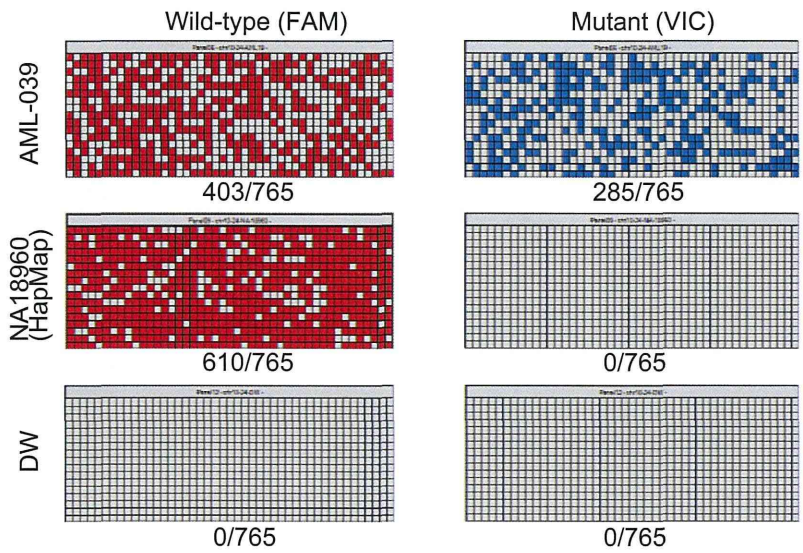
Supplementary Figure 5

a

Mutations	Sample	Alleles	Labelling	Forward (5'->3')	Reverse (5'->3')
SMC3 (1099C>G)	MDS-060	wild-type	FAM	TGGCT**C**AAGCTAC	TCCCATTTGCTTCTATTTTGTGTAGA
		mutant	VIC	TGGCT**G**AAGCTAC	CTGCTTTGCATAAAGATCCGTTCT
SMC3 (1175A>T)	AML-039	wild-type	FAM	AGAAGAAAGGG**A**TAAGTG	GGGTCGAGGAAGCCAGTTTAC
		mutant	VIC	AGAAGAAAGGG**T**TAAGTG	GCAATCTGTCTTTTCTTGTCATTAATAGC

b

c



d

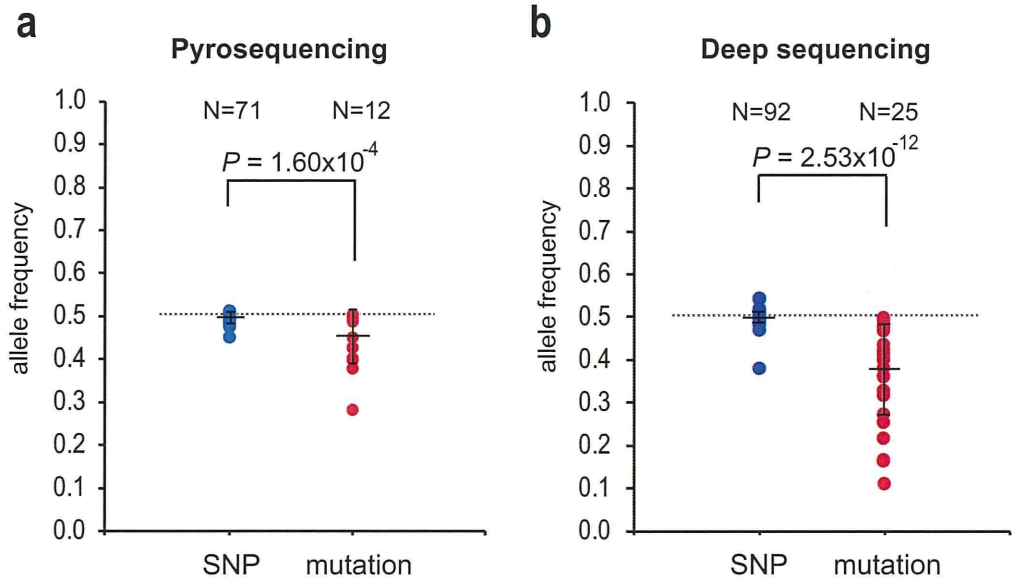
Mutations		WT			Mut			WT/Mut (99%CI)
		DW	NA18960	Tumor	DW	NA18960	Tumor	
SMC3 (1099C>G)	# Positive wells	0	616	4793	0	0	4489	1.100 (1.041-1.162)
	# Total wells	1530	765	9180	1530	765	9180	
SMC3 (1175A>T)	# Positive wells	0	610	788	0	0	588	1.492 (1.295-1.724)
	# Total wells	765	765	1530	765	765	1530	

Measurements of frequencies of mutant alleles by digital PCR

In digital PCR on the Fluidigm system, the presence or absence of the wild-type (WT) and mutant (Mut) is determined by allele-specific PCR using differentially labeled (FAM- or VIC-) primers (**a**) in each of a large number of micro-wells, in which samples are properly diluted so that each well contains less than one molecule on average to allow for determination of the concentration of each allele according to the numbers of observed positive wells, assuming a Poisson's distribution. Representative assays are shown in **b** and **c**. Each PCR showed discrete distributions of signals between productive and non-productive (**b** and **c**, top and middle panels) and no false positive amplifications were observed in samples with no DNA (DW) for both wild-type and mutant-specific PCRs (**c**, bottom panels) as well in normal DNA (NA18960 from the HapMap panel) for mutant-specific PCRs (**c**, right middle panel). The results of measurements are summarized in **d**, where total number of positive and negative wells for each allele-specific PCR for control and test samples, together with observed WT/Mut allelic ratios with their 99% confidential intervals are shown.

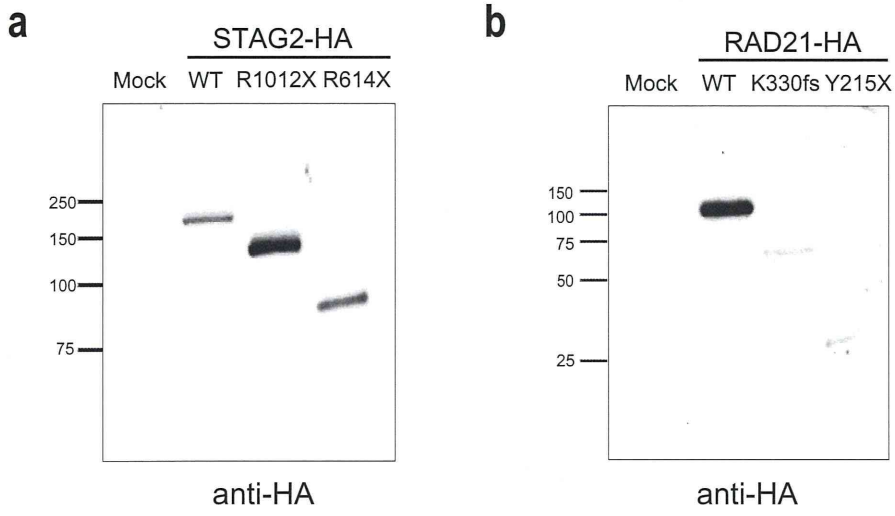
Nature Genetics: doi:10.1038/ng.2731

Supplementary Figure 6



Measurements of frequencies of mutant alleles by pyrosequencing and deep sequencing
Frequencies of mutant alleles together with those of SNP alleles were measured by pyrosequencing **(a)** and deep sequencing **(b)**. The mean ± S.D. value of measured allele frequencies are also indicated. All mutant and SNP loci did not show copy number alteration or copy neutral loss-of-heterozygosity in SNP array karyotyping. In the absence of chromosomal gains or deletions, the difference in the frequency between mutant and wild-type alleles suggests the somatic origin of mutations. In both assays, the mutant alleles showed significantly lower allele frequencies compared to known SNP alleles (Mann-Whitney's U test), supporting the hypothesis that many, if not all, of the mutations are not SNPs but represent somatic changes.

