Spectrum of diverse genomic alterations define non–clear cell renal carcinoma subtypes

Steffen Durinck[1,2,14], Eric W Stawiski[1,2,14], Andrea Pavía-Jiménez[3–5,14], Zora Modrusan[1,14], Payal Kapur[3,6,14], Bijay S Jaiswal[1], Na Zhang[1], Vanina Toffessi-Tcheuyap[3–5], Thong T Nguyen[1], Kanika Bajaj Pahuja[1], Ying-Jiun Chen[1], Sadia Saleem[3,4], Subhra Chaudhuri[1], Sherry Heldens[1], Marlena Jackson[1], Samuel Peña-Llopis[3–5], Joseph Guillory[1], Karen Toy[1], Connie Ha[1], Corissa J Harris[1], Eboni Holloman[3–5], Haley M Hill[3–5], Jeremy Stinson[1], Celina Sanchez Rivers[1], Vasantharajan Janakiraman[1], Weiru Wang[7], Lisa N Kinch[3,8,9], Nick V Grishin[3,8,9], Peter M Haverty[2], Bernard Chow[1], Julian S Gehring[2], Jens Reeder[2], Gregoire Pau[2], Thomas D Wu[2], Vitaly Margulis[3,10], Yair Lotan[3,10], Arthur Sagalowsky[3,10], Ivan Pedrosa[3,11,12], Frederic J de Sauvage[13], James Brugarolas[3–5] & Somasekar Seshagiri[1]

To further understand the molecular distinctions between kidney cancer subtypes, we analyzed exome, transcriptome and copy number alteration data from 167 primary human tumors that included renal oncocytomas and non–clear cell renal cell carcinomas (nccRCCs), consisting of papillary (pRCC), chromophobe (chRCC) and translocation (tRCC) subtypes. We identified ten significantly mutated genes in pRCC, including *MET*, *NF2*, *SLC5A3*, *PNKD* and *CPQ*. *MET* mutations occurred in 15% (10/65) of pRCC samples and included previously unreported recurrent activating mutations. In chRCC, we found *TP53*, *PTEN*, *FAAH2*, *PDHB*, *PDXDC1* and *ZNF765* to be significantly mutated. Gene expression analysis identified a five-gene set that enabled the molecular classification of chRCC, renal oncocytoma and pRCC. Using RNA sequencing, we identified previously unreported gene fusions, including *ACTG1-MITF* fusion. Ectopic expression of the *ACTG1-MITF* fusion led to cellular transformation and induced the expression of downstream target genes. Finally, we observed upregulation of the anti-apoptotic factor *BIRC7* in MiTF-high RCC tumors, suggesting a potential therapeutic role for BIRC7 inhibitors.

Kidney cancer accounts for ~60,000 new cases and ~13,000 deaths annually in the United States[1]. About 85% of kidney cancers are RCCs, which arise from the renal epithelium. Clear cell RCC (ccRCC), which constitutes 75% of RCC cases, is the best characterized kidney cancer subtype[2–4]. The remaining 25% of RCCs, broadly classified as nccRCC, represent distinct tumor subtypes, including pRCC (10–15%) and chRCC (4–5%)[5–8]. Presurgical diagnosis of chRCC on the basis of imaging studies is challenging, as tumors often show overlapping features with those of ccRCC and renal oncocytoma[9–11], a benign kidney epithelial tumor with an incidence rate of ~5% (refs. 5–8). Similarly, in needle core biopsies, chRCC is at times difficult to distinguish from renal oncocytoma. Thus, diagnosis remains a challenge and is confounded by the presence of mixed tumors that show features of both renal oncocytomas and chRCCs[5,6,12–15]. Other nccRCC types include collecting duct (<1%), tRCC (rare)[16]

and medullary (rare) subtypes. About 4–5% of tumors remain unclassified[5,8,17]. Although infrequent, tRCCs tend to affect adolescents and young adults and are particularly devastating. Although several drugs have recently been approved for metastatic RCC, registration trials involved almost exclusively patients with ccRCC, and there are no treatments with demonstrated efficacy in nccRCC subtypes[17].

An understanding of the genetic basis of nccRCCs has come from familial studies where germline mutations have been identified in *MET* in pRCC (type I), *FH* in pRCC (type II) and *FLCN* in the chRCC and renal oncocytoma tumor types[18,19]. In addition to predisposing germline mutations in *VHL* in ccRCC, pathogenic variants in *TSC1*, *TSC2*, *PTEN*, *SDHB*, *SDHC*, *SDHD*, *MITF* and *BAP1* have been associated with predisposition to kidney cancer subtypes[19–22]. Although somatic mutations in *MET* in sporadic pRCC[23,24] and translocations involving the microphthalmia family (MiTF) members *TFE3* and

TFEB in tRCCs[25] are known, alterations driving other sporadic forms of nccRCC remain to be identified.

To gain a better understanding of the nccRCC subtypes, we have applied next-generation sequencing technologies to characterize the exomes, transcriptomes and SNP-based genome-wide copy number alterations of multiple primary nccRCCs. We identified significantly mutated genes that include *MET* in pRCC and *TP53* in chRCC. Using RNA sequencing (RNA-seq)-based expression analysis, we have identified a set of five genes that enable the molecular classification of chRCC, renal oncocytoma and pRCC subtypes. Additionally, we describe the identification of previously unreported gene fusions, including a transforming *ACTG1-MITF* fusion.

RESULTS

Deep sequencing analysis of nccRCC samples
In this study, we have analyzed 167 human primary nccRCCs and their matched normal samples. The nccRCC samples, classified on the basis of morphological diagnosis, included 67 pRCCs, 49 chRCCs (36 classic, 12 eosinophilic and 1 renal oncocytic neoplasm favoring chromophobe eosinophilic), 35 renal oncocytomas (31 renal oncocytomas and 4 renal oncocytic neoplasms favoring oncocytoma), 8 unclassified RCCs, 6 tRCCs and 2 samples with largely sarcomatoid dedifferentiation (**Table 1**, **Supplementary Fig. 1** and **Supplementary Table 1**). DNA and RNA were isolated concomitantly from the same tissue sample to enable integrative genomic analyses[26]. Details of the samples and data analysis can be found in **Table 1**, the Online Methods, **Supplementary Figures 2** and **3**, and **Supplementary Table 2**.

nccRCC mutation profiles
Exome sequencing and analysis of pRCC samples identified an average of 61 (±31) protein-coding alterations per sample (**Fig. 1a**). This number is significantly higher than the average of 45 (±19) protein-coding alterations observed in The Cancer Genome Atlas (TCGA) ccRCC data set[4] ($P = 3.6 \times 10^{-7}$). Differences were also seen in the mutational load between subtypes (**Supplementary Tables 3–7**). For example, chRCC and renal oncocytoma tumors had, on average, a significantly lower number of somatic mutations than pRCC tumors (**Fig. 1b–d** and **Supplementary Table 8**). The somatic mutations identified in pRCC, chRCC and renal oncocytoma are summarized in **Supplementary Tables 3, 4** and **8**.

We compared all the protein-altering changes identified in this study with those reported in the Catalogue of Somatic Mutations in Cancer (COSMIC)[27] and found that 90% (4,280/4,751) of them were new somatic changes (**Supplementary Tables 3, 9** and **10**). We evaluated a subset of these mutations using mass spectrometry genotyping and observed a validation rate of 92% (164/178; **Supplementary Table 3**). Sanger sequencing confirmed somatic indels with a 96%

validation rate (127/132; **Supplementary Table 4**). Additionally, using RNA-seq data, we were able to confirm the expression of 1,412 somatic variants (Online Methods and **Supplementary Tables 3** and **4**).

The spectrum of base changes in tumors can provide insights into the underlying mutational processes that contribute to the alterations observed[28]. Using recently described methods (ref. 28 and J.S.G., B. Fischer and W. Huber, unpublished data), we identified five distinct mutational signatures[29] (S1, S2, S3, S4 and S5) in the cancer types analyzed (**Supplementary Figs. 4** and **5a**). In comparison to ccRCC where three signatures, S1, S3 and S4, were observed, the predominant mutational signature in pRCC consisted of S1 and S3 mutations (**Supplementary Fig. 4c**). S3 and S4 mutations were the major operative mutation signatures in chRCCs and renal oncocytomas (**Supplementary Fig. 4c**). In unsupervised hierarchical clustering of the mutation types at the sample level, the pRCCs predominantly clustered together, whereas the chRCCs and renal oncocytomas formed a separate group (**Supplementary Fig. 5**). Further, in clustering by mutational signatures at the level of cancer type, kidney cancer formed a distinct cluster, consistent with the shared mutation signatures among subtypes (**Supplementary Fig. 4d**).

Mutated genes and their significance
In pRCC, chRCC and renal oncocytoma, we identified somatic protein-altering mutations in 2,364, 781 and 509 genes, respectively. We assessed the impact of protein-altering single-nucleotide variants on gene function using SIFT[30], PolyPhen[31] and Condel[32], finding that 53% (1,140/2,164) of the mutations in pRCC, 50% (346/698) of the mutations in chRCC and 48% (211/441) of the mutations in renal oncocytoma were likely to result in functionally relevant alterations according to at least 2 of the 3 methods used for functional assessment (**Supplementary Table 3**). In contrast, in the normal samples, only 14% (204,537/1,501,072) of the protein-altering germline variants identified were predicted to result in altered function (**Supplementary Fig. 6**), suggesting an enrichment for pathogenic variants in the tumors. We further examined the data for enrichment of deleterious mutations by performing a simulation (Online Methods), finding that deleterious mutations, with the exception of those in renal oncocytoma ($P = 0.548$), were significantly enriched in pRCC ($P < 0.0001$) or trending toward significance in chRCC ($P = 0.075$), similar to ccRCC ($P = 0.086$; **Supplementary Fig. 7**).

