mutation[25,26]. TET2 catalyzes the conversion of 5-methylcytosine to 5-hydroxymethlcytosine, which is now believed to be a critical step in DNA demethylation. A recent study indicated that TET2 also mediates histone O-GlcNAcylation during gene transcription[27]. No *TET2* mutations have been reported in non-hematopoietic tumors, except for rare mutations in colorectal cancers (5/214 examined; 2.3%)[28]. *TET2* was mutated in 6 of 106 ccRCC cases (5.7%). Except for one frameshift mutation, five were missense mutations, of which four affected the cysteine-rich or catalytic domain (**Supplementary Fig. 13a**). In copy number analysis using SNP arrays (**Fig. 4**), *TET2* was also located within the significantly deleted regions at 4q24 ($n = 11$; 10.4%) (**Fig. 4f**). In combination, *TET2* mutations and deletions accounted for 17 ccRCC cases (16.0%), with no case having biallelic inactivation that indicated a haploinsufficiency effect of *TET2* on the pathogenesis of ccRCC. KEAP1 is a key component of another cullin–RING ubiquitin ligase complex that is involved in oxidative stress responses by regulating the ubiquitination of the KEAP1-bound NRF2 transcription factor (also known as NEF2L2) (**Fig. 2a,** right)[29]. Frequent *KEAP1* and *NRF2* mutations that abrogate their physical interaction were originally reported in squamous cell carcinoma of the lung and in other solid cancers[30–32], with KEAP1-mediated NRF2 degradation compromised, resulting in deregulated transcriptional activity of the abnormally accumulated NRF2. Of note, compromised KEAP1-mediated NRF2 degradation is also caused by abnormally accumulated fumarate in congenital fumarate hydratase deficiency[33,34], which predisposes to type 2 papillary renal cell carcinoma (pRCC2), and also by somatic mutations in *NRF2* and *CUL3* in sporadic cases with pRCC2 (ref. 35). The current study confirmed that mutually exclusive mutations in *KEAP1* ($n = 5$), *NRF2* ($n = 1$) and *CUL3* ($n = 1$) are also found in the clear-cell subtypes of RCC (6.6%) (**Supplementary Fig. 14**), together with deletions in the *CUL3* locus at 2q36 ($n = 11$; 10.4%) (**Fig. 4d**), with no case having biallelic inactivation in this pathway. *MTOR* was also a newly identified recurrent mutational target and was mutated in 6 of 106 ccRCC cases (5.7%), although a single case with an activating *MTOR* mutation was previously reported[17]. Together with mutations in *PTEN* ($n = 2$), *PIK3CA* ($n = 5$), *PIK3CG* ($n = 2$), *RPS6KA2* ($n = 3$), *TSC1* and *TSC2* (ref. 36) ($n = 2$), and other genes, a total of 28 cases (26%) had mutations that involved phosphoinositide 3-kinase (PI3K)-AKT-mTOR signaling. Except for 3 known tumor suppressor genes—*PTEN*, *TSC1* and *TSC2*—27 mutations were found in 13 genes that are thought to functionally act as oncogenes. In fact, none of the 27 mutations were nonsense, frameshift or splice-site changes, which was highly unexpected from the observed overall frequencies of these types of mutations in ccRCC ($P = 0.00946$, Fisher's exact test), suggesting that these mutations largely act as oncogenes. These mutations were mutually exclusive, except for in two cases that had both *PTEN* and *AKT2* mutations. *FGFR4* was within the significantly amplified region at 5q35 ($n = 69$), and, in total, 81 cases (76%) had genetic alterations in this pathway (**Fig. 5a**). These findings provide additional rationale for the effectiveness of mTOR inhibitors in ccRCC.

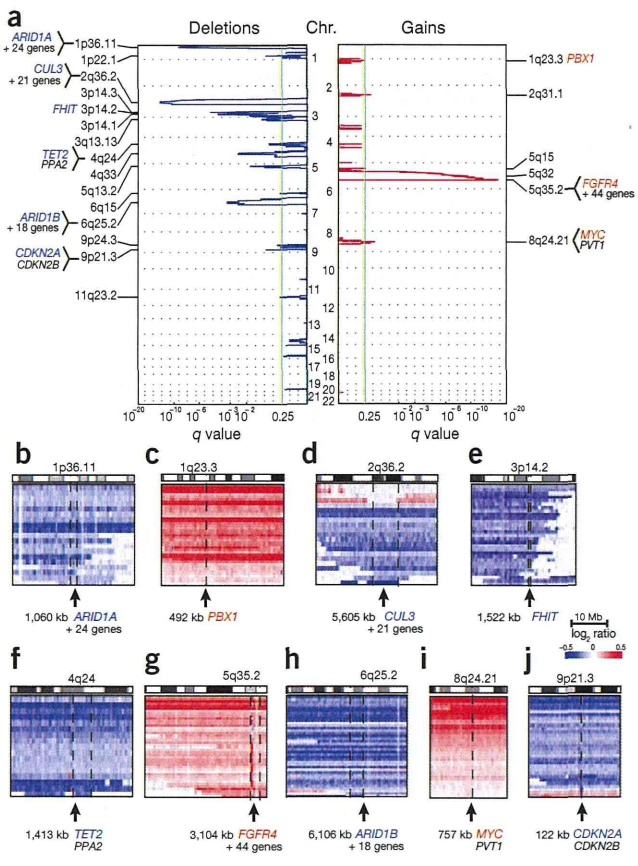


Figure 4 Significant copy number alterations in 240 ccRCC specimens. (**a**) Regions showing statistically significant increase or decrease in genomic copy number were detected using the GISTIC algorithm based on SNP array analysis. For each *q*-value peak, putative gene targets are listed. A dashed line represents the centromere of each chromosome. Red and blue lines indicate *q* value for gains and deletions, respectively. (**b–j**) Log-ratio copy number heatmaps are shown for gene targets at 1p36.11 (**b**), 1q23.3 (**c**), 2q36.2 (**d**), 3p14.2 (**e**), 4q24 (**f**), 5q35.2 (**g**), 6q25.2 (**h**), 8q24.21 (**i**) and 9p21.3 (**j**).

Copy number lesions and significantly affected pathways

We also performed SNP array–based copy number analysis for the 240 ccRCC specimens to identify candidate target genes involved in pathogenesis. Most copy number lesions involved large chromosomal segments, as found in LOH at 3p (94%), gain of 5q (65%), gain of 7q (41%), loss of 8p with or without loss of 8q (20%), LOH at 9p (15%), LOH at 14q (27%) and LOH at 18q (11%) (**Supplementary Fig. 15a**). Even though there was a close correlation between monosomy 8 and *TCEB1* mutation, total or partial loss of chromosome 8 was also found in cases with wild-type *TCEB1*, in which common loss of 8p seemed to be relevant to ccRCC pathogenesis. Hyperploid tumors (defined by ploidy of >2.5) accounted for 17.5% ($n = 42$) of the cases and had a significantly higher rate of metastasis ($P = 2.98 \times 10^{-5}$, Cox proportional hazards model) and poor prognosis ($P = 3.93 \times 10^{-2}$, Cox proportional hazards model). In hyperploid tumors, copy numbers in LOH involving 3p, 9p and 14q were largely neutral, suggesting that these tumors had evolved from diploid tumors with typical deletions of the relevant chromosome segments (**Supplementary Fig. 15b,c**). Supporting this notion was the fact that mutations in the 3p target genes in cases with UPD at 3p showed higher allele frequencies than those for mutations within 2*N* regions (**Supplementary Fig. 16**). Using GISTIC 2.0 analysis, significant focal gains and deletions ($q < 0.25$) were found at 20 loci (5 gains and 15 losses) that involved known tumor suppressors and oncogenes, including *ARID1A* (1p36.11), *CUL3* (2q36.2), *FHIT* (3p14.2), *TET2* (4q24), *ARID1B* (6q25.2), *CDKN2A* and *CDKN2B* (9p21.3), *PBX1* (1q23.3), *FGFR4* (5q35.2) and *MYC* (8q24) (**Fig. 4**).

