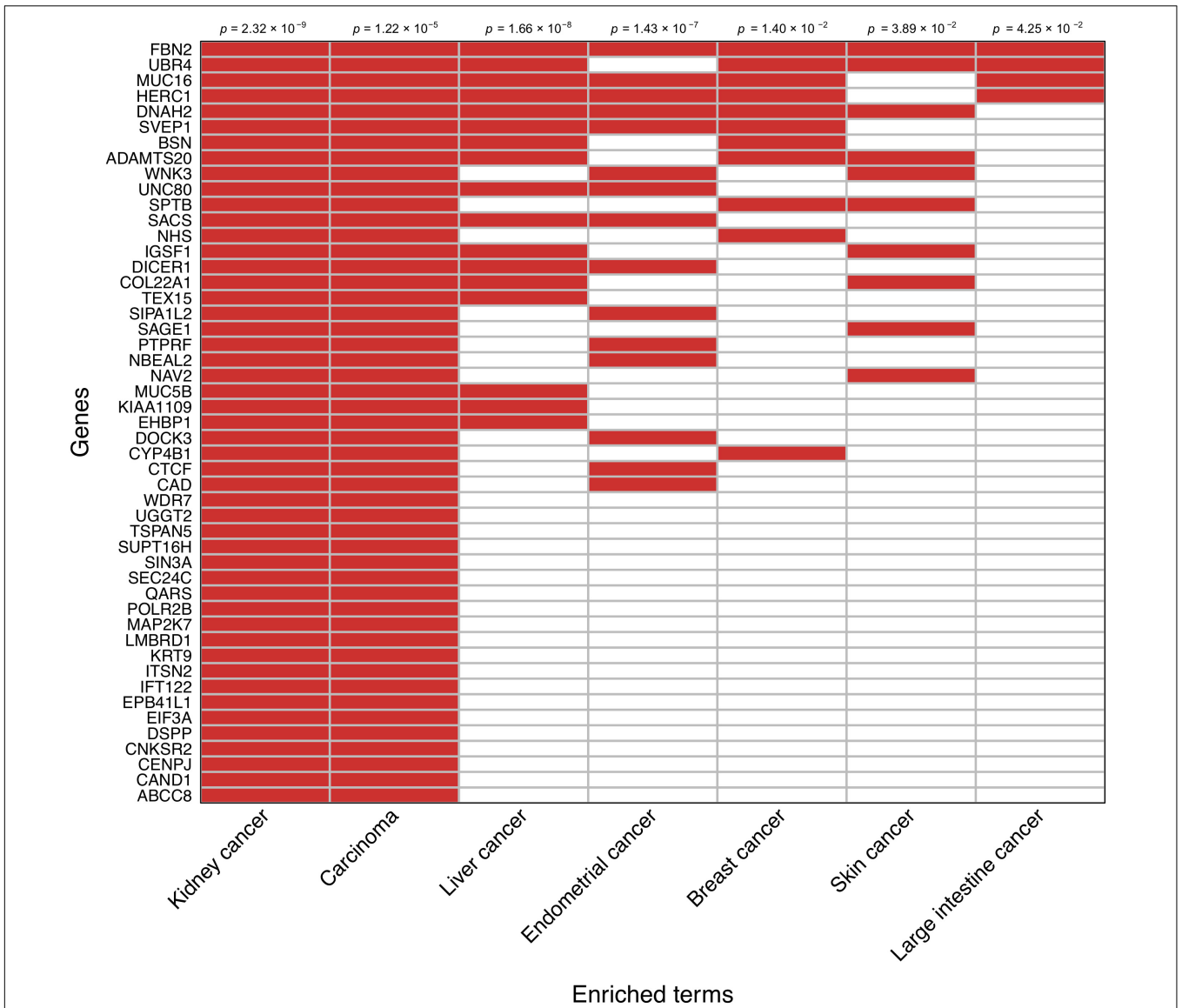


FIGURE 3 | Clustergram representing the association of the subset of 49 prioritized genes that have direct links with kidney cancer. Relationships with other cancer types are also shown. Significance values are shown on the top.