We assessed the relevance of the mutated genes by applying a *q*-score metric[33] and ranking the significantly mutated nccRCC genes (**Fig. 2a–c**). In pRCC, we identified ten significantly mutated genes (*q* score ≥ 1; false discovery rate (FDR) ≤ 10%; **Supplementary Table 11a**) that included *MET*, *SLC5A3*, *NF2*, *PNKD*, *CPQ*, *LRP2*, *CHD3*, *SLC9A3R1*, *SETD2* and *CRTC1* (**Fig. 2a**, **Supplementary Fig. 8** and **Supplementary Table 11a**). We found *MET* mutations

Table 1 Sample summary

| Subtype	Number of samples	Exome			Targeted sequencing			RNA-seq			SNP array		
		Tumor	Normal	Paired (tumor + matched normal)	Tumor	Normal	Paired (tumor + matched normal)	Tumor	Normal	Paired (tumor + matched normal)	Tumor	Normal	Paired (tumor + matched normal)
Papillary	67	46	48	46	19	16	16	64	50	48	48	44	44
Chromophobe	49	49	47	47	–	–	–	46	33	32	26	27	26
Oncocytoma	35	34	33	33	–	–	–	35	29	29	24	24	23
Unclassified	8	8	8	8	–	–	–	7	6	5	6	6	6
Translocation	6	6	4	4	–	–	–	5	3	3	6	4	4
Sarcomatoid	2	2	2	2	–	–	–	2	2	2	2	2	2
Total	167	145	142	140	19	16	16	159	123	119	112	107	105
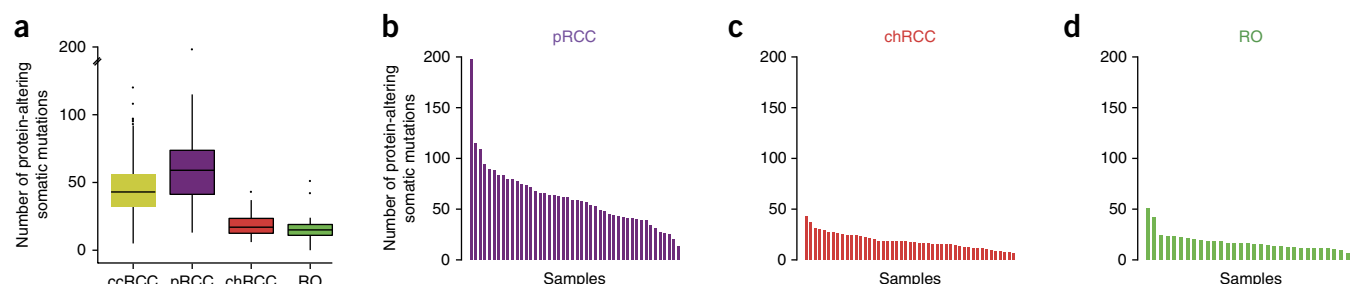
Figure 1 Somatic mutations in nccRCC. (**a**) Box plot of the number of protein-altering somatic mutations in each tumor subtype in comparison to ccRCC (TCGA data; ref. 4). The median value is shown as a line, with the whiskers extending from the highest value within 1.5 times the interquartile range of the third quartile to the lowest value within 1.5 times the interquartile range of the first quartile. Number of samples used: ccRCC, 417; pRCC, 46; chRCC, 47; renal oncocytoma (RO), 33. (**b**–**d**) Bar graphs showing the number of protein-altering somatic mutations observed in each sample for pRCC (**b**), chRCC (**c**) and renal oncocytoma (**d**).

in 15% (10/65) of the pRCC samples analyzed, in comparison to the previously reported rate of 8% (9/121) for mutations in this gene[23,24]. Except for a single mutation mapping to the extracellular domain (ECD), all the mutations affected the kinase domain of MET (**Fig. 2d**). Four previously unreported mutations of oncogenic relevance, including three kinase domain alterations (p.Val1088Ala, p.Ile1095Thr and the recurrent p.Phe1218Ile substitution) and an ECD alteration, p.Asp153Tyr, were identified (**Fig. 2d**). Mapping these alterations onto the crystal structures of the MET kinase domain and β chain (Protein Data Bank (PDB), 1R0P and 1SHY; **Fig. 2e** and **Supplementary Figs. 9** and **10**) suggested that they are potentially activating[34–36].

We tested the transforming ability of the MET mutants by stably expressing them in NIH3T3 cells and found that they promoted a significant increase in anchorage-independent growth when compared to ectopically expressed wild-type MET (**Fig. 2f** and **Supplementary Fig. 11a**). Consistent with this finding, all MET mutants showed elevated phosphorylation in comparison to wild-type MET, suggesting that they are constitutively active (**Fig. 2g** and **Supplementary Fig. 11b**). Cells expressing MET Asp153Tyr showed increased proliferation in comparison to those expressing wild-type MET at the concentrations of hepatocyte growth factor (HGF) ligand tested (**Supplementary Fig. 11c**). MET mutations in pRCC were mutually exclusive with mutations in *NF2*, *PTEN*, *TSC1* and *MTOR*, although they were not statistically significant given their low frequencies (**Supplementary Fig. 12**). *SLC5A3* (also known as *SMIT*), encoding a Na+/myoinositol cotransporter, was among the significantly mutated genes; its role in cancer is unknown. In this gene, we observed three loss-of-function frameshift mutations and a single point mutation that was predicted to be damaging (**Supplementary Fig. 8d**). We note that *Slc5a3* knockout mice show upregulation of genes involved in the mitochondrial electron transfer chain[37], indicating a role for *SLC5A3* mutations in mitochondrial alterations that might contribute to the development or maintenance of pRCC.

In chRCC, we found *TP53*, *PTEN*, *FAAH2*, *PDHB*, *PDXDC1* and *ZNF765* to be significantly mutated (**Fig. 2b** and **Supplementary Table 11a**). *PDHB* encodes the E1β subunit of the pyruvate dehydrogenase complex (PDHc) that catalyzes the conversion of pyruvate to acetyl-CoA[38]. This factor provides the primary link between glycolysis and the tricarboxylic acid (TCA) cycle. Human germline *PDHB* (E1β) mutations lead to lactic acidosis and heterogeneous neurological dysfunction[39]. Of the two mutations identified in *PDHB* in chRCC, one would lead to a truncated protein (p.Phe222fs*35) and the second would result in the substitution of arginine at codon 105 with leucine (p.Arg105Leu) (**Supplementary Fig. 13a**). Although the consequence of the p.Arg105Leu alteration needs functional

validation, we note that this substitution occurs at a highly conserved position (**Supplementary Fig. 13b**) that was also reported to be mutated (p.Arg105Gln) in an individual with lactic acidosis[40]. *PDHB* alterations, like the fumarate hydratase (*FH*) and succinate dehydrogenase[22] (*SDHB*, *SDHC* and *SDHD*) mutations known in kidney cancer[41], potentially allow the tumor to favor glycolysis over oxidative phosphorylation for energy production.

Among the nccRCC samples analyzed, *TP53* mutations were found to be significantly enriched in the chRCC classic subtype ($P = 2.3 \times 10^{-5}$). *ARID1A*, a known tumor-suppressor gene, and *PRKAG2* (encoding the AMP kinase (AMPK) γ subunit), although not achieving statistical significance, were recurrently mutated in chRCC. In addition, chRCC samples had mutations in *TSC1*, *TSC2* or *MTOR*, which might signal addiction to the mTORC1 pathway and responsiveness to mTORC1 inhibitors. In renal oncocytoma, the *q*-score analysis identified *ERCC2*, a nucleotide excision repair pathway gene, and *C2CD4C*, a C2 calcium-dependent domain–containing protein, as significantly mutated (**Fig. 2c**,**h** and **Supplementary Table 11a**). However, given the low mutation rates in renal oncocytoma, sequencing of a larger number of samples in combination with functional evaluation would be needed to fully ascertain the roles of the mutated genes.

PRKAG2 encodes one of the three γ subunits of AMPK (**Fig. 2i**,**j** and **Supplementary Fig. 14**), which functions as a key sensor of cellular metabolism[42]. The presence of an inhibitory pseudosubstrate sequence within the AMPK γ subunit was previously reported[43]. Mutations mapping to the pseudosubstrate sequence lead to constitutive activation of AMPK[43]. Among the mutations identified in *PRKAG2* in chRCC, the encoded p.Ile388Val substitution was present in the pseudosubstrate sequence within the second cystathionine-β synthase (CBS) motif (**Supplementary Fig. 14a**), indicating that the p.Ile388Val substitution might be activating.

A second mutation in *PRKAG2*, encoding p.Arg531Gln, was observed in two tumor samples. In a recent chRCC data set[44], we found a mutation encoding p.Arg299Gln in *PRKAG1* (**Supplementary Fig. 14c**). This mutation is analogous to the *PRKAG2* mutation encoding p.Arg531Gln previously reported in sporadic lethal congenital glycogen storage cardiomyopathy (GSC)[45]. Arg531 is a conserved residue located within non-exchangeable AMP-binding site 4 (CBS4) of AMPK that makes contact with AMP[46], and the p.Arg531Gln substitution is predicted to affect AMP binding (**Fig. 2j**). A previous study demonstrated elevated activity for this mutant, consistent with the dominant nature of the mutation in GSC[45]. A mouse transgenic model expressing the PRKAG2 Arg531Gly mutant showed a cardiac phenotype and glycogen accumulation in the heart, although no elevation in mutant AMPK activity was detected[47]. Previous studies have
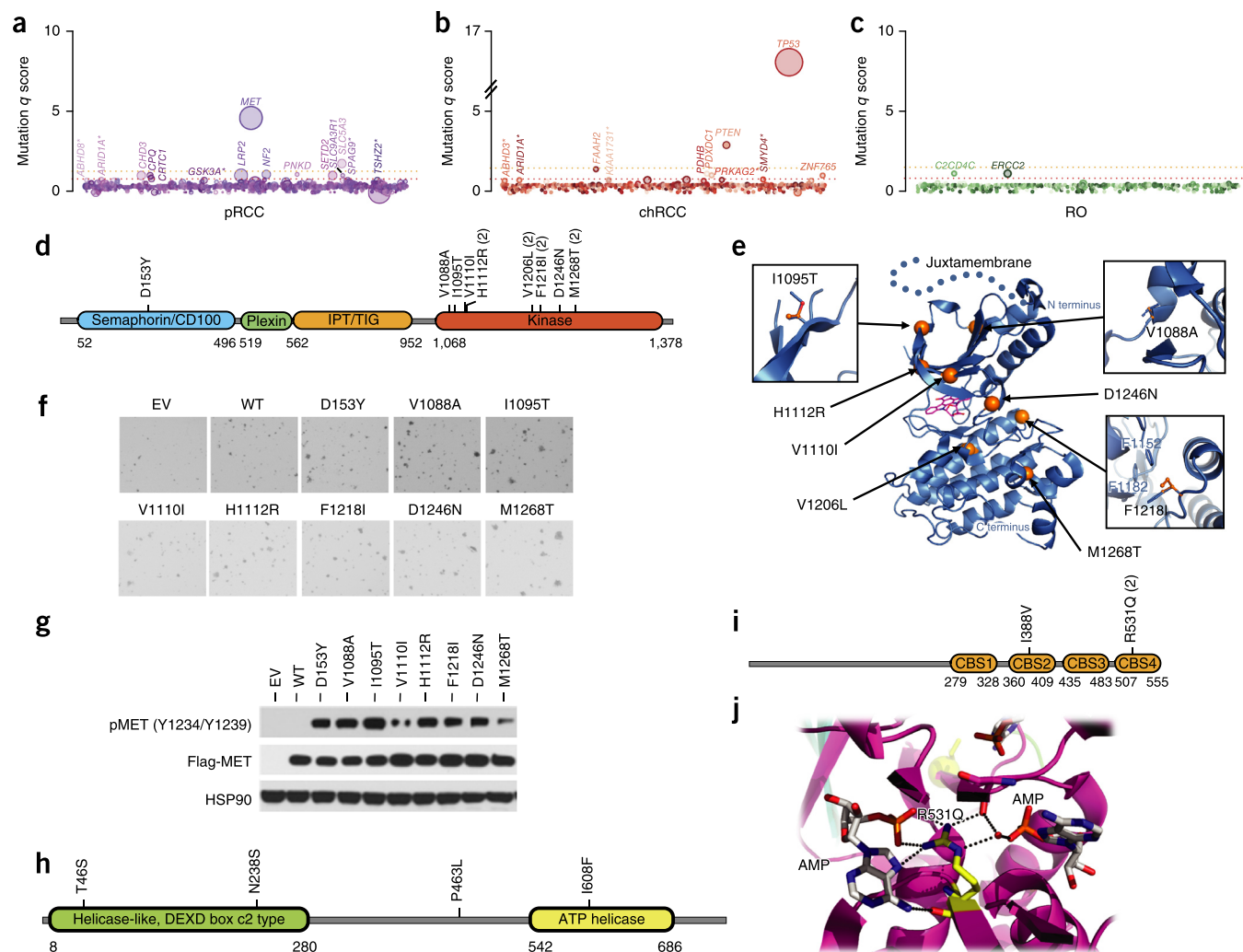
Figure 2 Significantly mutated genes in nccRCC. (**a–c**) Genes evaluated for significance on the basis of q score are shown for pRCC (**a**), chRCC (**b**) and renal oncocytoma (**c**). Each gene is represented as a circle, where the size of the circle is proportional to the observed mutation frequency. Genes are arranged from left to right along the x axis in alphabetical order. Genes with a significant q score appear above the dotted red line (FDR < 0.15). An asterisk next to a gene name indicates genes with a q score of 0.15 < FDR < 0.2. The dotted orange line represents FDR = 0.05. (**d**) Schematic depicting the MET alterations. (**e**) MET alterations mapped onto a crystal structure for the MET kinase domain (PDB, 1R0P). Blue ribbons represent the MET protein. Mutated residues are shown as orange spheres. K-252, an inhibitor, bound to the active site (PDB, 1R0P) is shown as magenta sticks. The dotted line extending beyond the N terminus of the kinase domain is not part of the crystal structure but is included to illustrate the likely location of the juxtamembrane domain. Accompanying panels show the positions of previously unreported alterations of oncogenic relevance. (**f**) Anchorage-independent colony growth of NIH3T3 cells stably expressing wild-type (WT) or mutant MET. EV, empty vector. (**g**) Protein blot analysis of the phosphorylation status of MET mutants stably expressed in NIH3T3 cells. pMET, phosphorylated MET. (**h**) Schematic depicting ERCC2 alterations. (**i,j**) PRKAG2 alterations depicted on Pfam domains (**i**) or the protein structure (PDB, 2V8Q) (**j**). Accession codes for the protein sequences used for the schematics can be found in **Supplementary Table 3**.