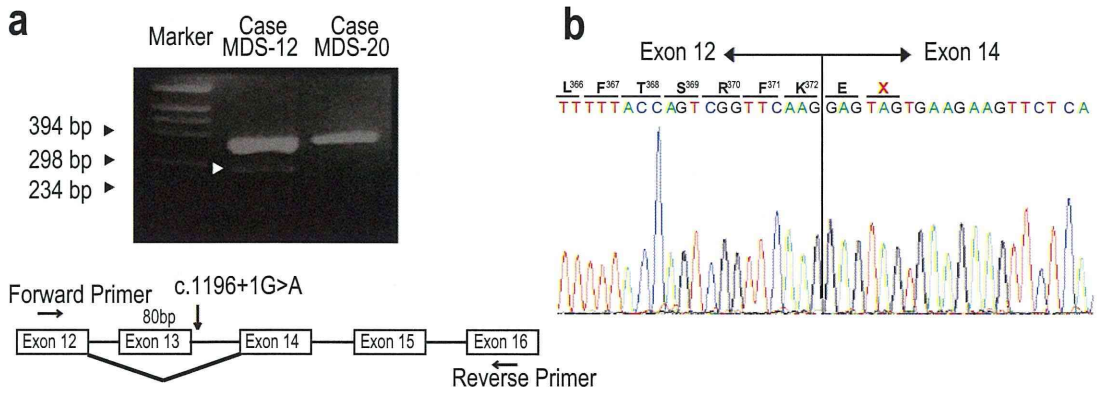
Supplementary Figure 7



Expression of truncating *STAG2* and *RAD21* alleles in HeLa cells
Anti-HA immunoblots of total lysates from HeLa cells transduced with HA-tagged wild-type and truncating alleles of *STAG2* (**a**) and *RAD21* (**b**) showing truncated STAG2 and RAD21 proteins having expected molecular weights, respectively. Transduced truncating mutants are indicated on top of each panel.

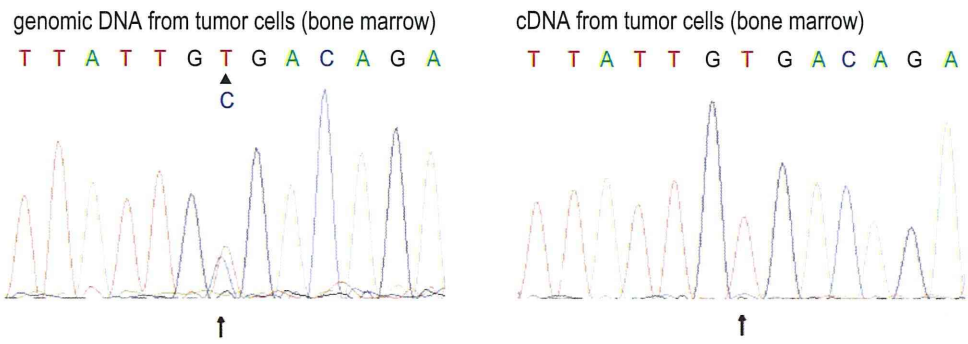
Nature Genetics: doi:10.1038/ng.2731

Supplementary Figure 8



Aberrant splicing of STAG2 transcript caused by a splice site mutation in MDS-12
(a). RT PCR analysis of transcripts from MDS-12 shows an aberrant PCR product having a smaller fragment length (indicated by an while arrowhead) in addition to a major product, which is also present in a control case (MDS-20). The difference in intensity of both bands is thought to reflect the low tumor content in this sample, which was also indicated from the lower mutation peak in the genomic sequencing (Supplementary Fig.1b, lower panel). The primer design is shown at the bottom. The position of the splice site mutation is shown by a vertical arrow. **(b).** Sanger sequencing of the aberrant RT-PCR product confirmed the aberrant splicing that causes skipping of exon 13, leading to a premature stop codon. Primer sequences are provided in **Supplementary Table 17**.

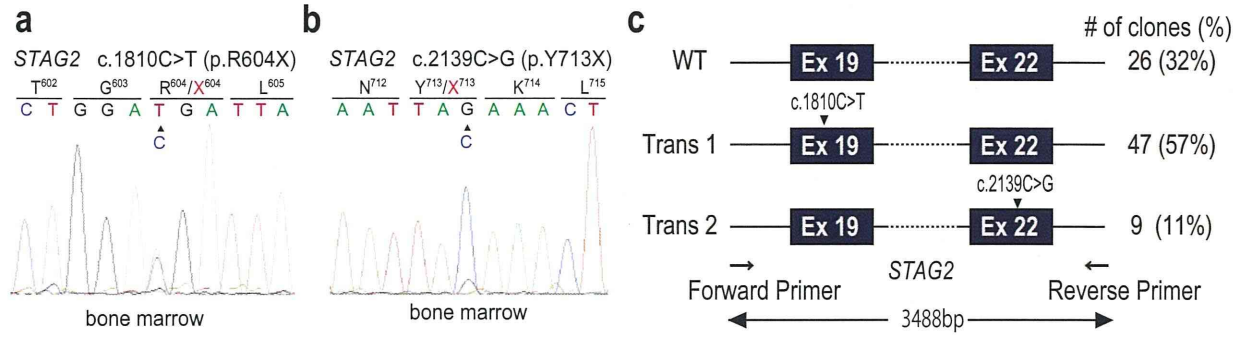
Supplementary Figure 9



Expression of *STAG2* from the mutated allele in case CMML-036
Results of Sanger sequencing of genomic DNA and cDNA from a female patient (CMML-036) carrying a heterozygous *STAG2* mutation (c.1840C>T). While genomic DNA from tumor cells shows comparable C (wild-type) and T (mutated) signals (left) at the nucleotide position 1840, the tumor-derived cDNA displays a T signal (right), indicating that the *STAG2* message was expressed exclusively from the mutated rather than the wild-type allele on the inactivated X chromosome. Primer sequences are provided in **Supplementary Table 17**.