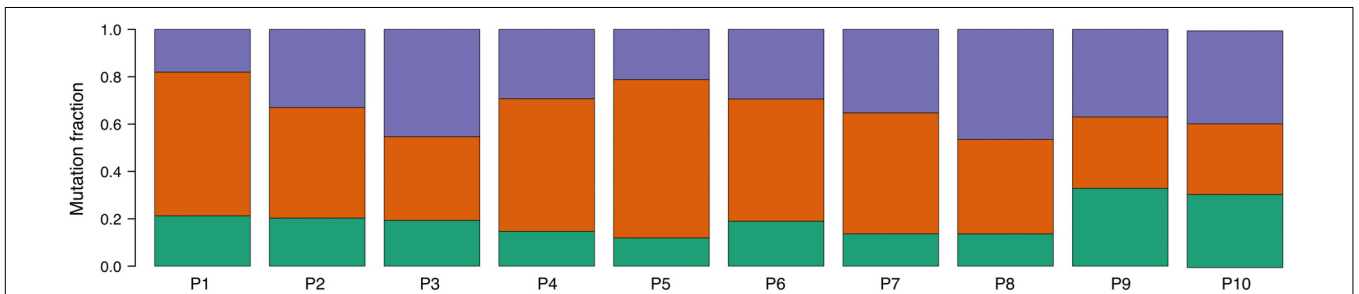


FIGURE 4 | Proportion of COSMIC signatures displaying significant activity in each patient tumor. Color code correspondence is: green, signature 1; red, signature 5; and purple, signature 12.

differences per gene in the somatic mutation burden. Only two out of the 138 prioritized genes were nominally significant in the TCGA-KIRC dataset, namely *SIPA1L2* ($p = 0.036$), which encodes the signal-induced proliferation-associated 1 like 2, and *EIF3A* ($p = 0.048$), which encodes the eukaryotic translation initiation factor 3 subunit A. Both of these genes have been previously linked to unfavorable prognosis in other cancers (Minato and Hattori, 2009; Yin et al., 2018).

DISCUSSION

Despite previous attempts to reveal molecular prognostic signatures of ccRCC based on multiomics data (The Cancer Genome Atlas Research Network, 2013), this study constitutes the first exome-wide approach for revealing genes with differential accumulation of somatic mutations in relation to cancer-associated mortality in ccRCC patients. Previous studies assessing links with prognosis or responses have either used targeted approaches directed to a limited set of genes that commonly accumulate mutations (Tennenbaum et al., 2017) or to evidence associations with treatment responses to the therapies (Fay et al., 2016; Miao et al., 2018). At most, these studies revealed that recurrent mutations in *PBRM1*, one of the well-known ccRCC genes, might have implications in the treatment responses. In contrast, our study enabled prioritizing 138 genes based on a refined set of putative somatic SNVs, where 49 of those genes had previously been related to kidney cancer according to the literature, and two genes (*SIPA1L2* and *EIF3A*) were validated in independent datasets from TCGA. Our study has also yielded evidence suggestive of the activity of COSMIC signature 12 in kidney cancer, in addition to two well-established endogenous mutational processes.

SIPA1L2 is a renal cancer biomarker according to The Human Protein Atlas⁴, as its expression associates with a significant unfavorable prognosis. Members of this family encode mitogen activators of GTPase activity for Ras-related Rap regulatory proteins, contributing as essential regulators in cell cycle and of metastasis in various types of solid cancers. Overexpression of its mouse ortholog *in vitro* alters primary mammary tumor cell morphology and adhesive properties (Zhang Y. et al., 2015), inducing detachment of cells from the matrix, while it increases the metastatic capacity *in vivo* (Park et al., 2005). Consistently, primary tumors with metastasis express more protein than nonmetastatic tumors for a variety of solid tumors (Park et al., 2005; Minato and Hattori, 2009). Therefore, its role in ccRCC prognosis may well be linked to this influential activity in cancer invasiveness. With respect to *EIF3A*, it encodes the largest subunit of eIF3 complex involved in translational regulation, cell growth and cancer. The eIF3a subunit itself is suggested to regulate a subset of mRNAs including factors required for development and differentiation. Its expression is ubiquitous, although at higher levels in adult proliferating tissues and in a number of cancer types, suggesting that it is

required for cell proliferation and that is a negative regulator of cell differentiation (Saletta et al., 2010). In fact, in some cancer types, there is a correlation between higher eIF3a expression and the metastatic ability (Bachmann et al., 1997). The expression of some of the eIF3 subunits, including eIF3a, have been associated with cancer prognosis and therapeutic response (Saramäki et al., 2001; Buttitta et al., 2005; Yin et al., 2018). Interestingly, eIF3a was also involved in fibrosis in humans, particularly in renal fibrotic tissue, by regulating the TGF- β 1/SMAD3 signaling pathway (Zhang Y.-F. et al., 2015). Based on this evidence, eIF3a may be associated with ccRCC prognosis by its central role in the maintenance of the malignant status of cells.