Significantly affected pathways in ccRCC were further investigated by searching for statistically overrepresented gene families that were somatically mutated and expressed and/or showed copy number abnormalities (Online Methods). In addition to PI3K-AKT-mTOR signaling, significantly affected pathways (false discovery rate

(FDR) < 0.01) included p53 signaling and cell cycle checkpoints, mRNA processing and the SWI/SNF complex (**Supplementary Fig. 13b** and **Supplementary Table 9**). *TP53* was mutated in only 3 cases (2.8%), but, considering these cases together with those with mutations in *ATM* ($n = 2$), *CHEK2* ($n = 1$), *MDM2* ($n = 1$) and *E2F3* ($n = 1$), copy number deletions in *CDKN2A* at 9p21 ($n = 17$; including 4 homodeletions) and gains in *MYC* at 8q24 ($n = 24$), a total of 42 cases (40%) had genetic alterations involved in p53 signaling and/or cell cycle checkpoints (**Fig. 5b**). The pathways involved in mRNA processing were not previously implicated in ccRCC. Frequent mutations in splicing machinery were reported in myelodysplastic syndromes (MDS) and in other hematopoietic neoplasms, with mutations affecting spliceosome components involved in recognizing the 3′ splice site[37]. In contrast, in ccRCC, mutated genes that were involved in mRNA processing were related to later steps in splicing, including the release of introns, 3′ end processing and mRNA export to the cytoplasm (**Supplementary Fig. 13c**).

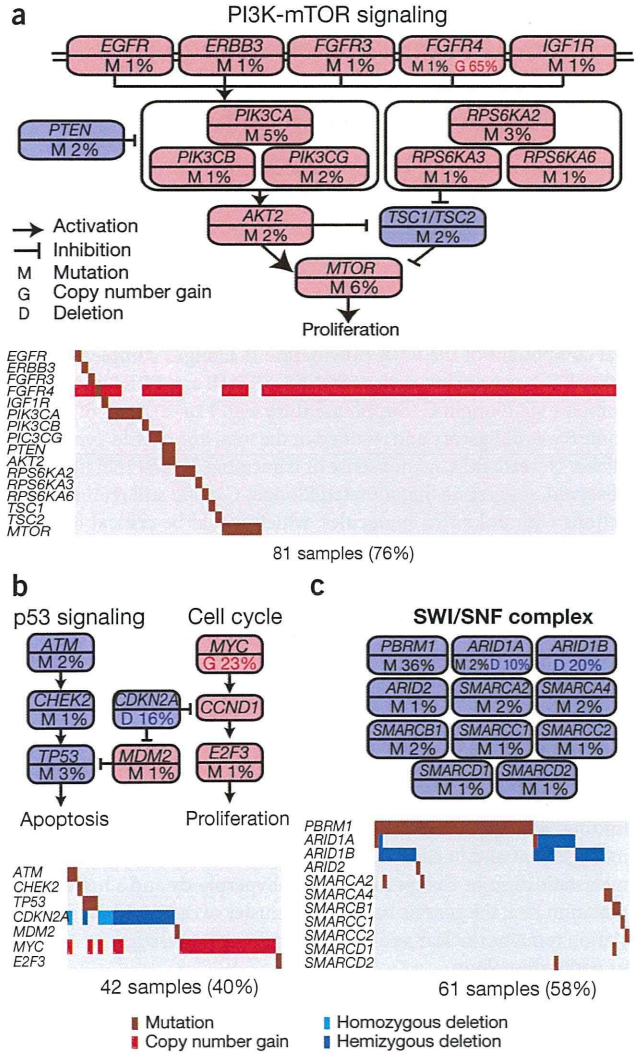


Figure 5 Significantly mutated pathways for 106 ccRCC specimens. (**a**–**c**) PI3K-mTOR signaling (**a**), p53 signaling (**b**) and the SWI/SNF complex (**c**) are commonly altered in ccRCC. Alterations are defined as somatic mutations and DNA copy number changes identified by GISTIC 2.0 analysis. Genes considered to be oncogenes or tumor suppressor genes are colored pink and blue, respectively.

PBRM1 was the second most commonly mutated gene in ccRCC and encodes a component of the SWI/SNF complex that regulates chromatin structure through ATP-dependent nucleosome remodeling. Other components of the SWI/SNF complex, including *ARID1A* and *ARID1B*, were within significantly deleted regions at 1p36 and 6q25, respectively. Taking into consideration these cases and those with mutations in *ARID1A* ($n = 2$), *SMARCA2* ($n = 2$), *SMARCA4* ($n = 2$), *SMARCB1* ($n = 2$) and other genes, a total of 61 cases (58%) had mutations and/or deletions in 11 different components of the SWI/SNF complex (**Fig. 5c**). These changes were not mutually exclusive, and 5 cases had ≥2 mutations. Multiple components of the SWI/SNF complex might be mutated in single cases in various types of cancer[38]. Notably, these five ccRCC cases had significantly worse prognosis compared to individuals with less than two mutations in the SWI/SNF complex (HR = 5.40, 95% CI = 1.56–14.5; $P = 0.0113$; Cox proportional hazards model), indicating that multiple mutations in the SWI/SNF complex could lead to aggressive phenotype in ccRCC.

Integrated molecular analysis of ccRCC

RNA sequencing was also performed in 100 ccRCC cases (**Supplementary Table 10**). In total, 44 fusion transcripts were identified in 25 specimens (**Supplementary Table 11**), of which about one-third ($n = 14$) were in frame. The lack of recurrent lesions largely obscured the significance of these fusions, except for *SETD2-QRICH1* and *NONO-TFE3* fusions. *TFE3* on Xp11 participates in promiscuous gene fusions with several partners, including *NONO*[39]. *TFE3*-containing fusions have been implicated in RCCs that are histologically distinct from typical ccRCCs (RCCs with Xp11 translocation)[39,40]. The current case with a *NONO-TFE3* fusion was positive for TFE3 expression in immunohistochemistry but showed histology otherwise indistinguishable from that typical for ccRCC cases[41] (**Supplementary Fig. 17**).

As previously reported, gene expression profiling of 101 ccRCC cases identified 2 major clusters—ccA and ccB—which were characterized by upregulated angiogenic factors and enhanced expression of genes involved in cell cycle progression, respectively (**Supplementary Fig. 18a–c**)[42]. In accordance with recent reports[10,20], Gene Set Enrichment Analysis (GSEA) showed discrete expression profiles in *BAP1*- and *PBRM1*-mutated tumors. *PBRM1*-mutated tumors were enriched for upregulated expression of the gene set with a hypoxia signature, whereas *BAP1*-mutated cases were associated with downregulated expression of target genes of polycomb repressive complex 2 (PRC2) (**Supplementary Fig. 19**).

