Next-Generation Sequencing of Translocation Renal Cell Carcinoma Reveals Novel RNA Splicing Partners and Frequent Mutations of Chromatin-Remodeling Genes

Gabriel G. Malouf, Xiaoping Su, Hui Yao, Jianjun Gao, Liangwen Xiong, Qiuming He, Eva Compérat, Jérôme Couturier, Vincent Molinier, Bernard Escudier, Philippe Camparo, Denaha J. Doss, Erika J. Thompson, David Khayat, Christopher G. Wood, Willie Yu, Bin T. Teh, John Weinstein, and Nizar M. Tannir

Abstract

**Purpose:** MITF/TFE translocation renal cell carcinoma (TRCC) is a rare subtype of kidney cancer. Its incidence and the genome-wide characterization of its genetic origin have not been fully elucidated.

**Experimental Design:** We performed RNA and exome sequencing on an exploratory set of TRCC (n = 7), and validated our findings using The Cancer Genome Atlas (TCGA) clear-cell RCC (ccRCC) dataset (n = 460).

**Results:** Using the TCGA dataset, we identified seven TRCC (1.5%) cases and determined their genomic profile. We discovered three novel partners of MITF/TFE (LUC7L3, KHSRP, and KHDARB2) that are involved in RNA splicing. TRCC displayed a unique gene expression signature as compared with other RCC types, and showed activation of MITF; the transforming growth factor β1 and the PI3K complex targets. Genes differentially spliced between TRCC and other RCC types were enriched for MITF and ID2 targets. Exome sequencing of TRCC revealed a distinct mutational spectrum as compared with ccRCC, with frequent mutations in chromatin-remodeling genes (six of eight cases, three of which were from the TCGA). In two cases, we identified mutations in INO80D, an ATP-dependent chromatin-remodeling gene, previously shown to control the amplitude of the S phase. Knockdown of INO80D decreased cell proliferation in a novel cell line bearing LUC7L3–TFE3 translocation.

**Conclusions:** This genome-wide study defines the incidence of TRCC within a ccRCC-directed project and expands the genomic spectrum of TRCC by identifying novel MITF/TFE partners involved in RNA splicing and frequent mutations in chromatin-remodeling genes. Clin Cancer Res; 20(15); 1–12. ©2014 AACR.

Introduction

Translocation renal cell carcinoma (TRCC) is a rare subtype of kidney cancer that was added to the World Health Organization (WHO) classification in 2004 (1) and is histologically and genomically a heterogeneous disease (2, 3). TRCC is characterized by translocations involving the genes for transcription factors E3 (TFE3) and EB (TFEB; refs. 1–4). TFE3 and TFEB belong to the microphthalmia transcription factor/transcription factor E (MITF/TFE) family of basic helix-loop-helix leucine zipper (bHLH-LZ), and are often called MITF TRCC. Although the TFEB gene has been reported to fuse exclusively with the Alpha gene, leading to t(6;11)(p21;q21) translocation, the TFE3 gene (Xp11.2) has been found to rearrange with at least five different partners: PRCC (1q21), ASPSCR1 (17q25), SFPQ (1p34), NONO (Xq12), and CLTC (17q23; refs. 1–6). The breakpoints of those translocations differ according to the TFE3 partner, and all TFE3 fusion proteins contain the bHLH-LZ and transcriptional activation domains of TFE3 (7). Unlike TFE3, the TFEB gene rearranges with Alpha, an intronless gene, leading to a translocation that preserves the

**Authors’ Affiliations:** Departments of ^1Medical Oncology and ^2Pathology, Groupe Hospitalier Pitié-Salpêtrière, Assistance Publique Hôpitaux de Paris, Faculty of Medicine Pierre et Marie Curie, Institut Universitaire de Cancérologie GRC5, University Paris 6; ^3Department of Genetics, Institut Curie; ^4Department of Pathology, Hôpital Saint Joseph, Paris; ^5Department of Medical Oncology, Institut Gustave Roussy, Villejuif; ^6Centre de Pathologie, Amiens, Picardie, France; Departments of ^7Bioinformatics and Computational Biology and ^8Genitourinary Medical Oncology, Institut Gustave Roussy, Villejuif; ^9DNA Analysis Facility, Department of Genetics, ^10Division of Surgery, Department of Urology, The University of Texas MD Anderson Cancer Center, Houston, Texas; and ^11Cancer and Stem Cell Biology Program, Duke-NUS Graduate Medical School, Singapore, Singapore

**Note:** Supplementary data for this article are available at Clinical Cancer Research Online (http://clincancerres.aacrjournals.org/).

**Corresponding Authors:** Nizar M. Tannir, University of Texas MD Anderson Cancer Center, 1155 Pressler Street, Unit 1374, Houston, TX 77030-3721, Phone: 713-563-7265; Fax: 713-745-0422; E-mail: ntannir@mdanderson.org; and Gabriel G. Malouf, Department of Medical Oncology, Hôpital de la Pitié-Salpêtrière, 43, Boulevard de l’Hôpital, 75013, Paris, France. Phone: +33 1 42 16 04 61; Fax: +33 1 42 16 04 23; gabriel.malouf@psl.aphp.fr

doi: 10.1158/1078-0432.CCR-13-3036

©2014 American Association for Cancer Research.
Translational Relevance

We performed exome and RNA sequencing on seven MITF/TFE translocation renal cell carcinoma (TRCC) tumors and validated our findings in seven of 460 (1.5%) clear-cell RCC cases from The Cancer Genome Atlas (TCGA). We discovered three novel partners of MITF/TFE (LUC7L3, KHSRP, and KHDRBS2) that are involved in RNA splicing. TRCC displayed a unique gene expression signature with activation of MITF, the transforming growth factor β1, and the PI3K complexes.

