Supplementary Appendix

This appendix has been provided by the authors to give readers additional information about their work.

Supplement to: Gerlinger M, Rowan AJ, Horswell S, et al. Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. N Engl J Med 2012;366:883-92.

Intratumor Heterogeneity and Darwinian Selection Revealed by Renal Cancer Sequencing

Supplementary Material Table of Contents

Contents

Author Contributions

MGe and CSw designed the study, MGe, AR, NM, BS-D, EG, GC, CSa, BP, SB, DJ, KR, CL, JL, LP, MN and MG gathered the data, MGe, SH, DE, PM, EG, AS, PT, IV, NMc, JD, GS, AR, AE, ZS, AF and CSw analyzed the data, CSw, MGe, PAF and SH

vouch for the data and the analysis, MGe and CSw wrote the paper, all authors approved and decided to publish the paper.

Materials and Methods

E-PREDICT Trial and Patient Consent

The E-PREDICT translational clinical trial (EUDRACT number: 2009-013381-54, http://public.ukcrn.org.uk/Search/StudyDetail.aspx?StudyID=10710, open to recruitment since January 2010, Principle Investigator: JL) of pre- and post-nephrectomy Everolimus treatment in patients presenting with metastatic renal cell carcinoma was approved by the Royal Marsden Hospital Research Ethics Committee. The first 4 patients (001-004) enrolled in this trial gave written informed consent for study participation and for the translational analyses presented here.

Patient characteristics:

Patient	Age	Sex	Histological Diagnosis	Fuhrman Grade	Stage
001	75	M	Clear cell renal cell carcinoma	1-4	T3N2M1
002	59	M	Clear cell renal cell carcinoma	1	T3N0M1
003	64	F	Clear cell renal cell carcinoma	1-4	T3N0M1
004	57	F	Clear cell renal cell carcinoma	1-2	T2N0M1

Patients underwent a pre-treatment core biopsy of the primary tumor for pathological confirmation of a renal cell carcinoma and molecular analyses. Patients were treated with Everolimus 10 mg once daily for 6 weeks and underwent a cytoreductive nephrectomy after a 1 week washout period. Everolimus treatment was resumed after wound healing was complete and continued until disease progression. Computed tomographic staging of thorax, abdomen and pelvis was performed before Everolimus treatment started, before nephrectomy and in 8 weekly intervals from the time patients re-started Everolimus after nephrectomy.

Tumor and Normal Tissue Processing

Eight to ten 10x5x5 mm samples, representing the spatial extent and macroscopic heterogeneity of the nephrectomy specimen, were harvested from each nephrectomy specimen and two samples from the metastectomy specimens from patient 001. Samples were macrodissected to minimize stromal contamination and half of each sample was snap frozen in liquid nitrogen within 45 minutes after clamping of the renal artery. The other half was formalin fixed and embedded in paraffin in order to confirm the pathological diagnosis and to assess tumor and stromal cell content. Sample harvesting was performed according to strict SOPs in all cases and documented by photography. In patient 001, 60% of cells in R5 were estimated to be tumor cells whereas all other sections contained more than 75% tumor cells. Specimens from patient 002 had lower relative tumor cell contents (median 52% on histopathological review) because very large grade 1 tumour cells were surrounded by a high proportion of small stromal cells. Microdissection was not undertaken in

order to avoid random selection of intra-regional subclones and whole genome amplification which could have caused an increased error rate. From patient 002, core biopsies were obtained from a liver metastasis which remained stable on treatment for 5 months until progression was detected on a routine re-staging CT scan. One core biopsy was snap frozen in liquid nitrogen and a second biopsy was paraffin embedded and confirmed viable clear cell renal cell carcinoma. DNA and RNA were purified using the Qiagen (Crawley, UK) DNeasy and the RNeasy extraction kits following the manufacturer's instructions. Germ line DNA was extracted from peripheral blood mononuclear cells from whole blood (patients 001-004) and also from normal kidney tissue (patient 002) by the Qiagen DNeasy kit. Nucleic acid yield and quality was determined spectrophotometrically using a NanoDrop (Thermo Scientific) analyzer. The quality and integrity of the RNA and DNA was examined after agarose gel electrophoresis. Macrodissection and sampling was performed by GS, MGe, MN and AR, histopathological review by GS and DNA and RNA extraction by MN and AR.