DNA methylation profiling based on hierarchical clustering identified 3 distinct clusters that were characterized by different DNA methylation levels, including clusters with high, intermediate and low mean methylation levels, as determined using a total of 1,288 differentially methylated genes (**Fig. 6a, Online Methods** and **Supplementary Fig. 20**). Functional annotation analysis using The Database for Annotation, Visualization and Integrated Discovery (DAVID)[43] showed over-representation of homeobox genes in the 1,228 differentially methylated genes (**Supplementary Table 12**), and GSEA of the differentially methylated genes in the high and low-intermediate subgroups showed a marked enrichment in the cluster with high methylation of the gene set regulated by PRC2 and of the genes undergoing methylation at histone H3 lysine 27 (H3K27) (**Fig. 6b**). The cluster with high methylation was also characterized by a higher *BAP1* mutation rate ($P = 5.22 \times 10^{-3}$) and lower *BAP1* ($P = 2.11 \times 10^{-2}$) and higher *EZH2* ($P = 1.57 \times 10^{-5}$) expression levels (**Fig. 6c,d**). Notably, *BAP1* mutation ($P = 4.86 \times 10^{-3}$), decreased *BAP1* expression ($P = 7.46 \times 10^{-3}$) and increased *EZH2* expression ($P = 8.10 \times 10^{-3}$) were shown to be significantly associated with increased methylation of PRC2 target genes (**Supplementary Fig. 21**). These

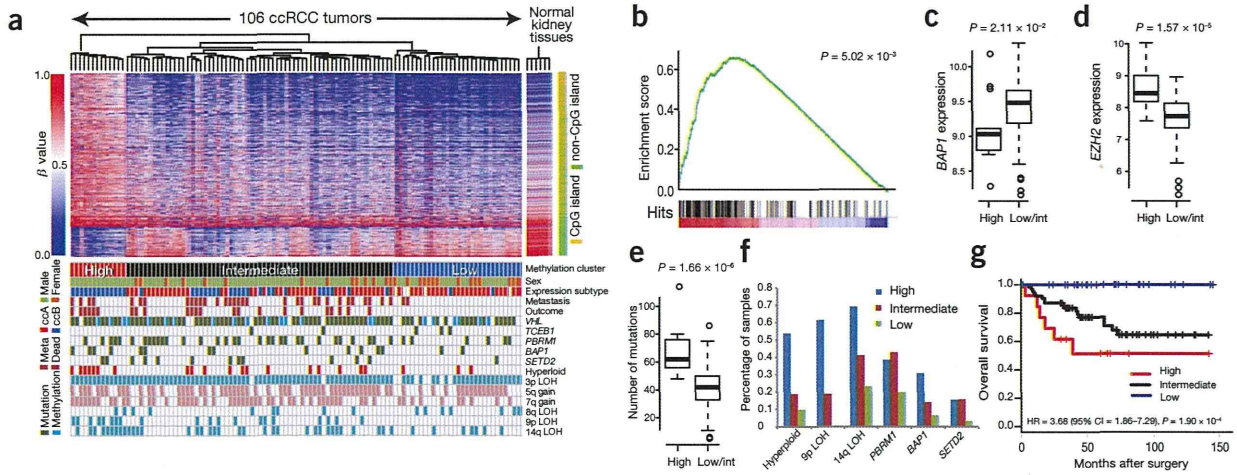


Figure 6 Correlations between DNA methylation and other genetic lesions. (**a**) Integrated view of DNA methylation clustering combined with mutation status of common driver genes, gene expression profiles and genome copy numbers. β values are for DNA methylation. Probes corresponding to CpG islands and non-CpG islands are shown on the right. Clinical outcomes and the presence-absence of metastasis are also shown. Meta, cases with distant metastasis. (**b**) GSEA showing significant enrichment of PRC2-regulated genes in differentially methylated genes in the subgroup with high DNA methylation. Hits displayed below the graph show where the members of the gene set appear in the ranked list of genes. (**c**–**g**) Integrated genetic and epigenetic analyses showed close correlation between DNA methylation status and *BAP1* expression (**c**), *EZH2* expression (**d**), total number of somatic mutations (**e**), hyperploidy, LOH at 9p, LOH at 14q and *PBRM1*, *BAP1* and *SETD2* mutations (**f**) and clinical outcome (**g**). *P* values were calculated using the *t* test in **c**–**e**, the Cochran-Armitage trend test in **f** and the log-rank test in **g**. Box plots show median, 25% and 75% quartile ranges.

findings strongly indicate that the cluster with high methylation was closely related to deregulated PRC2 activity.

Other prominent features of the cluster with high methylation included a higher number of somatic mutations ($P = 1.66 \times 10^{-6}$), more hyperploid cases ($P = 1.57 \times 10^{-3}$), cases with LOH at 14q ($P = 2.0 \times 10^{-4}$) and 9p ($P = 1.57 \times 10^{-3}$), enrichment of the ccB gene expression profile ($P = 2.46 \times 10^{-10}$) (**Fig. 6a,e,f**) and strong association with higher probability of metastasis (HR = 1.96, 95% CI = 1.17–3.30; $P = 1.11 \times 10^{-2}$) and poor overall survival (HR = 3.68, 95% CI = 1.86–7.29; $P = 1.90 \times 10^{-4}$) (**Fig. 6g**), which combined with the deregulated PRC2 profile, indicate that deregulation of polycomb-mediated gene silencing may contribute to aggressive tumor phenotype and poor clinical outcome.

DISCUSSION

Our comprehensive molecular study involving 106 ccRCC cases provided new insights into ccRCC genetics and biology and identified potential therapeutic targets. Similar to what has been observed in other cancers, the development of ccRCC seems to be shaped by the acquisition of a number of somatic gene mutations and chromosomal lesions that predominantly affect a handful of genes (*VHL*, *PBRM1*, *SETD2* and *BAP1*) and chromosome regions (3p, 5q, 9p and 14q), with many less frequent mutations additionally occurring in key targets. Pathway analysis of gene mutations disclosed several functional gene pathways, including PI3K-AKT-mTOR signaling, the KEAP1-NRF2 apparatus and mRNA processing, that were commonly affected by multiple, less frequent mutations.

A cardinal feature of ccRCC is a very high frequency of *VHL* inactivation caused by gene deletion, mutation and/or silencing via promoter methylation, leading to HIF accumulation. From this perspective, the discovery of *TCEB1* mutations and the obligatory loss of chromosome 8 in 36–42% of ccRCC cases with intact *VHL* was among the most notable findings of this study, demonstrating a new mechanism for inactivation of the VHL complex during ccRCC pathogenesis. These changes not only abolished the recruitment of VHL to the CUL2–RING ubiquitin ligase complex (CRL2), resulting in HIF accumulation, but could also compromise the recruitment of Elongin A, which is an essential component of the RNA polymerase II Elongin complex[21], and of other BC-box proteins, such as SOCS3, FEM1B and LRR1, to the CRL2 complex via Elongin C (**Supplementary Fig. 11b–e**). Loss of chromosome 8 was obligatory and resulted in the total loss of wild-type *TCEB1* alleles. Nevertheless, no nonsense or truncating *TCEB1* mutations were observed, suggesting that mutated Elongin C could still retain its interactions with unknown molecules, which would be critical for tumor cell viability.

By integrating multiple layers of different comprehensive analyses, we unmasked unique correlations between somatic mutations, DNA methylation, gene expression and copy number alterations, which were also closely linked to the clinical behaviors of tumors. In particular, correlation of the cluster with hypermethylation with hyperploid status, LOH at 9p and poor prognosis was conspicuous and was in stark contrast to the correlation of the cluster with low methylation with less frequent hyperploidy or LOH at 9p and an excellent prognosis. Clinically, the discovery of the cluster of cases with hypermethylation was a notable finding, and special attention should be paid to these cases in their management and in early detection or prevention of recurrent and/or metastatic disease. Except for frequent hyperploidy and a higher *BAP1* mutation rate, the genetic basis of this cluster of cases with hypermethylation remains unclear and will require further investigation to clarify its pathophysiology.

URLs. dbSNP, http://www.ncbi.nlm.nih.gov/projects/SNP/; 1000 Genomes Project, http://www.1000genomes.org/; RepeatMasker, http://www.repeatmasker.org/; Genomonfusion (in Japanese), http://genomon.hgc.jp/rna/; CNAG/AsCNAR, http://www.genome.umin.jp/; GISTIC 2.0, http://www.broadinstitute.org/cgi-bin/cancer/publications/pub_paper.cgi?mode=view&paper_id=216&p=t; Gene Set Enrichment Analysis (GSEA), http://www.broadinstitute.org/gsea/; MSigDB,

http://www.broadinstitute.org/gsea/msigdb/index.jsp; the European Genome-phenome Archive, https://www.ebi.ac.uk/ega/.