Genes differentially spliced between TRCC and other RCC types were enriched for MITF targets, suggesting a putative role for RNA splicing in kidney carcinogenesis. Exome sequencing revealed mutations in the chromatin-remodeling gene INO80D. Our study expands the spectrum of TRCC and raises potential therapeutic implications.

full-length TFE3 coding region, which becomes dysregulated by the Alpha gene promoter (8).

TRCC represents 15% of RCC in patients younger than 40 years (9, 10). The incidence of TRCC varies between 1% and 6% according to previously published studies, several of which used morphology and TFE3 expression alone to screen for RCC cases with translocation (10–12). However, the true incidence of TRCC in an unselected cohort of pathologically confirmed clear-cell RCC (ccRCC) remains to be determined. Next-generation sequencing highly improved our understanding of ccRCC, known to bear inactivation of the von Hippel-Lindau (VHL) tumor-suppressor gene, located on chromosome 3p arm (13). Large-scale screening has identified several new cancer genes that include mutations in the SWI/SNF family gene PBRM1 (14) and BAP1 (15), as well as mutations in chromatin remodelers such as KDM6A (16), KDM5C, and SETD2 (17). To date, the genetic basis and origins of TRCC remain poorly understood on a genome-wide scale.

Compared with ccRCC, which displays gene expression profiles composed of two main transcriptomic subsets named ccA and ccB (18), the transcriptomic signature of TRCC remains obscure. Although in a previous study tumors of the MITF/TFE family were shown to display a unique gene expression signature (19), their full transcriptomic signature remains unknown.

In this study, we report the incidence and describe the genomic profile of translocations of the MITF/TFE family identified in the ccRCC TCGA dataset. We also describe the genetic basis of other mutations in RCC, including the identification of frequent mutations in genes involved in chromatin remodeling, particularly the INO80D gene.

Materials and Methods

Primary TFE3-related TRCCs

After obtaining informed consent from patients according to approved research protocols at the MD Anderson Cancer Center (MDACC) and Institut Curie, fresh tissue specimens were obtained at the time of nephrectomy and stored at −80°C until DNA and RNA extraction were carried out. The clinicopathologic characteristics of these cases are summarized in Supplementary Table S1. Genomic DNA with matched normal DNA (adjacent normal kidney tissue) was available from three cases (Supplementary Table S1). RNA from seven controls (three papillary RCC, one ccRCC, and three normal kidney samples) was included. RNA was available for seven cases (Supplementary Table S1). DNA extraction was performed using the DNeasy Blood & Tissue Kit (Qiagen) according to the manufacturer’s instructions. RNA extraction was performed using the RNaseasy Kit (Qiagen) according to the manufacturer’s instructions.

RNA sequencing

Total RNA for each sample was converted into a library of template molecules for sequencing on the Illumina HiSeq 2000 according to the NuGen Ovation RNA-Seq System V2 protocol. The details are reported in Supplementary Materials and Methods.

Mapping/alignment

We checked the quality of the sequencing data by using HTSeq package. The raw paired-end reads in FASTQ format were then aligned to the human reference genome, GRCh37/hg19, using MOSAIK alignment software (20). MOSAIK works with paired-end reads from Illumina HiSeq 2000 and uses both a hashing scheme and the Smith–Waterman algorithm to produce gapped optimal alignments and to map exon junction-spanning reads with a local alignment option for RNA-Seq. The resulting pair-wise alignments were then consolidated into a multiple sequence alignment (assembly) and saved as a standard bam file.

Identification of differentially expressed genes from RNA-Seq

The details are presented in Supplementary Materials and Methods.

Fusion detection from RNA-Seq

A modified version of VirusSeq that implements a greedy algorithm with a robust statistical model was implemented and used in gene fusion discovery for RNA-seq data (21). The TRCC cases were confirmed by another bioinformatic pipeline called FusionSeq (22). Specifically, MOSAIK was used to align paired-end reads to human genome reference (hg19). A given paired-end read alignment was then quantified in terms of the genomic location (L) of the aligned read pair, the distance (D) between the aligned read pair of the fragment (insert), and the orientation (O) of the read pair, and the confidence that each read is uniquely aligned. The specific pattern in (L, D, O) space was used as a constraint to define the discordant read pair. For example, a discordant read pair may have an exceptionally long D
spanning a region in the reference genome. All discordant reads were then annotated using the genes defined in University of California, Santa Cruz refFlat file, and clustered the ones that support the same fusion event (e.g., PML–RARα). Finally, each fusion candidate was defined and selected as the discordant read clusters in which a statistical model-based algorithm with greedy strategy was implemented to accurately detect the boundaries of discordant read clusters and in silico fusion junctions. Here, in silico fusion junction is the nucleotide-level genomic coordinate on either side of the gene fusion and is not necessary to be at the ends of known exons. Specifically, the boundary for each discordant read cluster of candidate fusion was estimated on the basis of discordant read mapping locations and orientations with fragment length distribution (e.g., within mean plus three SDs, μ+3σ) as a constraint of cluster size. The cluster size of discordant reads was measured by using reads’ genomic location excluding introns if mapped reads are located across adjacent exons in a candidate fusion gene. Furthermore, an in silico sequence by using the consensus of reads within discordant read clusters for each fusion candidate was generated to help PCR primer design, which facilitates quick PCR validations.