shown a link between glycogen storage disease and hepatocellular carcinoma[48–50]. Interestingly, kidney proximal tube–specific expression of a mutant hypoxia-inducible factor (HIF)-1α protein that is not targeted for degradation by pVHL (a frequently inactivated protein in ccRCC) promotes the accumulation of glycogen[51]. Although AMPK is primarily thought to have a tumor-suppressor role owing to the loss-of-function mutations identified in its upstream activator STK11 (also known as LKB1) in cancers, recent evidence suggest that, under energy-limiting conditions, AMPK can support cell survival through inhibition of the ACC1 and ACC2 enzymes involved in fatty acid biosynthesis and thereby help conserve nicotinamide adenine dinucleotide phosphate (NADPH)[52]. Consistent with the fact that kidney cancers have been associated with prominent metabolic alterations[19] and the role of AMPKs in energy sensing, the mutations

identified in PRKAG2 along with those in PDHB potentially identify a subtype of chRCC in which metabolic deregulation contributes to pathogenesis.

In addition to the mutational hotspots observed in TP53, MET and PRKAG2 (**Supplementary Table 9**), mutational meta-analysis using data from COSMIC[27] identified 78 additional hotspot mutations in 60 genes (**Supplementary Table 10**), which included drug targets such as IDH2, JAK2 and MTOR. We further assessed the significance of the mutated genes at the pathway level using Reactome[53,54] and found that, whereas the chRCC samples were enriched for pathways that involve TP53 and metabolism, the pRCC samples had statistically significant enrichment for pathways involving MET (semaphorin pathways) and glucose transport (FDR < 0.1; **Supplementary Table 11b**).
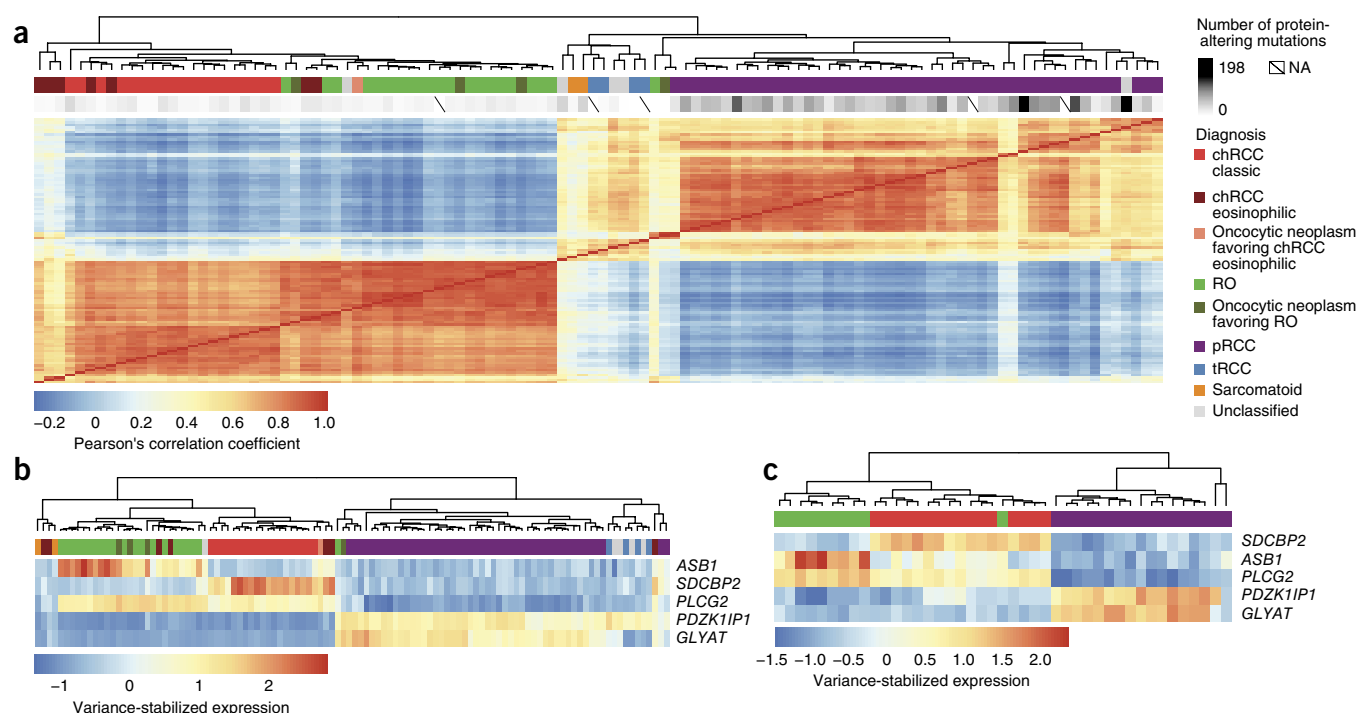
Figure 3 RNA-seq–based classification of nccRCCs. (**a**) Unsupervised clustering of the samples by sample correlation matrix using the 400 genes with the most variable expression. NA, not available. (**b**) Clustering of the samples on the basis of a minimal set of five differentially expressed genes. (**c**) Validation cohort clustering shown using the five-gene set.

Expression analysis identifies distinct nccRCC subtypes

Unsupervised clustering of the 400 genes with the most variable expression on the basis of RNA-seq analysis resulted in sample clusters that reflected the major histological groups. Among the different histological subtypes, the tRCC samples and a majority of the unclassified samples clustered with pRCC samples, indicating that these tumors might engage similar signaling pathways (**Fig. 3a** and **Supplementary Fig. 15**).

In routine clinical practice, distinguishing chRCC, a malignant neoplasm, from renal oncocytoma, a benign neoplasm, by histology alone can be challenging[6,55,56]. We found that renal oncocytoma shared more expression features with chRCC than pRCC. To further understand the subtypes, we performed pairwise differential expression analyses that involved the different tumor subtypes and also their matched normal samples (**Supplementary Table 12a–f**).

Analysis of differential expression in chRCC versus renal oncocytoma identified *ADAP1*, *SDCBP2*, *HOOK2*, *BAIAP3* and *SPINT1* as the top five genes with differential expression, all of which had a high level of expression in chRCC (2.5 to 5.3 log$_2$ fold increase in expression in comparison to renal oncocytoma; $P < 1.89 \times 10^{-51}$ to 1.007×10^{-101}). We also identified *ITGB3*, *MINOS1-NBL1* and *ASB1* as being upregulated in renal oncocytoma in comparison to chRCC (**Supplementary Table 12**). Also among our top differentially expressed genes were *AP1M2*, *TJP3*, *TMC4*, *CLDN7*, *MAL2*, *PROM2*, *KRT7*, *PRSS8* and *HOOK2*, which were previously reported in a differential expression analysis comparing chRCC and renal oncocytoma[57]. Although some of the genes identified in our analysis have been implicated in tumorigenesis, their exact role in cancer remains to be understood[58–62].

We tested the usefulness of RNA-seq data in classifying tumors into the nccRCC subtypes by using the 25 most differentially expressed genes from a 2-way comparison of each tumor subtype and identified a minimal set of 5 genes—*ASB1*, *GLYAT*, *PDZK1IP1*, *PLCG2* and *SDCBP2*—that were sufficient to separate the pRCC, chRCC and renal oncocytoma subtypes (**Fig. 3b** and **Supplementary Fig. 16**). We confirmed the usefulness of the 5 genes in nccRCC subtype stratification on a validation cohort consisting of 16 chRCC, 10 renal oncocytoma and 17 pRCC tumors. A leave-one-out cross-validation using a k nearest-neighbors classifier correctly classified 41 of the 43 samples (95.3%; **Fig. 3c**). We further confirmed the use of the 5 genes in nccRCC subtype stratification using an independent expression data set[63] consisting of 4 renal oncocytoma, 4 chRCC and 19 pRCC samples, finding that these genes clustered the nccRCC samples into groups that matched the histological subtypes (**Supplementary Fig. 17**).

Given the presence of *PRKAG2* mutations in chRCC and the potential for this gene as a drug target, we assessed the expression of *PRKAG2* in the different subtypes (**Supplementary Fig. 18**). *PRKAG2* was among the 10 most upregulated genes in renal oncocytoma (3.59 log$_2$ fold increase; $P < 2.03 \times 10^{-100}$) and chRCC (3.36 log$_2$ fold increase; $P < 5.29 \times 10^{-90}$) samples in comparison to pRCC samples (**Supplementary Table 12**). Further, we found that expression of *ASCL1* and *SCL25A5* was well correlated with expression of *PRKAG2* (**Supplementary Fig. 18**). *ASCL1* encodes an acyl-coenzyme A synthetase capable of increasing the AMP/ATP ratio[64] and thus potentially activating AMPK. *SLC25A5* is a nuclear-encoded gene that is responsible for mitochondrial ADP and ATP transport and is known to have a role in metabolism[65].