Multi region exome sequencing

All samples from patient 001 were normalized to 3µg of DNA and sheared to 150–200bp using a Covaris S2 (Covaris, Woburn, MA, USA), following the SureSelect Human All Exon 50Mb kit protocol run parameters[1]. In four samples (R1, R8, M1 and germ line DNA), both the SureSelect Human All Exon 50Mb and NimbleGen SeqCap EZ Whole Exome v2.0 [1](Roche NimbleGen, Inc.) kits were used in parallel for exon capture. Using the BioAnalyser 2100 (Agilent, Santa Clara, CA, USA), the sheared samples

were validated before library preparation was continued. The SureSelect Human All Exon 50Mb protocol was followed for library preparation [1]. The sheared DNA samples were end repaired, poly A tailed and Illumina Paired End Adapters were ligated. Library PCR of 4 cycles was performed using the Herculase II Fusion DNA Polymerase[1]. The SureSelect protocol was followed for the in-solution hybridisation of the libraries[1-3]. Excessive evaporation could cause bias in bate hybridisation and any capture library post incubation resulting in under 20ul total volume was remade. After a final quality control using the BioAnalyser 2100 (Agilent), the capture libraries were ready for Flowcell cluster formation on a cluster station[4,5] and then 72bp Paired End Sequenced by synthesis on the Genetic Analyser IIx (Illumina)[4-7]. Two lanes were required per region to provide an average read depth in excess of 30-fold. Generation of the ultra deep sequencing datasets of R4 and R9 required an entire 8 lane flow cell, each. Patient 001 sequencing experiments were performed by NM. Patient 002 samples were captured by the SureSelect Human All Exon 50Mb kit and 75bp Paired End Sequenced as described [8] on an Illumina HiSeq analyzer by PAF. Two samples were multiplexed per lane to achieve coverage in excess of 30-fold.

Bioinformatics analyses

Patient 001 data: 72bp paired raw reads were extracted from Eland export files (including reads which did not pass the Eland "chastity" filter) to fastq format and subsequently aligned to hg19 using bwa 0-5.9 [9] with a seed

length of 72 bases. All other settings were left as default (in particular, up to 3 mismatches were allowed per read). Aligned sam files were converted to bam format, sorted, indexed, replicates merged and reconverted to pileup format using samtools 0.1.16[10] , with variant calling performed by SNVMix2 0.11.8 [11].

Calls were then filtered using the following criteria for each variant position: Minimum of 10x coverage in the germline sample with zero non-reference reads; Minimum of 10x coverage in the tumor sample in which the variant is detected; Minimum of 95% confidence of a non-reference call given by SNVMix2; Variant position must not be listed in dbSNP (v132)[12];Variant position must be annotated as exonic by RefSeq (Release 45) or, in the case of a splice-site variant, lie within the two flanking base-pairs of an exonic region; Synonymous/non-synonymous calls were made using dbSNP[12]; Filtering was performed with R 2.13.0[13], using the Bioconductor package ShortRead[14] to determine coverage levels from sorted bam files and various in-house parsers written in C. We saw no evidence of a systematic bias towards an increased proportion of alternate reads in the Nimblegen captured samples vs. the Agilent captured samples (paired, one sided Wilcoxon test, p = 0.9951)

Non-synonymous mutations, called in at least one region and for which wild type or mutant calls had been made by SNVMix2 for at least 50% of the sequenced tumor regions, were selected for manual review. Mutations which were called in poorly aligned reads or which showed a characteristic profile suggestive of a sequencing error[15] were excluded from further

analysis. Small insertions and deletions (INDELs) were called with a modified version of Pindel as described previously[8]. INDELs present in the blood were excluded from further analysis. A region was called positive for a point mutation or INDEL if more than 1 read within the region contained the variant. Clonal ordering analysis was performed as described [16].

All genomic positions carrying a somatic mutation detected in the multi-region sequencing analysis of patient 001 (Supplementary Table 1) were reviewed in the ultra-deep exome sequencing data for R4 and R9. A mutation was called if more than 1% of the reads contained the variant.

Patient 002 multi-region sequencing data was processed and variants were called on the CAVEMAN and PINDEL pipeline as described[8]. INDELS or nucleotide variants which could be detected in the germline samples were excluded from the analysis. Non-synonymous mutations were identified and filtered as described [8] and mutations in poorly aligned regions or those showing an Illumina sequencer specific error profile [15] were removed manually. A region was called positive for a point mutation or INDEL if more than 1 read within the region contained the variant.

Patient 001 data were collated, processed and analysed by SH, AS, PM, MGe and patient 002 data by PAF, DJ, IV, AB, KR, PT and MGe.