Validation of fusion transcripts

Validation of identified fusion transcripts was performed by PCR using custom primers for amplification. Sanger sequencing was performed on a 3730xl DNA Analyzer (Applied Biosystems) using BigDye Terminator v3 chemistry (Applied Biosystems). Fusion sequences were verified using BLAST (NIH) and SeqScape Software v2.5 (Applied Biosystems).

Analysis of differentially spliced exons

To analyze genes that have differential usage of exons, we used DESeq, version 1.4.0 (23). Specifically, for each gene, a generalized linear model (GLM) was fit to detect the differential expression of its exons among four tissue types and to adjust the overall gene expression level and the batch effect. The exon-wise false discovery rate (FDR) was controlled by the Benjamini–Hochberg method. For exons with FDR < 0.01, pairwise comparisons were performed: TRCC versus normal kidney tissue, TRCC versus ccRCC, and TRCC versus papillary RCC. The Holm method was used to calculate the adjusted P values for pairwise comparisons.

Unsupervised classification of gene expression

The variance stabilizing transformations implemented in DESeq package were performed on the count data to conduct sensible distance calculation. For each gene, the dispersion was calculated to measure its variance among samples, and thus the 2,000 genes with highest dispersions were selected for clustering analysis. To remove the systematic difference between the MDACC and TCGA samples, the median expression values of each batch per gene were scaled to the same level. Hierarchical clustering analysis was performed using the Pearson correlation coefficient as the distance metric and Ward’s linkage rule. Principal component analysis was also applied to investigate the multivariate pattern. The consensus clustering algorithm with the hierarchical clustering method was used to perform clustering analysis for 460 TCGA ccRCC samples. The expressions of VHL, hypoxia inducible factor 1α (HIF1A), and HIF2A were compared as stated in “Identification of differentially expressed genes from RNA-Seq” (Supplementary Materials and Methods).

Exome sequencing

Exome capture was performed using Agilent SureSelect Human All Exon 50 Mb according to the manufacturer’s instructions. The technical details and mutation detection are presented in Supplementary Materials and Methods.

Transduction of human INO80D shRNA lentiviral particles into HCR-59 cells

The RCC cell lines generated in our laboratory at MDACC are karyotyped to ensure their human origin, and undergo cytokeratin profiling to confirm their epithelial origin, and are yearly tested for mycoplasma. The cell line has been tested and authenticated using DNA fingerprinting fewer than 2 months after resuscitation. We previously reported on a novel cell line (HCR-59), which we derived from the primary tumor of a 20-year-old female patient (3), and showed t(X;17q) by spectral karyotyping, and confirmed TFE3 fusion by fluorescence in situ hybridization (FISH). HCR-59 cells were seeded in 12-well plates and infected with human INO80D shRNA or control shRNA lentiviral particles when the cells reached about 50% of confluence, as per the manufacturer’s instructions (Santa Cruz Biotechnology). One day after transduction, the medium was changed to remove untransduced materials. Puromycin dihydrochloride was used to select and maintain stably transduced clones according to the manufacturer’s brochure.

Real-time PCR analysis of mRNA level of human INO80D from stably transduced clones

RNA was extracted from amplified single clones using RNeasy Plus Mini Kit (Qiagen). First-strand cDNA was synthesized using the SuperScript III First-Strand Synthesis System (Invitrogen), and then used for PCR using SYBR Green PCR Master Mix (Applied Biosystems). Beta-actin was used as an internal control.

Western blot analysis of INO80D

To detect INO80D protein expression level from cells transduced with human INO80D shRNA and control shRNA, Western blotting was performed and protein expression levels of INO80D were quantified using IRDye 680LT, 800CW as secondary antibodies. All images were generated on Odyssey with a scan intensity setting of 3.5/5 (700/800 nm), and sensitivity of 5. Beta-actin was used as an internal control.
Cell proliferation study

To analyze the effect of INO80D on the proliferation of HCR-59 cells, the INO80D shRNA–transduced clones were tested for proliferation using CellTiter 96 Aqueous One Solution Cell Proliferation Assay (Promega). Briefly, HCR-59 cells were cultured in 96-well plates in a humidified, 5% CO₂ atmosphere for 0 to 5 days, and 20 μL of CellTiter 96 Aqueous One Solution Reagent was added to each well containing cells in 100 μL of culture medium. The plates were then incubated at 37°C for 2 hours before detection of absorbance at 490 nm using a 96-well plate reader.

