Supplementary Information

Integrated molecular analysis of clear cell renal cell carcinoma

Yusuke Sato, Tetsuichi Yoshizato, Yuichi Shiraishi, Shigekatsu Maekawa, Yusuke Okuno, Takumi Kamura, Teppei Shimamura, Aiko Sato-Otsubo, Genta Nagae, Hiromichi Suzuki, Yasunobu Nagata, Yasunobu Nagata, Kenichi Yoshida, Ayana Kon, Yutaka Suzuki, Kenich Chiba, Hiroko Tanaka, Atsushi Niida, Akihiro Fujimoto, Tatsuhiko Tsunoda, Teppei Morikawa, Daichi Maeda, Haruki Kume, Sumio Sugano, Masashi Fukayama, Hiroyuki Aburatani, Masashi Sanada, Satoru Miyano, Yukio Homma and Seishi Ogawa

Nature Genetics: doi:10.1038/ng.2699

Table of contents:

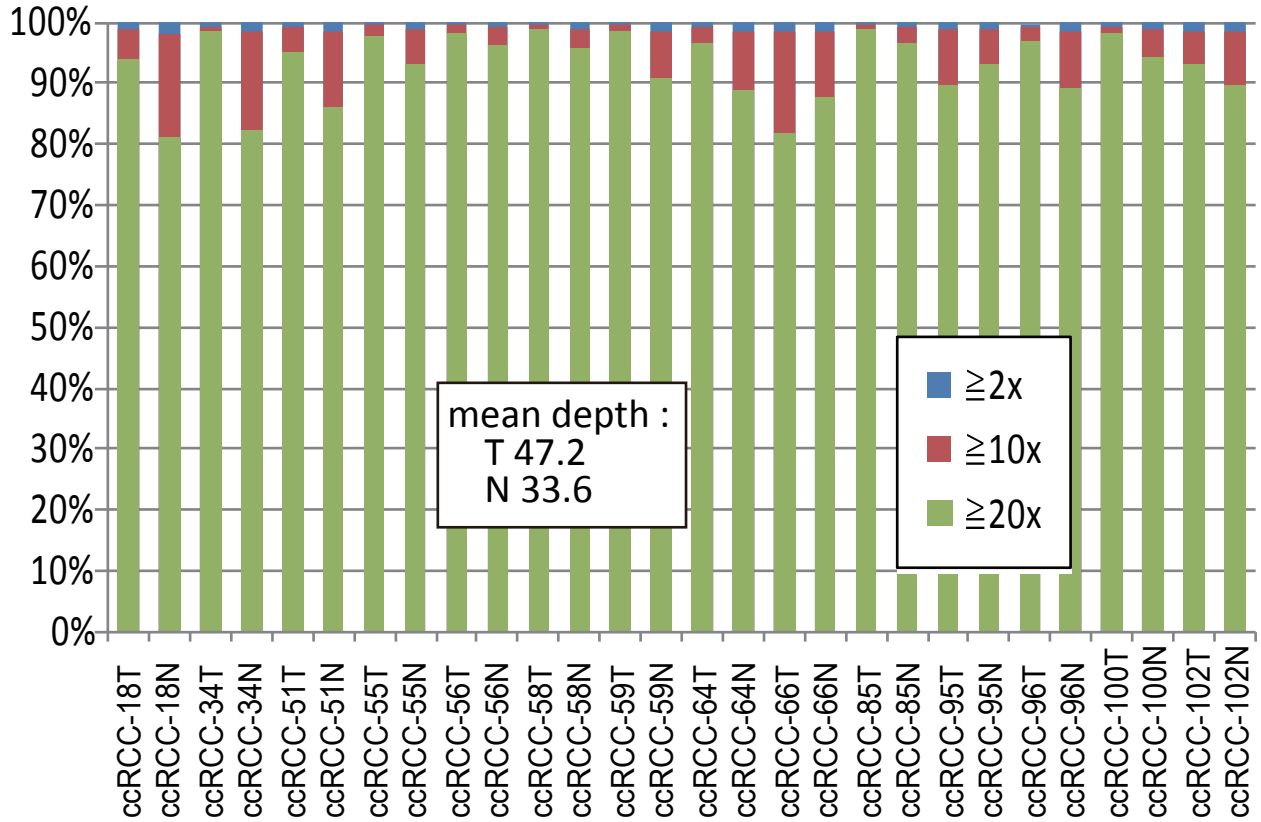
Nature Genetics: doi:10.1038/ng.2699

2. Supplementary Tables

3. Supplementary Note**……p.28**

Nature Genetics: doi:10.1038/ng.2699

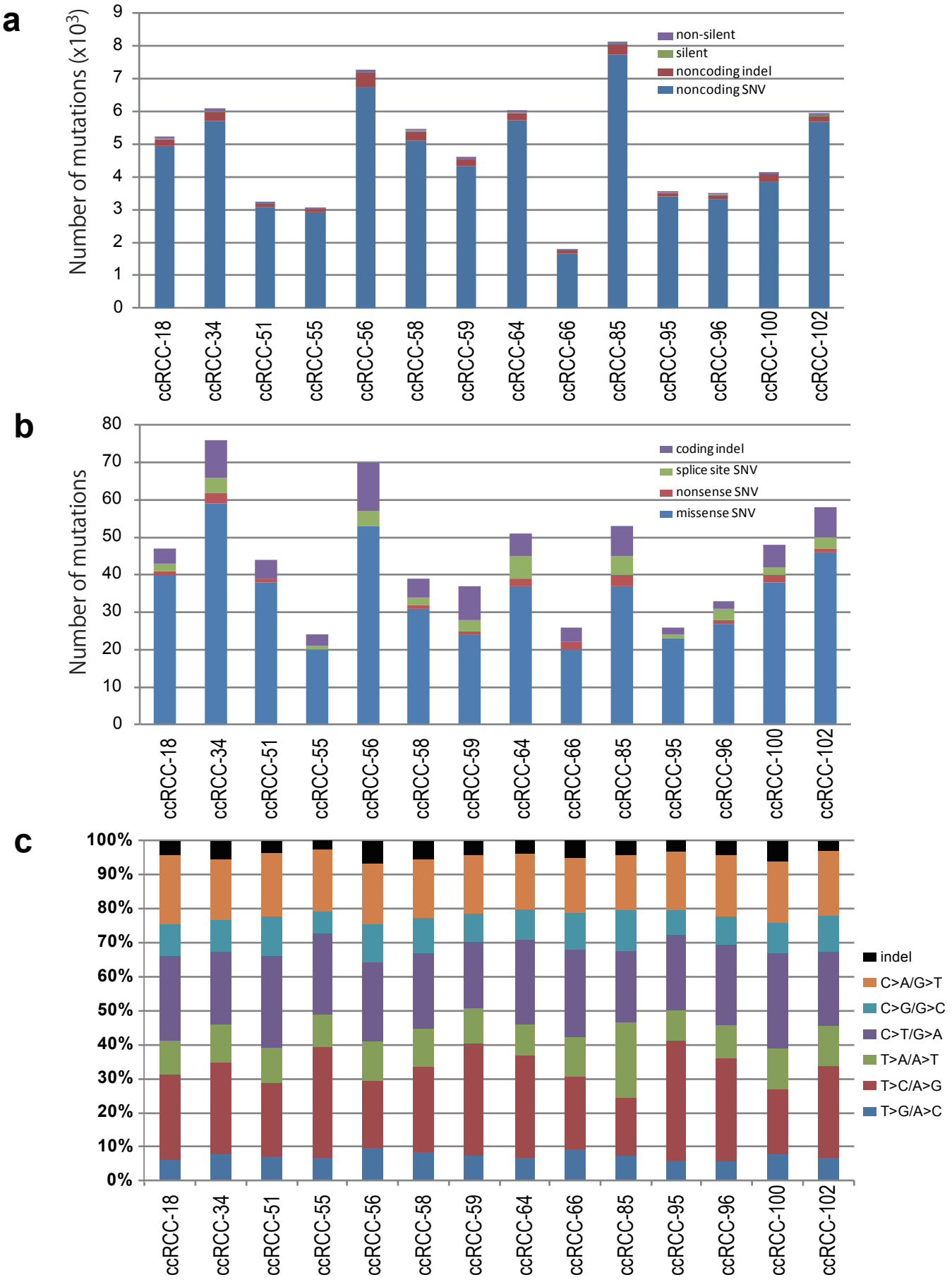
Supplementary Figure 1



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

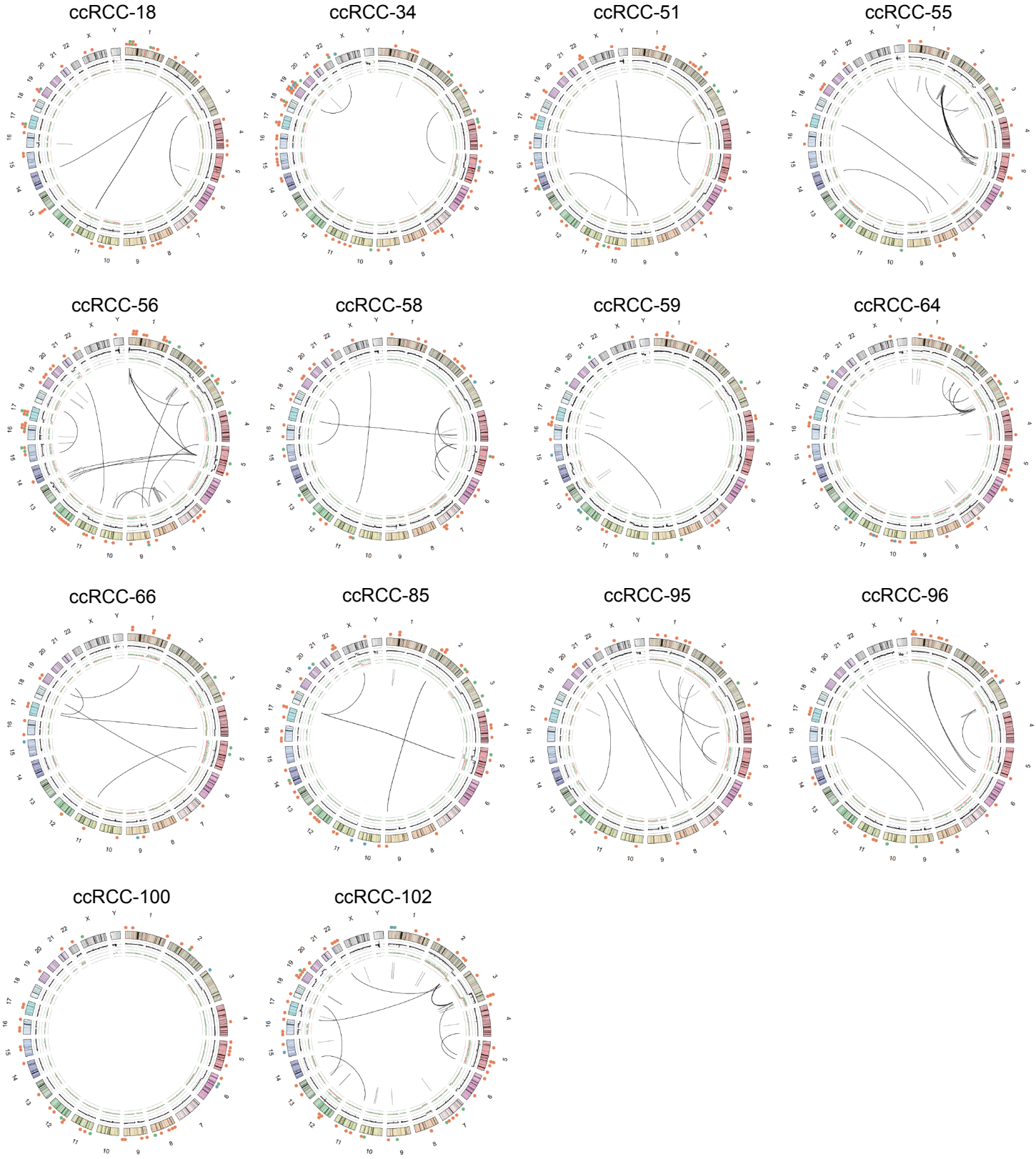
Supplementary Figure 2



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

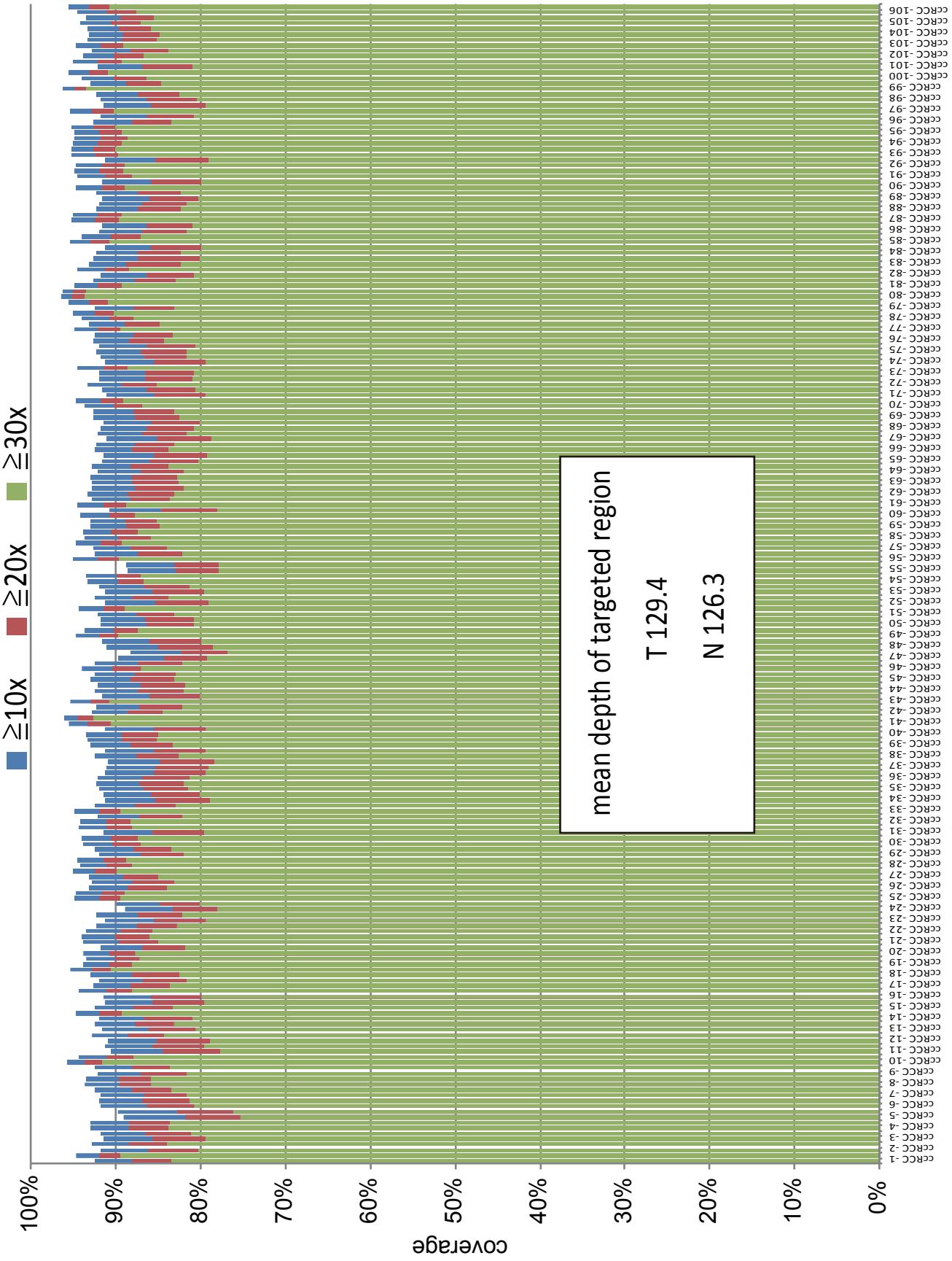
Supplementary Figure 3