Despite we were unable to validate their association with ccRCC mortality, there were other prioritized genes among the top ranked that are interesting candidates for further study, namely *GPR155*, *INPP5K*, *KRT7*, *CYP4B1* and the mucin encoding genes (*MUC5B*, *MUC12*, and *MUC16*). Most of these have been associated with tissue remodeling and cancer processes (Imaoka et al., 2000; Hedberg Oldfors et al., 2015; Giunchi et al., 2016; Umeda et al., 2017; Renshaw and Gould, 2018; Sarlos et al., 2018). Interestingly, albeit previous studies have argued that some of the MUCs should be excluded as cancer genes based on biological relevance (Lawrence et al., 2013), there is much evidence supporting the link between *MUC12* and *MUC16* and cancer development and evolution (Matsuyama et al., 2010; Yin et al., 2013; Felder et al., 2014; Wang et al., 2018). The results for *MUC5B* are less clear, since there is not a direct evidence in the literature of a link with oncogenic processes, but only with susceptibility and survival in pulmonary fibrosis through germline regulatory variants (Seibold et al., 2011; Fingerlin et al., 2013; Noth et al., 2013; Peljto et al., 2013; Allen et al., 2017). Curiously, a recent WES study by Lata et al. (2018), aimed at providing diagnosis of adult probands with CKD of unknown cause, identified causal germline mutations in *PARN* [poly(A)-specific ribonuclease], another pulmonary fibrosis susceptibility gene (Stuart et al., 2015). In agreement with these results, some studies have argued in support of pathogenic similarities between pulmonary fibrosis and cancer (Vancheri, 2015). Under such scenario, it could be speculated that coding mutations in *MUC5B* and *PARN* may play a role in oncogenesis in lung and kidney tissues.

One of the most notable strengths of this study is that it focuses on a homogeneous patient population, with all patients being included at stage IV and, similarly treated. Besides, the combination of high-depth WES of matched tumor-normal sample pairs from each patient, and the multiple filtering routines performed after variant calling, enabled the derivation of a high-confidence set of somatic variants. This led to findings which, for two of the resulting prioritized genes, were validated in an independent dataset from a much larger ccRCC patient population. The limited number of genes that validate in TCGA-KIRC can be explained by the limited sample size analyzed in this study, and by the systematic analytic differences between the two studies. Notwithstanding, we also acknowledge some salient limitations in the study. First, the study evaluated a very small patient sample and focused on the high-mortality risk

⁴<https://www.proteinatlas.org/ENSG00000116991-SIPA1L2/pathology>

spectrum of ccRCC cases that may not be representative of the full disease spectrum. Despite the similarity in demographic and clinical data of patients irrespective of the mortality, the limited sample size precluded a formal survival analysis accounting for time-dependent risks while adjusting for potential confounders. Therefore, the results should be taken with caution, as there may be alternative patient variables that can explain the associations found. Second, we only sequenced a single specimen at the point of patient diagnosis and did not analyze WES of metastasis in remote tissues. As a consequence, our capacity to identify candidate genes linked to ccRCC survival was limited to those at the pre-treatment stage, hindering the possibility of identifying additional genes as the tumors responded to therapy or the effect of mutations in metastases. Third, because of the capture design of WES, we were unable to assess the association of noncoding variants with patient mortality, as has been previously suggested for regulatory variants in the telomerase reverse transcriptase encoding gene *TERT* (Casuscelli et al., 2017). Fourth, structural variants are common in ccRCC (Thiesen et al., 2017) and have been associated with poorer prognosis (Chen et al., 2009; Moore et al., 2012). However, we did not explore the implications of these on patient mortality because we anticipated that they would be highly uncertain in our data. The reasons for this are: (i) the algorithms used for calling SNVs and small indels are not well suited for the detection structural variation, which require dedicated algorithms; (ii) WES has limited sensibility for the detection of structural variants; (iii) there is an extensive lack of agreement between the structural variant callers, implying that their calls necessitate a consensus of multiple algorithms; and (iv) we lacked WES data from a pool of references to serve as controls for the AmpliSeq technology to be used for calling structural variants. Finally, we did not adjust for multiple hypothesis testing because our analyses were exploratory in nature: our interest lay in maximizing the utility of the available data for this small patient sample, to identify a sufficient number of prioritized genes as to power our gene set enrichment analysis. Hence, we tolerated the existence of false positives as a reasonable trade-off for an enhanced gene prioritization, which was necessary for the enrichment analysis. As such, our results should be regarded as hypothesis-generating findings.