Mutation frequencies in region R4

Somatic mutations identified with the standard sequencing depth in regions R4 and R9 were further investigated with ultra-deep sequencing. Mutation frequencies were defined as the ratio of mutant reads against

the total number of reads spanning the variant position (supplementary table 3). Greater than 40-fold coverage was achieved in the ultra-deep sequencing dataset for all mutations previously reported to be present in R4/R9. Data analysis was performed by SH, PM and MGe.

SNP-array analysis and DNA segmentation

DNA was processed and hybridized to Illumina Human Omni 2.5 SNP arrays according to the manufacturer's protocol. Not enough DNA was available to perform SNP array analysis for regions R8 and M from patient 002 and region R7 from patient 004. Illumina's GenomeStudio software was used to obtain B allele frequencies (BAF) and normalized logarithmic probe intensities (log R ratios) from the raw output data. Because of weak call rates (<0.94), SNP data for tumor regions R1, R3 and R5 from patient 001 were excluded from the analysis and region R5 from patient 004 was removed based on high noise to signal ratios in the BAF. Mirrored BAF (mBAF) were generated by reflecting the BAF values along the 0.5 axis. Non-informative homozygous calls were removed from each sample by filtering out SNP with mBAF values >= 0.9 in the blood or >= 0.95 in the tumor sample. In addition, outliers were removed by triplet filtering as described [17] with a 0.8 threshold for outlier detection. SNPs on chromosomes X and Y were excluded from the analysis. Segmentation of the smoothed filtered mBAF values for each sample was generated using the DNAcopy R package[18] with alpha, the parameter for statistical significance, set to 0.01. A minimum consistency matrix was obtained by merging the segmentations from all samples to the shortest possible segments. Segments consisting of less than 10 probes were removed. The

filtered minimum consistency matrix was used for the further analysis of all samples. Data were processed and analysed by DE and PM.

Detection of allelic imbalance

In each sample *s*, we defined $Norm_s$ the mean value for allelic balance (same number of A and B alleles) as the first mode of the kernel density estimate of all segment means, weighted by the number of SNP probes in each segment. Allelic imbalance was detected by comparing all filtered SNPs in each minimum consistent region to all filtered SNPs in the equivalent segment from the control sample (blood). We used a two-sample t-test to evaluate if the means of the SNP probes mBAF in the segments are different between the tumor and control samples. To account for shifts in the allelic balance mode, we tested the alternative hypothesis that the difference of the means is significantly greater than the difference between $Norm_s$ from the tumour sample and $Norm_{blood}$. A p-value threshold of $1e^{-5}$ was used to define statistical significance. Data were processed and analysed by DE and PM.

Analysis of chromosome 3p signal intensities in patient 001

The normalized logarithmic signal intensities (log R ratios) of tumour regions with reasonable call rates were smoothed and segmented with the DNAcopy R package. The “sdundo” segmentation parameter was set to 2 standard deviations for all regions besides R4. Due to poor segmentation results with this parameter setting, no “sdundo” was chosen for R4 and the significance level “alpha” was set to 0.1 for this tumour region. Standard DNAcopy parameters were used for smoothing and

segmentations of log R ratios otherwise. To define strong loss, loss and regions of no copy number loss for each sample separately, we searched for all minima in the weighted density estimate of the log R ratios of each tumour sample and defined all log R ratios smaller than a density minimum at a logR value of approximately -0.5 as strong loss, all logR ratios smaller than a minimum at approximately -0.2 as loss and all higher logR ratios as no copy number loss. Data were processed and analysed by DE and PM.

mRNA Expression Analysis

mRNA Expression profiles from fresh frozen specimens from each region of the primary and metastasis were generated by Affymetrix human Gene 1.0 expression arrays according to the manufacturers protocol and normalized using the RMA method from the oligo R package[19]. The data are publicly available on GEO (accession number GSE31232). The samples were clustered using hierarchical clustering on the normalized expression values of probes corresponding to ccA/ccB prognostic signature genes [20]. Data were processed and analysed by DE and PM.

mTOR Functional Analysis

Full length mTOR cDNA with a C-terminal tGFP tag in the PrecisionShuttle vector was purchased from Origine and the n7292 T>C mutation was introduced with the Agilent QuickChangeXL kit as described in the product literature.