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

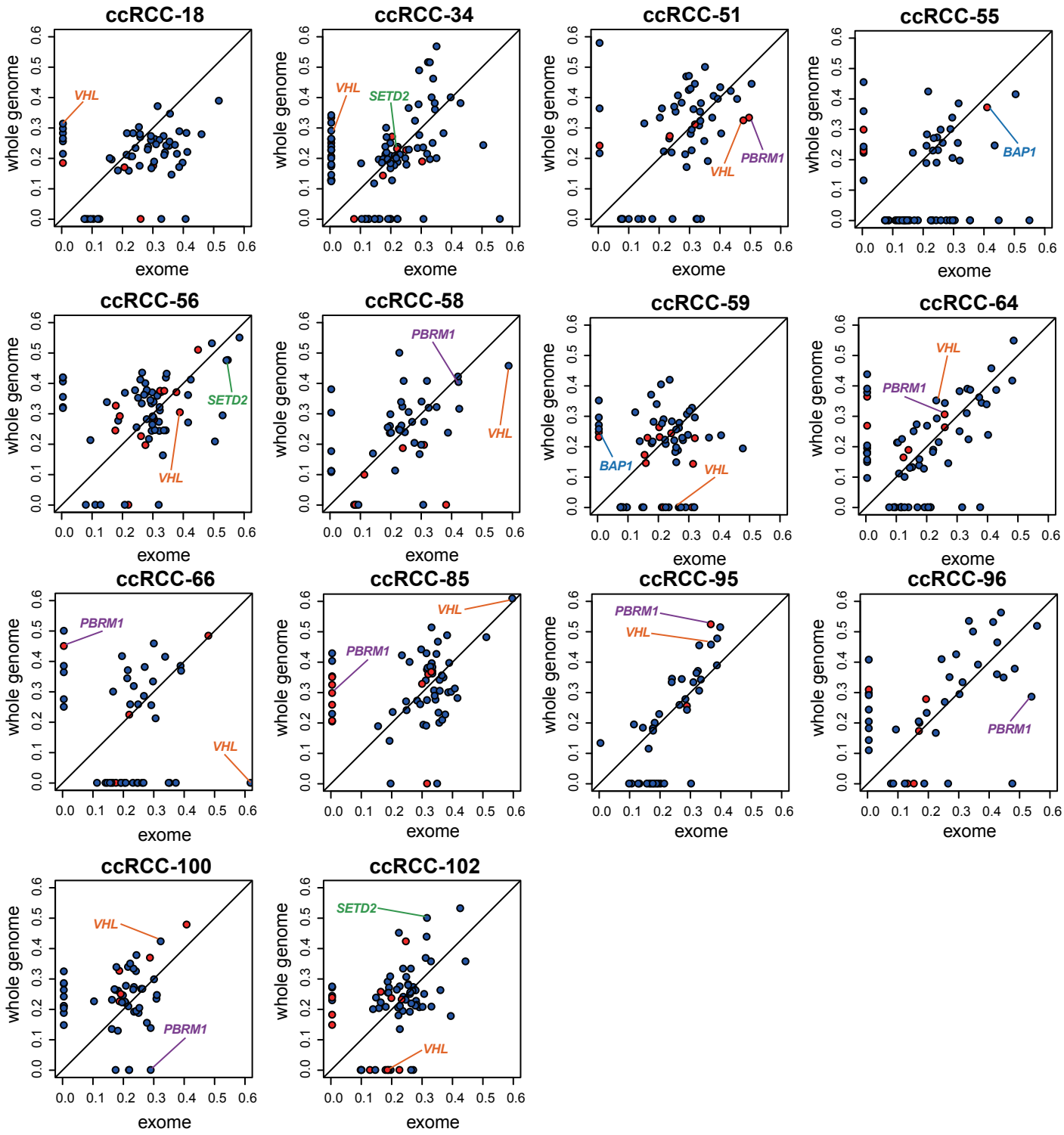


Mean coverage by whole-exome sequencing for paired tumor (T)/normal (N) DNA from 106 ccRCC cases.

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

Nature Genetics: doi:10.1038/ng.2699

Supplementary Figure 5

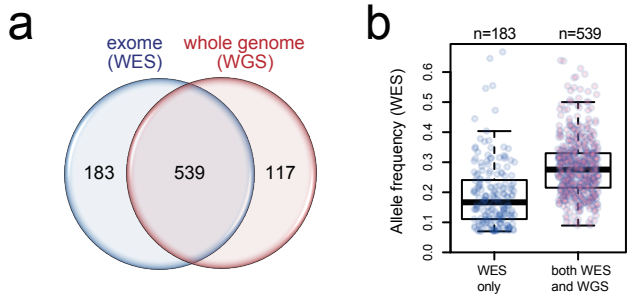


Diagonal plots of validated coding sequence mutations detected by whole-genome and/or exome sequencing for 14 ccRCC specimens analyzed with both platforms

Observed allele frequencies from both sequencing platforms are plotted for each mutation. Major driver mutations are indicated. SNVs are shown in blue whereas indels are shown in red.

Nature Genetics: doi:10.1038/ng.2699

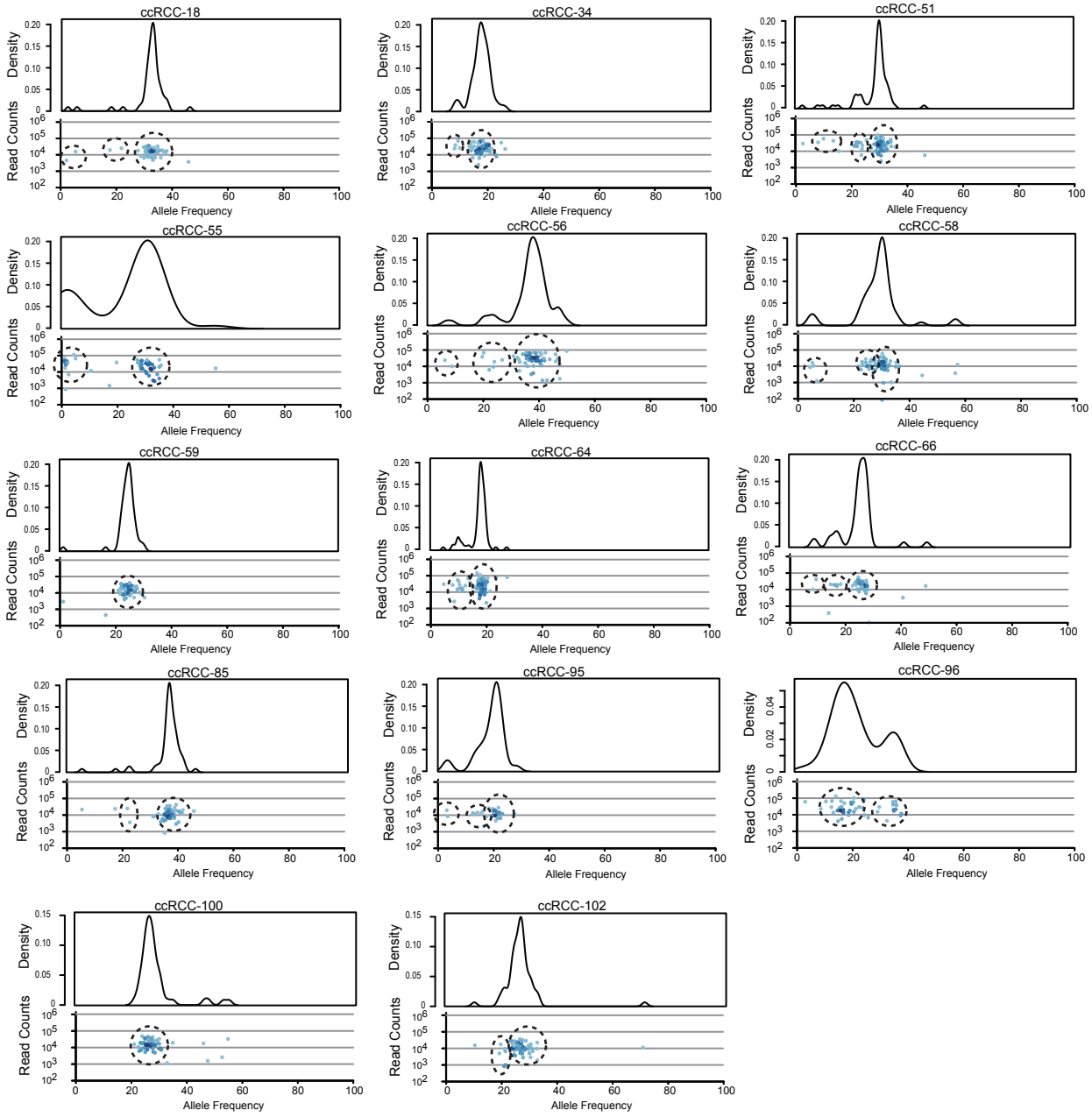
Supplementary Figure 6



Lower allele frequencies of somatic mutations detected only by whole-exome sequencing for 14 ccRCC specimens

(**a**) Number of confirmed somatic mutations detected by whole-genome and/or exome sequencing. (**b**) Comparisons of allele frequencies between validated somatic mutations detected by whole-exome sequencing only and by both whole-genome and exome sequencing.