CONCLUSION

In this study, we identified and validated two genes that are recurrently altered in ccRCC tumors and that associate with patient mortality. These have been previously suggested as biomarkers of cancer prognosis and participate in molecular pathways linked to tumor development and progression. Additionally, we provide evidence suggestive of the activity of COSMIC mutational signature 12 in kidney cancer, hinting at a potentially incomplete understanding of the mutational processes that are involved in this kind of malignancy. Independent validation studies achieving larger statistical power are needed to better evaluate the impact on ccRCC patient mortality.

ETHICS STATEMENT

This study was carried out in accordance with the recommendations of the Ethics Committee for Clinical Research from the Hospital Universitario Nuestra Señora de Candelaria in accordance with the Declaration of Helsinki. The protocol was approved by the Ethics Committee for Clinical Research from the Hospital Universitario Nuestra Señora de Candelaria.

AUTHOR CONTRIBUTIONS

AB-O, AM-A, BG-G, and CF wrote the manuscript and the **Supplementary Files**. BG-G and SL-L performed DNA extractions and quantifications, as well as the preparation of samples for sequencing. CH-P, M-d-CM, and MM performed data collection and the record of patient and sample features. AB-O and JL-S conceived and implemented software procedures. AM-A, CF, and JL-S performed the statistical tests on detected somatic variants. CF provided a general supervision of the project, giving guidelines for each step. All the authors provided insights, corrections and approved the final version of the manuscript.

FUNDING

This study was funded by Fundación CajaCanarias (SALUCAN11) and Area Tenerife 2030 from Cabildo de Tenerife (CGIEU0000219140). AM-A was supported by a fellowship from CajaSiete-ULL. BG-G was supported by a fellowship from the Canarian Agency for Research, Innovation and Information Society (ACIISI, Grant No. TESIS2015010057) co-funded by the European Social Fund (ESF).

ACKNOWLEDGMENTS

We would like to thank Joaquín Martínez-Muñoz for the helpful technical support at the early stages of this project.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2019.00439/full#supplementary-material>

FIGURE S1 | Overview of the variant calling pipeline. Aligned BAM files are input to generate a final set of genomic variants for subsequent analysis.

FIGURE S2 | Proportion of significant results in the gene set enrichment analysis for “kidney cancer” and “carcinoma” as a result of the random sampling of gene sets from the prioritized gene set.

TABLE S1 | Whole-exome sequencing summary metrics for each sample. Average for both normal and tumor samples is shown (minimum and maximum values in parenthesis).

TABLE S2 | Ranking of prioritized genes sorted by significance level.