Primer sequences used for QickChange protocol:

Forward primer: CTTGCTGAACTGGAGGCCGATGGACACAAATACCA

Reverse primer : TGGTATTTGTGTCCATCGGCCTCCAGTTCAGCAAG

The entire mutated mTOR cDNA was sequenced to confirm the presence of the mutation and to rule out further mutations. The Caki1 RCC cell line was transiently transfected using Fugene HD (Promega) according to the manufacturer's protocol. 24 hours post-transfection, cells were washed twice in PBS and either starved overnight in serum-free media or returned to complete media. After 48 hours cells were lysed in ice cold lysis buffer (50 mM Tris-HCl (pH 7.4), 150 mM NaCl, 2.5 mM EDTA, 1 % Triton X-100, 0.25 % IPEGAL and 0.25 % Sodium Deoxycholate) supplemented with complete protease inhibitors (Roche) and phosphatase inhibitor cocktail set III (Calbiochem) and cleared by centrifugation. Protein levels were adjusted using Bradford reagent (Biorad) to ensure equal loading. Sample buffer (2 % (w/v) SDS, 200 mM Tris-HCl (pH 6.8) 40 % glycerol, 4% (v/v) 2-mercaptoethanol) was added and the samples were heated prior to resolving by SDS-PAGE. The samples were transferred to PVDF membranes (Millipore). The membranes were incubated with the appropriate antibodies, washed with TRIS buffered saline containing 0.1 % Tween and incubated with horse-radish peroxidase conjugated secondary antibodies. After washing, blots were visualized with Immobilon Western (Millipore). Antibodies against Phospho-S6 Ribosomal Protein (Ser235/236), S6 Ribosomal Protein and phospho-4E-BP1 (Thr37/46) were from Cell Signaling. HRP-conjugated anti-actin was obtained from Sigma. The anti-TurboGFP antibody is from Evrogen. Secondary anti-rabbit

antibodies were from Dako. Experiments were conceived and carried out by MGe, CSa and EG.

Regional Ploidy Profiling Analysis

A 50uM formalin fixed paraffin embedded section of tissue was placed in a microfuge tube and xylene was added to remove the paraffin wax and serially rehydrated through 100%, 95%, 70% and 50% ethanol for 5 minutes at room temperature and washed twice with distilled water. A suspension of nuclei was made by incubating the tissue in a 0.5% pepsin solution (Sigma, UK) prepared in 0.9% saline pH 1.5. Incubation is carried out at 37°C for 30 minutes. The nuclei were washed once with PBS, stained with propidium iodide and analysed using the Calibur 1 FACS machine and CellQuest software (Becton Dickinson). The DNA Index of the aneuploid peak, where present, was calculated as the G1 peak of the aneuploid population divided by the G1 peak of the normal diploid cells. Experiments were conceived and performed by AR.

Immunohistochemistry and Functional Analyses

Rabbit anti-phospho-S6 ribosomal protein (Ser235/236), rabbit anti-tri-methyl-histone H3 (Lys36), rabbit anti-phospho-Akt (Ser473) and rabbit anti-phospho-4EBP1 antibodies (Thr37/46) (Cell Signaling: #2211, #9763, #4060, #2855), were used for immunohistochemistry on paraffin sections. Antigens were unmasked by microwaving in Tris-EDTA pH9 (pS6) and citrate pH6 (Lys36, pAkt, 4E-BP) and incubated with primary antibodies at 1:25 (pAkt), 1:50 (pS6), 1:100 (H3K36-3Me) and 1:800 (p4E-BP) respectively. After incubation in biotinylated secondary antibody and

Avidin Biotin Complex, slides were developed in DAB substrate (all from Vector). Experiments were conceived and carried out by MGe, EG, GS and BS-D

Validation of candidate somatic mutations

Candidate mutations were validated and verified using Sanger sequencing technology. Mutations found in genes recurrently altered in RCC and additional mutations chosen randomly from the groups of ubiquitous, shared and private mutations were forwarded for validation. Unless specified otherwise (e.g. for private mutations), routine validation for patient 001 was performed for R3, R4, R9 and M2a and germline DNA from PBMCs. Because of limited DNA availability from PreP and PreM, a limited set of shared mutations was validated (mTOR, SOX9, ALKBH8, SETD2, KDM5C-splice site). For patient 002, routine validation was performed in R3, R4, R7, R9, M and germline DNA from PBMCs. Oligonucleotides were designed to span the mutations using Primer3 software (Rozen & Skaletsky). The genomic region of interest was amplified using the polymerase chain reaction, incorporating Big Dye Terminators (BDT v3.1, Life Technologies) followed by capillary separation on the Applied Biosystems 3730xl Genetic Analyser and manual inspection of all sequencing traces. Experiments were conceived, performed and analysed by AR and GC.