Supplementary Figure 7

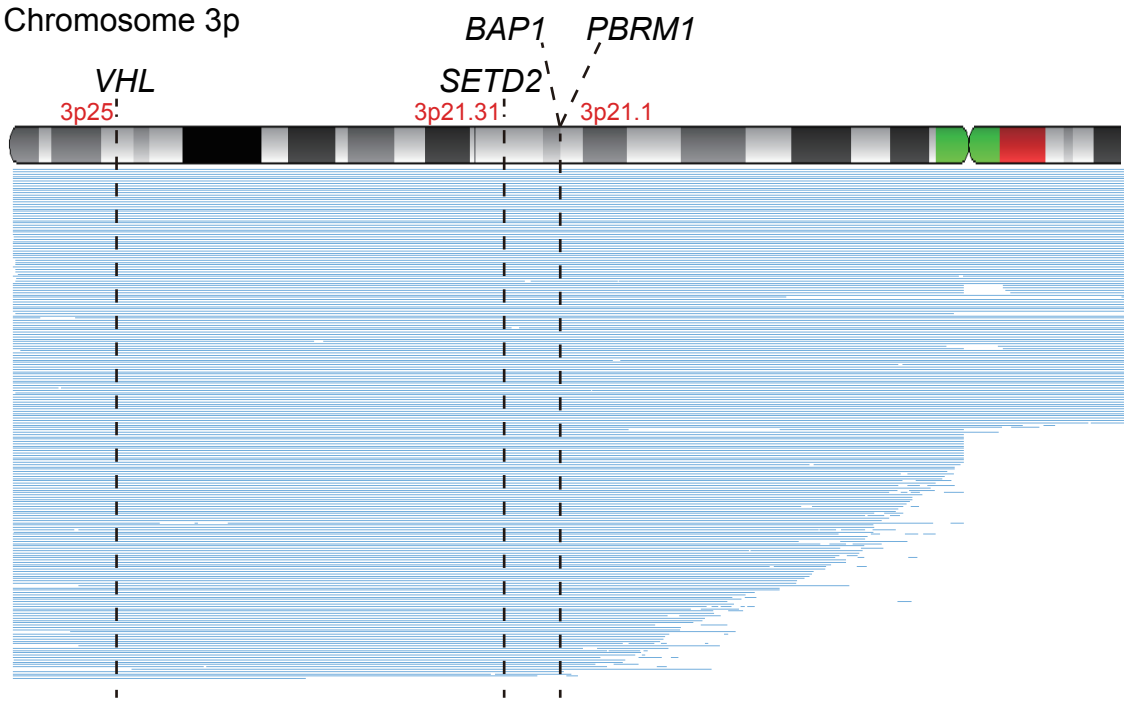


Intratumoral heterogeneity in ccRCC cases

Kernel density estimations of clonal populations based on allele frequencies of observed somatic mutations using deep sequencing (top panels).The frequency of each variant allele is plotted against the total number of sequencing reads that covered the corresponding nucleotide positions (bottom panels).

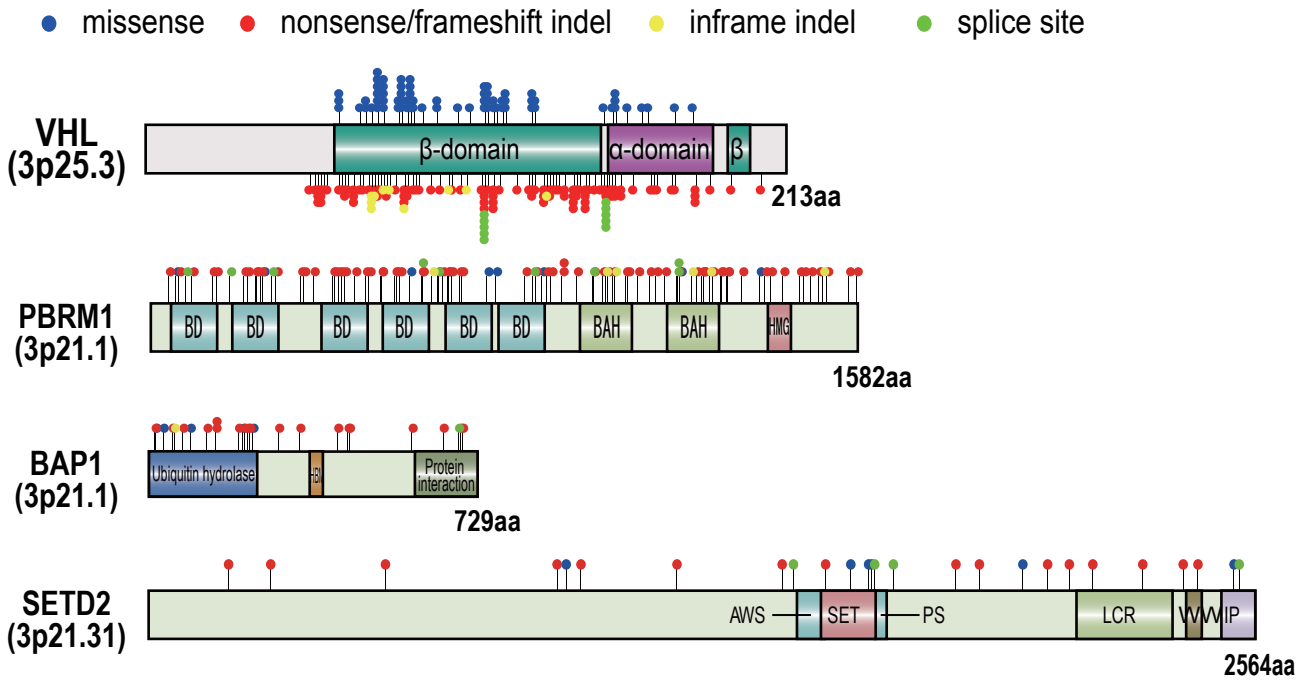
Nature Genetics: doi:10.1038/ng.2699

Supplementary Figure 8



LOH mapping in the 3p arm in ccRCC
LOH in the 3p arm found in a total of 226 ccRCC cases by SNP array analysis. The genetic loci of the 4 major targets of 3p LOH are indicated in which the *PBRM1* locus demarcates the common LOH regions.

Supplementary Figure 9

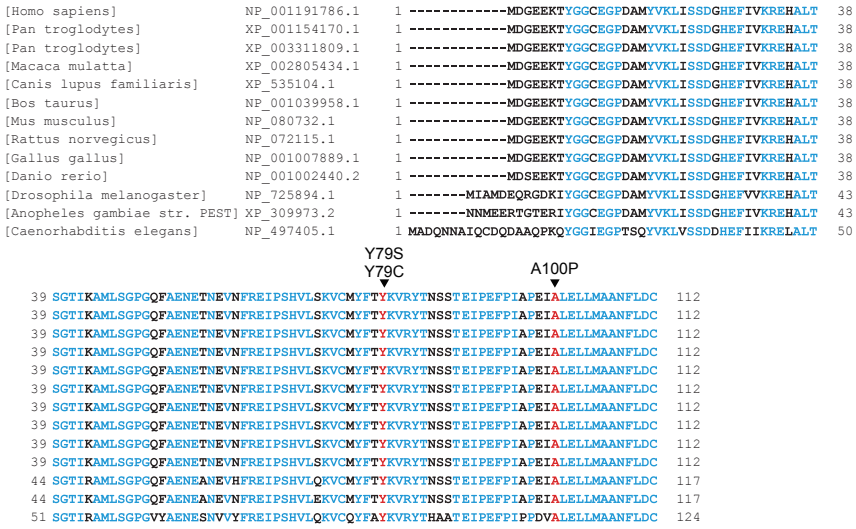


Mutation distributions in 3p targets, including *VHL*, *PBRM1*, *SETD2* and *BAP1*
Mutations of *VHL, PBRM1, SETD2* and *BAP1* in a cohort of 240 ccRCC cases. Types of mutations are distinguished by the indicated colors.

Nature Genetics: doi:10.1038/ng.2699

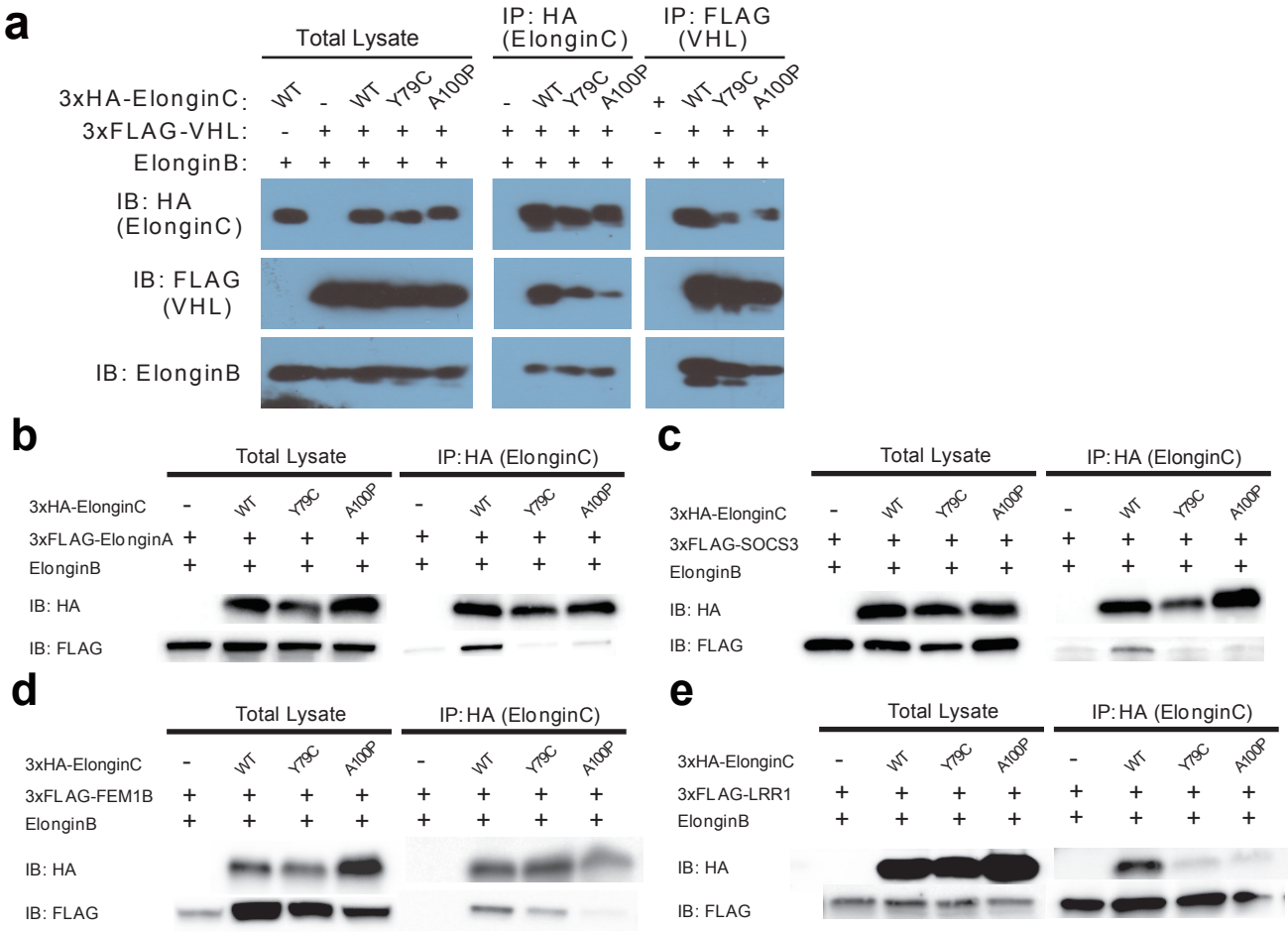
Supplementary Figure 10

Elongin C (*TCEB1*)



Amino-acid sequence alignments for Elongin C (TCEB1) from different species
Completely conserved amino acids among all species are indicated in blue; mutational hot spots are shown in red.