We also assessed copy number changes in the nccRCC samples using SNP arrays. The pRCCs showed frequent amplification of the entire chromosomes 3, 7, 12, 16, 17 and 20 (**Supplementary Fig. 19**), consistent with previous reports[66]. Although chRCCs (classic) showed frequent loss of chromosomes 1, 2, 6, 8, 10, 13, 17 and 21, the chRCC eosinophilic subtype appeared to be almost completely diploid. Renal oncocytomas showed very few copy number alterations, with chromosome 1 deletion being the most frequent (**Supplementary Fig. 19**). A pRCC sample (1216T) showed amplification of a 490-kb
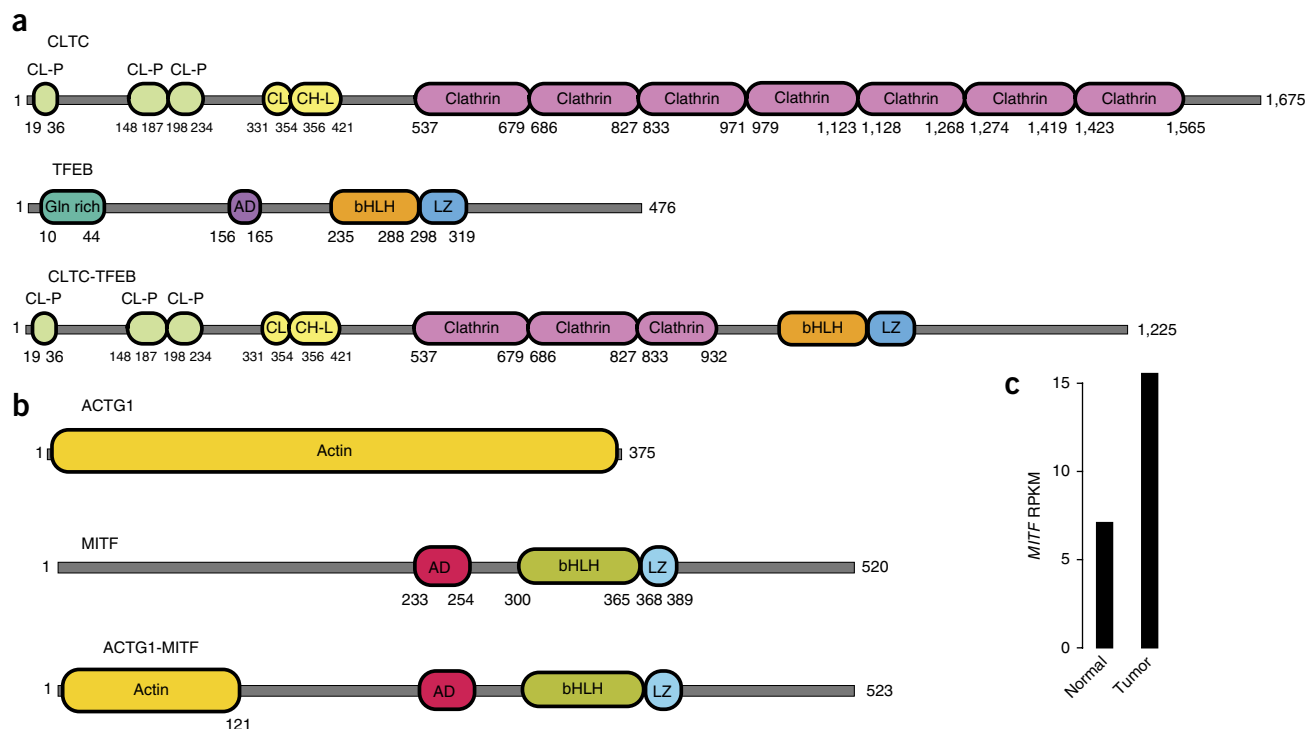
Figure 4 Proteins encoded by the *CLTC-TFEB* and *MITF* gene fusions. (**a**) Schematic of the CLTC-TFEB fusion protein resulting from the *CLTC-TFEB* fusion transcript. (**b**) Schematic of the ACTG1-MITF fusion protein predicted from the *ACTG1-MITF* fusion transcript. (**c**) *MITF* expression (RPKM, reads per kilobase of target per million mapped reads) in the tumor harboring the *MITF* fusion. CL-P, clathrin propel; CL, clathrin link; CH-L, clathrin H link; Gln rich, glutamine rich; AD, activation domain; bHLH, basic helix-loop-helix domain; L, leucine zipper. Accession codes for the proteins depicted: CLTC, NP_004850.1; TFEB, P19484.3 (NP_001258873); ACTG1, NP_001186883.1; MITF, NP_937802.1.

region on chromosome 6 that included *TFEB* (**Supplementary Fig. 20a**). We confirmed the amplification event using FISH analyses (**Supplementary Fig. 21a**). Consistent with amplification, this sample had the highest level of *TFEB* expression in comparison to all other samples (**Supplementary Fig. 20b**), indicating that, in addition to previously known *TFEB* translocations in tRCCs, amplification might be a cancer-relevant *TFEB* alteration. Given this finding, we reevaluated this sample and found characteristics of low-grade RCC with papillary and oncocytic features.

We observed that ~70% of the pRCC samples studied had amplification of the entire chromosome 7, and one pRCC sample (16864T) had amplification of a 22-Mb region within chromosome 7. The minimally amplified region within chromosome 7 contained *MET* (**Supplementary Fig. 20c**). Consistent with amplification, this sample showed the highest level of *MET* expression (**Supplementary Fig. 20d**), indicating that amplification, in addition to mutation, might be relevant for MET activation in pRCC pathogenesis. In general, we observed a higher level of *MET* expression in pRCC samples, in particular, in samples with chromosome 7 amplifications ($P = 1.78 \times 10^{-6}$; **Supplementary Fig. 20d**). Our data are consistent with previous observations of chromosome 7 amplification[67] and trisomy[68].

MITF gene fusion in nccRCC

The MiTF basic helix-loop-helix (bHLH) transcription factors *TFE3*, *TFEB*, *TFEC* and *MITF*[69] are deregulated in cancers, and translocations involving *TFE3* and *TFEB* are known in tRCC[70]. *TFEB* translocations are low-frequency events and are often missed in the clinic. We analyzed our RNA-seq data to evaluate its usefulness for the discovery of new and known fusions in tRCC (**Supplementary Table 13**). Of the samples classified as tRCC on the basis of morphology and TFE3

immunohistochemistry, we found evidence for the previously reported *ASPSCR1-TFE3* fusion[25,70,71] and *PRCC-TFE3* fusion[71]. We confirmed this result further using FISH analyses (**Supplementary Fig. 21b,c**). We did not detect fusion events involving *TFE3* (or other MiTF members) in two tRCC samples (14336T and PtS1T), even though they showed elevated *TFE3* expression at the RNA and protein (immunohistochemistry) levels (**Supplementary Fig. 22**). Further, we did not find evidence for *TFE3* amplification in these samples, suggesting the involvement of an alternate mechanism leading to upregulated expression of this gene. However, in one tRCC sample (14336T) lacking *TFE3* fusion, we identified a fusion involving *MIDN*, encoding midnolin, a nucleolar protein, and *SBNO2*, encoding strawberry notch homolog 2, a DEXD/H helicase family corepressor. *MIDN* and *SBNO2* are located on opposite strands of chromosome 19p13.3 (**Supplementary Fig. 23**). The observed fusion is likely due to a genomic inversion at 19p13.3 that, when transcribed and spliced, places the noncoding exon of *MIDN* at the 5′ end of the second exon of *SBNO2*. This event results in a transcript corresponding to full-length *SBNO2* under the control of the *MIDN* promoter (**Supplementary Fig. 23**). The sample with the *MIDN-SBNO2* fusion had the second highest level of *SBNO2* expression (**Supplementary Fig. 24**). Further, the second tRCC sample (PtS1T) lacking *TFE3* fusion had the highest level of *SBNO2* expression, although the exact mechanism leading to upregulation of *SBNO2* remains to be determined (**Supplementary Fig. 24**). Recently, SBNO2 has been shown to have a critical role in bone homeostasis through the activation of MITF[72]. However, whether SBNO2 can modulate the levels and transcriptional activity of other MiTF members, including TFE3, requires further investigation.

We found an unreported gene fusion involving *CLTC* and *TFEB* (*CLTC-TFEB*) in an nccRCC sample (8432T; **Fig. 4a** and

Figure 5 *ACTG1-MITF* gene fusion promotes anchorage-independent growth. (**a**) Expression of MITF target genes in HEK293T cells expressing wild-type MITF or the ACTG1-MITF fusion protein. The values shown are from three biological replicates (error bars, s.e.m.; **P < 0.01, ***P < 0.001, two-tailed Student's t test). (**b**) Stability of the MITF fusion protein over time in HEK293T cells transfected with the indicated constructs after cycloheximide treatment, assessed using protein blotting. Cells at time 0 h were not treated with cycloheximide. (**c**) Protein blot showing the expression of Flag-tagged ACTG1, MITF and ACTG1-MITF fusion proteins in NIH3T3 cell expressing the indicated constructs. HSP90 was used as a loading control. (**d**) Representative images depicting colony formation by NIH3T3 cells stably expressing the indicated constructs. (**e**) Quantification of the number of colonies (>300 μm in diameter) shown in **d**. Data shown are mean values ±s.e.m. from three biological replicates (***P < 0.0001, two-tailed Student's t test).
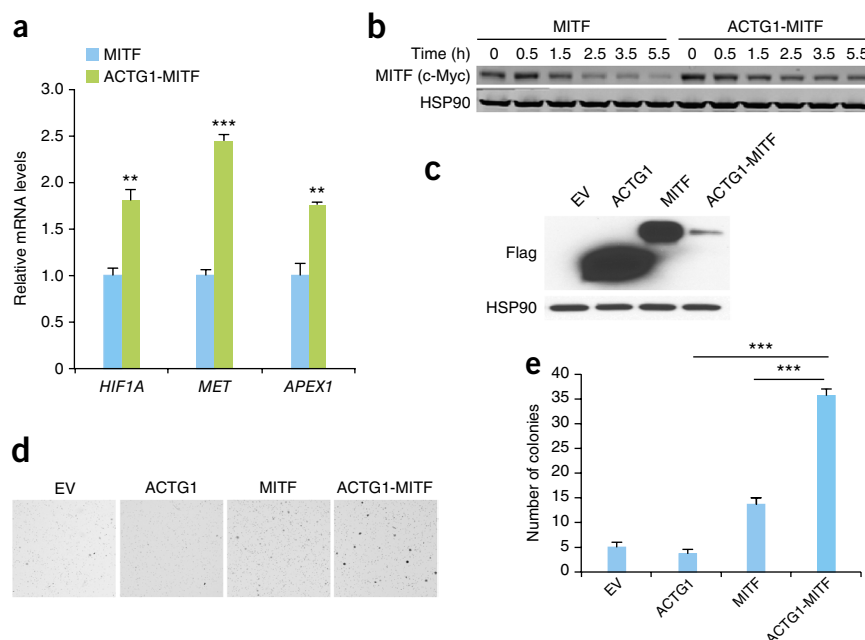
Supplementary Fig. 25a) that was designated as unclassified. We confirmed *TFEB* translocation in this sample using FISH analyses (**Supplementary Fig. 21d**). The *CLTC-TFEB* gene fusion encodes an in-frame fusion protein containing the bHLH domain of TFEB, as observed in other known TFEB fusions[70], indicating that the resulting protein is likely functional. This unclassified sample was found to cluster closely with the tRCC and sarcomatoid subtypes (**Fig. 3a**). Similarly, we found *PRCC-TFE3* fusion in another unclassified sample (20825T1). On the basis of the presence of fusions, both 8432T and 20825T1 were reclassified as tRCCs after additional pathological review.