REFERENCES

- Alexandrov, L. B., Jones, P. H., Wedge, D. C., Sale, J. E., Campbell, P. J., Nik-Zainal, S., et al. (2015). Clock-like mutational processes in human somatic cells. *Nat. Genet.* 47, 1402–1407. doi: 10.1038/ng.3441
- Alexandrov, L. B., Nik-Zainal, S., Wedge, D. C., Aparicio, S. A., Behjati, S., Biankin, A. V., et al. (2013). Signatures of mutational processes in human cancer. *Nature* 500, 415–421. doi: 10.1038/nature12477
- Alexandrov, L. B., and Stratton, M. R. (2014). Mutational signatures: the patterns of somatic mutations hidden in cancer genomes. *Curr. Opin. Genet. Dev.* 24, 52–60. doi: 10.1016/j.gde.2013.11.014
- Allen, R. J., Porte, J., Braybrooke, R., Flores, C., Fingerlin, T. E., Oldham, J. M., et al. (2017). Genetic variants associated with susceptibility to idiopathic pulmonary fibrosis in people of European ancestry: a genome-wide association study. *Lancet Respir. Med.* 5, 869–880. doi: 10.1016/S2213-2600(17)30387-9
- Aulchenko, Y. S., Ripke, S., Isaacs, A., and van Duijn, C. M. (2007). GenABEL: an R library for genome-wide association analysis. *Bioinformatics* 23, 1294–1296. doi: 10.1093/bioinformatics/btm108
- Bachmann, F., Bänziger, R., and Burger, M. M. (1997). Cloning of a novel protein overexpressed in human mammary carcinoma. *Cancer Res.* 57, 988–994.
- Becker, R. A., Chambers, J. M., and Wilks, A. R. (1988). *The New S Language. A Programming Environment for Data Analysis and Graphics*. Pacific Grove, CA: Wadsworth & Brooks.
- Buttitta, F., Martella, C., Barassi, F., Felicioni, L., Salvatore, S., Rosini, S., et al. (2005). Int6 expression can predict survival in early-stage non-small cell lung cancer patients. *Clin. Cancer Res.* 11, 3198–3204. doi: 10.1158/1078-0432.CCR-04-2308
- Cai, L., Yuan, W., Zhang, Z., He, L., and Chou, K.-C. (2016). In-depth comparison of somatic point mutation callers based on different tumor next-generation sequencing depth data. *Sci. Rep.* 6:36540. doi: 10.1038/srep36540
- Casuscelli, J., Becerra, M. F., Manley, B. J., Zabor, E. C., Reznik, E., Redzematovic, A., et al. (2017). Characterization and impact of tert promoter region mutations on clinical outcome in renal cell carcinoma. *Eur. Urol. Focus* doi: 10.1016/j.euf.2017.09.008 [Epub ahead of print].
- Chen, E. Y., Tan, C. M., Kou, Y., Duan, Q., Wang, Z., Meirelles, G. V., et al. (2013). Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics* 14:128. doi: 10.1186/1471-2105-14-128
- Chen, M., Ye, Y., Yang, H., Tamboli, P., Matin, S., Tannir, N. M., et al. (2009). Genome-wide profiling of chromosomal alterations in renal cell carcinoma using high-density single nucleotide polymorphism arrays. *Int. J. Cancer* 125, 2342–2348. doi: 10.1002/ijc.24642
- Damiati, E., Borsani, G., and Giacomuzzi, E. (2016). Amplicon-based semiconductor sequencing of human exomes: performance evaluation and optimization strategies. *Hum. Genet.* 135, 499–511. doi: 10.1007/s00439-016-1656-8 doi: 10.1007/s00439-016-1656-8
- Fay, A. P., de Velasco, G., Ho, T. H., Van Allen, E. M., Murray, B., Albiges, L., et al. (2016). Whole-exome sequencing in two extreme phenotypes of response to VEGF-targeted therapies in patients with metastatic clear cell renal cell carcinoma. *J. Natl. Compr. Canc. Netw.* 14, 820–824. doi: 10.6004/jnccn.2016.0086
- Felder, M., Kapur, A., Gonzalez-Bosquet, J., Horibata, S., Heintz, J., Albrecht, R., et al. (2014). MUC16 (CA125): tumor biomarker to cancer therapy, a work in progress. *Mol. Cancer* 13:129. doi: 10.1186/1476-4598-13-129
- Fingerlin, T. E., Murphy, E., Zhang, W., Peljto, A. L., Brown, K. K., Steele, M. P., et al. (2013). Genome-wide association study identifies multiple susceptibility loci for pulmonary fibrosis. *Nat. Genet.* 45, 613–620. doi: 10.1038/ng.2609
- Flanigan, R. C., Salmon, S. E., Blumenstein, B. A., Bearman, S. I., Roy, V., McGrath, P. C., et al. (2001). Nephrectomy followed by interferon alfa-2b compared with interferon alfa-2b alone for metastatic renal-cell cancer. *N. Engl. J. Med.* 345, 1655–1659. doi: 10.1056/NEJMoa003013
- Fujita, S., Masago, K., Okuda, C., Hata, A., Kaji, R., Katakami, N., et al. (2017). Single nucleotide variant sequencing errors in whole exome sequencing using the ion proton system. *Biomed. Rep.* 7, 17–20. doi: 10.3892/br.2017.911
- Giunchi, F., Fiorentino, M., Vagnoni, V., Capizzi, E., Bertolo, R., Porpiglia, F., et al. (2016). Renal oncocytosis: a clinicopathological and cytogenetic study of 42 tumours occurring in 11 patients. *Pathology* 48, 41–46. doi: 10.1016/j.pathol.2015.11.009