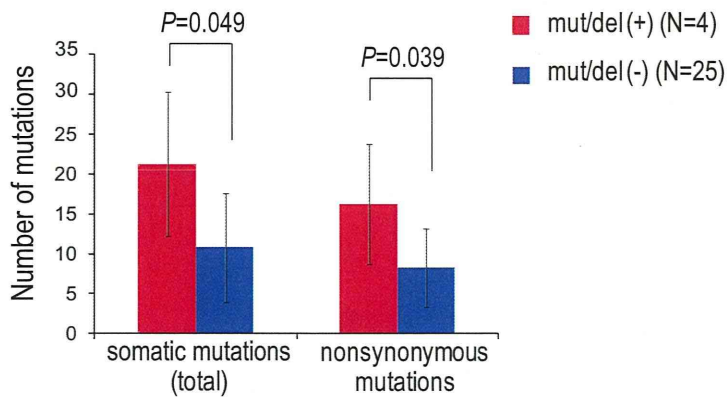
Nature Genetics: doi:10.1038/ng.2731

Supplementary Figure 10



Two independent *STAG2* mutations in a male case (MDS-176)
Sanger sequencing showing two *STAG2* mutations, 1810C>T **(a)** and 2139C>G **(b)**, found in a male case (MDS-176). The presence of the wild-type signals in both chromatogram indicates that these mutations were somatic.
(c) To determine the phase of these two mutations, a 3488 bp *STAG2* gene fragment containing both mutation sites was amplified by long PCR using primeSTAR Max DNA Polymerase (Takara), which was subcloned into TOPO TA cloning vector and sequenced for the relevant mutation sites. The distribution of mutations supports the model that two mutations are harbored in different subclones, one major (1810C>T) and the other minor (2139C>G) populations. This is consistent with the a smaller peak for 2139C>G (b) compared to 1810C>T. The presence of two independent STAG2 mutant alleles in a single case underscores the important driver role of these mutations. Primer sequences are shown in **Supplementary Table 17**.

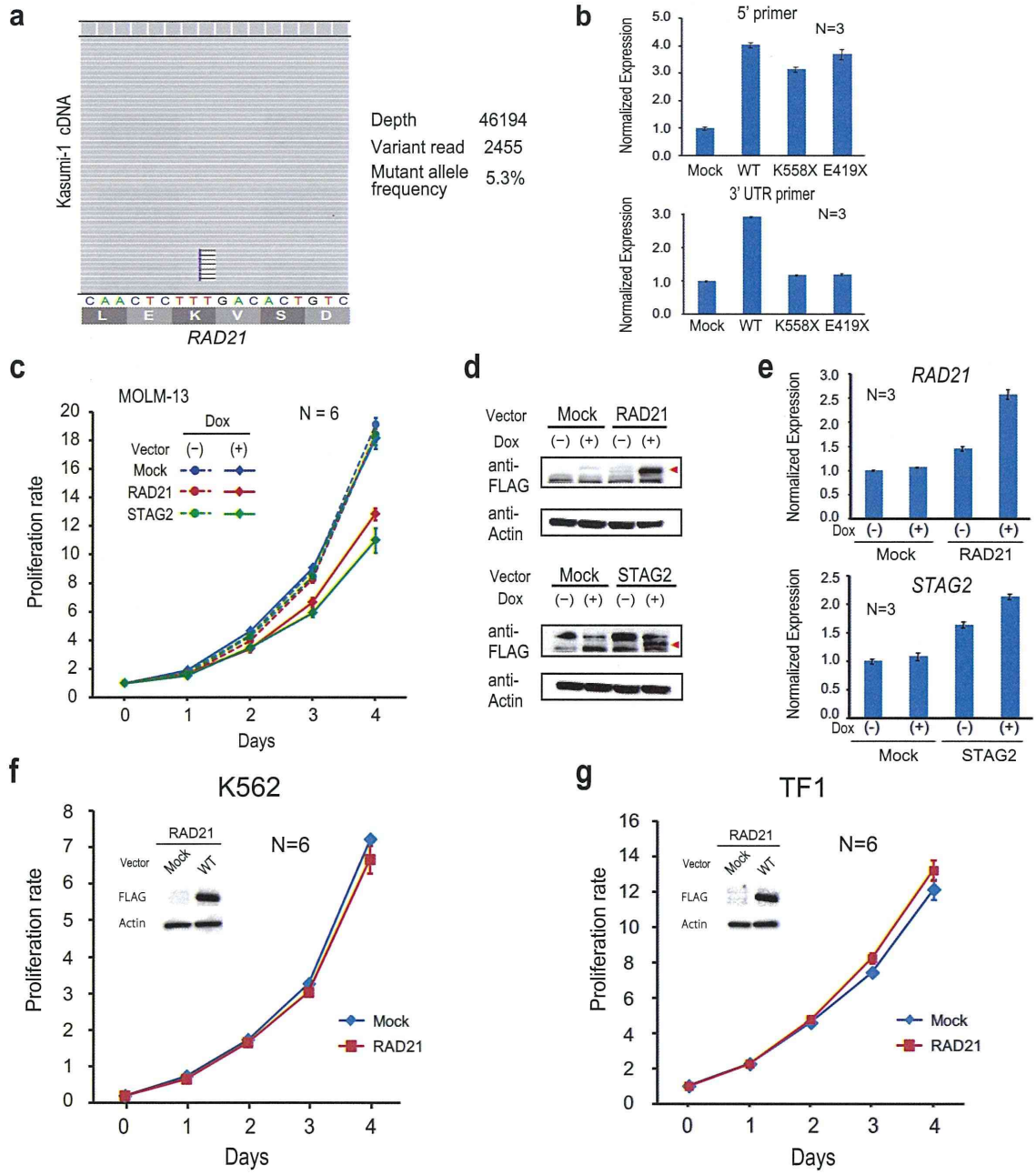
Supplementary Figure 11



Number of mutations in the discovery cohort
The impact of cohesin mutations on the number of coexisting somatic mutations was tested by Mann-Whitney U test based on the result of whole exome study. Cohesin-mutated/deleted cases (red) show significantly higher numbers of total somatic mutations and non-silent mutations as determined by whole exome sequencing, compared to those having no cohesin mutations/deletions (blue). Means±SD are shown.