We identified an *ACTG1-MITF* gene fusion (**Fig. 4b**) in a pRCC sample (159T; this case also exhibited oncocytic and papillary features). Although ccRCC-predisposing germline mutations in *MITF* have been discovered, to our knowledge, no gene fusions involving *MITF* in nccRCC have been reported[70]. We validated the *ACTG1-MITF* fusion and confirmed that it was somatic (**Supplementary Fig. 25b**). The fusion protein encoded by *ACTG1-MITF* is about the same size as wild-type MITF, with the first 118 amino acids of MITF replaced by the N-terminal 121 amino acids of ACTG1 (**Fig. 4b**). We found that

the tumor expressing the fusion had a higher level of *MITF* expression than the matched normal sample (**Fig. 4c**).

To further characterize the ACTG1-MITF fusion protein, we transfected a cDNA encoding the fusion protein into HEK293T cells and tested the expression of known MITF target genes[73,74]. Expression of the ACTG1-MITF fusion resulted in a significant induction in *HIF1A*, *MET* and *APEX1* transcript levels in comparison to wild-type MITF (*P* < 0.01; **Fig. 5a** and **Supplementary Fig. 26**). Given that RCC-predisposing mutation in *MITF* is thought to function by increasing its stability[75], we assessed the stability of the ACTG1-MITF fusion protein by analyzing its turnover in cells. The ACTG1-MITF protein was more stable than wild-type MITF (**Fig. 5b**). Further, we tested the transforming ability of the ACTG1-MITF fusion by stably expressing it or wild-type MITF in NIH3T3 cells and assessing anchorage-independent growth (**Fig. 5c–e**). Cells expressing ACTG1-MITF had a significantly higher number of anchorage-independent colonies than cells expressing wild-type protein (**Fig. 5d,e**). Taken together, these data suggest that the MITF fusion, like the TFE3 and TFEB fusions, can contribute to tumorigenesis in nccRCC.

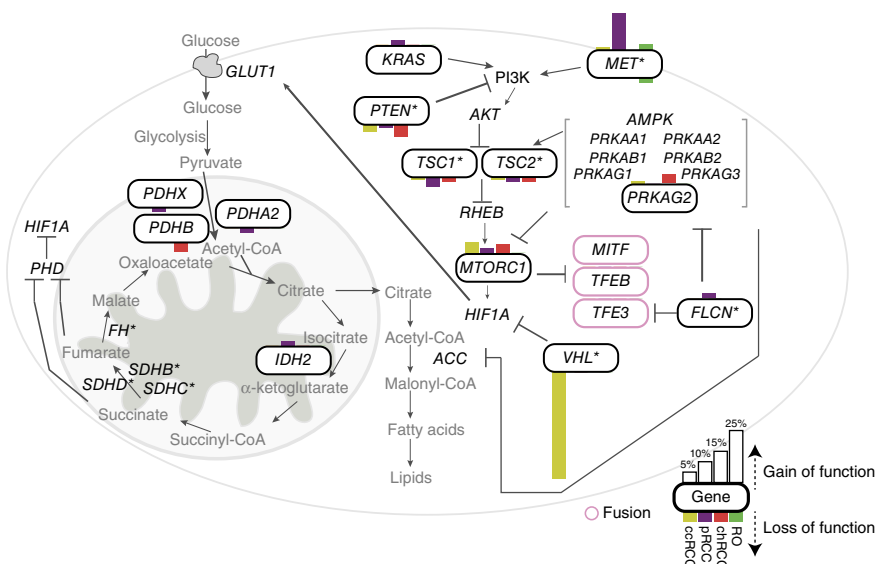The MiTF proteins homo- or heterodimerize with other family members in various

Figure 6 Integrated analysis of alterations in key pathways in nccRCC subtypes. Shown are mutated genes and their mutational frequencies in the indicated RCC subtypes for MET signaling and metabolism pathways. Mutated genes in the pathway are shown inside a curved-edge rectangle. Bars on top of the curved-edge rectangle indicate mutations that are known or predicted to be activating, and bars on the bottom edge of the curved-edge rectangle indicate mutations known or predicted to be inactivating. The length of each bar is proportional to the frequency of the mutations observed by subtype in this study or the TCGA study (for ccRCC; ref. 4). An asterisk indicates genes with known RCC risk alleles. Enclosed in pink curved-edge rectangles are genes involved in fusions identified in this study.

combinations and bind similar DNA elements to modulate gene expression[5,76]. Thus, we assessed the samples with MiTF fusion or amplification for genes upregulated in common between them that could serve as drug targets. We found that a majority (6/7) of the samples with MiTF fusion or amplification had elevated *BIRC7* expression (**Supplementary Fig. 27**). *BIRC7*, encoding an anti-apoptotic protein, is an MITF target gene whose expression is known to be upregulated in several cancers[73,74]. Our data suggest that *BIRC7* expression might aid in the diagnosis of tRCC and other subtypes that overexpress MiTF family members. Small molecule BIRC7 inhibitors that sensitize cancer cells to apoptosis are in clinical development[73] and might prove effective in treating tumors positive for MiTF fusion or overexpressing MiTF, which currently remain intractable.

DISCUSSION

In this study, we have performed a comprehensive genomic analysis of human primary nccRCC samples using next-generation sequencing technologies. We find that pRCCs have a higher mutation rate than chRCCs and renal oncocytomas. Whereas *VHL*, *TCEB1*, *PTEN*, *PBRM1*, *SETD2*, *BAP1*, *KDM5C*, *MTOR*, *PIK3CA* and *TP53* were identified as frequently altered genes in ccRCC, the genes altered in nccRCC are distinct, although mutations were also found in *SETD2*, *PTEN*, *MTOR*, *PBRM1* and *TP53* and vary within each subtype[3,4,77].

Integrated analysis of altered genes that impinge on pathways involving MET and glucose metabolism (**Fig. 6** and **Supplementary Fig. 28**) showed MET deregulation in pRCC, chRCC and renal oncocytoma (**Fig. 6** and **Supplementary Fig. 29**). In particular, in pRCC, a majority of the tumors carried *MET* alterations that included mutation, amplification and/or overexpression (**Supplementary Figs. 28** and **29**). Given these observations, it would be relevant to stratify patients on the basis of *MET* alterations in trials involving MET inhibitors, as patients with pRCC who have germline *MET* mutations were responsive to a broad-spectrum tyrosine kinase inhibitor[67]. Analysis of chRCC identified frequent *TP53* mutations that were exclusive to the chRCC classic type, building on previous findings[44,78,79].

Using differentially expressed genes from RNA-seq data, we identified a five-gene set that can be used to stratify nccRCC tumors in the clinic, although further prospective validation would be needed. Further, we demonstrate the usefulness of RNA-seq data for the detection of known and new MiTF gene fusions. Our data suggest that the tRCC subtype might need to be broadened to incorporate tumors with *MITF* translocations, as well as possibly other modes of activation of MiTF family members, including direct activation through amplification, indirect activation through fusions involving genes such as *SBNO2* or yet to be defined mechanisms in fusion-negative *TFE3*-overexpressing samples. This suggests that tumors with MiTF gene fusions, encompassing the tRCCs, and/or MiTF gene (*MITF*, *TFE3* or *TFEB*) overexpression likely form a distinct nccRCC 'MiTF-high' subtype. Overall, these data emphasize the importance of integrative histogenomics in the diagnosis of samples with MiTF family alterations. Our efforts to identify drug targets in the MiTF-high subtype showed that a majority of these tumors express the anti-apoptotic protein BIRC7 and might be candidates for therapy involving an apoptosis-sensitizing BIRC7 inhibitor.

URLs. Ensembl, http://www.ensembl.org/; The Cancer Genome Atlas (TCGA), http://cancergenome.nih.gov/.

METHODS

Methods and any associated references are available in the online version of the paper.

Accession codes. Sequencing and genotype data for patients specifically consenting to have their genomic data in a public database were deposited in the European Genome-phenome Archive (EGA), which is hosted by the European Bioinformatics Institute (EBI), under accession EGAS00001000926.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

AUTHOR CONTRIBUTIONS

S.D. and E.W.S. performed the exome and RNA-seq analyses. T.T.N. performed the simulation analysis. E.W.S. and J.S.G. conducted the mutational signature studies. S.D. and P.M.H. performed copy number analysis. A.P.-J., V.T.-T., E.H., Z.M., H.M.H. and S.P.-L. were responsible for sample collection, annotation, processing, and DNA and RNA extraction. Z.M. oversaw the collection of the various data types. Z.M. and Y.-J.C. performed validation of the fusions. P.K. facilitated sample procurement, selected samples for DNA and RNA extraction, oversaw FISH analyses and served as the pathologist for the study. J.R. and G.P. processed the RNA-seq reads. T.D.W. provided support for gene fusion prediction. K.T., C.H., C.J.H. and C.S.R. prepared the sequencing libraries. N.Z., K.B.P., S.C. and B.S.J. performed biological validation studies. S. Saleem collected information about the cases. V.M., Y.L., A.S. and I.P. facilitated sample procurement. L.N.K. and N.V.G. performed *in silico* analyses. J.G., V.J. and J.S. collected sequencing data. J.G. performed mutation validation. B.C. analyzed targeted capture data. S.H. and M.J. performed Sanger sequencing to validate indels. W.W. predicted the structural consequences of *MET* mutations. F.J.d.S. provided organizational support. J.B. and S. Seshagiri conceived the study and designed the experiments. E.W.S., S.D., Z.M., P.K., J.B. and S. Seshagiri wrote the manuscript, which was reviewed and edited by the other coauthors.

COMPETING FINANCIAL INTERESTS

The authors declare competing financial interests: details are available in the online version of the paper.

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Siegel, R., Naishadham, D. & Jemal, A. Cancer statistics, 2013. *CA Cancer J. Clin.* **63**, 11–30 (2013).
2. Peña-Llopis, S. *et al*. BAP1 loss defines a new class of renal cell carcinoma. *Nat. Genet.* **44**, 751–759 (2012).
3. Sato, Y. *et al*. Integrated molecular analysis of clear-cell renal cell carcinoma. *Nat. Genet.* **45**, 860–867 (2013).
4. Cancer Genome Atlas Research Network. Comprehensive molecular characterization of clear cell renal cell carcinoma. *Nature* **499**, 43–49 (2013).
5. Srigley, J.R. *et al*. The International Society of Urological Pathology (ISUP) Vancouver Classification of Renal Neoplasia. *Am. J. Surg. Pathol.* **37**, 1469–1489 (2013).
6. Yusenko, M.V. Molecular pathology of renal oncocytoma: a review. *Int. J. Urol.* **17**, 602–612 (2010).
7. Amin, M.B. *et al*. Chromophobe renal cell carcinoma: histomorphologic characteristics and evaluation of conventional pathologic prognostic parameters in 145 cases. *Am. J. Surg. Pathol.* **32**, 1822–1834 (2008).
8. Eble, J.N., Sauter, G., Epstein, J.I. & Sesterhenn, I.A. *Pathology and Genetics of Tumours of the Urinary System and Male Genital Organs* (IARCPress, Lyon, France, 2004).
9. Schieda, N., Al-Subhi, M., Flood, T.A., El-Khodary, M. & McInnes, M.D. Diagnostic accuracy of segmental enhancement inversion for the diagnosis of renal oncocytoma using biphasic computed tomography (CT) and multiphase contrast-enhanced magnetic resonance imaging (MRI). *Eur. Radiol.* **24**, 2787–2794 (2014).
10. Vargas, H.A. *et al*. Renal cortical tumors: use of multiphasic contrast-enhanced MR imaging to differentiate benign and malignant histologic subtypes. *Radiology* **264**, 779–788 (2012).
11. Rosenkrantz, A.B. *et al*. MRI features of renal oncocytoma and chromophobe renal cell carcinoma. *AJR Am. J. Roentgenol.* **195**, W421–W427 (2010).