Supplementary Table 1: Sequencing coverage in patients 001 and 002

Median sequencing depth and percentages of the exon capture kit target region for which a minimal coverage of 10x and 30x has been achieved in each sequenced tumor region.

Sample	Median depth	10x	30x
patient 001 Germline (blood/PBMCs)	43x	89%	66%
patient 001 PreP	74x	91%	79%
patient 001 PreM	63x	91%	76%
patient 001 R1	81x	95%	85%
patient 001 R2	83x	92%	80%
patient 001 R3	45x	89%	67%
patient 001 R4	65x	89%	76%
patient 001 R5	71x	88%	76%
patient 001 R6	46x	88%	67%
patient 001 R7	36x	88%	59%
patient 001 R8	117x	96%	91%
patient 001 R9	60x	90%	74%
patient 001 M1	109x	96%	90%
patient 001 M2a	76x	91%	78%
patient 001 M2b	49x	88%	68%
patient 001 – average of tumor samples (not including Germline and regions excluded from analysis (R6 and R7))	**74x**	-	-
patient 002 Germline (normal kidney)	48x	87%	67%
patient 002 Germline (blood/PBMCs)	46x	87%	66%
patient 002 R1	65x	90%	76%
patient 002 R2	67x	90%	76%
patient 002 R3	62x	89%	74%
patient 002 R4	54x	88%	70%
patient 002 R5	76x	91%	80%
patient 002 R6	51x	87%	68%
patient 002 R7	59x	89%	74%

patient 002 R8	58x	91%	74%
patient 002 R9	67x	89%	75%
patient 002 M	71x	91%	78%
patient 002 – average of tumor samples (not including Germline and regions excluded from analysis (R2, R5, R8))	**61x**	-	-

Supplementary Table 2: Details of 133 somatic mutations identified in patient 001.

Del = deletion, Ins = insertion, FS = frameshift.

Gene	Chrom	Position	Nucleotide Variant	Amino Acid Change
IL12RB2	1	67795338	T>C	Y>H
BCAS2	1	115118221	C>T	V>I
IFI16	1	158984570	A>T	K>X
FCAMR	1	207134032	Del:C	FS
PLB1	2	28801019	T>A	N>K
ALS2CR12	2	202216087	G>T	P>H
C2orf21	2	210658553	C>A	T>N
VHL	3	10183769	Del:AG	FS
SGOL1	3	20215857	A>G	I>T
KLHL18	3	47385303	G>A	D>N
SSR3	3	156272853	T>C	Q>R
CLCN2	3	184072353	Del:C	FS
WHSC1	4	1902962	G>T	S>I
ATXN1	6	16306994	G>A	P>S
DOPEY1	6	83848337	A>C	K>Q
CCR6	6	167550439	T>A	L>M
INTS1	7	1512818	Del:T	FS
PTPRZ1	7	121653122	A>G	K>R
ZC3HC1	7	129666125	G>T	L>I
EXT1	8	118816994	C>G	Q>H
RALGDS	9	135984240	T>C	I>V
MSRB2	10	23399206	Del:A	FS
EIF4G2	11	10825745	Ins:A	FS
ANO5	11	22294457	Del:T	FS
C11orf68	11	65685051	C>T	G>D
MRPL51	12	6602286	Del:AA	FS
KDM2B	12	121867936	G>A	R>C
TOX4	14	21966415	G>C	P>R
NUSAP1	15	41667966	Del:A	FS
TCF12	15	57544686	G>A	C>Y
ZC3H18	16	88694377	Del:T	FS
DDX52	17	36002244	C>G	A>P
ZNF519	18	14106244	T>C	N>D
AKAP8	19	15471720	G>A	P>L
CYP4F3	19	15769201	Del:CCCAAAG	FS
KIAA0355	19	34843646	C>A	A>E
WDR62	19	36594487	T>C	S>P
KLK4	19	51412045	C>T	D>N