Nature Genetics: doi:10.1038/ng.2731

Supplementary Figure 12

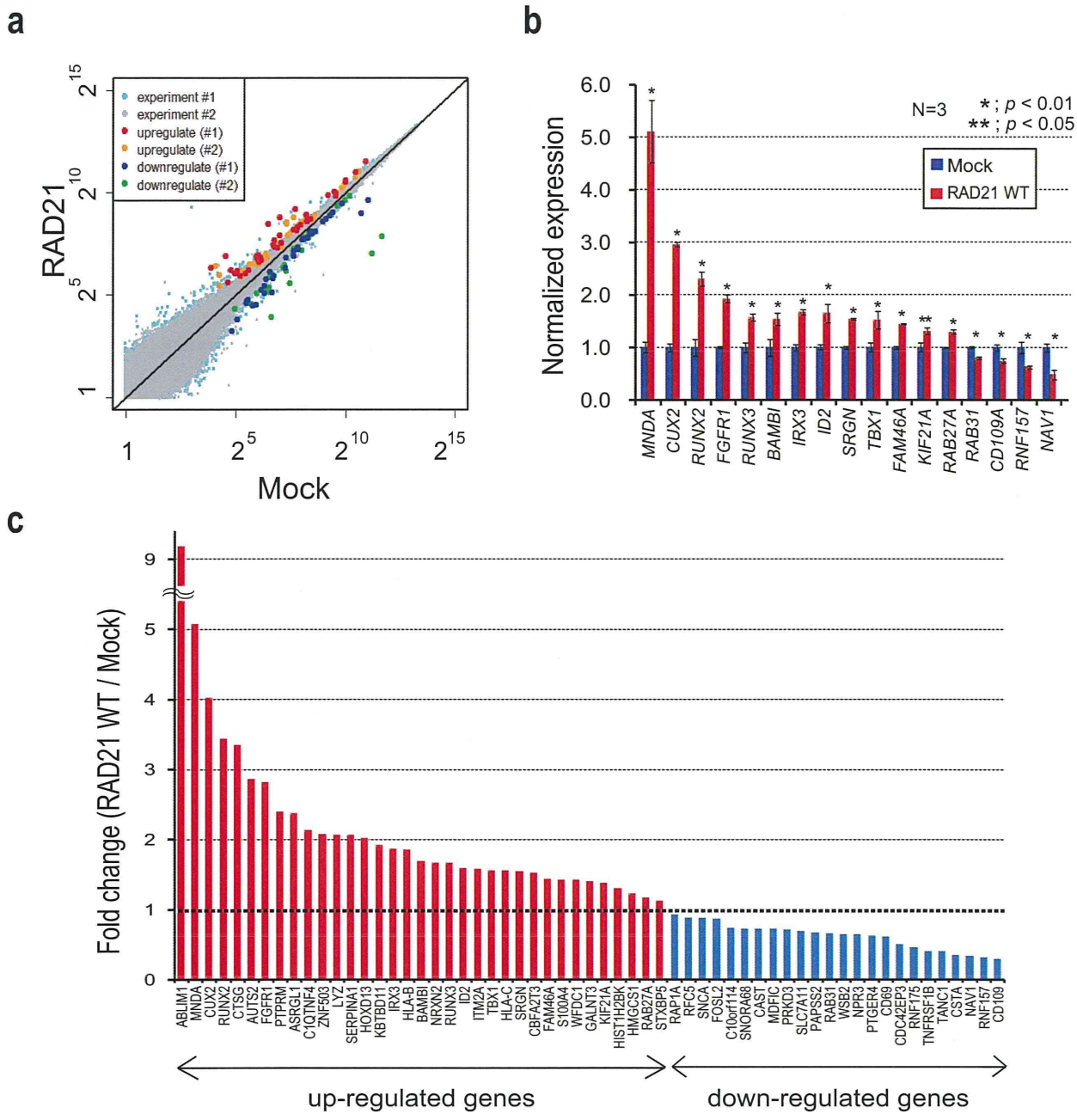


Forced expression of cohesin components in myeloid leukemic cell lines

(a) Deep sequencing of the Kasumi-1 transcripts identified the expression of the mutant allele accounting for only 5.3% of the total read counts, suggesting that the frameshift RAD21 mutation caused a premature termination codon and likely activated nonsense-mediated RNA decay pathway, and/or lead to low stability of any translated truncated protein. Thus, the frameshift RAD21 mutation in Kasumi-1 cell leads to loss-of-function (haploinsufficiency). A representative view of sequencing reads overlapping the frameshift insertion mutation in *RAD21*, visualized using the Integrative Genomics Viewer (IGV), is shown. **(b)** RT-PCR analysis of total (wild-type and mutant) RAD21 expression was performed using 5' primers (upper panel) and of full-length RAD21 expression using 3' UTR primers (lower panel) (N=3, means±SD). See also Figure 4c. **(c)** Effects of induced expression of wild-type RAD21 or STAG2 alleles on the proliferation of MOLM-13 with abnormal RAD21/STAG2 expression were determined (N=6). Means ± SD of the absorbance at 450 nm relative to the value at day 0 are plotted. A representative result from three independent experiments is shown. The induced expression of RAD21 and STAG2 are shown in **(d)**. **(e)** RT-PCR analysis of RAD21 and STAG2 expression (N=3, means±SD). Proliferation of K562 **(f)** and TF1 **(g)** cell lines stably transduced with the wild-type *RAD21* allele or a mock vector was measured (N=6, means±SD). A representative result from three independent experiments is shown. Western blotting of the transduced wild-type RAD21 allele are also shown. Primer sequences used for RT-PCR analysis are provided in **Supplementary Table 17**.

Nature Genetics: doi:10.1038/ng.2731

Supplementary Figure 13



Expression microarray analysis of RAD21- and mock- transduced Kasumi-1 cells
(a) Comparison of gene expression profiles between RAD21- and mock-transduced Kasumi-1 cells. Mean expression values in triplicate array measurements were plotted for two independent experiments. Significantly and reproducibly up-regulated (>1.2-fold) or down-regulated (<0.83-fold) genes are shown in the indicated colors. **(b)** The validation of the microarray analysis for the expression of 17 genes by qPCR. *GAPDH* normalized data are plotted (N=3, Means ± SD). *, $p < 0.01$; **, $p < 0.05$ (Student's t-test, one-sided). **(c)** The validation of the microarray data by RNA sequencing, in which altered gene expression was confirmed for 59 out of 63 genes that showed significant alteration in gene expression in two independent microarray experiments. Fold changes in gene expression in RNA sequencing is shown in up-regulated (N=35, red) and down-regulated (N=29, blue) genes.