METHODS

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

Accession code. Sequencing and genotype data have been deposited in the European Genome-phenome Archive (EGA) under accession EGAS00001000509.

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

ACKNOWLEDGMENTS

We thank Y. Mori, M. Nakamura, N. Mizota and S. Ichimura for their technical assistance. We also thank M. Nangaku and N. Takeda for fruitful discussion and comments. We thank T. Kitamura (University of Tokyo) for providing pMXs-puro, M. Onodera (National Center for Child Health and Development, Japan) for providing pGCDNsamIRESEGFP and R.C. Mulligan (Boston Children's Hospital) for providing 293gp cells. This work was supported by KAKENHI (22134006), the Industrial Technology Research Grant Program from the New Energy and Industrial Technology Development Organization (NEDO) (08C46598a) and the Japan Society for the Promotion of Science through the Funding Program for World-Leading Innovative R&D on Science and Technology, initiated by the Council for Science and Technology Policy.

AUTHOR CONTRIBUTIONS

Y. Sato, S. Maekawa, Y.N., H.S., Y. Suzuki, S.S., K.Y. and A.K. performed DNA sequencing. Y. Shiraishi, Y.O., K.C., H.T., A.F., T.T. and S. Miyano performed bioinformatics analyses of the sequencing data. T.Y., M.S. and T.K. performed the functional analyses of Elongin C mutants. Y. Sato, A.S.-O., A.N. and M.S. performed SNP array and expression array analyses. T.S., G.N. and H.A. performed methylation analysis. H.K. and Y.H. provided specimens and were also involved in planning the project. T.M., D.M. and M.F. confirmed histological diagnosis and performed immunostaining for HIF proteins. Y. Sato, T.Y., Y.O., A.S.-O. and S.O. generated figures and tables and wrote the manuscript. S.O. led the entire project. All authors participated in the discussion and interpretation of data and results.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Ferlay, J. *et al.* Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. *Int. J. Cancer* **127**, 2893–2917 (2010).
2. Rini, B.I., Campbell, S.C. & Escudier, B. Renal cell carcinoma. *Lancet* **373**, 1119–1132 (2009).
3. Ljungberg, B. *et al.* EAU guidelines on renal cell carcinoma: the 2010 update. *Eur. Urol.* **58**, 398–406 (2010).
4. Gnarra, J.R. *et al.* Mutations of the *VHL* tumour suppressor gene in renal carcinoma. *Nat. Genet.* **7**, 85–90 (1994).
5. Gallou, C. *et al.* Mutations of the *VHL* gene in sporadic renal cell carcinoma: definition of a risk factor for VHL patients to develop an RCC. *Hum. Mutat.* **13**, 464–475 (1999).
6. Schraml, P. *et al. VHL* mutations and their correlation with tumour cell proliferation, microvessel density, and patient prognosis in clear cell renal cell carcinoma. *J. Pathol.* **196**, 186–193 (2002).
7. Herman, J.G. *et al.* Silencing of the *VHL* tumor-suppressor gene by DNA methylation in renal carcinoma. *Proc. Natl. Acad. Sci. USA* **91**, 9700–9704 (1994).
8. Varela, I. *et al.* Exome sequencing identifies frequent mutation of the SWI/SNF complex gene *PBRM1* in renal carcinoma. *Nature* **469**, 539–542 (2011).
9. Dalgliesh, G.L. *et al.* Systematic sequencing of renal carcinoma reveals inactivation of histone modifying genes. *Nature* **463**, 360–363 (2010).
10. Peña-Llopis, S. *et al.* BAP1 loss defines a new class of renal cell carcinoma. *Nat. Genet.* **44**, 751–759 (2012).
11. Guo, G. *et al.* Frequent mutations of genes encoding ubiquitin-mediated proteolysis pathway components in clear cell renal cell carcinoma. *Nat. Genet.* **44**, 17–19 (2012).
12. Greenman, C. *et al.* Patterns of somatic mutation in human cancer genomes. *Nature* **446**, 153–158 (2007).
13. Guichard, C. *et al.* Integrated analysis of somatic mutations and focal copy-number changes identifies key genes and pathways in hepatocellular carcinoma. *Nat. Genet.* **44**, 694–698 (2012).
14. Huang, J. *et al.* Exome sequencing of hepatitis B virus–associated hepatocellular carcinoma. *Nat. Genet.* **44**, 1117–1121 (2012).
15. Sung, W.K. *et al.* Genome-wide survey of recurrent HBV integration in hepatocellular carcinoma. *Nat. Genet.* **44**, 765–769 (2012).
16. Fujimoto, A. *et al.* Whole-genome sequencing of liver cancers identifies etiological influences on mutation patterns and recurrent mutations in chromatin regulators. *Nat. Genet.* **44**, 760–764 (2012).
17. Gerlinger, M. *et al.* Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N. Engl. J. Med.* **366**, 883–892 (2012).
18 Hakimi, A.A. *et al.* Adverse outcomes in clear cell renal cell carcinoma with mutations of 3p21 epigenetic regulators *BAP1* and *SETD2*: a report by MSKCC and the KIRC TCGA Research Network. *Clin. Cancer Res.* **19**, 3259–3267 (2013).
19. Hakimi, A.A. *et al.* Clinical and pathologic impact of select chromatin-modulating tumor suppressors in clear cell renal cell carcinoma. *Eur. Urol.* **63**, 848–854 (2013).
20. Kapur, P. *et al.* Effects on survival of *BAP1* and *PBRM1* mutations in sporadic clear-cell renal-cell carcinoma: a retrospective analysis with independent validation. *Lancet Oncol.* **14**, 159–167 (2013).
21. Aso, T., Lane, W.S., Conaway, J.W. & Conaway, R.C. Elongin (SIII): a multisubunit regulator of elongation by RNA polymerase II. *Science* **269**, 1439–1443 (1995).
22. Kamura, T. *et al.* Activation of HIF1a ubiquitination by a reconstituted von Hippel-Lindau (VHL) tumor suppressor complex. *Proc. Natl. Acad. Sci. USA* **97**, 10430–10435 (2000).
23. Stebbins, C.E., Kaelin, W.G. Jr. & Pavletich, N.P. Structure of the VHL-ElonginC-ElonginB complex: implications for VHL tumor suppressor function. *Science* **284**, 455–461 (1999).
24. Takagi, Y., Pause, A., Conaway, R.C. & Conaway, J.W. Identification of elongin C sequences required for interaction with the von Hippel-Lindau tumor suppressor protein. *J. Biol. Chem.* **272**, 27444–27449 (1997).
25. Delhommeau, F. *et al.* Mutation in *TET2* in myeloid cancers. *N. Engl. J. Med.* **360**, 2289–2301 (2009).
26. Langemeijer, S.M. *et al.* Acquired mutations in *TET2* are common in myelodysplastic syndromes. *Nat. Genet.* **41**, 838–842 (2009).
27. Chen, Q., Chen, Y., Bian, C., Fujiki, R. & Yu, X. TET2 promotes histone O-GlcNAcylation during gene transcription. *Nature* **493**, 561–564 (2013).
28. The Cancer Genome Atlas Network. Comprehensive molecular characterization of human colon and rectal cancer. *Nature* **487**, 330–337 (2012).
29. Zimmerman, E.S., Schulman, B.A. & Zheng, N. Structural assembly of cullin–RING ubiquitin ligase complexes. *Curr. Opin. Struct. Biol.* **20**, 714–721 (2010).
30. Padmanabhan, B. *et al.* Structural basis for defects of Keap1 activity provoked by its point mutations in lung cancer. *Mol. Cell* **21**, 689–700 (2006).
31. Shibata, T. *et al.* Cancer related mutations in *NRF2* impair its recognition by Keap1-Cul3 E3 ligase and promote malignancy. *Proc. Natl. Acad. Sci. USA* **105**, 13568–13573 (2008).
32. Kim, Y.R. *et al.* Oncogenic *NRF2* mutations in squamous cell carcinomas of oesophagus and skin. *J. Pathol.* **220**, 446–451 (2010).
33. Adam, J. *et al.* Renal cyst formation in Fh1-deficient mice is independent of the Hif/Phd pathway: roles for fumarate in KEAP1 succination and Nrf2 signaling. *Cancer Cell* **20**, 524–537 (2011).
34. Kinch, L., Grishin, N.V. & Brugarolas, J. Succination of Keap1 and activation of Nrf2-dependent antioxidant pathways in FH-deficient papillary renal cell carcinoma type 2. *Cancer Cell* **20**, 418–420 (2011).
35. Ooi, A. *et al. CUL3* and *NRF2* mutations confer an NRF2 activation phenotype in a sporadic form of papillary renal cell carcinoma. *Cancer Res.* **73**, 2044–2051 (2013).
36. Kucejova, B. *et al.* Interplay between pVHL and mTORC1 pathways in clear-cell renal cell carcinoma. *Mol. Cancer Res.* **9**, 1255–1265 (2011).
37. Yoshida, K. *et al.* Frequent pathway mutations of splicing machinery in myelodysplasia. *Nature* **478**, 64–69 (2011).
38. Kadoch, C. *et al.* Proteomic and bioinformatic analysis of mammalian SWI/SNF complexes identifies extensive roles in human malignancy. *Nat. Genet.* **45**, 592–601 (2013).
39. Clark, J. *et al.* Fusion of splicing factor genes *PSF* and *NonO* (*p54^{nrb}*) to the *TFE3* gene in papillary renal cell carcinoma. *Oncogene* **15**, 2233–2239 (1997).
40. Ross, H. & Argani, P. Xp11 translocation renal cell carcinoma. *Pathology* **42**, 369–373 (2010).
41. Kuroda, N. *et al.* Review of renal carcinoma associated with Xp11.2 translocations/*TFE3* gene fusions with focus on pathobiological aspect. *Histol. Histopathol.* **27**, 133–140 (2012).
42. Brannon, A.R. *et al.* Molecular stratification of clear cell renal cell carcinoma by consensus clustering reveals distinct subtypes and survival patterns. *Genes Cancer* **1**, 152–163 (2010).
43. Huang, W., Sherman, B.T. & Lempicki, R.A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4**, 44–57 (2009).