- Gori, K., and Baez-Ortega, A. (2018). sigfit: flexible bayesian inference of mutational signatures. *bioRxiv* [Preprint]. doi: 10.1101/372896
- Hedberg Oldfors, C., Dios, D. G., Linder, A., Visuttijai, K., Samuelson, E., Karlsson, S., et al. (2015). Analysis of an independent tumor suppressor locus telomeric to Tp53 suggested Inpp5k and Myo1c as novel tumor suppressor gene candidates in this region. *BMC Genet.* 16:80. doi: 10.1186/s12863-015-0238-4
- Heng, D. Y. C., Xie, W., Regan, M. M., Harshman, L. C., Bjarnason, G. A., Vaishampayan, U. N., et al. (2013). External validation and comparison with other models of the international metastatic renal-cell carcinoma database consortium prognostic model: a population-based study. *Lancet Oncol.* 14, 141–148. doi: 10.1016/S1470-2045(12)70559-4
- Hudes, G., Carducci, M., Tomczak, P., Dutcher, J., Figlin, R., Kapoor, A., et al. (2007). Temsirolimus, interferon alfa, or both for advanced renal-cell carcinoma. *N. Engl. J. Med.* 356, 2271–2281. doi: 10.1056/NEJMoa066838
- Iglesias, A., Anyane-Yeboah, K., Wynn, J., Wilson, A., Truitt Cho, M., Guzman, E., et al. (2014). The usefulness of whole-exome sequencing in routine clinical practice. *Genet. Med.* 16, 922–931. doi: 10.1038/gim.2014.58
- Imaoka, S., Yoneda, Y., Sugimoto, T., Hiroi, T., Yamamoto, K., Nakatani, T., et al. (2000). CYP4B1 is a possible risk factor for bladder cancer in humans. *Biochem. Biophys. Res. Commun.* 277, 776–780. doi: 10.1006/bbrc.2000.3740
- Kong, S. W., Lee, I.-H., Liu, X., Hirschhorn, J. N., and Mandl, K. D. (2018). Measuring coverage and accuracy of whole-exome sequencing in clinical context. *Genet. Med.* 20, 1617–1626. doi: 10.1038/gim.2018.51
- Kuleshov, M. V., Jones, M. R., Rouillard, A. D., Fernandez, N. F., Duan, Q., Wang, Z., et al. (2016). Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* 44, W90–W97. doi: 10.1093/nar/gkw377
- Lata, S., Marasa, M., Li, Y., Fasel, D. A., Groopman, E., Jobanputra, V., et al. (2018). Whole-exome sequencing in adults with chronic kidney disease: a pilot study. *Ann. Intern. Med.* 168, 100–109. doi: 10.7326/M17-1319
- Lawrence, M. S., Stojanov, P., Polak, P., Kryukov, G. V., Cibulskis, K., Sivachenko, A., et al. (2013). Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* 499, 214–218. doi: 10.1038/nature12213
- Mandriota, S. J., Turner, K. J., Davies, D. R., Murray, P. G., Morgan, N. V., Sowter, H. M., et al. (2002). HIF activation identifies early lesions in VHL kidneys: evidence for site-specific tumor suppressor function in the nephron. *Cancer Cell* 1, 459–468. doi: 10.1016/S1535-6108(02)00071-5
- Manley, B. J., Zabor, E. C., Casuscelli, J., Tennenbaum, D. M., Redzematovic, A., Becerra, M. F., et al. (2017). Integration of recurrent somatic mutations with clinical outcomes: a pooled analysis of 1049 patients with clear cell renal cell carcinoma. *Eur. Urol. Focus* 3, 421–427. doi: 10.1016/j.euf.2016.06.015
- Martincorena, I., Raine, K. M., Gerstung, M., Dawson, K. J., Haase, K., Van Loo, P., et al. (2017). Universal patterns of selection in cancer and somatic tissues. *Cell* 171, 1029–1041.e21. doi: 10.1016/j.cell.2017.09.042
- Matsuyama, T., Ishikawa, T., Mogushi, K., Yoshida, T., Iida, S., Uetake, H., et al. (2010). MUC12 mRNA expression is an independent marker of prognosis in stage II and stage III colorectal cancer. *Int. J. Cancer* 127, 2292–2299. doi: 10.1002/ijc.25256
- McLaren, W., Gil, L., Hunt, S. E., Riat, H. S., Ritchie, G. R. S., Thormann, A., et al. (2016). The ensembl variant effect predictor. *Genome Biol.* 17:122. doi: 10.1186/s13059-016-0974-4
- Miao, D., Margolis, C. A., Gao, W., Voss, M. H., Li, W., Martini, D. J., et al. (2018). Genomic correlates of response to immune checkpoint therapies in clear cell renal cell carcinoma. *Science* 359, 801–806. doi: 10.1126/science.aan5951
- Mickisch, G. H., Garin, A., van Poppel, H., de Prijck, L., Sylvester, R., and European Organisation for Research and Treatment of Cancer [EORTC] Genitourinary Group. (2001). Radical nephrectomy plus interferon-alfa-based immunotherapy compared with interferon alfa alone in metastatic renal-cell carcinoma: a randomised trial. *Lancet* 358, 966–970. doi: 10.1016/S0140-6736(01)06103-7
- Minato, N., and Hattori, M. (2009). Spa-1 (Sipa1) and Rap signaling in leukemia and cancer metastasis. *Cancer Sci.* 100, 17–23. doi: 10.1111/j.1349-7006.2008.01011.x
- Mitchell, T. J., Turajlic, S., Rowan, A., Nicol, D., Farmery, J. H. R., O'Brien, T., et al. (2018). Timing the landmark events in the evolution of clear cell renal cell cancer: TRACERx renal. *Cell* 173, 611–623.e17. doi: 10.1016/j.cell.2018.02.020