Results

High accuracy of paired-end RNA sequencing for detection of MITF/TFE family fusion transcripts

To assess the sensitivity and accuracy of paired-end transcriptome sequencing in detection of MITF/TFE fusion transcripts, we performed paired-end RNA-Seq on six samples of TFE3-related TRCC, for which we previously reported karyotypes, fusion transcripts, and outcomes (3, 24, Supplementary Table S1), and seven control samples (three papillary RCC, one ccRCC, and three normal kidney tissues). All of the cases studied here were genetically confirmed. Using our bioinformatic pipeline, we were able to detect TFE3 translocations in all Xp11.2 RCC samples (Table 1) and also identified additional fusion transcripts that we validated (Supplementary Table S2). All TFE3 breakpoints identified here were previously known (Table 1). One TFE3-related RCC sample (RCC-T34) had a confirmed translocation involving the vascular endothelial growth factor A (VEGFA) in the 3’ position, in addition to the ASPSCR1–TFE3 fusion transcript. To our knowledge, this is the first report of VEGFA translocation in RCC (Supplementary Table S2). No fusion transcripts were identified in the control samples. Overall, we confirmed 10 of 12 translocations using reverse-transcription polymerase chain reaction (RT-PCR; Supplementary Table S2).

Identification of a novel LUC7L3–TFE3 fusion transcript

Because one of our identified TRCC tumor samples (RCC-T1) had TFE3 translocation confirmed by FISH, without a classic TFE3 partner identified, we performed paired-end RNA-Seq on this sample and revealed a novel fusion transcript involving LUC7L3–TFE3, leading to a balanced translocation (Fig. 1A). LUC7L3 is located on chromosome 17q (17q21.33), near two other TFE3 partners, ASPSCR1 and CLTC. This gene encodes a protein with a C-terminal half that is rich in arginine and glutamate residues (RE domain) and arginine and serine residues (RS domain). Interestingly, this protein is involved in RNA splicing via the RE and RS domains (Fig. 1A), thus expanding the spectrum of genes involved in RNA splicing and fusing with TFE3.

Spectral karyotyping performed on a novel cell line (HCR-59), derived from a patient with this novel translocation, revealed multiple interchromosomal translocations occurring within 17q (3), but only LUC7L3–TFE3 led to a fusion transcript. LUC7L3–TFE3 was further validated by RT-PCR in the primary sample (Fig. 1B and C) and the HCR-59 cell line (not shown).

Table 1. Genomic characterization of MITF/TFE TRCC

<table>
<thead>
<tr>
<th>Sample name</th>
<th>Sample origin</th>
<th>Fusion product</th>
<th>Mate number</th>
<th>TFE3 fusion site (amino acid)</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>RCC-T25</td>
<td>MDACC</td>
<td>NONO–TFE3</td>
<td>38</td>
<td>295–296</td>
<td>Previously identified fusion</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TFE3–NONO</td>
<td>21</td>
<td>295–296</td>
<td></td>
</tr>
<tr>
<td>RCC-T31</td>
<td>MDACC</td>
<td>PRCC–TFE3</td>
<td>6</td>
<td>178–179</td>
<td>Previously identified fusion</td>
</tr>
<tr>
<td>RCC-T32</td>
<td>MDACC</td>
<td>PRCC–TFE3</td>
<td>7</td>
<td>295–296</td>
<td>Previously identified fusion</td>
</tr>
<tr>
<td>RCC-T33</td>
<td>MDACC</td>
<td>PRCC–TFE3</td>
<td>3</td>
<td>295–296</td>
<td></td>
</tr>
<tr>
<td>RCC-T34</td>
<td>MDACC</td>
<td>ASPSCR1–TFE3</td>
<td>2</td>
<td>295–296</td>
<td>Previously identified fusion</td>
</tr>
<tr>
<td>RCC-T35</td>
<td>MDACC</td>
<td>SFPO–TFE3</td>
<td>26</td>
<td>260–261</td>
<td>Previously identified fusion</td>
</tr>
<tr>
<td>RCC-T1</td>
<td>MDACC</td>
<td>TFE3–LUC7L3</td>
<td>9</td>
<td>77–78</td>
<td>Fusion identified by RNA-Seq</td>
</tr>
<tr>
<td></td>
<td></td>
<td>LUC7L3–TFE3</td>
<td>1</td>
<td>77–78</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>TFE3–SFPO</td>
<td>162</td>
<td>39–40</td>
<td>Fusion identified by RNA-Seq</td>
</tr>
<tr>
<td>K345601A</td>
<td>TCGA</td>
<td>SFPO–TFE3</td>
<td>123</td>
<td>39–40</td>
<td>Fusion identified by RNA-Seq</td>
</tr>
<tr>
<td>K475601A</td>
<td>TCGA</td>
<td>SFPO–TFE3</td>
<td>296</td>
<td>39–40</td>
<td>Fusion identified by RNA-Seq</td>
</tr>
<tr>
<td>K475801</td>
<td>TCGA</td>
<td>SFPO–TFE3</td>
<td>5</td>
<td>39–40</td>
<td>Fusion identified by RNA-Seq</td>
</tr>
<tr>
<td>K545801</td>
<td>TCGA</td>
<td>SFPO–TFE3</td>
<td>89</td>
<td>295–296</td>
<td>Fusion identified by RNA-Seq</td>
</tr>
<tr>
<td>K570501</td>
<td>TCGA</td>
<td>SFPO–TFE3</td>
<td>2</td>
<td>295–296</td>
<td>Fusion identified by RNA-Seq</td>
</tr>
<tr>
<td>K568101A</td>
<td>TCGA</td>
<td>KHSRP–TFE3</td>
<td>19</td>
<td>178–179</td>
<td>Fusion identified by RNA-Seq</td>
</tr>
<tr>
<td>K331301A</td>
<td>TCGA</td>
<td>TFEB–KHDRBS2</td>
<td>485</td>
<td>397–398/TFEB</td>
<td>Fusion identified by RNA-Seq</td>
</tr>
</tbody>
</table>
Identification of MITF/TFE translocation in ccRCC of TCGA