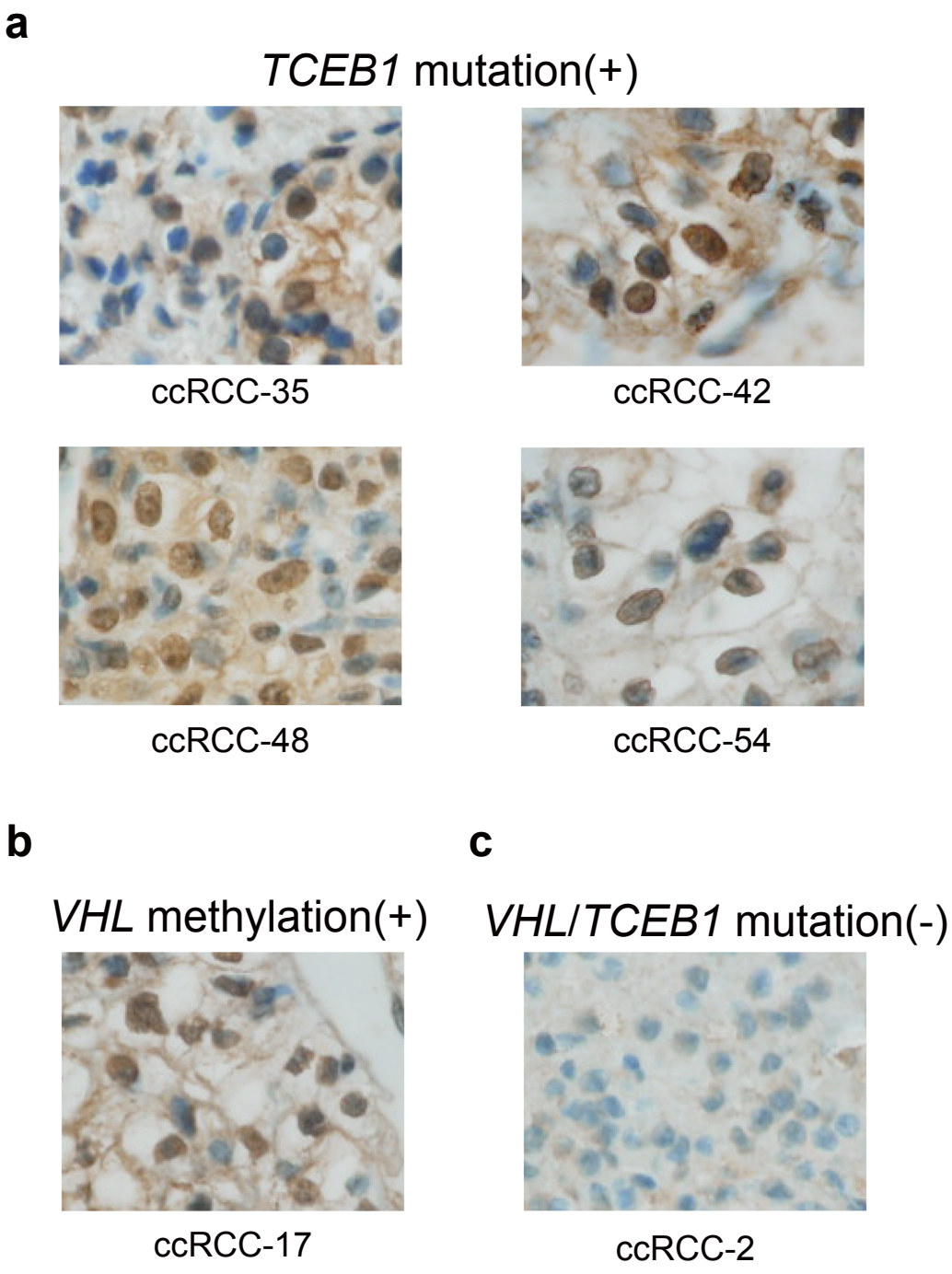
Supplementary Figure 11



Compromised binding of mutant Elongin C (TCEB1) to VHL and other BC-box proteins
(**a**) Western blotting for indicated components of the VHL complex in total cell lysates (left panels), and precipitated with anti-HA (Elongin C) (middle panels) and anti-FLAG (VHL) in lysates from 293T cells transduced with the indicated mock, wild-type or mutant 3xHA-tagged Elongin C, Elongin B and 3xFLAG-tagged VHL. (**b-e**) Western blot analyses of BC-box proteins in total cell lysates (left panels) and precipitated with anti-HA (Elongin C) or anti-FLAG (BC-box protein) (right panels) in lysates from 293T cells transduced with either mock, wild-type, mutant 3xHAtagged Elongin C, Elongin B and 3xFLAG-tagged Elongin A (**b**), SOCS3 (**c**), FEM1B (**d**) and LRR1 (**e**).

Nature Genetics: doi:10.1038/ng.2699

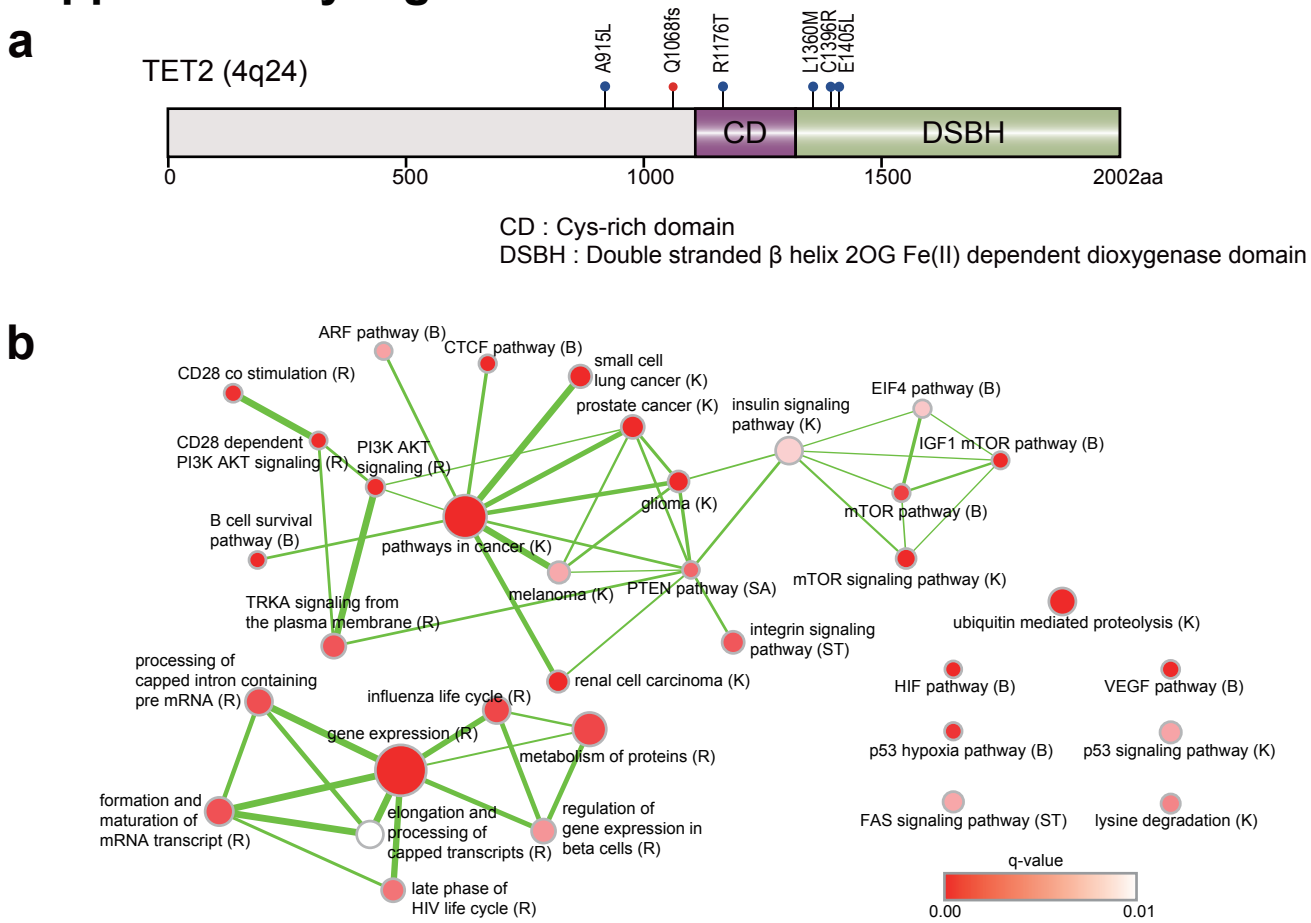
Supplementary Figure 12



HIF1α expression in IHC
Increased HIF expression was confirmed in *TCEB1* mutated tumors (**a**) as well as a tumor with *VHL* promoter methylation (**b**), but not in a tumor without *VHL*/*TCEB1* alterations (**c**).

Nature Genetics: doi:10.1038/ng.2699

Supplementary Figure 13



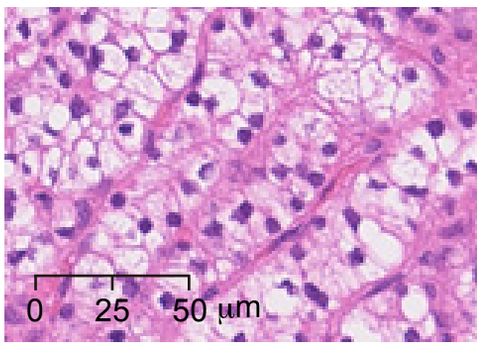
Significantly mutated genes/pathways in 106 exome cases

(**a**) TET2 mutations in ccRCC. Type and position of TET2 mutations are indicated. (**b**) Significantly mutated pathways are shown in red circles based on the significance level (*q* values) as indicated by the color gradient. (B):Biocarta (K):KEGG (R):Reactome (ST):Signaling Transduction KE (**c**) Somatic mutations observed in the apparatus for mRNA processing.