12. Young, A.N. Editorial comment from Dr Young to Molecular pathology of renal oncocytoma: a review. *Int. J. Urol.* **17**, 612–613 (2010).
13. Picken, M.M. Editorial comment from Dr Picken to Molecular pathology of renal oncocytoma: a review. *Int. J. Urol.* **17**, 613–614 (2010).
14. Yusenko, M.V. Molecular pathology of chromophobe renal cell carcinoma: a review. *Int. J. Urol.* **17**, 592–600 (2010).
15. Osunkoya, A.O. Editorial comment to Molecular pathology of chromophobe renal cell carcinoma: a review. *Int. J. Urol.* **17**, 600–601 (2010).
16. Macher-Goeppinger, S. *et al.* Molecular heterogeneity of TFE3 activation in renal cell carcinomas. *Mod. Pathol.* **25**, 308–315 (2012).
17. Bellmunt, J. & Dutcher, J. Targeted therapies and the treatment of non–clear cell renal cell carcinoma. *Ann. Oncol.* **24**, 1730–1740 (2013).
18. Hagenkord, J.M., Gatalica, Z., Jonasch, E. & Monzon, F.A. Clinical genomics of renal epithelial tumors. *Cancer Genet.* **204**, 285–297 (2011).
19. Linehan, W.M. & Ricketts, C.J. The metabolic basis of kidney cancer. *Semin. Cancer Biol.* **23**, 46–55 (2013).
20. Popova, T. *et al.* Germline *BAP1* mutations predispose to renal cell carcinomas. *Am. J. Hum. Genet.* **92**, 974–980 (2013).
21. Farley, M.N. *et al.* A novel germline mutation in *BAP1* predisposes to familial clear-cell renal cell carcinoma. *Mol. Cancer Res.* **11**, 1061–1071 (2013).
22. Ricketts, C.J. *et al.* Succinate dehydrogenase kidney cancer: an aggressive example of the Warburg effect in cancer. *J. Urol.* **188**, 2063–2071 (2012).
23. Schmidt, L. *et al.* Germline and somatic mutations in the tyrosine kinase domain of the *MET* proto-oncogene in papillary renal carcinomas. *Nat. Genet.* **16**, 68–73 (1997).
24. Schmidt, L. *et al.* Novel mutations of the *MET* proto-oncogene in papillary renal carcinomas. *Oncogene* **18**, 2343–2350 (1999).
25. Argani, P. *et al.* Xp11 translocation renal cell carcinoma in adults: expanded clinical, pathologic, and genetic spectrum. *Am. J. Surg. Pathol.* **31**, 1149–1160 (2007).
26. Peña-Llopis, S. & Brugarolas, J. Simultaneous isolation of high-quality DNA, RNA, miRNA and proteins from tissues for genomic applications. *Nat. Protoc.* **8**, 2240–2255 (2013).
27. Forbes, S.A. *et al.* COSMIC (the Catalogue of Somatic Mutations in Cancer): a resource to investigate acquired mutations in human cancer. *Nucleic Acids Res.* **38**, D652–D657 (2010).
28. Nik-Zainal, S. *et al.* Mutational processes molding the genomes of 21 breast cancers. *Cell* **149**, 979–993 (2012).
29. Pfeifer, G.P. Mutagenesis at methylated CpG sequences. *Curr. Top. Microbiol. Immunol.* **301**, 259–281 (2006).
30. Ng, P.C. & Henikoff, S. Accounting for human polymorphisms predicted to affect protein function. *Genome Res.* **12**, 436–446 (2002).
31. Ramensky, V., Bork, P. & Sunyaev, S. Human non-synonymous SNPs: server and survey. *Nucleic Acids Res.* **30**, 3894–3900 (2002).
32. González-Pérez, A. & López-Bigas, N. Improving the assessment of the outcome of nonsynonymous SNVs with a consensus deleteriousness score, Condel. *Am. J. Hum. Genet.* **88**, 440–449 (2011).
33. Kan, Z. *et al.* Diverse somatic mutation patterns and pathway alterations in human cancers. *Nature* **466**, 869–873 (2010).
34. Wang, W. *et al.* Structural characterization of autoinhibited c-Met kinase produced by coexpression in bacteria with phosphatase. *Proc. Natl. Acad. Sci. USA* **103**, 3563–3568 (2006).
35. Miller, M. *et al.* Structural basis of oncogenic activation caused by point mutations in the kinase domain of the *MET* proto-oncogene: modeling studies. *Proteins* **44**, 32–43 (2001).
36. Gandino, L., Longati, P., Medico, E., Prat, M. & Comoglio, P.M. Phosphorylation of serine 985 negatively regulates the hepatocyte growth factor receptor kinase. *J. Biol. Chem.* **269**, 1815–1820 (1994).
37. Toker, L. *et al.* Inositol-related gene knockouts mimic lithium's effect on mitochondrial function. *Neuropsychopharmacology* **39**, 319–328 (2014).
38. Patel, M.S., Nemeria, N.S., Furey, W. & Jordan, F. The pyruvate dehydrogenase complexes: structure-based function and regulation. *J. Biol. Chem.* **289**, 16615–16623 (2014).
39. Imbard, A. *et al.* Molecular characterization of 82 patients with pyruvate dehydrogenase complex deficiency. Structural implications of novel amino acid substitutions in E1 protein. *Mol. Genet. Metab.* **104**, 507–516 (2011).
40. Quintana, E. *et al.* PDH E1β deficiency with novel mutations in two patients with Leigh syndrome. *J. Inherit. Metab. Dis.* **32**, 339–343 (2009).
41. Linehan, W.M., Srinivasan, R. & Schmidt, L.S. The genetic basis of kidney cancer: a metabolic disease. *Nat. Rev. Urol.* **7**, 277–285 (2010).
42. Steinberg, G.R. & Kemp, B.E. AMPK in health and disease. *Physiol. Rev.* **89**, 1025–1078 (2009).
43. Scott, J.W., Ross, F.A., Liu, J.K. & Hardie, D.G. Regulation of AMP-activated protein kinase by a pseudosubstrate sequence on the γ subunit. *EMBO J.* **26**, 806–815 (2007).
44. Davis, C.F. *et al.* The somatic genomic landscape of chromophobe renal cell carcinoma. *Cancer Cell* **26**, 319–330 (2014).
45. Burwinkel, B. *et al.* Fatal congenital heart glycogenosis caused by a recurrent activating R531Q mutation in the γ2-subunit of AMP-activated protein kinase (*PRKAG2*), not by phosphorylase kinase deficiency. *Am. J. Hum. Genet.* **76**, 1034–1049 (2005).
46. Carling, D., Mayer, F.V., Sanders, M.J. & Gamblin, S.J. AMP-activated protein kinase: nature's energy sensor. *Nat. Chem. Biol.* **7**, 512–518 (2011).

47. Davies, J.K. *et al.* Characterization of the role of γ2 R531G mutation in AMP-activated protein kinase in cardiac hypertrophy and Wolff-Parkinson-White syndrome. *Am. J. Physiol. Heart Circ. Physiol.* **290**, H1942–H1951 (2006).
48. de Moor, R.A. *et al.* Hepatocellular carcinoma in glycogen storage disease type IV. *Arch. Dis. Child.* **82**, 479–480 (2000).
49. Manzia, T.M. *et al.* Glycogen storage disease type Ia and VI associated with hepatocellular carcinoma: two case reports. *Transplant. Proc.* **43**, 1181–1183 (2011).
50. Calderaro, J. *et al.* Molecular characterization of hepatocellular adenomas developed in patients with glycogen storage disease type I. *J. Hepatol.* **58**, 350–357 (2013).
51. Fu, L., Wang, G., Shevchuk, M.M., Nanus, D.M. & Gudas, L.J. Generation of a mouse model of von Hippel–Lindau kidney disease leading to renal cancers by expression of a constitutively active mutant of HIF1. *Cancer Res.* **71**, 6848–6856 (2011).
52. Jeon, S.M., Chandel, N.S. & Hay, N. AMPK regulates NADPH homeostasis to promote tumour cell survival during energy stress. *Nature* **485**, 661–665 (2012).
53. Milacic, M. *et al.* Annotating cancer variants and anti-cancer therapeutics in reactome. *Cancers (Basel)* **4**, 1180–1211 (2012).
54. Croft, D. *et al.* Reactome: a database of reactions, pathways and biological processes. *Nucleic Acids Res.* **39**, D691–D697 (2011).
55. Nagashima, Y., Kuroda, N. & Yao, M. Transition of organizational category on renal cancer. *Jpn. J. Clin. Oncol.* **43**, 233–242 (2013).
56. Beleut, M. *et al.* Integrative genome-wide expression profiling identifies three distinct molecular subgroups of renal cell carcinoma with different patient outcome. *BMC Cancer* **12**, 310 (2012).
57. Rohan, S. *et al.* Gene expression profiling separates chromophobe renal cell carcinoma from oncocytoma and identifies vesicular transport and cell junction proteins as differentially expressed genes. *Clin. Cancer Res.* **12**, 6937–6945 (2006).
58. Venkateswarlu, K. & Cullen, P.J. Molecular cloning and functional characterization of a human homologue of centaurin-α. *Biochem. Biophys. Res. Commun.* **262**, 237–244 (1999).
59. Zimmermann, P. The prevalence and significance of PDZ domain–phosphoinositide interactions. *Biochim. Biophys. Acta* **1761**, 947–956 (2006).
60. Weimer, J.M., Chattopadhyay, S., Custer, A.W. & Pearce, D.A. Elevation of Hook1 in a disease model of Batten disease does not affect a novel interaction between ankyrin G and Hook1. *Biochem. Biophys. Res. Commun.* **330**, 1176–1181 (2005).
61. Palmer, R.E. *et al.* Induction of BAIAP3 by the EWS-WT1 chimeric fusion implicates regulated exocytosis in tumorigenesis. *Cancer Cell* **2**, 497–505 (2002).
62. Cheng, H., Fukushima, T., Takahashi, N., Tanaka, H. & Kataoka, H. Hepatocyte growth factor activator inhibitor type 1 regulates epithelial to mesenchymal transition through membrane-bound serine proteinases. *Cancer Res.* **69**, 1828–1835 (2009).
63. Yusenko, M.V. *et al.* High-resolution DNA copy number and gene expression analyses distinguish chromophobe renal cell carcinomas and renal oncocytomas. *BMC Cancer* **9**, 152 (2009).
64. Righi, L., Rapa, I., Votta, A., Papotti, M. & Sapino, A. Human achaete-scute homolog-1 expression in neuroendocrine breast carcinoma. *Virchows Arch.* **460**, 415–421 (2012).
65. Palmieri, F. The mitochondrial transporter family SLC25: identification, properties and physiopathology. *Mol. Aspects Med.* **34**, 465–484 (2013).
66. Monzon, F.A. *et al.* Whole genome SNP arrays as a potential diagnostic tool for the detection of characteristic chromosomal aberrations in renal epithelial tumors. *Mod. Pathol.* **21**, 599–608 (2008).
67. Choueiri, T.K. *et al.* Phase II and biomarker study of the dual MET/VEGFR2 inhibitor foretinib in patients with papillary renal cell carcinoma. *J. Clin. Oncol.* **31**, 181–186 (2013).
68. Lager, D.J., Huston, B.J., Timmerman, T.G. & Bonsib, S.M. Papillary renal tumors. Morphologic, cytochemical, and genotypic features. *Cancer* **76**, 669–673 (1995).
69. Steingrímsson, E., Copeland, N.G. & Jenkins, N.A. Melanocytes and the microphthalmia transcription factor network. *Annu. Rev. Genet.* **38**, 365–411 (2004).
70. Haq, R. & Fisher, D.E. Biology and clinical relevance of the micropthalmia family of transcription factors in human cancer. *J. Clin. Oncol.* **29**, 3474–3482 (2011).
71. Malouf, G.G. *et al.* Next-generation sequencing of translocation renal cell carcinoma reveals novel RNA splicing partners and frequent mutations of chromatin-remodeling genes. *Clin. Cancer Res.* **20**, 4129–4140 (2014).
72. Maruyama, K. *et al.* Strawberry notch homologue 2 regulates osteoclast fusion by enhancing the expression of DC-STAMP. *J. Exp. Med.* **210**, 1947–1960 (2013).
73. Dynek, J.N. *et al.* Microphthalmia-associated transcription factor is a critical transcriptional regulator of melanoma inhibitor of apoptosis in melanomas. *Cancer Res.* **68**, 3124–3132 (2008).
74. Hoek, K.S. *et al.* Novel MITF targets identified using a two-step DNA microarray strategy. *Pigment Cell Melanoma Res.* **21**, 665–676 (2008).
75. Yokoyama, S. *et al.* A novel recurrent mutation in *MITF* predisposes to familial and sporadic melanoma. *Nature* **480**, 99–103 (2011).
76. Hemesath, T.J. *et al.* Microphthalmia, a critical factor in melanocyte development, defines a discrete transcription factor family. *Genes Dev.* **8**, 2770–2780 (1994).
77. Hakimi, A.A., Pham, C.G. & Hsieh, J.J. A clear picture of renal cell carcinoma. *Nat. Genet.* **45**, 849–850 (2013).
78. Gad, S. *et al.* Mutations in *BHD* and *TP53* genes, but not in *HNF1β* gene, in a large series of sporadic chromophobe renal cell carcinoma. *Br. J. Cancer* **96**, 336–340 (2007).
79. Contractor, H., Zariwala, M., Bugert, P., Zeisler, J. & Kovacs, G. Mutation of the p53 tumour suppressor gene occurs preferentially in the chromophobe type of renal cell tumour. *J. Pathol.* **181**, 136–139 (1997).