To characterize the functional translocations in ccRCC and identify fusion transcripts in a large dataset of kidney cancer, we analyzed paired-end transcriptome sequencing of 460 ccRCC samples profiled by TCGA. Using a threshold detection of a minimum of four tags per fusion junction, we identified fusion transcripts in 87 tumors (18.7%) with a median translocation number of one per tumor (range, 1–12). Aiming to select the translocations that are most likely oncogenic, we used stringent criteria of a minimum of six tags identified per fusion junction, with at least one tag spanning the fusion junction. These criteria were based on the assumption that oncogenic fusion transcripts should be overexpressed. Using the criteria of six or more tags per fusion junction, we identified 123 translocations, including 84 (68.3%) intrachromosomal and 39 (31.7%) extrachromosomal translocations (Supplementary Tables S3 and S4). These were related to 81 tumors (17.6%; Fig. 2), with a median of one translocation per tumor (range, 1–10; Fig. 3A).

To analyze whether the fusion genes have functional relevance, we used Database for Annotation Visualization and Integrated Discovery (DAVID). Functional annotations revealed exclusive enrichment of genes involved in chromosomal rearrangements ($P = 4 \times 10^{-4}$; FDR < 0.04), which were related to 13 fusion genes and had occurred in 19 patients (~23.5% of all patients with translocation; Supplementary Table S5). These genes were TFE3 ($n = 6$), TFEB ($n = 1$), SFPQ ($n = 5$), FHIT ($n = 2$), SLC9A9 ($n = 2$), AFF1 ($n = 1$), MKL1 ($n = 1$), LHFP ($n = 1$), ELL ($n = 1$), JAK2 ($n = 1$), DCX ($n = 1$), EP300 ($n = 1$), and LNP1 ($n = 1$). Six of these genes are known to be protooncogenes (TFEB, TFE3, MKL1, ELL, JAK2, and AFF1). Thus, TRCC represents the most frequently occurring common translocation found in TCGA’s RCC dataset and represents 1.5% (7 of 460) of all ccRCC cases (Fig. 2).

Interestingly, tumors bearing fusion transcripts with a partner previously reported to be involved in translocation have an overall higher rate of translocation as compared with those with newly identified translocations, with translocations per tumor having mean values of 2.8 versus 1.2, respectively ($P < 0.0003$; Fig. 3B). After excluding the six tumors with a translocation rate of four or more per tumor, DAVID analysis found that samples with fusion genes were enriched for fusions occurring within chromosome 3 ($n = 20; P = 4.6 \times 10^{-4}$) and chromosome 5 ($n = 17; P = 1.7 \times 10^{-3}$). The majority of those fusions were intrachromosomal translocations ($n = 24$); there were only five cases with $t(3;5)$ translocation. Two of those five cases involved the FHIT gene, known to be involved in hereditary RCC; in our cohort, FHIT fused with FAM172A and CAMK4 genes. No gene fusion transcripts were identified in 62 cases of normal kidney samples extracted from tissue adjacent to ccRCC.

Figure 1. Characterization of the novel fusion transcript LUC7L3–TFE3. A, genomic breakpoint mapped to base resolution, with arrows indicating breakpoints; and schematic of chimeric protein. e, exon. B, Sanger sequencing validation of the fusion transcript. C, RT-PCR validation of the fusion transcript.
Genomic profile of MITF/TFE family TRCC

All but one TFE3-related TRCC identified in TCGA data were found to fuse with the SFPQ gene (Table 1). These results were confirmed by FusionSeq, another modular framework for gene fusion detection (22). Three TRCC samples showed novel breakpoint junctions that occurred within the first intron, leading to fusion proteins that contained almost the entire TFE3 protein (Table 1). We also identified a novel partner of TFE3, which is the KHSRP gene. This novel fusion protein includes almost the entire KHSRP protein at the C-terminal part and the TFE3 protein at the N-terminal part (Fig. 4A). Of note, KHSRP regulates the maturation of a group of miRNAs in addition to its role in promoting

Figure 2. Workflow process of the fusion detection algorithm implemented and applied to data from MDACC and TCGA.

Figure 3. Distribution of identified TRCCs from TCGA. A, distribution of the number of fusion transcripts identified in cases with at least one identified fusion transcript. B, comparison of the number of fusion transcripts per sample according to whether translocation involves previously known or novel genes. Mann–Whitney non parametric test was used for this analysis.
mRNA decay, leading to the integration of specific regulatory processes for protein expression. Furthermore, we discovered a novel fusion transcript involving TFEB–KHDRBS2 genes (Fig. 4B). This fusion transcript is unique in two ways:

1. The TFEB gene was located at the 5' end of the novel fusion protein, which retained almost the entire TFEB coding protein, in particular the helix-loop-helix domain. This has also been found in Alpha–TFEB translocation (8). Furthermore, KHDRBS2 contains a KH-type splicing regulatory domain that is involved in RNA splicing, as does KHSRP, and the majority of TFE3 partners.