Nature Genetics: doi:10.1038/ng.2731

nature genetics

Integrated molecular analysis of clear-cell renal cell carcinoma

Yusuke Sato[1,2,11], Tetsuichi Yoshizato[1,11], Yuichi Shiraishi[3,11], Shigekatsu Maekawa[1,2,11], Yusuke Okuno[1,11], Takumi Kamura[4], Teppei Shimamura[3], Aiko Sato-Otsubo[1], Genta Nagae[5], Hiromichi Suzuki[1], Yasunobu Nagata[1], Kenichi Yoshida[1], Ayana Kon[1], Yutaka Suzuki[6], Kenichi Chiba[3], Hiroko Tanaka[7], Atsushi Niida[3], Akihiro Fujimoto[8], Tatsuhiko Tsunoda[8], Teppei Morikawa[9], Daichi Maeda[9], Haruki Kume[2], Sumio Sugano[6], Masashi Fukayama[9], Hiroyuki Aburatani[5], Masashi Sanada[1,10], Satoru Miyano[3,7], Yukio Homma[2] & Seishi Ogawa[1,10]

Clear-cell renal cell carcinoma (ccRCC) is the most prevalent kidney cancer and its molecular pathogenesis is incompletely understood. Here we report an integrated molecular study of ccRCC in which ≥100 ccRCC cases were fully analyzed by whole-genome and/or whole-exome and RNA sequencing as well as by array-based gene expression, copy number and/or methylation analyses. We identified a full spectrum of genetic lesions and analyzed gene expression and DNA methylation signatures and determined their impact on tumor behavior. Defective VHL-mediated proteolysis was a common feature of ccRCC, which was caused not only by *VHL* inactivation but also by new hotspot *TCEB1* mutations, which abolished Elongin C–VHL binding, leading to HIF accumulation. Other newly identified pathways and components recurrently mutated in ccRCC included PI3K-AKT-mTOR signaling, the KEAP1-NRF2-CUL3 apparatus, DNA methylation, p53-related pathways and mRNA processing. This integrated molecular analysis unmasked new correlations between DNA methylation, gene mutation and/or gene expression and copy number profiles, enabling the stratification of clinical risks for patients with ccRCC.

Renal cell carcinomas (RCCs) constitute 2–3% of all adult malignancies, with 271,000 new cases and 116,000 related deaths estimated worldwide in 2008 (ref. 1). RCC can be histologically classified into several subtypes, among which ccRCC is the most common, accounting for 70–80% of all kidney cancers[2]. Although immunomodulation using interferon-α and VEGF and/or mTOR inhibition has been applied as systemic therapy for patients with locally advanced or metastatic disease[3], complete surgical resection remains the only curative treatment for ccRCC, except for high-dose interleukin-2, which is used for only limited cases[3]. Genetically, ccRCC is characterized by a very high frequency of biallelic *VHL* inactivation caused by allelic deletion or loss of heterozygosity (LOH) on chromosome 3p (>90%)[4] along with gene mutation (~50%)[5,6] or promoter hypermethylation (5–10%)[7]. In addition, recent whole-exome and targeted sequencing studies have identified frequent recurrent mutations in genes involved in chromatin modification, such as *PBRM1* (ref. 8), *SETD2* (ref. 9), *KDM5C*[9], *KDM6A*[9] and *BAP1* (refs. 10,11), as well as in those involved in the ubiquitin-mediated proteolysis pathway[11]. However, in previous studies, gene mutations were comprehensively investigated for entire coding sequences in only a limited number of cases, and other genetic or epigenetic lesions, including structural abnormalities and DNA methylation, have not been addressed in a comprehensive manner. Thus, knowledge about genetic and/or epigenetic alterations in ccRCC is most likely still incomplete. For example, a subset of ccRCC cases has no detectable *VHL* alterations, and pathogenesis in this subset is poorly characterized compared to that in *VHL*-mutated ccRCC cases in which accumulated hypoxia-inducible factors (HIFs) have a critical role.

Here we performed an integrated molecular study of ccRCC in which ≥100 ccRCC specimens were simultaneously analyzed by whole-genome and/or whole-exome and RNA sequencing in conjunction with microarray-based gene expression, DNA methylation and genomic copy number analyses and immunohistochemistry (Online Methods and **Supplementary Table 1**). An extended cohort of 240 ccRCC specimens, including 106 discovery specimens, was analyzed by SureSelect-based targeted deep sequencing to validate and clarify the effects of major genetic lesions. In addition to previously described common

[1]Cancer Genomics Project, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan. [2]Department of Urology, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan. [3]Laboratory of DNA Information Analysis, Human Genome Center, Institute of Medical Science, The University of Tokyo, Tokyo, Japan. [4]Division of Biological Science, Graduate School of Science, Nagoya University, Nagoya, Japan. [5]Genome Science Division, Research Center for Advanced Science and Technology, The University of Tokyo, Tokyo, Japan. [6]Department of Medical Genome Sciences, Graduate School of Frontier Sciences, The University of Tokyo, Tokyo, Japan. [7]Laboratory of Sequence Analysis, Human Genome Center, Institute of Medical Science, The University of Tokyo, Tokyo, Japan. [8]Center for Genomic Medicine, RIKEN, Yokohama, Japan. [9]Department of Pathology, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan. [10]Department of Pathology and Tumor Biology, Graduate School of Medicine, Kyoto University, Kyoto, Japan. [11]These authors contributed equally to this work. Correspondence should be addressed to S.O. (sogawa-tky@umin.ac.jp).

Received 15 April; accepted 18 June; published online 24 June 2013; doi:10.1038/ng.2699

mutational targets, we identified new mutated genes and pathways that are involved in the pathogenesis of ccRCC, including potentially druggable molecular targets. We also identified unique correlations between mutations, gene expression, DNA methylation and copy number profiles. Our study highlights the role of integrated genome, transcriptome and methylome analyses in clarifying tumor biology and identifying potential therapeutic targets in human cancers.

RESULTS

Whole-genome and whole-exome sequencing

The mean coverage by whole-genome sequencing for paired tumor-normal DNA from 14 ccRCC specimens was 47.2× and 33.6× with 95% and 90% of the entire genome analyzed with ≥20 independent reads on average, respectively (**Supplementary Fig. 1**). A total of 71,424 somatic changes, including 68,273 single-nucleotide variants (SNVs) and 3,151 insertion and/or deletion polymorphisms (indels), were detected in 14 cases (1.7 per megabase per sample) with a true positive rate (TPR) of 99% (630 of the 634 tested were confirmed) (**Supplementary Fig. 2a** and **Supplementary Table 2**). The spectrum of SNVs was over-represented by T>C/A>G transitions followed by C>T/G>A transitions and C>A/G>T transversions. C>T/G>A transitions are predominant in most cancer types[12], whereas T>C/A>G transitions and C>A/G>T transversions were characteristic of ccRCC and have also been reported in hepatocellular carcinoma[13–16] (**Supplementary Fig. 2c**). The mean number of structural variations per case was 12 (range of 0–35) with no apparent breakpoint cluster regions (**Supplementary Fig. 3**).