- Moore, L. E., Jaeger, E., Nickerson, M. L., Brennan, P., De Vries, S., Roy, R., et al. (2012). Genomic copy number alterations in clear cell renal carcinoma: associations with case characteristics and mechanisms of VHL gene inactivation. *Oncogenesis* 1:e14. doi: 10.1038/oncis.2012.14
- Motzer, R. J., Hutson, T. E., Cella, D., Reeves, J., Hawkins, R., Guo, J., et al. (2013). Pazopanib versus sunitinib in metastatic renal-cell carcinoma. *N. Engl. J. Med.* 369, 722–731. doi: 10.1056/NEJMoa1303989
- Noth, I., Zhang, Y., Ma, S.-F., Flores, C., Barber, M., Huang, Y., et al. (2013). Genetic variants associated with idiopathic pulmonary fibrosis susceptibility and mortality: a genome-wide association study. *Lancet Respir. Med.* 1, 309–317. doi: 10.1016/S2213-2600(13)70045-6
- Park, Y.-G., Zhao, X., Lesueur, F., Lowy, D. R., Lancaster, M., Pharoah, P., et al. (2005). Sip1 is a candidate for underlying the metastasis efficiency modifier locus Mtes1. *Nat. Genet.* 37, 1055–1062. doi: 10.1038/ng1635
- Peljto, A. L., Zhang, Y., Fingerlin, T. E., Ma, S.-F., Garcia, J. G. N., Richards, T. J., et al. (2013). Association between the MUC5B promoter polymorphism and survival in patients with idiopathic pulmonary fibrosis. *JAMA* 309, 2232–2239. doi: 10.1001/jama.2013.5827
- R Development Core Team (2008). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.
- Renshaw, A. A., and Gould, E. W. (2018). Ancillary studies in fine needle aspiration of the kidney. *Cancer Cytopathol.* 126(Suppl. 8), 711–723. doi: 10.1002/cncy.22029
- Riley, R. D., Snell, K. I. E., Ensor, J., Burke, D. L., Harrell, F. E. Jr., Moons, K. G. M., et al. (2018). Minimum sample size for developing a multivariable prediction model: Part I - continuous outcomes. *Stat. Med.* 38, 1262–1275. doi: 10.1002/sim.7993
- Rimmer, A., Phan, H., Mathieson, I., Iqbal, Z., Twigg, S. R. F., WGS500 Consortium, et al. (2014). Integrating mapping-, assembly- and haplotype-based approaches for calling variants in clinical sequencing applications. *Nat. Genet.* 46, 912–918. doi: 10.1038/ng.3036
- Saletta, F., Suryo Rahmanto, Y., and Richardson, D. R. (2010). The translational regulator eIF3a: the tricky eIF3 subunit! *Biochim. Biophys. Acta* 1806, 275–286. doi: 10.1016/j.bbcan.2010.07.005
- Samarakoon, P. S., Sorte, H. S., Kristiansen, B. E., Skodje, T., Sheng, Y., Tjønnfjord, G. E., et al. (2014). Identification of copy number variants from exome sequence data. *BMC Genomics* 15:661. doi: 10.1186/1471-2164-15-661
- Saramäki, O., Willi, N., Bratt, O., Gasser, T. C., Koivisto, P., Nupponen, N. N., et al. (2001). Amplification of EIF3S3 gene is associated with advanced stage in prostate cancer. *Am. J. Pathol.* 159, 2089–2094. doi: 10.1016/S0002-9440(10)63060-X
- Sarlos, D. P., Peterfi, L., Szanto, A., and Kovacs, G. (2018). Shift of keratin expression profile in end-stage kidney increases the risk of tumor development. *Anticancer. Res.* 38, 5217–5222. doi: 10.21873/anticancer.12845
- Seibold, M. A., Wise, A. L., Speer, M. C., Steele, M. P., Brown, K. K., Loyd, J. E., et al. (2011). A common MUC5B promoter polymorphism and pulmonary fibrosis. *N. Engl. J. Med.* 364, 1503–1512. doi: 10.1056/NEJMoa1013660