2. In addition to the TFEB–KHDRBS2 fusion transcript, this sample of TRCC showed one of the highest numbers of tags spanning the fusion junction in the entire TCGA dataset (Supplementary Table S3). In particular, this tumor bears nine additional fusion transcripts, consistent with rearrangements occurring within 28.9 Mb (chr6: 33,748,810–62,638,252; Fig. 4C and D). We termed this process a "translocator phenotype."

Likewise, the RCC sample with KHSRP–TFE3 fusion transcript had nine additional translocations occurring within a region spanning approximately 17.3 Mb on chromosome 21 within a region spanning approximately 54.8 Mb on chromosome X and consistent with a "translocator phenotype."

To rule out the possibility of aberrant splicing that may explain the "translocator phenotype," we downloaded whole-exome sequencing data from TCGA for both KHSRP–TFE3 and TFEB–KHDRBS2 cases (there are no whole-genome sequencing data for both cases), and performed translocation detection. Interestingly, we were able to detect the breakpoint for KHSRP–TFE3 with whole-exome data, as the breakpoint is located within exon 4 of TFE3 (Supplementary Fig. S1); thus, we could exclude the
possibility that KHSRP–TFE3 is related to an aberrant splicing. Unfortunately, we were not able to detect the break-point for TFEB–KHDRBS2 with whole-exome data, as the breakpoint does not occur within any exon of both fusion partners.

Pathway analysis of MITF/TFE TRCC

To assess whether TRCC displays a specific transcriptional profile, we analyzed our seven TRCC samples, and seven cases (six TFE3-related and one TFEB-related) that we identified in the dataset from TCGA, and compared them with 17 ccRCC samples (one from MDACC and 16 from TCGA), 19 papillary RCC samples (three from MDACC and 16 from TCGA), and 16 normal kidney tissue samples (two from MDACC and 14 from TCGA). After removing the batch effect, unsupervised clustering revealed that TRCC displays a unique gene expression signature, as compared with those of ccRCC and papillary RCCs, except for two TFE3-related RCCs that were found to be minor clones (Fig. 5A). These data were confirmed by principal component analysis (not shown). Thus, MITF/TFE family translocations identified in the RCC cohort from TCGA have transcriptional profiles that are similar to those of the pediatric and young adult patients profiled independently through our sample collection, and share the same active and distinct pathways as those of other RCC subtypes.

Compared with the gene expression profile of normal kidney tissue, Ingenuity Pathway Analysis (IPA) revealed that besides the expected activation of MITF ($P = 1.1 \times 10^{-9}$), TRCC had predicted activation of the following upstream regulators: transforming growth factor β1 (TGFβ1; $P = 6.6 \times 10^{-17}$), lipopolysaccharides ($P = 6.07 \times 10^{-14}$), tumor necrosis factor ($P = 4.2 \times 10^{-11}$), and the PI3K complex ($P = 6.9 \times 10^{-15}$). Compared with ccRCC, IPA revealed that TRCC showed activation of the MITF ($P = 3.7 \times 10^{-8}$) and estrogen-related receptor α (ESRRA; $P = 2.99 \times 10^{-4}$) pathways, and inhibition of the HIF2α ($P = 1.9 \times 10^{-10}$) and VEGFA ($P = 3.7 \times 10^{-3}$) pathways.

Figure 5. A, unsupervised clustering of MITF/TFE TRCC, ccRCC, papillary RCC, and normal kidney tissue revealing that TRCC displays a unique transcriptomic profile as compared with normal kidney and other RCC types. Note that the two possible minor clones among the TFE3-related TRCC cases are clustered with other ccRCC samples. B, unsupervised clustering of TRCC showing two subgroups with no samples clustered according to the TFE3 partner. C, unsupervised clustering of gene expression of the 460 ccRCCs extracted from TCGA, showing the TRCCs clustered together within the ccB group, except for the two cases that we defined as being TFE3 minor clones.
Analysis of differentially spliced genes in MITF/TFE TRCC

Because the majority of TFE3 and TFEB partners were involved in RNA splicing, we investigated whether any genes were differentially spliced between TRCC and normal kidney tissue, ccRCC, or papillary RCC, respectively. Using the following thresholds (adjusted P-values of pairwise comparisons < 0.01 and fold change > 2), we identified 86 genes that were differentially spliced in respective comparisons of TRCC and each of the three other groups (not shown). Interestingly, for those three comparisons, IPA showed four pathways that were consistently differentially spliced (not shown), which included MITF and inhibitor of DNA binding 2 (ID2) targets. Thus, TRCCs are not only characterized by activation of the MITF target genes but also by alterations of their RNA splicing. We found consistent differential splicing of the PMEL gene, which encodes a premelanosine protein. Differential exon usage between TRCC versus normal kidney tissue, papillary RCC or ccRCC was present within exons 4, 26, 27, 41, and 42 (Supplementary Fig. S2). We also identified differentially spliced genes belonging to the HNF4A, TP53, MGEA5, and HRAS pathways, suggesting that oncogenic fusion proteins may affect splicing in pathways deregulated by tumor-suppressor genes and oncogene targets.