ONLINE METHODS

Samples, DNA and RNA preps. We analyzed 167 human primary nccRCC samples and corresponding normal tissue in this study. The nccRCC samples included 67 pRCCs, 49 chRCCs, 35 renal oncocytomas, 8 unclassified type carcinomas, 6 tRCCs and 2 samples with sarcomatoid dedifferentiation (**Supplementary Table 1**).

Samples used in the study were typically obtained from patients undergoing surgery for a renal mass at hospitals affiliated with University of Texas Southwestern: Saint Paul University Hospital, Parkland Memorial Hospital and Zale Lipshy University Hospital. These hospitals include tertiary care referral centers (Saint Paul and Zale Lipshy) as well as a county hospital (Parkland Memorial) and serve a wide variety of patients of multiple ancestry groups, including European-descent, Hispanic, African-American, Asian and South Asian individuals. Patients were excluded if they were known to have HIV, hepatitis B virus (HBV), hepatitis C virus (HCV) or tuberculosis infections.

This study was conducted with institutional review board (IRB) approval and written patient informed consent. The protocol allows for the collection of discarded tissue samples as well as blood.

Patients were asked to specifically consent to placement of their genomic information in a protected publicly accessible database, and deidentified genomic information has been deposited for those subjects explicitly allowing it on the consent form.

Human tissue samples were deidentified before their shipment and analysis and are not considered human subject research under US Department of Health and Human Services regulations and related guidance (45 CFR Part 46). Tumor and adjacent normal kidney samples were frozen fresh in liquid nitrogen and stored at −80 °C. Perpendicular sections immediately flanking fragments of 1–3 mm in thickness were reviewed by a pathologist (P.K.) for all frozen tumor and normal tissues to confirm the diagnosis and tumor content[26]. Basic demographic information for the patient samples in the study, where available, is included in **Supplementary Table 1**. Tissue processing as well as simultaneous extraction of high-quality genomic DNA and total RNA from the same samples was performed as previously described[26].

Exome capture and sequencing. We analyzed the exomes of 140 nccRCC samples and their patient-matched normal samples to assess the mutational burden. In addition, we also obtained exome data for an additional five nccRCC samples that lacked a matched normal sample (**Table 1** and **Supplementary Table 1**). Exome capture was performed using the Agilent SureSelect Human All Exon kit (50Mb). Exome capture libraries were sequenced on the HiSeq 2000 platform (Illumina) to generate 2 × 75-bp paired-end data. A targeted mean coverage of 82× with 94% of bases covered at ≥10× was achieved for exome libraries (**Supplementary Fig. 2** and **Supplementary Table 2**).

Target enrichment and sequencing. Targeted sequencing was performed for 19 pRCC tumor samples (including 16 tumor-normal pairs) using NuGEN's proprietary Single Primer Enrichment Technology (SPET). SPET was performed using the Ovation Cancer Panel kit (344 genes; 9079-32) and some additional content incorporated through NuGEN's custom design service (**Supplementary Table 6a**). Briefly, fragmented DNA was end repaired and ligated to barcoded adaptors. After bead purification, up to 32 samples were combined, and probes designed to targeted regions were annealed and extended. After bead purification, libraries were PCR amplified and purified again. SPET libraries were sequenced on the MiSeq platform (Illumina) to generate 2 × 150-bp paired-end data. We obtained an average coverage of 100× (range 60–200×).

RNA-seq. We obtained RNA-seq data for 159 tumor samples (119 with data for tumor-normal pairs). RNA-seq libraries were prepared using the TruSeq RNA Sample Preparation kit (Illumina). The libraries were multiplexed three per lane and sequenced on the HiSeq 2000 platform to obtain, on average, ~68 million paired-end (2 × 75-bp) reads per sample (**Supplementary Fig. 3**).

Sequence data processing. All sequencing reads were evaluated for quality using the Bioconductor ShortRead package[80]. To confirm that all samples were identified correctly, all exome and RNA-seq data variants that overlapped with Illumina HumanOmni2.5-8 array data were compared and checked for consistency. An all-against-all sample comparison was carried out on germline variants to confirm matched tumor-normal pairing before additional data analysis.

Variant calling. Sequencing reads were mapped to the UCSC human genome (GRCh37/hg19) using Burrows-Wheeler Aligner (BWA) software[81] set to default parameters. Local realignment, duplicate marking and raw variant calling were performed as described previously[82]. Calling of somatic variants for a tumor and its matched normal BAM file was performed using Strelka[83]. We used a Strelka variant quality score of ≥1 to filter the variants. Known germline variants represented in dbSNP Build 131 (ref. 84) or 6,515 previously published normal exomes[85] but not represented in COSMIC v62 (ref. 27) were filtered out for all samples. In addition, germline variants that were present in both the tumor and normal samples were removed. To evaluate the performance of this algorithm, we randomly selected 178 protein-altering variants and validated them using Sequenom nucleic acid technology, as described previously[33]. Of these variants, 92% (164/178) were validated as somatic. All variants that were invalidated by Sequenom analysis were removed from the final set. Variants labeled "VALIDATED:RNA-seq" had confirmed expression in the RNA-seq data (**Supplementary Table 3**). In addition to the variant filtering using dbSNP described above, unpaired samples had their initial called variants filtered against normal variants from this data set as well as normal samples from a previously published colon data set[86]. The effect of all nonsynonymous somatic mutations on gene function was predicted using PolyPhen[31], SIFT[30] and Condel[32]. All variants were annotated using Ensembl (release 63).

Evaluation of mutations using simulation. We generated a database of all possible nonsynonymous mutations (~70 million) within our exome targets. On the basis of substitution, C:G>G:C, C:G>A:T, C:G>T:A, T:A>A:T, T:A>C:G or T:A>G:C, the variants were classified into six mutation types. We applied PolyPhen[31], SIFT[30] and Condel[32] to assess the functional impact of each mutation. We classified a mutation as deleterious when at least two of the three methods employed showed that it had an adverse functional impact.

We performed Monte Carlo simulations to assess whether the observed mutations differed from randomly generated mutations. We first randomly selected 100 exome samples (with replacement) for each subtype (pRCC, chRCC, renal oncocytoma and ccRCC (TCGA data)). Then, for each of the 100 samples, from the precomputed database of 70 million variants, we randomly selected the same number of nonsynonymous mutations of each type as was observed in each of the samples. We repeated this process 1,000 times. To evaluate the significance of each mutation observed, we compared the frequency of the observed deleterious mutation with the frequency of deleterious mutations obtained from 1,000 simulations.

Mutational signatures. As described recently[87], we analyzed the nccRCC exome sequence data for the frequencies of the 96 possible mutation types (**Supplementary Fig. 4a**). We included in the analysis TCGA exome data for 2,437 samples from 8 other cancer types, as provided by the SomaticCancerAlterations Bioconductor package (**Supplementary Fig. 4d**). For the identification of mutational signatures[87], we applied the framework of the SomaticSignatures package, version 1.1.11 (J.S.G., B. Fischer and W. Huber, unpublished data). A set of five common signatures was detected by analyzing eight TCGA studies and our three nccRCC subtypes, using non-negative matrix factorization with the Brunet algorithm[88].

Mutational significance and hotspot meta-analysis. We evaluated the mutational significance of genes using Mutational Significance in Cancer (MuSIC)[89]. We calculated q scores by taking the negative $\log_{10}$ value of the likelihood ratio test q values produced by MuSIC. A minimum mean RPKM of 0.26 (corresponding to the cutoff for the bottom quartile of all genes) derived from RNA-seq data was used to filter out genes that were not expressed or were expressed at very low levels, as described previously[90], as they have been shown to have a higher background mutation rate[91]. We also treated the chRCC classic and eosinophilic types as a single group for q-score calculation. We did not attempt to compute q scores for tRCC, as the sample size was small ($n = 6$ tumor-normal pairs). We further filtered for significant genes that were found

to have multiple mutations in only a single sample. Analysis was limited to paired tumor-normal whole-exome sequencing samples.

Mutation meta-analysis to identify mutational hotspots was performed by combining mutations from our study with those that have been reported to occur in at least two samples in COSMIC[27]. Mutations were checked against 6,515 previously published normal exomes[85] such that no mutations present at a frequency of greater than 0.1% in the population were reported.