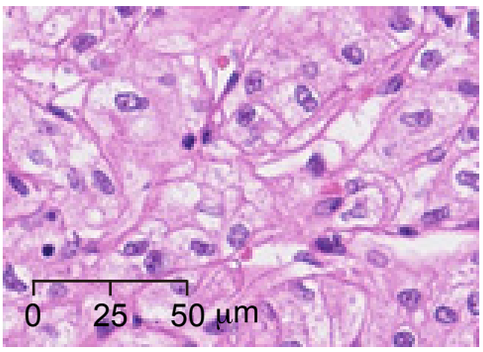
Nature Genetics: doi:10.1038/ng.2699

Supplementary Figure 14

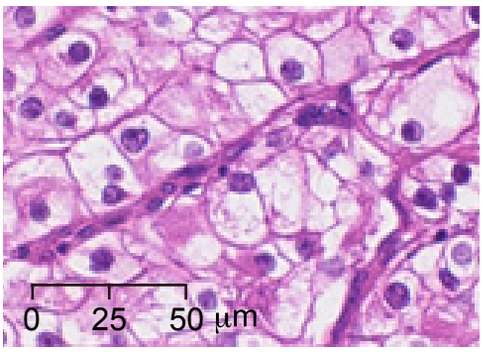
KEAP1 mutation



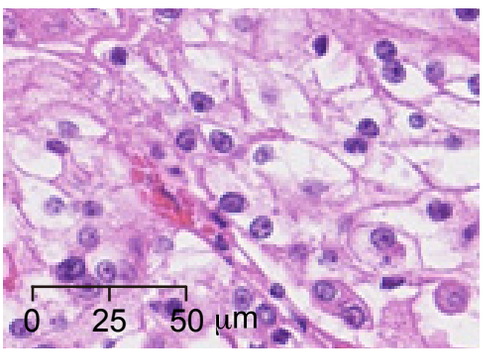
ccRCC-11



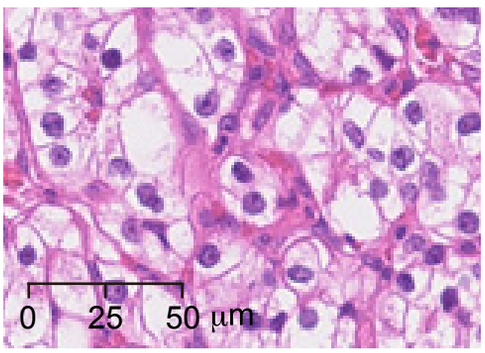
ccRCC-15



ccRCC-34

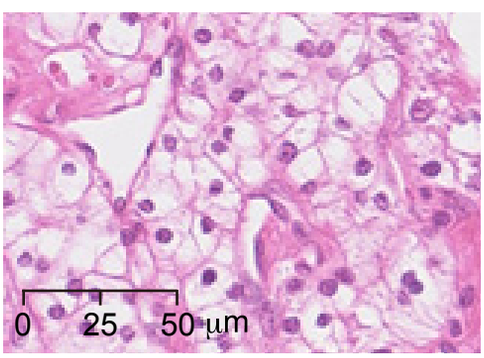


ccRCC-46



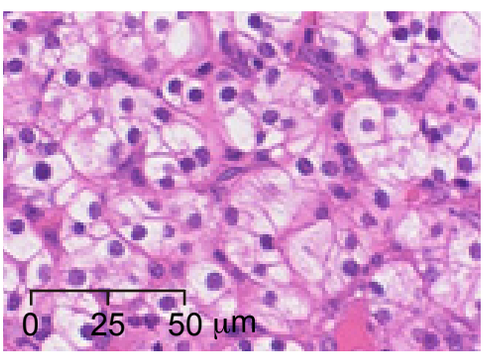
ccRCC-51

NRF2 mutation



ccRCC-82

CUL3 mutation

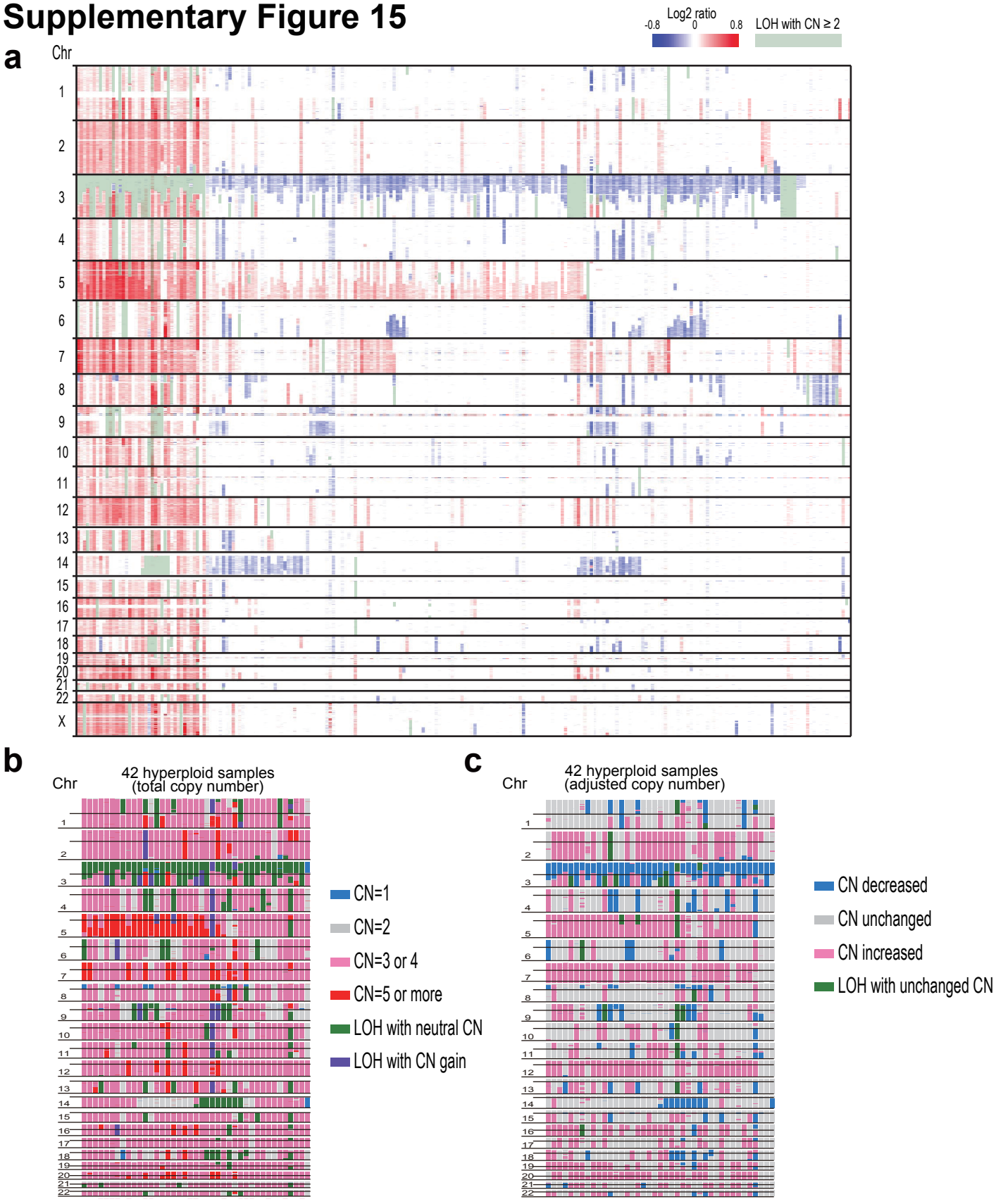


ccRCC-18

Histologies of tumors having *KEAP1*/*NRF2*/*CUL3* mutations.
All cases were confirmed as clear cell RCC with no papillary components on HE staining.

Nature Genetics: doi:10.1038/ng.2699

Supplementary Figure 15

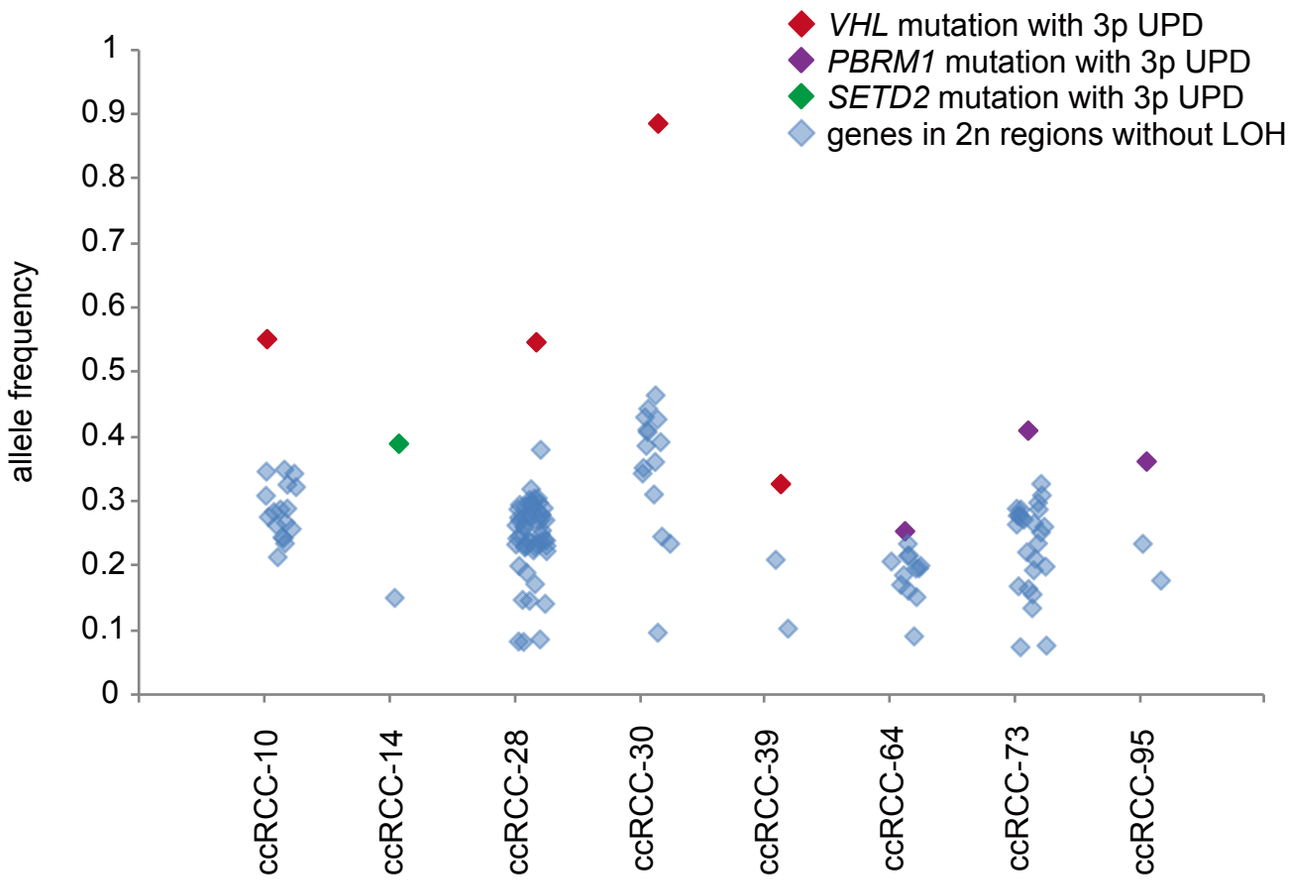


Copy number profiles of 240 ccRCC specimens

(**a**) Genomic copy number determined by SNP array analysis are shown by a color gradient based on CNAG output for 240 ccRCC specimens. Regions showing copy neutral LOH are overlayed in light green. Samples were clustered based on major copy number lesions, including hyperploidy, 3p loss, 5q gain, 14q LOH, 9p LOH and other abnormalities. (**b**) Genomic copy numbers inferred by HMM based analysis of 42 hyperploid samples are shown by the indicated colors. (**c**) The copy number plots for hyperploid cases (b) were transformed by calculating relative copy numbers to the base line copy number (=3), in which the copy number status was either increased (CN > 3, pink), decreased (CN < 3, blue), or unchanged (CN = 3, gray). This relative copy number profile was an essentially identical to that for diploid samples (a), characterized by losses of 3p, 4q, 6q, 9p,9q, 14q and gains of 5q and 7q, suggesting that these hyperploid tumors were most likely progressed from diploid tumors as a relatively late event.