Integrative analysis of TRCC and gene expression reveals enrichment of translocations within the cc2B subgroup

We considered whether TRCC has a distinct signature according to its fusion partners. Although the number of samples was small, unsupervised clustering performed on all samples in the MDACC and TCGA datasets revealed two subgroups that clustered independently from the partners of TRCC (Fig. 5B). Another interesting consideration was whether the MITF-related TRCC identified in the data from TCGA formed a distinct subgroup. We performed hierarchical unsupervised consensus clustering on the entire TCGA cohort. Our results were consistent with previous results by Brannon and colleagues (18), which showed that ccRCC can be divided in two dominant types, ccA and ccB. In our analyses, five of the seven samples of MITF-related TRCC were associated with ccB cluster, whereas the two samples of TFE3-related TRCC (minor clones) were associated with the ccA cluster (Fig. 5C).

We then considered whether the expression of VHL, HIF1A, HIF2A, or EPAS1 differed between TRCC and normal kidney tissue, and found no difference for VHL expression, suggesting that carcinogenesis of TRCC is independent of VHL (not shown).

Mutational landscape of MITF/TFE family TRCC displays frequent mutations of chromatin remodelers

We used exome sequencing to assess the mutational status of three TFE3-related TRCC cases for which matched normal samples were available at MDACC (one collected from a subject younger than 18 years; Supplementary Table S1), and two cases without matched normal samples available (1 TFE3- and 1 TFEB-related TRCC). We did not identify any recurrent mutation in our cohort. Moreover, no VHL, PBRM1, or BAP1 mutations were identified in any of our cases, suggesting different mechanisms for the initiation and progression of these tumors, as compared with ccRCC. Because chromatin-remodeling genes were recently shown to be frequently altered in ccRCC, we decided to look for mutations in these genes in TRCC. As a result, three of the five cases had mutations in chromatin-remodeling genes (Supplementary Table S6). One case (RCC-T29) had a confirmed frameshift mutation within the SMARC2 gene, a member of the SWI/SNF family likely leading to a deleterious protein. The second case (RCC-T2) had a missense mutation of KDM5C gene, which acts as a histone H3 lysine 4 demethylase and is frequently inactivated in ccRCC (17). The third case (RCC-T1) had a missense and a frameshift mutation of INO80D gene, a chromatin remodeler belonging to the INO80 complex. Both KDM5C and INO80D mutations were predicted by SIFT and mutation assessor to lead to deleterious proteins. These mutations were all validated by Sanger sequencing at the DNA level. Furthermore, INO80D mutation was also validated at the RNA level, as RNA was available for RCC-T1 case (Fig. 6A).

We then decided to validate our findings in the TCGA cohort. We extracted from COSMIC database the list of confirmed somatic missense mutations related to three of the seven MITF/TFE family TRCC cases identified in the TCGA, and for which data were available. The K570501 was the only case with VHL mutation. None of the two other cases had VHL, PBRM1, or BAP1 mutations. Interestingly, all three cases had mutations in chromatin-remodeling genes, including CHD9 in K570701, CHD7 and INO80D in K568101, and MLL3 in K554601. Thus, 75% of TRCCs (six of eight cases) display mutations in chromatin-remodeling genes.

INO80D knockdown affects cell proliferation

Because we identified missense mutations in INO80D in two TRCC cases, we decided to assess the frequency of INO80D mutations in ccRCC using COSMIC database. Among 334 cases assessed, only two cases had missense-substitution mutations, and one of those is the TRCC case K570701. Because the INO80 chromatin-remodeling complex has not been implicated in cancer, we decided to analyze the effect on proliferation of this ATP-dependent chromatin-remodeling factor, which has been demonstrated to control the amplitude of the S phase (25). We used the HCR-59 cell line that we generated from RCC-T1 (3), which bears the INO80D mutation at both DNA and RNA levels. Interestingly, this mutation was predicted to be deleterious by Provean (cutoff = 2.5) and SIFT (cutoff = 0.05). Moreover, SNP array data showed no gain or loss of copy number in chromosome 2q where INO80D resides (not shown). Abrogation of INO80D expression via shRNA knockdown in the HCR-59 cell
line showed decreased proliferation as a consequence of INO80D loss (Fig. 6B–D).

Discussion

To our knowledge, this is the first genome-wide analysis using RNA-Seq and exome sequencing to analyze in depth the genomic abnormalities of TRCC. Our analysis reveals important findings. First, MITF/TFE was the most frequent recurrent translocation identified in the TCGA cohort with an observed 1.5% incidence within a ccRCC-directed project. We believe that TRCC may have been deselected from the ccRCC TCGA cohort, resulting in ascertainment bias. Consequently, the true incidence of TRCC among ccRCC might be higher than the 1.6% reported by Komai and colleagues (10). Second, the spectrum of TFE3–TFEB fusion transcripts we identified in adults differed from those of historic series, such as five of the seven cases we identified from TCGA were related to SFPQ–TFE3 fusion genes. Moreover, we identified novel breakpoints for SFPQ–TFE3 translocation, suggesting that breakpoints of translocations may vary among different patient groups.