ONLINE METHODS

Subjects and materials. Paired tumor-normal DNA was isolated from 240 ccRCC specimens and subjected to comprehensive molecular analyses after written informed consent was obtained. Matched normal specimens for germline controls were obtained from adjacent normal kidney tissue or from peripheral blood specimens (**Supplementary Note**). No subjects received preoperative treatments, including immunotherapies or molecular targeted therapies. Histopathological specimens were reviewed to confirm that the tumor specimens were histologically consistent with ccRCC (**Supplementary Figs. 14** and **22**). This study was approved by the ethics committee of the Graduate School of Medicine at the University of Tokyo.

Whole-genome sequencing. For whole-genome sequencing, genomic DNA was sonicated to generate approximately 400-bp fragments and was analyzed by HiSeq 2000 with the 100-bp paired-end read option according to the manufacturer's protocol.

FASTQ sequences generated by CASAVA 1.8 were aligned to the human reference genome (hg19) using Burrows-Wheeler Aligner (BWA)[44] version 0.5.8. We attempted to realign unmapped or poorly mapped reads using BLAT[45]. For both SNVs and indels, the variant bases found in ≥7 reads in tumor samples and in ≤1 (SNVs) or 0 (indels) reads in germline samples were designated as variants. Within noncoding regions, variants found on average in 1% of the total reads in 13 unrelated germline samples were excluded from further analysis owing to the high probability that they represented false positives. Synonymous variants and variants found in either an in-house SNP database constructed from exome sequence data from 98 germline samples, dbSNP131 or the 1000 Genomes Project database were excluded. We also excluded SNVs that were in a tandem repeat region identified using tandem repeats finder[46] and indels that were in RepeatMasker.

To detect structural variations, soft-clipped sequences that could be mapped to unique genomic positions were collected. Structural variation candidates were called if they had >4 supporting read pairs in total and at least 1 read pair from each shore of the breakpoint. Structural variations having a contig sequence that could be aligned to an alternate assembly of the hg19 genome with >93% identity were excluded as false positives. Structural variations with a read depth of >150 at either breakpoint were considered to be in a repeat element and were excluded.

Whole-exome and RNA sequencing. Whole-exome sequencing was performed using target capture with SureSelect v.4 followed by sequencing as previously described[37]. Mutation calling used the EBcall algorithm[47]. For targeted sequencing, 3 mg of whole-genome DNA was amplified using the REPLI-g kit (Qiagen) and sonicated to generate a peak target size of 200 bp. Captured targets were sequenced using the Illumina HiSeq2000 platform with the 100-bp paired-end read option. A custom bait (SureSelect) library for a panel of eight genes was used for target capture, and high-throughput sequencing was performed using the 100-bp paired-end read option. Libraries for RNA sequencing were generated using the TruSeq RNA Sample Preparation kit (Illumina) and were analyzed using the Illumina HiSeq 2000 platform with 100-bp paired-end reads according to the manufacturer's protocol. Gene fusion detection was performed using Genomon-fusion.

To detect significantly mutated genes, the background mutation rate was calculated for each gene on the basis of data from the current whole-exome sequencing study, taking into consideration the effect of the replication timing of the gene during DNA replication on the background mutation rate as previously described[48,49]. Mutations showing significantly higher mutation frequencies compared to the corresponding background mutation rate were adopted as significantly mutated genes, with Benjamini-Hochberg correction applied.

Analysis of significantly mutated pathways. To detect significantly mutated pathways, each pathway registered in the Kyoto Encyclopedia of Genes and Genomes (KEGG), BioCarta, Reactome, Sigma-Aldrich and Signaling Transduction KE was tested with the PathScan package[50] on the basis of the background mutation rate observed in whole-exome sequencing data. We excluded somatic mutations that were not observed in RNA sequencing because we expected that they would make only a small contribution to pathogenesis. Benjamini-Hochberg correction was applied.

PCR-based deep sequencing. To validate somatic mutations and estimate mutant allele frequencies, we conducted deep sequencing using the Illumina MiSeq platform. We randomly selected SNVs and indels called through our pipeline, finding that TPRs were 99% (476/477) for coding SNVs, 99% (93/94) for noncoding SNVs, 96% (29/30) for coding indels and 97% (32/33) for noncoding indels in whole-genome sequencing. In whole-exome sequencing, TPRs were 96% (504/525) for SNVs and 96% (55/57) for indels. Primer sequences are shown in **Supplementary Tables 13** and **14**.

Microarray analyses. Global DNA methylation profiles were analyzed using the Infinium HumanMethylation450 BeadChip (Illumina) according to the manufacturer's instructions. To determine DNA methylation profiles, the following filtering steps were adopted to select probes for unsupervised clustering analysis. We first removed probes that were designed for sequences on the X and Y chromosomes. Second, we selected probes with variance ranked in the top 1% of the remaining probes. We then performed unsupervised hierarchical clustering with 3,562 probes, identifying 3 distinct clusters. The β values of 1,672 probes, containing 1,228 genes, that were differentially methylated in the high and low-intermediate groups were represented graphically using a heatmap. Genome-wide analysis of DNA copy number was conducted for 240 tumor-normal specimens using the Affymetrix GeneChip Human Mapping 250K NspI Array according to the manufacturer's protocol. Microarray data were analyzed to determine total and allele-specific copy numbers using CNAG/AsCNAR[51,52]. Significant focal copy number alterations were identified using GISTIC 2.0. Gene expression analysis used 500 ng of total RNA extracted from each tumor ($n = 101$) using the Human Gene Expression 4x44K v2 Microarray (Agilent) according to the manufacturer's instructions. GSEA was performed for both gene expression and DNA methylation using the curated gene sets (c2) acquired from MSigDB.