Nature Genetics: doi:10.1038/ng.2699

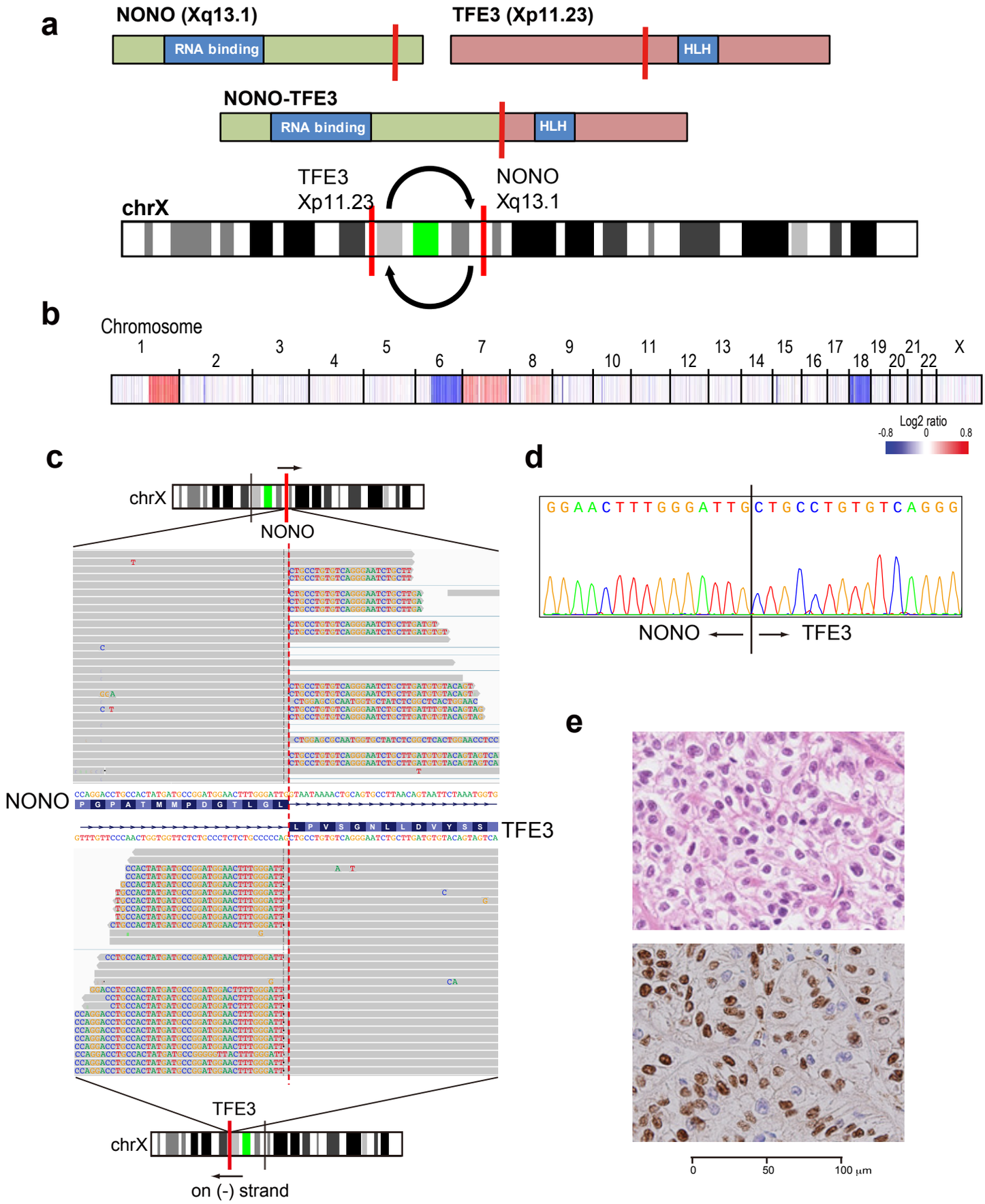
Supplementary Figure 16



Higher allele frequencies of mutations in the 3p target genes in cases with 3p UPD. 8 out of 25 cases with 3p UPD had mutations in either *VHL*, *PBRM1* or *SETD2*, which were analyzed by enough depths (> 50x) with whole exome sequencing for accurate estimation of allele frequency and also had one or more mutations in copy number 2n regions. Allele frequencies of 3p target in 3p UPD were higher than those of other mutations within 2n regions in all 3p UPD cases.

Nature Genetics: doi:10.1038/ng.2699

Supplementary Figure 17

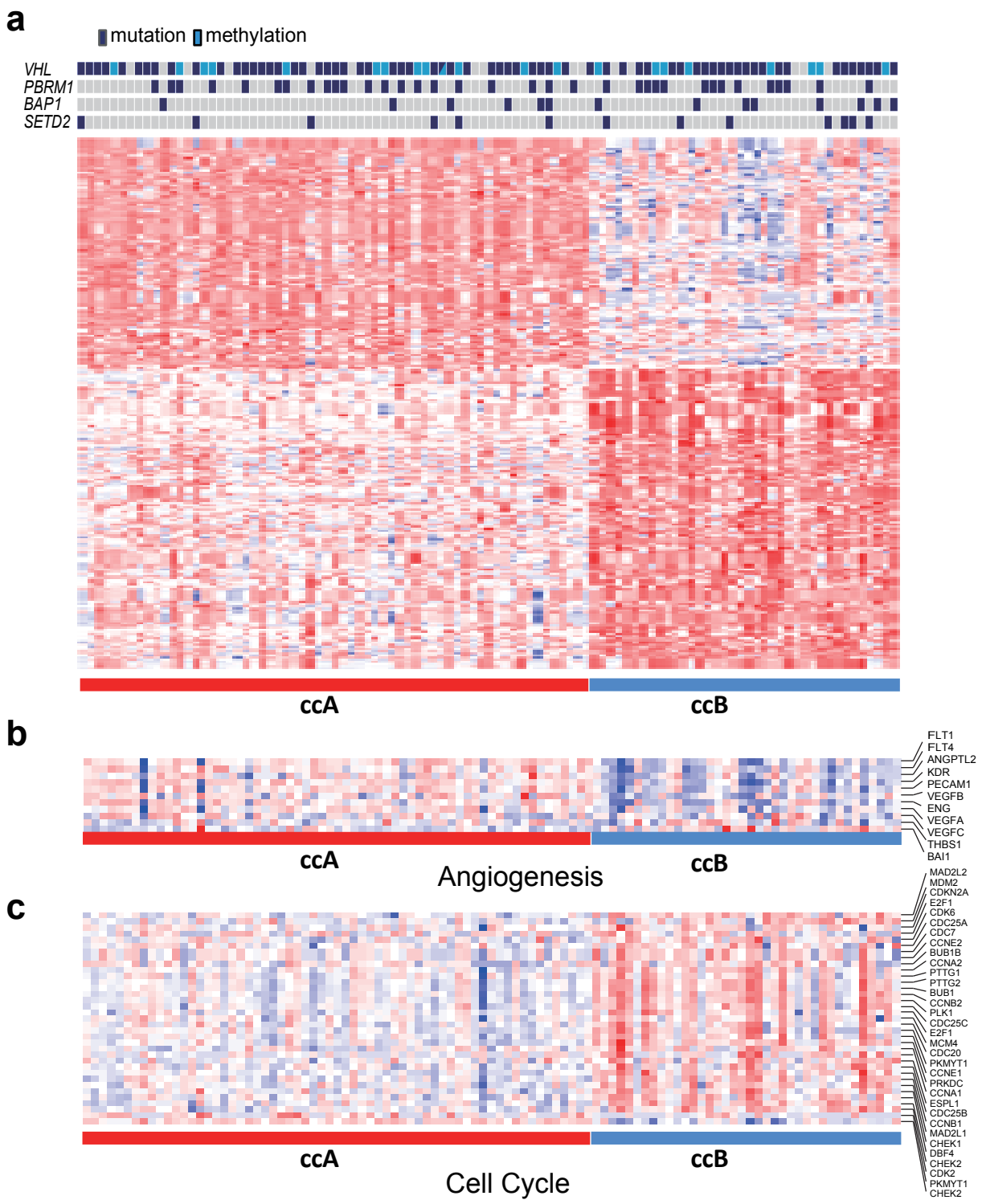


***NONO-TFE3* fusion transcript**

NONO-TFE3 fusion transcript was found in single case (**a**), in which copy number alterations characteristic to ccRCC such as 3p LOH and 5q gain were lacked (**b**). The junction sequence of fusion transcript was showed with IGV viewer (**c**) and confirmed with sanger sequencing (**d**). (**e**) The tumor was positive for TFE3 in IHC (lower panel) but hardly distinguishable from other ccRCC cases on HE staining (upper panel).

Nature Genetics: doi:10.1038/ng.2699

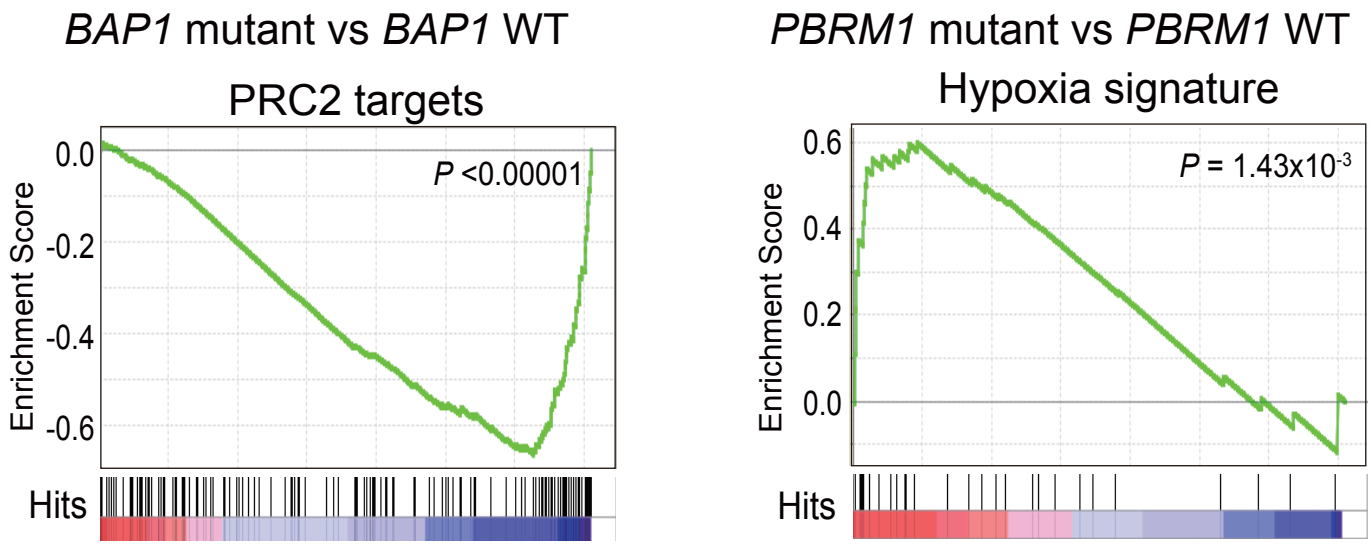
Supplementary Figure 18



ccA and ccB clusters identified from the expression profiles of 101 ccRCC samples
K-means clustering for 101 ccRCC specimens based on the expression of genes showing 2 major gene expression clusters: ccA and ccB (**a**). These were discriminated by the expression of genes involved in angiogenesis (**b**) and cell cycle progression (**c**). Mutation status of *VHL, PBRM1, BAP1* and *SETD2* is indicated in the top panels.

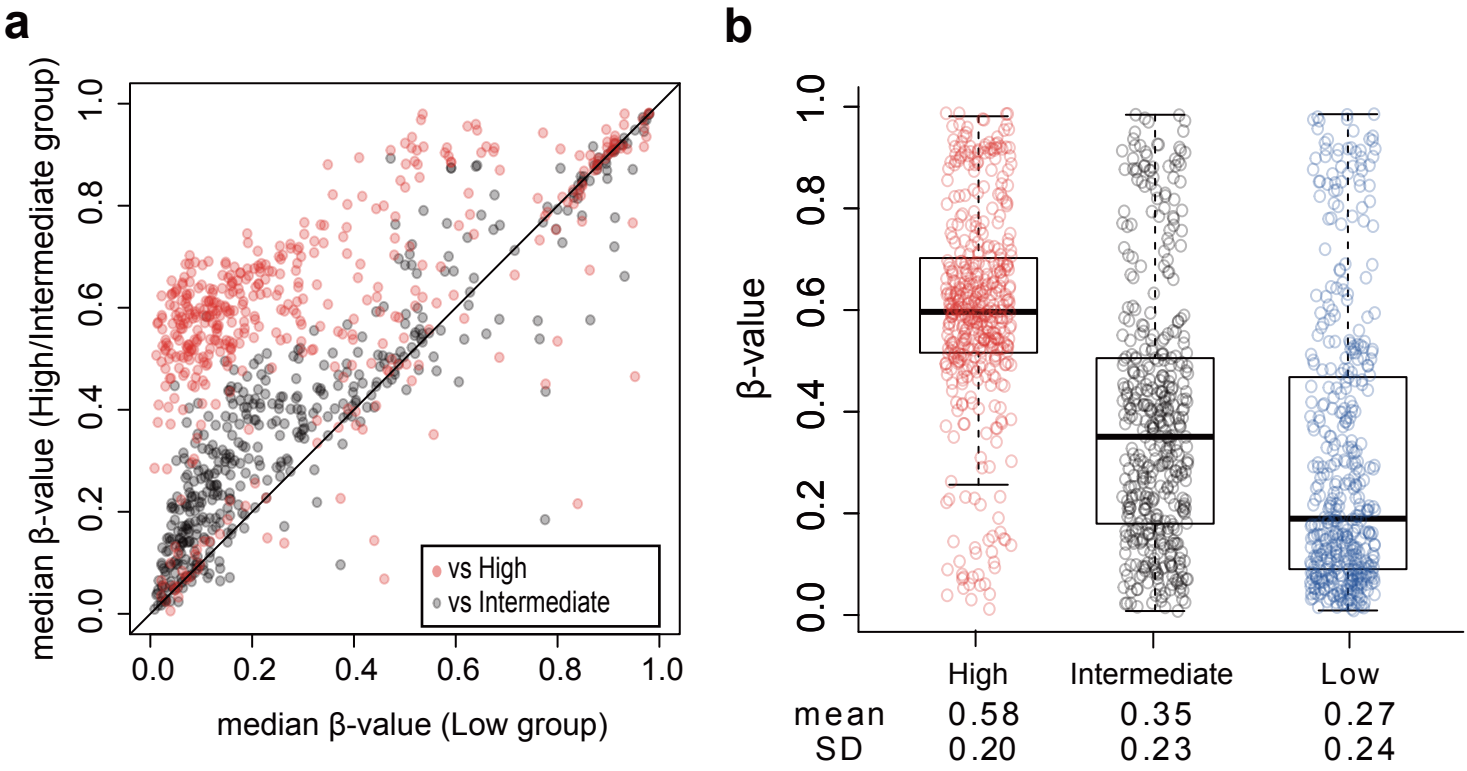
Nature Genetics: doi:10.1038/ng.2699

Supplementary Figure 19



Different expression signatures between *BAP1*- and *PBRM1*-mutated tumors. Gene set enrichment analysis showed *BAP1*-mutated tumors showed significantly down-regulated expression of the PRC2 target genes, whereas *PBRM1*-mutated tumors were enriched for an up-regulated expression of gene set of hipoxia signature.