By identifying two novel partners of TFE3, LUC7L3 and KHSRP, which are involved in RNA splicing, we expanded the spectrum of translocations. In addition, we found the TFEB gene to fuse at the 3' end with KHDRBS2, which is another gene involved in RNA splicing. As for Alpha–TFEB translocation, the entire TFEB protein is preserved in this translocation, and it is likely that carcinogenesis in this type of RCC is related to TFEB deregulation, as suggested by other investigators (8). Interestingly, the two cases with KH-domain translocations showed multiple fusion transcripts, which represented unique patterns as compared with those of the other ccRCCs. We labeled those tumors as having "translocator phenotype." For TFE3–KHSRP, using exome sequencing, we were able to demonstrate that breakpoint occurs indeed in exon 4. For TFE3-related TRCC, those fusion transcripts were generated within a short genomic region in chromosome 6. Further analyses are required to better understand the genetic basis of those multiple fusion transcripts. Although we have not been able to validate the TFE3 fusion by RT-PCR (TCGA cohort), evidence suggests that this fusion transcript is highly likely to exist. This is based on the high number of tags identified by RNA-Seq with 79 tags spanning the fusion junction, clustering of TFE3- and TFE3-related TRCC, which showed clustering of TFE3-related TRCC with other TRCC cases and finally MITF and TGFB1 pathway activation. A plausible explanation for the "translocator phenotype" could be a predisposing factor for the oncogenic translocation, such as chemotherapy in childhood, as postulated by Argani and colleagues (26). It can also be related to chromothripsis, a process by which clusters of thousands of rearrangements occur in confined genomic regions (27). Unfortunately, we were not able to get whole-genome data from the TCGA to investigate this process.

Another interesting finding relates to the transcriptomic profiling of TRCC, which revealed that the majority of cases belonged to the ccb transcriptomic group. IPA revealed TGFB1 and PI3K complex activations. We previously reported clinical activity of VEGF and mammalian target of rapamycin (mTOR)–directed therapies in TRCC (28, 29).
An important question is whether there is a difference in gene splicing between TRCC and normal kidney tissue and other RCC types (30). Using RNA-Seq, we analyzed gene splicing and discovered that MITF targets are differentially spliced in TRCC. We speculate that by binding to MITF targets, the fusion proteins may activate MITF targets and affect their splicing; as is the case for the PMEL gene, which encodes for a premelanosome protein that is regulated by MITF. Additional experiments are needed to further explore this process.

Our analysis describes the landscape of mutations for TRCC. Although no recurrent hotspot mutations were identified, the spectrum of mutations in TRCC differs from those of other RCC types, which are characterized by mutations of VHL, PBRM1, or BAP1. It is interesting to highlight that six of eight TRCC cases had mutations in chromatin-remodeling genes, particularly, mutations in the INO80D chromatin-remodeling gene. To our knowledge, this is the first report of INO80D mutations in cancer. Whether these mutations can lead to translocations, by altering a DNA-repair process, remains to be determined. It is interesting to note that knockdown of INO80D, previously shown to control the amplitude of the S phase (25), decreased cell proliferation in HCR-59 cell line bearing LUC7L3–TFE3 translocation. We postulate that INO80D mutations may play a role in promoting an aggressive phenotype of TRCC.

In summary, we identified novel partners of TFE3 and TFEB, which are all involved in RNA splicing. We analyzed the genetic landscape for somatic mutations, and defined the gene expression signature as well as altered signaling pathways in this disease. We believe that our findings provide a framework for future therapeutic interventions.

Disclosure of Potential Conflicts of Interest
No potential conflicts of interest were disclosed.

Authors’ Contributions

Conception and design: G.G. Malouf, H. Yao, L. Xiong, B. Escudier, C.G. Wood, B.T. Teh, N.M. Tannir


Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): G.G. Malouf, J. Gao, L. Xiong, Q. He, E. Compérat, J. Couturier, V. Moliné, P. Camargo, D.J. Doss, E. Thompson, C.G. Wood, B.T. Teh, N.M. Tannir

Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): G.G. Malouf, X. Su, H. Yao, L. Xiong, Q. He, B. Escudier, C.G. Wood, W. Yu, J. Weinstein, N.M. Tannir

Writing, review, and/or revision of the manuscript: G.G. Malouf, X. Su, H. Yao, J. Gao, L. Xiong, J. Couturier, V. Moliné, D.J. Doss, D. Khayat, C.G. Wood, N.M. Tannir

Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): G.G. Malouf, D. Khayat, C.G. Wood, N.M. Tannir

Study supervision: D. Khayat, C.G. Wood, N.M. Tannir

Grant Support
This work was supported in part by grants from Kidney Cancer Research Group, NIH/NCI award number P30CA016672, and the Genitourinary Cancers Program of the Cancer Center Support Grant shared resources at the MD Anderson Cancer Center.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received November 7, 2013; revised May 7, 2014; accepted May 15, 2014; published OnlineFirst June 4, 2014.

References


Next-Generation Sequencing of Translocation Renal Cell Carcinoma Reveals Novel RNA Splicing Partners and Frequent Mutations of Chromatin-Remodeling Genes

Gabriel G. Malouf, Xiaoping Su, Hui Yao, et al.

Clin Cancer Res  Published OnlineFirst June 4, 2014.

Updated version  Access the most recent version of this article at:
doi:10.1158/1078-0432.CCR-13-3036

Supplementary Material  Access the most recent supplemental material at:
http://clincancerres.aacrjournals.org/content/suppl/2014/06/04/1078-0432.CCR-13-3036.DC1

E-mail alerts  Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions  To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions  To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.