Table 1 Significantly mutated genes in whole-exome analysis of 106 ccRCCs

Gene	Missense mutations	Nonsense, indel or splicing mutations	Total mutations	Samples	Passenger probability (*P* value)	*q* value
VHL	19	23	42	42	1.32×10^{-102}	1.03×10^{-99}
PBRM1	4	24	28	28	2.63×10^{-36}	1.02×10^{-33}
BAP1	3	5	8	8	1.82×10^{-9}	4.71×10^{-7}
TCEB1	5	0	5	5	7.07×10^{-9}	1.37×10^{-6}
SETD2	5	7	12	12	2.06×10^{-8}	3.20×10^{-6}
FPGT	4	1	5	3	1.13×10^{-7}	1.46×10^{-5}
MUDENG	6	1	7	2	3.38×10^{-7}	3.75×10^{-5}
KEAP1	3	2	5	5	5.95×10^{-5}	5.78×10^{-3}
TET2	7	1	8	6	5.59×10^{-5}	4.83×10^{-3}
MUC4	6	0	6	6	1.02×10^{-4}	7.91×10^{-3}
MLLT10	3	0	3	3	2.30×10^{-4}	1.62×10^{-2}
MSGN1	3	0	3	2	2.85×10^{-4}	1.85×10^{-2}
KRT32	3	1	4	4	2.21×10^{-4}	1.32×10^{-2}
M6PR	1	2	3	3	2.77×10^{-4}	1.54×10^{-2}
RPL14	3	0	3	2	3.90×10^{-4}	2.02×10^{-2}
GRB7	4	0	4	4	4.20×10^{-4}	2.04×10^{-2}
TP53	1	2	3	3	3.85×10^{-4}	1.76×10^{-2}
CSMD3	8	1	9	8	7.08×10^{-4}	3.06×10^{-2}
DNHD1	3	1	4	3	6.44×10^{-4}	2.64×10^{-2}
PIK3CA	5	0	5	5	6.90×10^{-4}	2.68×10^{-2}
NLRP12	3	0	3	3	8.93×10^{-4}	3.31×10^{-2}
VMO1	2	0	2	2	9.89×10^{-4}	3.49×10^{-2}
OR4C13	2	1	3	3	1.10×10^{-3}	3.72×10^{-2}
KCNMA1	4	1	5	5	1.24×10^{-3}	4.00×10^{-2}
LMAN2L	1	2	3	2	1.69×10^{-3}	5.24×10^{-2}
MTOR	7	0	7	6	1.44×10^{-3}	4.31×10^{-2}
ZNF536	5	0	5	5	1.63×10^{-3}	4.70×10^{-2}
YIPF3	2	1	3	2	1.57×10^{-3}	4.36×10^{-2}

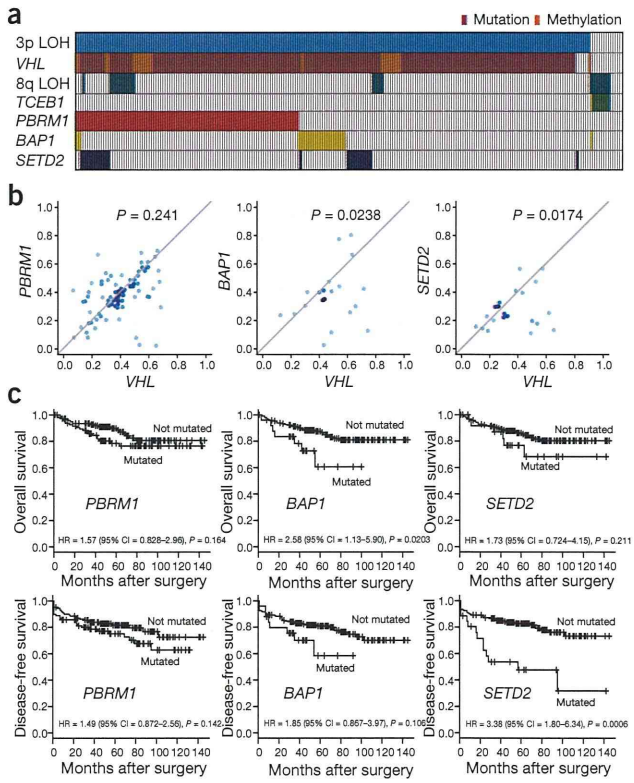


Figure 1 Mutations in 3p target genes and their impact on survival. (**a**) Distribution of common gene mutations and LOH in 240 ccRCC specimens. (**b**) Diagonal plots of observed mutant allele frequencies for *VHL* (*x* axes) and *PBRM1*, *BAP1* and *SETD2* (*y* axes). *P* values were calculated using the paired *t* test (two-sided). (**c**) Effects of common gene mutations on overall survival (top) and disease-free survival (bottom) for 240 ccRCC cases. *P* values were calculated using the log-rank test.

On average, 47 non-silent mutations were identified per case (**Supplementary Fig. 2b**), which accounted for approximately 0.92% of all somatic mutations. The numbers of mutations within coding, intronic (regulatory) and intergenic sequences were roughly proportional across the 14 cases, indicating that mutations were largely random events. To identify the complete spectrum of driver gene targets, we analyzed a total of 106 paired ccRCC specimens by whole-exome sequencing (SureSelect v.4, Agilent Technologies) in which approximately 89% of the target sequences were covered by ≥20 independent reads (**Supplementary Fig. 4**). A total of 5,171 non-silent somatic mutations (48.8 per tumor) were detected with a TPR of 96% (559 of the 582 tested were confirmed). These consisted of 4,234 missense mutations, 232 nonsense mutations, 140 splice-site mutations, 557 indels and 8 read-through changes (**Supplementary Table 3**).

In the 14 specimens that were analyzed by both whole-genome and whole-exome sequencing, 539 of the 839 non-silent mutations (64%) were identified with both platforms. However, reflecting its higher coverage (129×), whole-exome sequencing more efficiently captured the subclonal mutations harbored by a subset of the tumor population, which consequently had lower allele frequencies (**Supplementary Figs. 5** and **6**). Only whole-genome sequencing captured 117 mutations, for which coverage depths were lower in whole-exome than in whole-genome sequencing in most cases (n = 96), even though the mean