Multivariate analysis. To evaluate the impact of the mutation status of *PBRM1*, *BAP1* and *SETD2* on overall survival and disease-free survival, we performed multivariate analysis of the three genes for an extended cohort of 240 ccRCC cases.

Immunostaining for HIF-1α, HIF-2α and TFE3. Immunohistochemistry used mouse monoclonal antibody against HIF-1α (1:300 dilution; clone H1alpha67, NB100-105, Novus Biomedicals) and rabbit polyclonal antibody against HIF-2α (1:1,000 dilution; NB100-122, Novus Biomedicals). Cases were considered positive for expression when >5% of tumor cells showed nuclear immunoreactivity. Expression-positive cases were further classified on the basis of the intensity of nuclear immunoreactivity: 1+, mild; 2+, moderate; 3+, strong. Immunohistochemistry for TFE3 was performed as previously described[53] (sc-5958, Santa Cruz Biotechnology).

doi:10.1038/ng.2699

Plasmid construction. Full-length *TCEB1* cDNA was obtained by PCR amplification of DNA extracted in normal human colon and was cloned adjacent to a sequence encoding three copies of a hemagglutinin (HA) tag at the N terminus of the protein. This construct was subcloned into pMXs-puro[54], pGCDNsamIRESEGFP[55] and pCI-neo (Promega). cDNA encoding the Tyr79Cys and Ala100Pro mutants was obtained from *in vitro* mutagenesis using wild-type *TCEB1* cDNA as the template and the QuikChange Site-Directed Mutagenesis kit (Stratagene). In the pGCDNsamIRESEGFP-*TCEB1* constructs, four synonymous nucleotide substitutions, c.210A>C (p.Leu70Leu) and c.211_213TCG>AGC (p.Ser71Ser), were introduced to protect the transcripts from targeting by siRNA designed for endogenous *TCEB1* transcripts using PrimeSTAR HS DNA Polymerase (Takara Bio) according to the manufacturer's protocol. Human *TCEB2* cDNA was obtained with the same method as for *TCEB1* and was cloned into pCI-neo. For *VHL*, *FEM1B*, *TCEB3*, *LRR1* and *SOCS3*, each cDNA was tagged with three copies of sequence encoding a Flag epitope (Flag tag) at its N terminus and was cloned into pcDNA3/Puro[56].

Cell culture and gene transfer. HeLa, HEK 293T and 293gp cells were maintained in DMEM (Gibco) supplemented with 10% FCS and 1% penicillin and streptomycin in a humidified atmosphere with 5% CO_2 or 10% CO_2 at 37 °C. HeLa cells were engineered to stably express wild-type or mutant *TCEB1* or indicated empty vectors (mock) using retrovirus-mediated gene transfer (pMXs-puro-*TCEB1* or pGCDNsamIRESEGFP-*TCEB1*)[37]. HeLa cells transfected with pMXs-puro were subjected to puromycin selection. Cells transduced with pGCDNsamIRESEGFP were sorted for green fluorescent protein (GFP) signal using a flow cytometer (BD FACSAria III, Becton Dickinson). HEK 293T cells were transiently cotransfected with the indicated genes by the calcium phosphate transfection method. Cells were used for experiments after 48 h in culture.

RNA interference. Synthetic siRNAs were obtained from Takara Bio (**Supplementary Table 7b**). HeLa cells were transfected with each siRNA using Lipofectamine 2000 (Invitrogen) according to the manufacturer's instructions. Cells were used for experiments after incubation for 48 h.

Quantitative RT PCR. RNA (500 ng) extracted from the indicated cells using the RNA RNeasy Mini kit (Qiagen) was subjected to reverse transcription using the ReverTra Ace qPCR RT kit (Toyobo) according to the manufacturer's protocol. Quantitative expression levels of mRNA were measured as described previously[37]. Primers used for quantitative RT-PCR are listed in **Supplementary Table 7c**.

Antibodies. Antibodies used for immunoblot analysis are described in **Supplementary Table 15**. Antibodies were from commercial companies except for antibody against Elongin B[57].

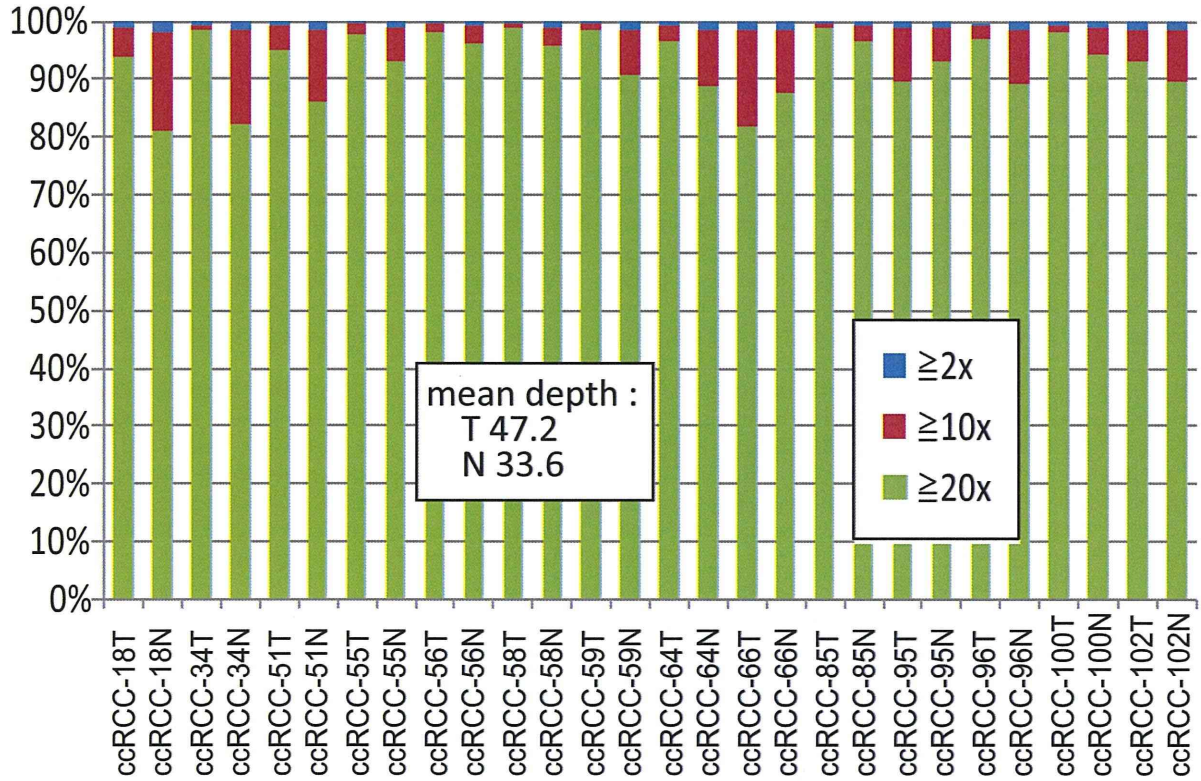
Protein extraction and immunoblot analysis. Cells were lysed in RIPA buffer (Santa Cruz Biotechnology). Lysates were subjected to SDS-PAGE. Proteins were separated and electrophoretically transferred to polyvinylidene difluoride (PVDF) membranes. Membranes were incubated with the indicated antibodies, and proteins were detected using Immobilon Western Chemiluminescent HRP Substrate (Millipore).

Immunoprecipitation. Cells transfected with the indicated vectors were lysed. For endogenous VHL stabilization, HeLa cells transfected with pMXs-puro-*TCEB1* were treated with 10 mM MG132 for 6 h before harvesting. HEK 293T cells cotransfected with constructs encoding HA-TCEB1, untagged TCEB2 and Flag-tagged binding proteins were subjected to immunoprecipitation using an antibody to HA or Flag, and immunoblotting was performed as described.

44. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
45. Kent, W.J. BLAT—the BLAST-like alignment tool. *Genome Res.* **12**, 656–664 (2002).
46. Benson, G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* **27**, 573–580 (1999).
47. Shiraishi, Y. *et al.* An empirical Bayesian framework for somatic mutation detection from cancer genome sequencing data. *Nucleic Acids Res.* **41**, e89 (2013).
48. Hellmann, I. *et al.* Why do human diversity levels vary at a megabase scale? *Genome Res.* **15**, 1222–1231 (2005).
49. Stamatoyannopoulos, J.A. *et al.* Human mutation rate associated with DNA replication timing. *Nat. Genet.* **41**, 393–395 (2009).
50. Wendl, M.C. *et al.* PathScan: a tool for discerning mutational significance in groups of putative cancer genes. *Bioinformatics* **27**, 1595–1602 (2011).
51. Nannya, Y. *et al.* A robust algorithm for copy number detection using high-density oligonucleotide single nucleotide polymorphism genotyping arrays. *Cancer Res.* **65**, 6071–6079 (2005).
52. Yamamoto, G. *et al.* Highly sensitive method for genomewide detection of allelic composition in nonpaired, primary tumor specimens by use of Affymetrix single-nucleotide-polymorphism genotyping microarrays. *Am. J. Hum. Genet.* **81**, 114–126 (2007).
53. Tsuji, K., Ishikawa, Y. & Imamura, T. Technique for differentiating alveolar soft part sarcoma from other tumors in paraffin-embedded tissue: comparison of immunohistochemistry for TFE3 and CD147 and of reverse transcription polymerase chain reaction for *ASPSCR1-TFE3* fusion transcript. *Hum. Pathol.* **43**, 356–363 (2012).
54. Morita, S., Kojima, T. & Kitamura, T. Plat-E: an efficient and stable system for transient packaging of retroviruses. *Gene Ther.* **7**, 1063–1066 (2000).
55. Nabekura, T., Otsu, M., Nagasawa, T., Nakauchi, H. & Onodera, M. Potent vaccine therapy with dendritic cells genetically modified by the gene-silencing-resistant retroviral vector GCDNsap. *Mol. Ther.* **13**, 301–309 (2006).
56. Kamura, T. *et al.* VHL-box and SOCS-box domains determine binding specificity for Cul2-Rbx1 and Cul5-Rbx2 modules of ubiquitin ligases. *Genes Dev.* **18**, 3055–3065 (2004).
57. Garrett, K.P. *et al.* Positive regulation of general transcription factor SIII by a tailed ubiquitin homolog. *Proc. Natl. Acad. Sci. USA* **92**, 7172–7176 (1995).

doi:10.1038/ng.2699

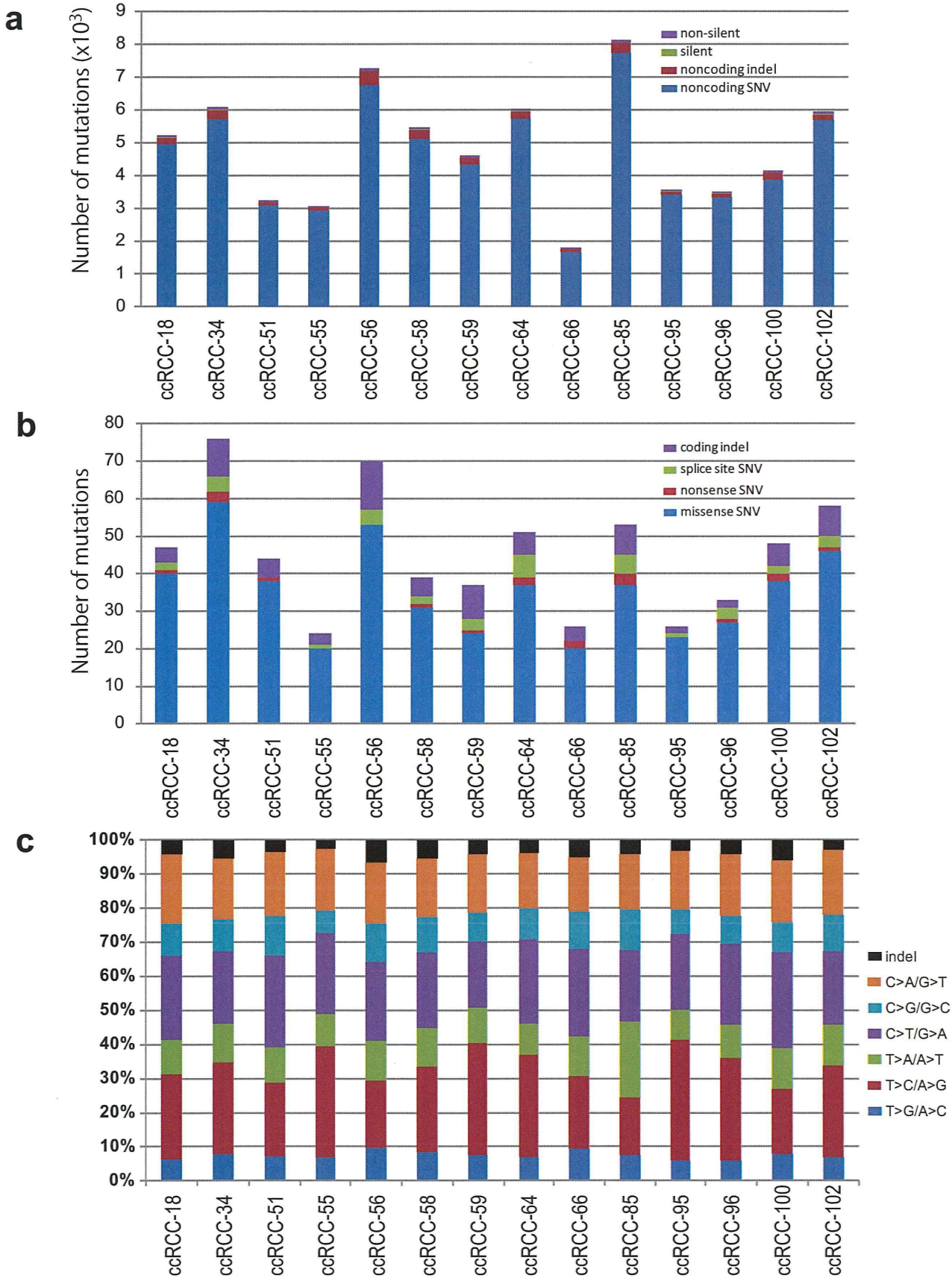
Supplementary Figure 1



Mean coverage by whole-genome sequencing for paired tumor (T)/normal (N) DNA from 14 ccRCC cases.
Genomic fractions analyzed by the indicated coverage are shown by colors.