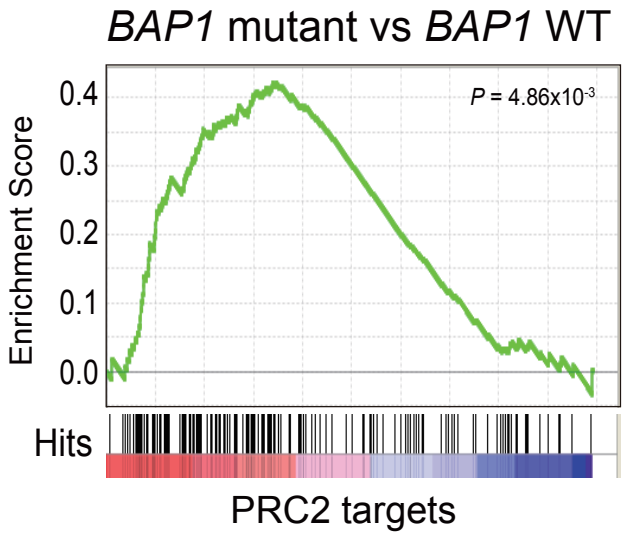
Supplementary Figure 20

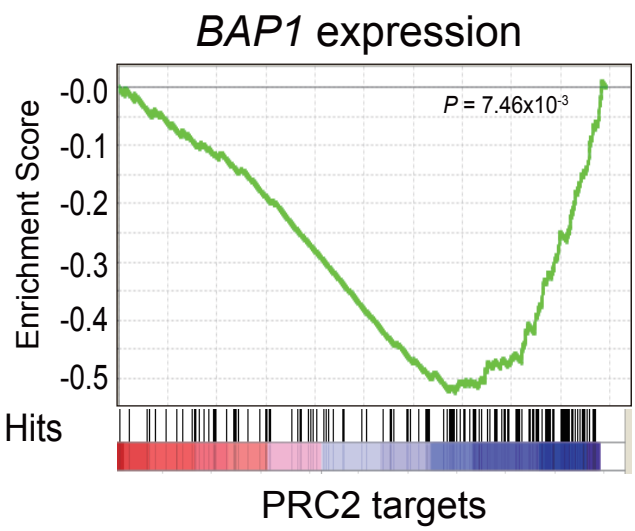


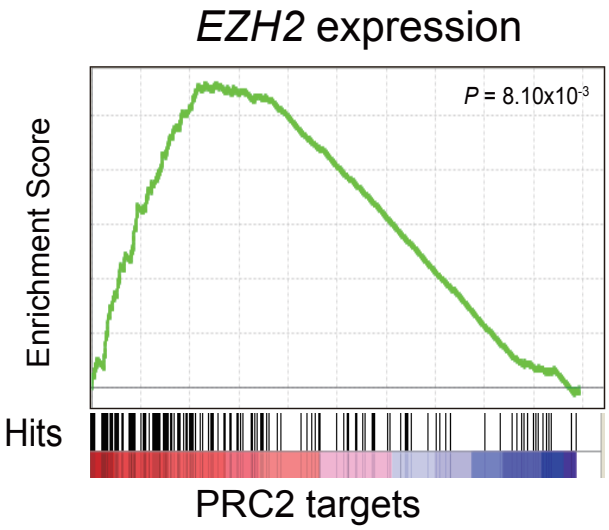
Difference of methylation level among high/intimidate/low methylation clusters
(**a**) Median methylation levels (β values) are plotted for Low and High (red circles) and Low and Intermediate (black circles) methylation clusters for CpG island probes selected for unsupervised clustering analysis in Infinium 450K arrays. (**b**) Distribution of median methylation values (β values) are plotted within each methylation cluster (High, Intermediate and Low).

Nature Genetics: doi:10.1038/ng.2699

Supplementary Figure 21





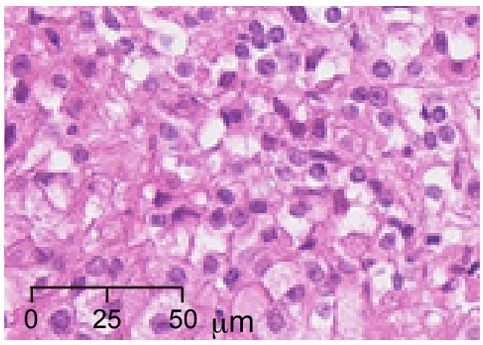


Association between methylation of PRC2 target genes and *BAP1* mutation/expression and *EZH2* expression

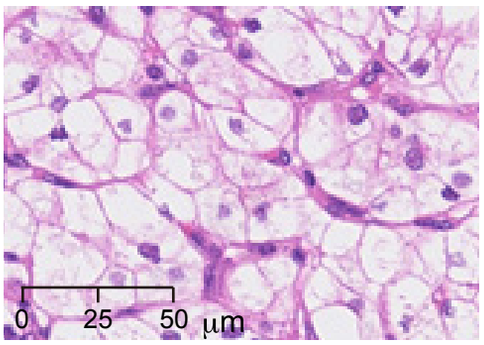
Gene set enrichment analysis showed *BAP1* mutation, decreased *BAP1* expression and increased *EZH2* expression were significantly associated with increased methylation of PRC2 target genes.

Nature Genetics: doi:10.1038/ng.2699

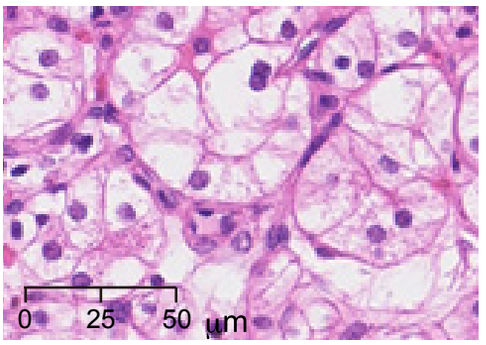
Supplementary Figure 22



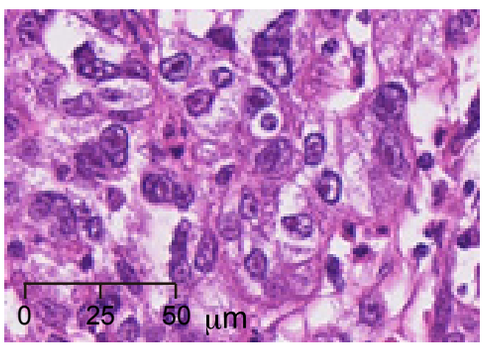
ccRCC-2



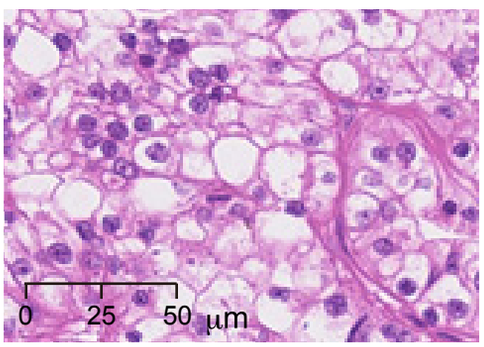
ccRCC-31



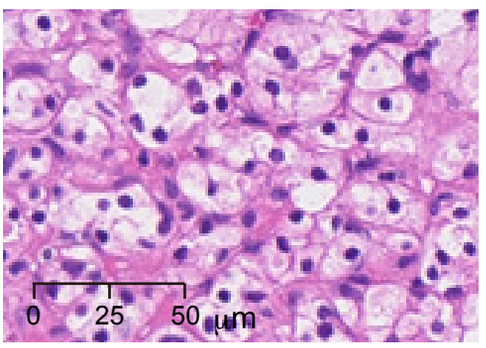
ccRCC-37



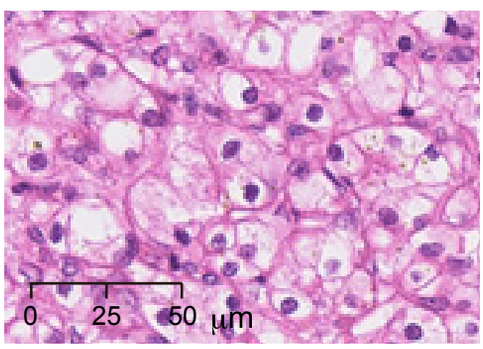
ccRCC-41



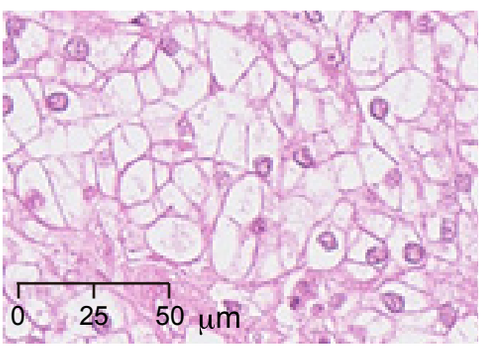
ccRCC-57



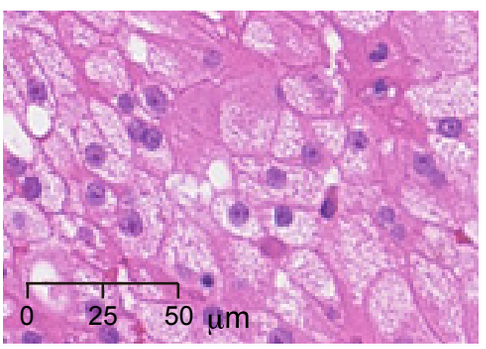
ccRCC-60



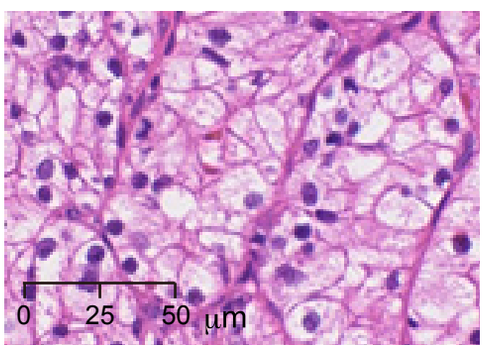
ccRCC-79



ccRCC-81



ccRCC-138



ccRCC-184

Histologies of tumors having no *VHL* or *TCEB1* alterations.
All cases were confirmed as clear cell RCC on HE staining .