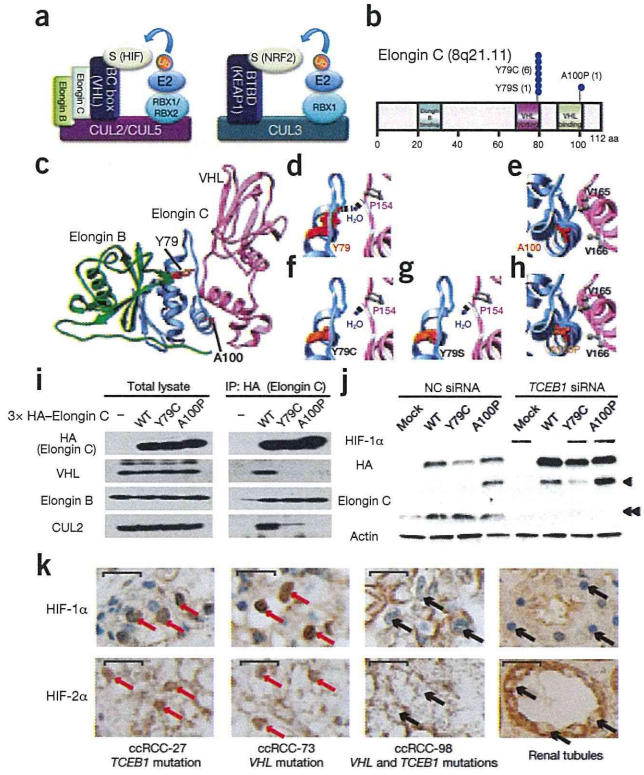


Figure 2 New *TCEB1* mutations and HIF accumulation. (**a**) Two examples of cullin–RING ubiquitin ligase system molecular assemblies using CUL2 or CUL5 (left) and CUL3 (right) that interact with the BC-box protein–Elongin C–Elongin B complex and BTB protein, respectively, to recruit substrate for ubiquitination and subsequent degradation. VHL and KEAP1 are examples of BC-box and BTB proteins, respectively, that recruit HIF and NRF2 proteins for ubiquitin-mediated degradation. (**b**) *TCEB1* mutations (8 of 240 tumors) affect the domains for binding to VHL in Elongin C. (**c**–**e**) Structure of the VHL complex comprising Elongin B, Elongin C and VHL (**c**), with the positions of mutated amino acids Tyr79 (**d**) and Ala100 (**e**) in Elongin C indicated. (**f**,**g**) A critical hydrogen bond between Tyr79 in Elongin C and Pro154 in VHL (**d**) was predicted to be abolished in the Tyr79Cys (**f**) and Tyr79Ser (**g**) Elongin C mutants. (**h**) Hydrophobic binding around Ala100 in Elongin C and Val165 and Val166 in VHL (**g**) could be compromised in the Ala100Pro Elongin C mutant. (**i**) Protein blotting for the indicated components of the VHL-CUL2 complex in total cell lysates (left) and in lysates after immunoprecipitation (IP) with antibody to HA (Elongin C) (right). Lysates were from HEK 293T cells transduced with mock, wild-type (WT) or mutant *TCEB1* constructs (encoding Tyr79Cys and Ala100Pro Elongin C). (**j**) Protein blotting for the effect of *TCEB1* (Elongin C) mutations on HIF accumulation using non-specific siRNA (left) or siRNA specific for endogenous *TCEB1* (right). Endogenous and exogenous Elongin C and HIF-1α were examined. Exogenous 3′ HA–tagged Elongin C was detected with an antibody to HA. Note that antibody to Elongin C could discriminate slower migrating exogenous protein (single arrowhead) from faster migrating endogenous protein (double arrowheads). (**k**) Immunohistochemical analysis of HIF-1α and HIF-2α expression in representative cases from primary ccRCC specimens with *TCEB1* mutations ($n = 5$), *VHL* mutations ($n = 92$) or without *TCEB1* or *VHL* mutations ($n = 9$) as well as in normal kidney tissue ($n = 1$). Red and black arrows indicate positive and negative nuclear immunoreactivity, respectively. Scale bars, 20 μm.

depths of coverage of the entire targeted regions were much higher in whole-exome sequencing (129×) than in whole-genome sequencing (~50×). The coverage in whole-exome sequencing was especially low (<8×) for 45 variants owing to high GC content (for 30 variants) or to no bait designing in targeted exome capture at all (for 15 variants). Subsequent deep sequencing of mutations identified intratumoral heterogeneity in 12 of the 14 ccRCC specimens (**Supplementary Fig. 7**), with the presence of heterogeneity was more explicitly demonstrated in a previous study using whole-exome and targeted deep sequencing combined with extensive multisite sampling from the same tumors and/or metastasized tumor blocks[17].

Recurrent mutations in 3p targets

In whole-genome and/or whole-exome sequencing of the 106 ccRCC specimens, recurrent mutations were observed in 777 genes, of which 28 were considered to be significantly mutated ($q < 0.05$) compared to background mutation rates (**Table 1** and **Supplementary Table 4**). Of the top five significantly mutated genes, *VHL*, *PBRM1*, *BAP1* and *SETD2* were all located within the common site of LOH at 3p between the 3p25 and 3p21 segments (**Supplementary Fig. 8**) and were considered to be the targets of the LOH at 3p found in more than 90% of ccRCC specimens.

Mutations of these common targets were further investigated in detail by deep and/or Sanger sequencing of the relevant genes in combination with assays for DNA methylation status and SNP array–based allele-specific copy number analysis in the extended cohort of paired tumor-normal DNA samples from 240 ccRCC cases (**Fig. 1a** and **Supplementary Fig. 9**). LOH at 3p was found in 226 specimens (94%), which was caused either by simple 3p loss ($n = 175$) or copy-neutral LOH (uniparental disomy, UPD; $n = 51$). There were no significant differences in the mutation rate of 3p target genes in cases with 3p loss and those with 3p UPD. Mutation or promoter hypermethylation was rarely found in cases without LOH at 3p. The vast majority of the 226 cases with LOH at 3p ($n = 221$; 97.8%) had the remaining *VHL* allele affected either by somatic mutation ($n = 197$; 16 nonsense mutations, 70 missense mutations, 100 indels and 11 splice-site mutations) or promoter methylation ($n = 24$). Inactivation of other genes was exclusively caused by gene mutation. Almost all mutations (147/149) involving *PBRM1* (98/98), *SETD2* (25/26) and *BAP1* (24/25) were found in a subset of *VHL*-inactivated cases (**Fig. 1a**) in which *SETD2* and *BAP1* mutations tended to show significantly lower allelic burdens than coexisting *VHL* mutations (**Fig. 1b**). This finding indicates that *SETD2* and *BAP1* mutations are likely to be acquired and selected from within pre-existing *VHL*- and/or *PBRM1*-mutated clones and contribute to tumor progression, as suggested in a recent report.

We next investigated the impact of these mutations on survival and tumor recurrence. In accordance with recent reports[18–20], there was no significant impact of *PBRM1* mutations on overall survival or disease-free survival in univariate analysis (**Fig. 1c**). In contrast, *BAP1* mutations, which were mutually exclusive with *PBRM1* mutations[10] ($P = 2.05 \times 10^{-3}$, Fisher's exact test), were significantly associated with shorter overall survival time (hazards ratio (HR) = 2.58, 95% confidence interval (CI) = 1.13–5.90; $P = 0.0203$), although their impact on relapsed disease was less prominent ($P = 0.106$). This effect could be partly owing to the effects of *SETD2* mutations in *BAP1* mutation–negative cases, as *SETD2*-mutated cases showed a very high relapse rate (HR = 3.38, 95% CI =1.80–6.34; $P = 6.00 \times 10^{-4}$) but did not necessarily have shorter overall survival times. Equivalent results were obtained in multivariate analysis in which mutations in all three 3p target genes were included (**Supplementary Table 5**).