- Sternberg, C. N., Davis, I. D., Mardiak, J., Szczylik, C., Lee, E., Wagstaff, J., et al. (2010). Pazopanib in locally advanced or metastatic renal cell carcinoma: results of a randomized phase III trial. *J. Clin. Oncol.* 28, 1061–1068. doi: 10.1200/JCO.2009.23.9764
- Stuart, B. D., Choi, J., Zaidi, S., Xing, C., Holohan, B., Chen, R., et al. (2015). Exome sequencing links mutations in PARN and RTEL1 with familial pulmonary fibrosis and telomere shortening. *Nat. Genet.* 47, 512–517. doi: 10.1038/ng.3278
- Tennenbaum, D. M., Manley, B. J., Zabor, E., Becerra, M. F., Carlo, M. I., Casuscelli, J., et al. (2017). Genomic alterations as predictors of survival among patients within a combined cohort with clear cell renal cell carcinoma undergoing cytoreductive nephrectomy. *Urol. Oncol.* 35, 532.e7–532.e13. doi: 10.1016/j.urolonc.2017.03.015
- The Cancer Genome Atlas Research Network (2013). Comprehensive molecular characterization of clear cell renal cell carcinoma. *Nature* 499, 43–49. doi: 10.1038/nature12222
- Thiesen, H.-J., Steinbeck, F., Maruschke, M., Koczan, D., Ziemis, B., and Hakenberg, O. W. (2017). Stratification of clear cell renal cell carcinoma (ccRCC) genomes by gene-directed copy number alteration (CNA) analysis. *PLoS One* 12:e0176659. doi: 10.1371/journal.pone.0176659
- Umeda, S., Kanda, M., Sugimoto, H., Tanaka, H., Hayashi, M., Yamada, S., et al. (2017). Downregulation of GPR155 as a prognostic factor after curative resection of hepatocellular carcinoma. *BMC Cancer* 17:610. doi: 10.1186/s12885-017-3629-2
- Vancheri, C. (2015). Idiopathic pulmonary fibrosis and cancer: do they really look similar? *BMC Med.* 13:220. doi: 10.1186/s12916-015-0478-1
- Velmurugan, K. R., Varghese, R. T., Fonville, N. C., and Garner, H. R. (2017). High-depth, high-accuracy microsatellite genotyping enables precision lung cancer risk classification. *Oncogene* 36, 6383–6390. doi: 10.1038/onc.2017.256
- Vittinghoff, E., and McCulloch, C. E. (2007). Relaxing the rule of ten events per variable in logistic and Cox regression. *Am. J. Epidemiol.* 165, 710–718. doi: 10.1093/aje/kwk052
- Wang, A., Wu, L., Lin, J., Han, L., Bian, J., Wu, Y., et al. (2018). Whole-exome sequencing reveals the origin and evolution of hepato-cholangiocarcinoma. *Nat. Commun.* 9:894. doi: 10.1038/s41467-018-03276-y
- Yin, H., Liang, Y., Yan, Z., Liu, B., and Su, Q. (2013). Mutation spectrum in human colorectal cancers and potential functional relevance. *BMC Med. Genet.* 14:32. doi: 10.1186/1471-2350-14-32
- Yin, J.-Y., Zhang, J.-T., Zhang, W., Zhou, H.-H., and Liu, Z.-Q. (2018). eIF3a: a new anticancer drug target in the eIF family. *Cancer Lett.* 412, 81–87. doi: 10.1016/j.canlet.2017.09.055
- Zhang, Y., Gong, Y., Hu, D., Zhu, P., Wang, N., Zhang, Q., et al. (2015). Nuclear SIPA1 activates integrin β 1 promoter and promotes invasion of breast cancer cells. *Oncogene* 34, 1451–1462. doi: 10.1038/onc.2014.36
- Zhang, Y.-F., Wang, Q., Luo, J., Yang, S., Wang, J.-L., and Li, H.-Y. (2015). Knockdown of eIF3a inhibits collagen synthesis in renal fibroblasts via inhibition of transforming growth factor- β 1/Smad signaling pathway. *Int. J. Clin. Exp. Pathol.* 8, 8983–8989.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Mendoza-Alvarez, Guillen-Guio, Baez-Ortega, Hernandez-Perez, Lakhwani-Lakhwani, Maeso, Lorenzo-Salazar, Morales and Flores. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.