Nature Genetics: doi:10.1038/ng.2699

Supplementary Table 5 Multivariate analysis of prognosis for 240 ccRCC cases

Factors		No. of cases	Overall survival		Desease free survival	
			Relative risk	*P* value	Relative risk	*P* value
PBRM1				0.075		0.115
	WT	142				
	mut	98	1.82 (0.940-3.55)		1.56 (0.896-2.73)	
BAP1				0.020		0.078
	WT	215				
	mut	25	3.11 (1.22-6.97)		2.13 (0.912-4.41)	
SETD2				0.266		0.001
	WT	214				
	mut	26	1.69 (0.637-3.78)		3.24 (1.66-5.93)	

Nature Genetics: doi:10.1038/ng.2699

Supplementary Table 6. Clinicopathological characteristics of *VHL* - and *TCEB1* -mutated tumors

Factors	Categories	VHL-mut	TCEB1-mut	*P* value
age		61.3 (27-91)	62.5 (42-77)	0.815
sex	M	168 (76.4%)	8 (100%)	0.218
	F	52 (23.6%)	0	
tumor diameter		4.2 (3-6.125)	3 (1.2-4.025)	0.0680
pT	1a	116 (52.7%)	5 (62.5%)	0.744
	1b	55 (25.0%)	2 (25.0%)	
	2	18 (8.2%)	0	
	3	29 (13.2%)	1 (12.5%)	
	4	2 (0.9%)	0	
N	0	211 (95.9%)	8 (100%)	0.230
	1	5 (2.3%)	0	
	2	4 (1.8%)	0	
M	0	196 (89.1%)	8 (100%)	1.00
	1	24 (10.9%)	0	
metastasis duaring	+	170 (77.3%)	8 (100%)	0.377
	-	50 (22.7%)	0	
Fuhrman grade	1	38 (17.4%)	2 (25.0%)	0.314
	2	130 (59.6%)	6 (75.0%)	
	3	45 (20.6%)	0	
	4	5 (2.3%)	0	
outcome	alive	185 (84.1%)	8 (100%)	0.277
	dead	35 (15.9%)	0	

Nature Genetics: doi:10.1038/ng.2699

Supplementary Table 7a. Knockdown efficacy of siRNA for TCEB1 mRNA

TCEB1 vector	-			mock			WT			Y79C			A100P		
siRNA	TCEB1	NC	-	TCEB1	NC	-	TCEB1	NC	-	TCEB1	NC	-	TCEB1	NC	-
endogenous *TCEB1*	0.028	0.79	0.87	0.021	0.91	0.93	0.033	0.89	1.10	0.028	1.02	0.99	0.026	0.97	0.82
GAPDH	0.99	0.92	0.95	1.09	1.01	1.14	0.85	0.92	1.28	1.20	1.03	1.09	1.02	1.05	0.91
endogenous *TCEB1* / *GAPDH*	0.028	0.86	0.91	0.019	0.90	0.82	0.039	0.96	0.86	0.024	0.99	0.91	0.026	0.93	0.91
ratio to nontreated-cell (%)	3.10	94.7	100	2.12	98.5	90.0	4.27	105	93.7	2.60	108	99.4	2.81	102	99.1

- : not treated
NC : negative control

Supplementary Table 7b. List of siRNA used for knocking down of TCEB1

TCEB1	5'-CUAUCGAAAGUAUGCAUGUTT-3'
Non-targeting negative control	5'-UCUUAAUCGCGGUAUAAGGC-3'

Supplementary Table 7c. List of primers used for quantitative RT PCR

	Forward primer	Reverse primer
endogenous *TCEB1*	GGCTGCGGGACTGACGAGAAAC	GACCTGGGCCACTCAACATGGC
GAPDH	ACTGGCATGGCCTTCCGTGT	ATGCCAGCCCCAGCGTCAAA

Nature Genetics: doi:10.1038/ng.2699

Supplementary Table 8. Results of immunostainig of HIF1α and HIF2α

sample ID	HIF1α	HIF2α	*VHL*	*TCEB1*
ccRCC-1	1+	1+	mutation	-
ccRCC-2	-	-	-	-
ccRCC-3	3+	2+	mutation	-
ccRCC-4	2+	2+	methylation	-
ccRCC-5	-	1+	mutation	-
ccRCC-6	1+	1+	mutation	-
ccRCC-7	2+	3+	mutation	-
ccRCC-8	1+	3+	methylation	-
ccRCC-9	1+	-	mutation	-
ccRCC-10	-	1+	mutation	-
ccRCC-11	1+	2+	mutation	-
ccRCC-12	1+	1+	mutation	-
ccRCC-13	2+	2+	mutation	-
ccRCC-14	1+	-	mutation	-
ccRCC-15	3+	-	mutation	-
ccRCC-16	1+	2+	mutation	-
ccRCC-17	3+	3+	methylation	-
ccRCC-18	2+	2+	mutation	-
ccRCC-19	3+	-	mutation	-
ccRCC-20	2+	-	mutation	-
ccRCC-21	2+	1+	methylation	-
ccRCC-22	-	1+	mutation	-
ccRCC-23	1+	1+	methylation	-
ccRCC-24	2+	-	mutation	-
ccRCC-25	1+	1+	methylation	-
ccRCC-26	-	-	mutation	-
ccRCC-27	2+	2+	-	mutation
ccRCC-28	1+	-	mutation	-
ccRCC-29	2+	1+	mutation	-
ccRCC-30	-	-	mutation	-
ccRCC-31	1+	-	-	-
ccRCC-32	2+	2+	methylation	-
ccRCC-33	2+	1+	mutation	-
ccRCC-34	1+	1+	mutation	-
ccRCC-35	2+	2+	-	mutation
ccRCC-36	2+	1+	mutation	-
ccRCC-37	3+	2+	-	-
ccRCC-38	2+	3+	methylation	-
ccRCC-39	-	-	mutation	-
ccRCC-40	1+	1+	mutation	-
ccRCC-41	-	1+	-	-
ccRCC-42	3+	1+	-	mutation
ccRCC-43	1+	-	methylation	-
ccRCC-44	3+	1+	methylation	-
ccRCC-45	-	1+	methylation	-
ccRCC-46	-	1+	mutation	-
ccRCC-47	2+	1+	mutation	-
ccRCC-48	2+	-	-	mutation
ccRCC-49	-	-	mutation	-
ccRCC-50	2+	1+	mutation	-
ccRCC-51	-	1+	mutation	-
ccRCC-52	2+	2+	mutation	-
ccRCC-53	2+	3+	methylation	-
ccRCC-54	1+	-	-	mutation
ccRCC-55	2+	1+	mutation	-
ccRCC-56	2+	1+	mutation	-
ccRCC-57	2+	-	-	-
ccRCC-58	1+	1+	mutation	-

Nature Genetics: doi:10.1038/ng.2699

ccRCC-59	1+	-	mutation	-
ccRCC-60	2+	2+	-	-
ccRCC-61	2+	2+	mutation	-
ccRCC-62	3+	-	mutation	-
ccRCC-63	2+	1+	mutation	-
ccRCC-64	-	1+	mutation	-
ccRCC-65	-	-	mutation	-
ccRCC-66	3+	2+	mutation	-
ccRCC-67	3+	-	mutation	-
ccRCC-68	1+	1+	mutation	-
ccRCC-69	2+	1+	mutation	-
ccRCC-70	3+	1+	methylation	-
ccRCC-71	3+	-	mutation	-
ccRCC-72	2+	1+	mutation	-
ccRCC-73	3+	3+	mutation	-
ccRCC-74	2+	1+	methylation	-
ccRCC-75	1+	1+	mutation	-
ccRCC-76	2+	1+	methylation	-
ccRCC-77	3+	1+	mutation	-
ccRCC-78	1+	-	mutation	-
ccRCC-79	2+	1+	-	-
ccRCC-80	2+	1+	mutation	-
ccRCC-81	1+	-	-	-
ccRCC-82	3+	1+	mutation	-
ccRCC-83	1+	-	mutation	-
ccRCC-84	1+	1+	methylation	-
ccRCC-85	3+	1+	mutation	-
ccRCC-86	2+	-	methylation	-
ccRCC-87	2+	1+	mutation	-
ccRCC-88	3+	1+	mutation	-
ccRCC-89	2+	-	methylation	-
ccRCC-90	2+	1+	methylation	-
ccRCC-91	-	1+	mutation	-
ccRCC-92	2+	1+	methylation	-
ccRCC-93	1+	1+	mutation	-
ccRCC-94	-	1+	mutation	-
ccRCC-95	3+	2+	mutation	-
ccRCC-96	3+	1+	methylation	-
ccRCC-97	1+	1+	mutation	-
ccRCC-98	-	-	-	-
ccRCC-99	2+	2+	mutation	-
ccRCC-100	1+	1+	mutation	-
ccRCC-101	1+	-	mutation	-
ccRCC-102	3+	1+	mutation	-
ccRCC-103	3+	2+	mutation	-
ccRCC-104	1+	1+	methylation	-
ccRCC-105	1+	1+	mutation	-
ccRCC-106	2+	1+	mutation	-

Nature Genetics: doi:10.1038/ng.2699

SupplementaryTable 9. Significantly enriched mutational pathways

Gene set	Number of bases in gene set	Number of mutation	*P* value	*q* value
pathways in cancer (K)	647781	131	< 1E-16	< 1E-16
VEGF pathway (B)	52089	48	8.79.E-17	3.87.E-14
HIF pathway (B)	25881	43	1.28.E-16	3.76.E-14
renal cell carcinoma (K)	117204	65	1.33.E-16	2.93.E-14
ubiquitin mediated proteolysis (K)	280992	73	4.85.E-15	8.53.E-13
CTCF pathway (B)	38646	20	8.61.E-10	1.26.E-07
prostate cancer (K)	161541	40	2.23.E-08	2.81.E-06
IGF1mTOR pathway (B)	36615	17	7.05.E-08	7.75.E-06
glioma (K)	114846	31	1.30.E-07	1.28.E-05
small cell lung cancer (K)	208560	44	3.74.E-07	3.29.E-05
mTOR signaling pathway (K)	96279	27	3.94.E-07	3.15.E-05
CD28 dependent PI3K AKT signaling (R)	36003	15	1.51.E-06	1.11.E-04
PI3K AKT signalling (R)	74460	22	1.96.E-06	1.33.E-04
gene expression (R)	671376	99	2.22.E-06	1.39.E-04
bcellsurvival pathway (B)	33207	14	2.91.E-06	1.71.E-04
CD28 co stimulation (R)	49233	16	1.50.E-05	8.24.E-04
p53hypoxia pathway (B)	50592	16	2.07.E-05	1.07.E-03
mTOR pathway (B)	52908	16	3.50.E-05	1.71.E-03
metabolism of proteins (R)	245715	43	4.31.E-05	2.00.E-03
influenza life cycle (R)	176829	34	4.49.E-05	1.98.E-03
formation and maturation of mRNA transcript (R)	225633	40	5.93.E-05	2.48.E-03
processing of capped intron containing pre mRNA (R)	249408	43	5.97.E-05	2.39.E-03
TRKA signalling from the plasma membrane (R)	211965	38	7.11.E-05	2.72.E-03
integrin signaling pathway (ST)	181737	34	7.52.E-05	2.76.E-03
PTEN pathway (SA)	29622	11	9.87.E-05	3.48.E-03
late phase of HIV life cycle (R)	178863	33	1.22.E-04	4.12.E-03
lysine degradation (K)	106773	23	1.48.E-04	4.82.E-03
regulation of gene expression in beta cells (R)	67380	17	1.76.E-04	5.53.E-03
ARF pathway (B)	32121	11	1.97.E-04	5.98.E-03
FAS signaling pathway (ST)	116727	24	2.12.E-04	6.21.E-03
p53 signaling pathway (K)	102570	22	2.15.E-04	6.11.E-03
melanoma (K)	96036	21	2.30.E-04	6.32.E-03
EIF4 pathway (B)	50982	14	2.76.E-04	7.36.E-03
insulin signaling pathway (K)	260913	42	2.95.E-04	7.64.E-03
elongation and processing of capped transcripts (R)	198138	34	3.55.E-04	8.91.E-03

(B) : Biocarta (K) : KEGG (R) : Reactome (SA) : SigmaAldrich (ST) : Signaling Transduction KE

Nature Genetics: doi:10.1038/ng.2699

Supplementary Table 15. List of antibodies for immunoblot analysis

	Target	Host	Company	Catalog Number	Dilution
Primary antibody	Elongin C	Mouse	BD Bioscience	610761	1:1000
	Elongin B	Rabbit	previously described		1:500
	HIF-1a	Mouse	BD Bioscience	610958	1:1000
	CUL2	Rabbit	Invitrogen	511800	1:500
	VHL	Mouse	BD Bioscience	556347	1:500
	Actin	Goat	Santa Cruz Biotechnology	sc-1616	1:4000
	HA	Rabbit	Covance	PRB-101C	1:4000
	HA	Mouse	Covance	MMS-101P	1:4000
	Flag	Rabbit	Sigma-Aldrich	F425	1:4000
	Flag	Mouse	Wako Pure Chemical Industries	018-22381	1:4000
Secondary HRP-conjugated antibody	Anti-mouse IgG	Sheep	GE Healthcare Life Science	NA931	1:8000
	Anti-rabbit IgG	Donkey	GE Healthcare Life Science	NA934	1:8000
	Anti-goat IgG	Goat	Santa Cruz Biotechnology	sc-2033	1:8000

Nature Genetics: doi:10.1038/ng.2699

Supplementary Note.

Tumor samples with matched normal tissue or blood were obtained at the time of surgery from the University of Tokyo. All patients provided informed consent as a part of the ethics committee of the Graduate School of Medicine, the University of Tokyo. Collected phonotypic data elements were deidentified.

Nature Genetics: doi:10.1038/ng.2699