TCEB1 mutations in ccRCC

Another highly significant mutational target was *TCEB1*, which encodes Elongin C, a 112-residue protein[21]. Elongin C was originally identified as a subunit of the heterotrimeric RNA polymerase II elongation factor complex (Elongin) that potently induces mRNA elongation but is also known to be a vital component of the VHL complex. In the latter complex,

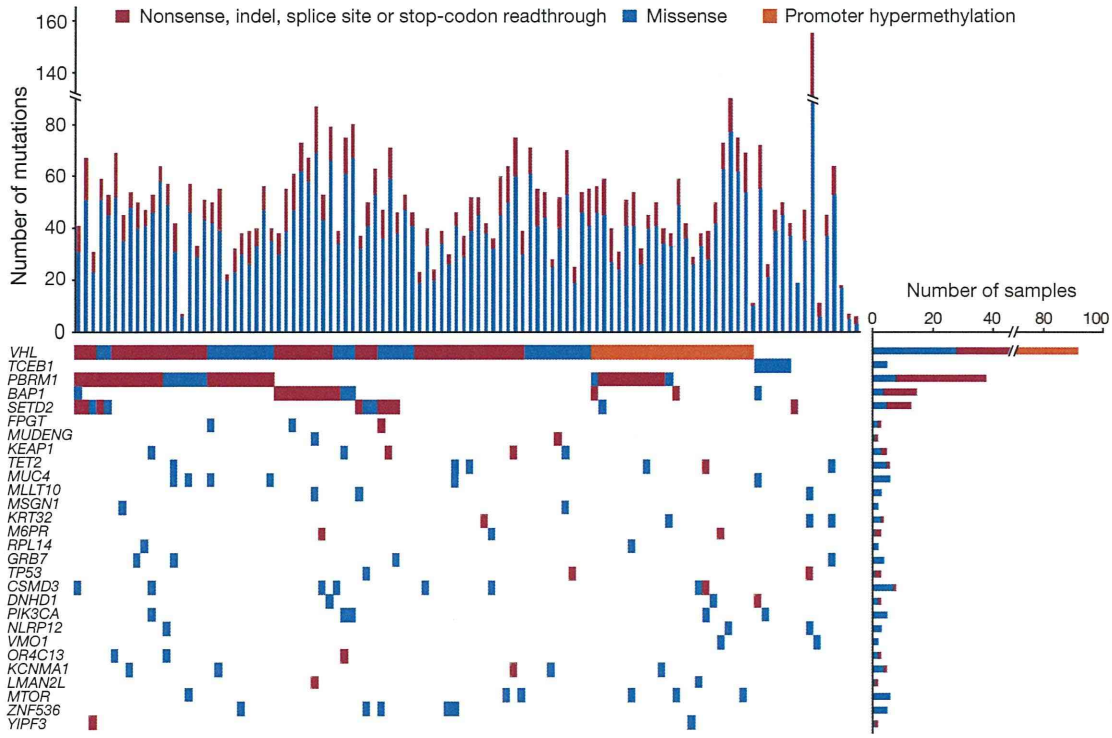


Figure 3 Significantly mutated genes and pathways for 106 ccRCC specimens. The number of somatic mutations in each case (top) and the number of cases that had alterations in significantly mutated genes (bottom right) are shown in a bar plot.

VHL, Elongin B, Elongin C and a catalytic RING subunit (RBX1), which binds ubiquitin-conjugated E2 component, are organized on a cullin scaffold protein (CUL2) to accomplish ubiquitination of VHL-bound HIF proteins (**Fig. 2a,** left)[22]. Targeted deep sequencing and methylation analysis for the entire cohort identified 8 mutations (3.3%) in *TCEB1*. No mutations were detected in other complex components, including *TCEB2* (encoding Elongin B), *CUL2* and *RBX1*. Together with *VHL* lesions, genetic and epigenetic alterations in the VHL complex accounted for 229 of the 240 ccRCC specimens (95.4%)in which *TCEB1* mutations and *VHL* lesions were completely mutually exclusive ($P = 2.50 \times 10^{-14}$, Fisher's exact test), further underscoring the critical role of VHL complex inactivation in the pathogenesis of ccRCC (**Fig. 1a**). There were no significant differences in the clinicopathological characteristics of cases with mutated *VHL* and those with mutated *TCEB1* (**Supplementary Table 6**).

Whereas *VHL* mutations and methylation were closely associated with LOH at 3p, *TCEB1* mutations were always accompanied by loss of chromosome 8 ($P = 3.03 \times 10^{-9}$, Fisher's exact test), leading to complete loss of wild-type *TCEB1* alleles (on 8q21) (**Fig. 1a**). However, in contrast to *VHL* mutations of which nonsense or frameshift alterations frequently result in complete loss of protein function, *TCEB1* mutations exclusively involved two conserved amino acids, Tyr79 ($n = 7$) and Ala100 ($n = 1$), with the former being a mutational hotspot (**Fig. 2b** and **Supplementary Fig. 10**). Notably, these two amino acids are positioned close together within the binding interface for the VHL protein (**Fig. 2c**): Tyr79 mediates a hydrogen bond with Pro154 of VHL via a water molecule (**Fig. 2d**), and Ala100 participates in hydrophobic interactions with Val165 and Val166 of VHL (**Fig. 2e**)[23,24]. Thus, mutations that affect these two amino acids (p.Tyr79Cys, p.Tyr79Ser and p.Ala100Pro) are predicted to abolish the interaction between Elongin C and VHL and to result in compromised ubiquitination and subsequent accumulation of HIF (**Fig. 2f–h**).

Indeed, when expressed in HEK 293T cells, wild-type Elongin C effectively coprecipitated with VHL and CUL2, whereas the interaction with VHL and CUL2 was markedly reduced with mutant Elongin C (Tyr79Cys and Ala100Pro; **Fig. 2i** and **Supplementary Fig. 11a**). We also examined the effects of mutant Elongin C on HIF-1α accumulation by exogenously expressing various Elongin C proteins in HeLa cells. As expected, HIF-1α accumulation was not observed after simply expressing these putative loss-of-function Elongin C mutants or the wild-type protein (**Fig. 2j**, left). However, when endogenous wild-type Elongin C expression was suppressed by small interfering RNA (siRNA) specific for the endogenous *TCEB1* transcripts (Online Methods and **Supplementary Table 7a**), both mock-transduced cells and mutant *TCEB1*–transduced cells but not wild-type *TCEB1*–transduced cells showed recognizable HIF-1α accumulation (**Fig. 2j**, right). These results suggest that the two *TCEB1* mutants (encoding Tyr79Cys and Ala100Pro) actually represent loss-of-function alleles with regard to VHL complex function and that biallelic inactivation is required for HIF-1α accumulation, explaining why *TCEB1* mutations were always accompanied by a deletion of the intact *TCEB1* allele on chromosome 8.

Finally, to confirm the oncogenic role of these *TCEB1* mutations in primary ccRCC, we investigated HIF-1α and HIF-2α expression in primary surgical ccRCC specimens with *TCEB1* mutations by immunohistochemistry. As with *VHL*-mutated tumors, all five tumors with *TCEB1* mutation exhibited increased HIF-1α expression in immunohistochemistry compared to normal kidney samples or tumors that lacked *TCEB1* and *VHL* mutations (**Fig. 2k**, **Supplementary Fig. 12** and **Supplementary Table 8**).

Other recurrent mutations

Other newly identified recurrent mutational targets included *TET2*, *KEAP1* and *MTOR* (**Fig. 3**). *TET2* encodes an α-ketoglutarate–dependent oxigenase frequently inactivated in myeloid malignancies by gene