RNA-seq data analysis. RNA-seq reads were aligned to human genome version NCBI GRCh37 using GSNAP[92]. Expression counts per gene were obtained by counting the number of reads aligned concordantly within a pair and uniquely to each gene locus as defined by NCBI and Ensembl gene annotations and RefSeq mRNA sequences. Analysis of differential gene expression was performed with edgeR[93], and DESeq2 (ref. 94) was used to compute the variance-stabilized expression values used in plotting expression heat maps. For differential expression analysis, only samples that clustered into their respective pathology classifications using unsupervised clustering of the 400 genes with the most variable expression were used. As the samples for oncocytic neoplasm (favoring oncocytoma) clustered with the oncocytomas, these samples were grouped together for calculations of differential gene expression. Similarly, the samples for oncocytic neoplasm (favoring chromophobe eosinophilic) were grouped with the chromophobe eosinophilic samples. The sample-by-sample correlation heat map (**Fig. 3a**) was clustered with complete linkage and using the maximum method in the R dist function as a distance metric for the 400 most variable genes. All gene-by-sample heat maps were clustered with complete linkage and using the Euclidian distance as the distance metric. Raw ccRCC RNA-seq data from TCGA were obtained and processed using the same methods applied to our data set (**Supplementary Fig. 15**). Variants in RNA-seq data were determined using the Genome Analysis Toolkit (GATK)[82].

Gene signature validation. To validate our five-gene classifier, we used a previously published nccRCC microarray data set. Affymetrix CEL files were downloaded from the NCBI Gene Expression Omnibus (GEO) corresponding to GEO accession GSE11151. Data were normalized using the robust multiarray average (RMA) method as implemented in the Bioconductor affy package. Only samples annotated as pRCC, chRCC and renal oncocytoma were used. Clustering was performed using a Euclidian distance matrix and complete linkage.

SNP array data generation and analysis. Illumina HumanOmni2.5-8 arrays were used to assay 112 samples (including 105 tumor-normal pairs) for genotype, DNA copy number and loss of heterozygosity (LOH) at ~2.5 million SNP positions. The data from these arrays were processed as described recently[90]. The PICNIC algorithm[95] was used to estimate normal contamination, ploidy and chromosomal segments with LOH. After adjusting the raw data for normal contamination, the cghFLasso algorithm[96] was used to obtain the final estimation and segmentation of total copy number. A subset of 2,228,703 high-quality SNPs was selected for all analyses.

Genomic regions with recurrent DNA copy gain and loss were identified by computing the frequency of a $\log_2$ copy number ratio of >0.45 or <−0.45 for gains and losses, respectively, for each tumor subtype. Given that our algorithm for estimating copy number ratios assumes segmental changes, we applied manual recentering to the chRCC data where we observed many whole-chromosome losses.

Gene fusion detection and validation. Putative gene fusions were identified using a computational pipeline we have developed called GSTRUCT-fusions[90]. Only fusion events that had at least three reads mapping to the fusion junction and were not found in any of the normal samples were included for further consideration. Furthermore, we removed events that included unannotated exons or fusion partners that had closely related sequences, as these are likely false positives. Validation of gene fusions was performed using RT-PCR with nccRCC tumor and matched normal samples, as previously described[90].

Pathway analysis. We evaluated the mutational significance of genes using the MuSIC[89] path-scan program with Reactome[53,54] pathways as implemented in the curated MSigDB v4.0 (ref. 97). Quilt plots depicting the integrated analysis

of mutation, gene expression, copy number variation and fusion data were limited to samples for which paired tumor-normal exome data were present.

Cells and plasmids. NIH3T3 and HEK293T cells obtained from the Genentech cell bank were maintained in DMEM supplemented with 10% FBS. Clones expressing C-terminally Myc/DDK-tagged MITF, ACTG1 and MET from the pCMV6 expression vector were purchased from Origene. A construct encoding an ACTG1-MITF fusion with a C-terminal Myc/DDK tag sequence was generated using splicing by overlap PCR and cloned into pCMV6. Expression vector containing *MET* served as a template for the generation of the *MET* mutants using the QuikChange II XL site-directed mutagenesis kit (Stratagene). NIH3T3 cells stably expressing wild-type MET, the MET mutants or ACTG1-MITF were generated using retroviral constructs as previously described[98].

FISH analysis. Sections (3 micron) of formalin-fixed, paraffin-embedded tissue were mounted on positively charged glass slides. Using the stained slide as a reference, target areas were etched with a diamond-tipped etcher onto the back of the unstained slide to be assayed. Pretreatment, hybridization and the following washes were performed according to the microwave method in the DAKO Histology FISH Accessory kit guide (SSK5799CE_001/EFG/LMA/2012). DNA probe sets for *TFE3* (Xp11.2) and *TFEB* (6p21.1) were obtained from Agilent Technologies. Selection of tissue and target areas on the slide after staining with hematoxylin and eosin was performed by a board-certified pathologist (P.K.).

MITF stability analysis. HEK293T cells (1×10^5 cells/well) were transfected with pCMV6-MITF (0.5 µg) and pCMV6-ACTG1-MITF (0.3 µg) using FuGene 6 according to the manufacturer's instructions (Roche). We used a lower amount of pCMV6-ACTG1-MITF in transfections as we found that more fusion protein was expressed than wild-type MITF from the vector. At 24 h after transfection, cells were treated with 50 µg/ml cycloheximide (Sigma) to block translation. Samples were processed and subjected to protein blotting as described previously[98]. MITF proteins were assessed by immunoblot with mouse antibody to c-Myc (1:1,000 dilution; Genentech, 9E10). HSP90 expression was assessed using rabbit antibody to HSP90 (1:5,000 dilution; Santa Cruz Biotechnology, sc-7947) and was used as a loading control. Expression was analyzed using appropriate secondary antibodies on a LI-COR Odyssey imager (LI-COR Biotechnology).

Protein blot analysis. Cell lysates from NIH3T3 cells (5×10^6) stably expressing either wild-type or mutant protein were prepared and used for protein blotting as described previously[98]. The phosphorylation status of MET was assessed using antibody to phosphorylated MET (Tyr1234/1235) (1:1,000 dilution; Cell Signaling Technology, 3077). Expression of MET, ACTG1, MITF, ACTG1-MITF and HSP90 in NIH3T3 lines stably expressing the constructs was assessed by protein blotting using antibody to Flag (1:2,000 dilution; Sigma, F1804) or HSP90 (1:5,000 dilution; Santa Cruz Biotechnology, sc-7947) as indicated. Immunoblotting was performed using appropriate secondary antibodies, as described previously[98].

Anchorage-independent growth. The assay was performed as previously described[98]. Briefly, 20,000 NIH3T3 cells stably expressing Flag-tagged wild-type MET, MET mutants, ACTG1, MITF or ACTG1-MITF were mixed with 0.35% agar in DMEM (high glucose) and plated in triplicate on 0.5% base agar in a 6-well plate. Plated cells were then overlaid with complete growth medium (1 ml) and incubated at 37 °C. The number of colonies formed in each plate was assessed using GelCount (Oxford Optronix) after 3 weeks. Student's *t* test (two-tailed) was used for statistical analyses to compare treatment groups in GraphPad Prism 5.00. A *P* value of <0.05 was considered statistically significant.

Cell growth assays. NIH3T3 cells stably expressing wild-type or Asp153Tyr mutant MET were plated in complete medium. After 24 h, the medium was replaced with serum-free medium and cells were treated with the indicated concentration of recombinant HGF (R&D Systems). Cell growth was measured after 3 d with the CellTiter-Glo luminescence cell viability kit (Promega) as

described previously[98]. Student's *t* test (two-tailed) was used for statistical analyses to compare treatment groups with GraphPad Prism 5.00. A *P* value of <0.05 was considered statistically significant (*P < 0.05 and **P < 0.01).

Quantitative PCR analysis. RNA (1 μg), isolated at 24 h after transfection from cells transfected with construct encoding MITF or MITF fusion protein using the RNeasy mini kit (Qiagen), was reverse transcribed to produce cDNA with SuperScript VILO Master Mix (Life Technologies). cDNA was then diluted and used for quantitative PCR with TaqMan Gene Expression Master Mix on the ViiA 7 Real-Time PCR System (Life Technologies). Primer and probe sets (20×) used for TaqMan Gene Expression Assays were obtained from Life Technologies. The primer and probe sets used include *GAPDH* (Hs02758991_g1), *MITF* (Hs01117294_m1), *HIF1A* (Hs00153153_m1), *MET* (Hs01565584_m1) and *APEX1* (Hs00959050_g1). Relative gene expression values were normalized against *GAPDH* expression and then further normalized to *MITF* mRNA levels in the transfected HEK293T cells.

80. Morgan, M. *et al.* ShortRead: a bioconductor package for input, quality assessment and exploration of high-throughput sequence data. *Bioinformatics* **25**, 2607–2608 (2009).
81. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
82. DePristo, M.A. *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43**, 491–498 (2011).
83. Saunders, C.T. *et al.* Strelka: accurate somatic small-variant calling from sequenced tumor-normal sample pairs. *Bioinformatics* **28**, 1811–1817 (2012).
84. Sherry, S.T. *et al.* dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.* **29**, 308–311 (2001).
85. Fu, W. *et al.* Analysis of 6,515 exomes reveals the recent origin of most human protein-coding variants. *Nature* **493**, 216–220 (2013).
86. Seshagiri, S. *et al.* Recurrent R-spondin fusions in colon cancer. *Nature* **488**, 660–664 (2012).
87. Alexandrov, L.B., Nik-Zainal, S., Wedge, D.C., Campbell, P.J. & Stratton, M.R. Deciphering signatures of mutational processes operative in human cancer. *Cell Rep.* **3**, 246–259 (2013).
88. Brunet, J.P., Tamayo, P., Golub, T.R. & Mesirov, J.P. Metagenes and molecular pattern discovery using matrix factorization. *Proc. Natl. Acad. Sci. USA* **101**, 4164–4169 (2004).
89. Dees, N.D. MuSiC: identifying mutational significance in cancer genomes. *Genome Res.* **22**, 1589–1598 (2012).
90. Rudin, C.M. *et al.* Comprehensive genomic analysis identifies *SOX2* as a frequently amplified gene in small-cell lung cancer. *Nat. Genet.* **44**, 1111–1116 (2012).
91. Lawrence, M.S. *et al.* Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* **499**, 214–218 (2013).
92. Wu, T.D. & Nacu, S. Fast and SNP-tolerant detection of complex variants and splicing in short reads. *Bioinformatics* **26**, 873–881 (2010).
93. Robinson, M.D., McCarthy, D.J. & Smyth, G.K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
94. Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biol.* **11**, R106 (2010).
95. Greenman, C.D. *et al.* PICNIC: an algorithm to predict absolute allelic copy number variation with microarray cancer data. *Biostatistics* **11**, 164–175 (2010).
96. Tibshirani, R. & Wang, P. Spatial smoothing and hot spot detection for CGH data using the fused lasso. *Biostatistics* **9**, 18–29 (2008).
97. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA* **102**, 15545–15550 (2005).
98. Jaiswal, B.S. *et al.* Somatic mutations in p85α promote tumorigenesis through class IA PI3K activation. *Cancer Cell* **16**, 463–474 (2009).