Nature Genetics: doi:10.1038/ng.2699

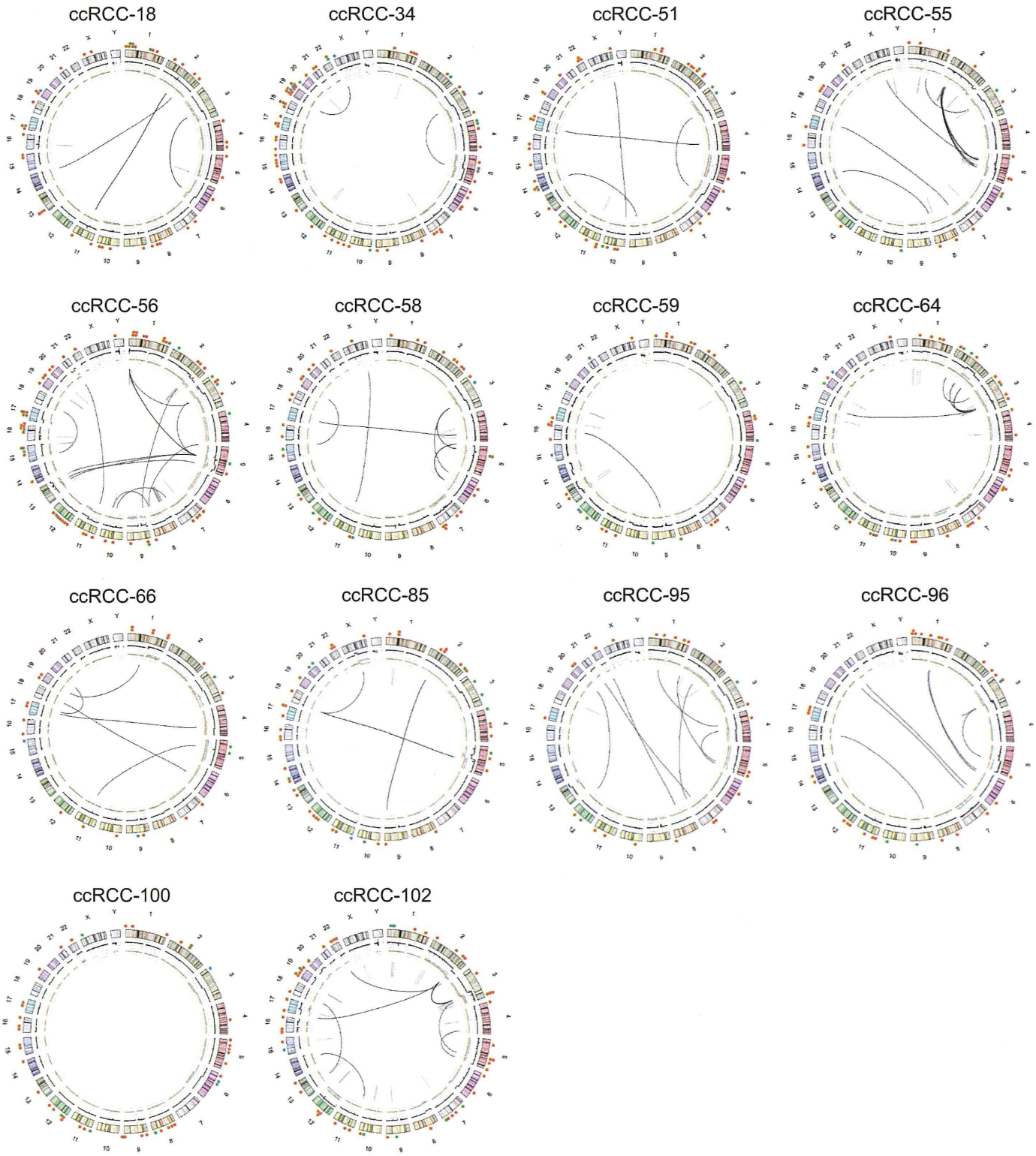
Supplementary Figure 2



Somatic mutations detected by whole-genome sequencing for 14 ccRCC specimens
(**a**) Total number of mutations within coding and non-coding regions. (**b**) Differential counts of non-silent mutations. (**c**) Spectrum of single-nucleotide substitutions.

Nature Genetics: doi:10.1038/ng.2699

Supplementary Figure 3

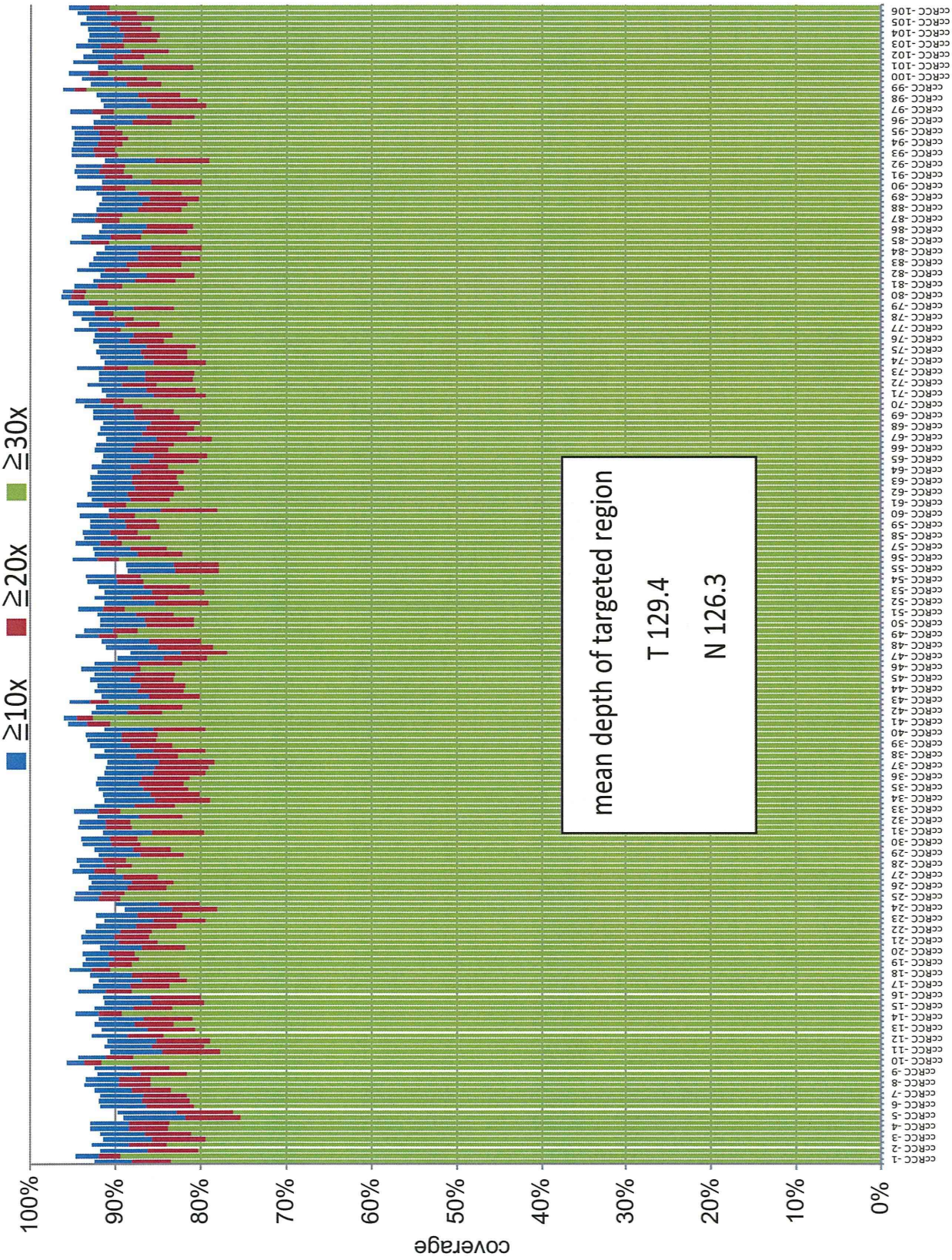


Circos plots of 14 ccRCC genomes

Locations of non-silent mutations, including missense (orange), nonsense (blue) and frameshift (green) mutations are indicated. Total (black) and allele-specific (red and green) genomic copy numbers and structural variations (transverse lines) are indicated in the inner circles. Sample IDs are shown at the top of each circos.

Nature Genetics: doi:10.1038/ng.2699

Supplementary Figure 4



Mean coverage by whole-exome sequencing for paired tumor (T)/normal (N) DNA from 106 ccRCC cases.

Fractions of target exome sequences analyzed by the indicated coverage are shown by different colors.

Nature Genetics: doi:10.1038/ng.2699