IGLON5	19	51827044	Del:C	FS
NLRP7	19	55451315	A>C	M>R
MAGEB16	X	35820817	Del:TGATG	FS
SESN2	1	28599163	C>G	H>Q
CCBL2	1	89434486	G>C	T>S
SETD2	3	47165518	Del:GA	FS
PLRG1	4	155467287	Ins:A	FS
CASP2	7	143001797	G>T	G>V
SSNA1	9	140084291	Del:T	FS
TH	11	2185608	C>T	R>H
PPFIA1	11	70224166	T>A	S>T
CDKN1B	12	12871131	Del:G	FS
WSCD2	12	108603968	G>A	V>M
ZNF780A	19	40581873	C>T	C>Y
PPP6R2	22	50878183	C>G	P>A
MTOR	1	11174383	A>G	L>P
UGT2A1	4	70460968	C>T	splice site
ABHD11	7	73151579	G>A	A>V
GALNT11	7	151805164	G>C	V>L
RIMBP2	12	130927064	C>T	G>E
PSMD7	16	74335551	G>A	splice site
CENPN	16	81058350	C>G	L>V
SOX9	17	70118864	C>A	L>M
NPHS1	19	36322595	G>A	S>F
RBFOX2	22	36157335	C>T	S>N
KDM5C	X	53222717	Del:C	FS
KDM5C	X	53222723	C>G	E>Q
SATL1	X	84362594	T>C	T>A
FLNA	X	153596345	Del:CCT	Del:E
ITGB3	17	45380104	G>T	A>S
LATS2	13	21562148	C>A	E>X
DIRAS3	1	68512683	C>T	A>T
NGEF	2	233791888	G>T	S>Y
ZNF493	19	21606235	T>G	I>M
SPATA21	1	16727227	G>A	Q>X
DDX58	9	32487509	A>T	N>K
DAPK1	9	90317948	G>C	C>S
ALKBH8	11	107424632	T>A	E>V
KL	13	33635697	Del:C	FS
ERCC5	13	103514868	C>A	H>N
DIO1	1	54371785	Del:AAC	Del:N
PIAS3	1	145578718	C>G	S>C
MR1	1	181021644	A>T	Q>L
C3orf20	3	14803014	G>C	R>P
SETD2	3	47161717	G>A	P>L
TNIK	3	170856133	Del:T	FS

LIAS	4	39466770	Del:A	FS
FBXO1	5	41934124	G>C	splice site
AKAP9	7	91690733	G>C	E>Q
ITIH5	10	7621865	C>A	R>L
WDR24	16	735684	G>A	A>V
MYH8	17	10299657	T>G	E>A
TOM1	22	35723300	Del:G	FS
SBF1	22	50900272	G>T	L>M
KDM5C	X	53228342	T>A	splice site
USP51	X	55514714	C>G	W>S
NAP1L3	X	92928275	G>A	S>L
ADAMTSL4	1	150532548	DeL:AGGAC	FS
DUSP12	1	161719926	T>G	L>W
SLC2A12	6	134328018	A>G	FS
RAB27A	15	55497741	T>A	L>F
CIB2	15	78416057	Del:C	FS
RPS8	1	45241793	G>T	R>L
FAM129B	9	130271358	C>A	C>F
PHF21B	22	45312250	G>T	S>R
HDAC6	X	48675014	C>A	L>M
MAP3K6	1	27688199	G>T	C>X
MAMLD1	X	149671727	T>G	W>G
RLF	1	40703301	G>A	R>Q
DNMT3A	2	25523017	C>A	K>N
HMG20A	15	77770822	A>C	T>P
ZNF521	18	22805039	T>A	H>L
MMAB	12	109998874	G>T	C>X
DACH2	X	86071070	A>G	N>S
SLC2A1	1	43394624	DEL:catga<6>cacaa	
TM7SF4	8	105361447	G>A	V>I
ANKRD26	10	27326806	C>G	L>F
CD44	11	35231528	G>C	Q>H
KRT4	12	53202570	C>T	R>H
KIAA1267	17	44116456	G>C	P>A
C3	19	6713210	G>C	I>M
ADAMTS10	19	8668659	C>T	R>H
IFNAR1	21	34717554	G>A	E>K
BCL11A	2	60688070	C>A	W>C
PLCL1	2	198948857	Del:A	FS
SETD2	3	47059231	T>A	splice site
KIAA1524	3	108285373	Del:A	FS
NRAP	10	115412717	T>A	K>X
HPS5	11	18306944	Del:G	FS
DIXDC1	11	111888597	Del:T	FS

LAMA3	18	21533020	T>C	I>T
CDH19	18	64235921	G>T	N>K
SUPT6H	17	27018057	Del:GGACAATTTCCCT	FS
WDR7	18	54444053	Del:CT	FS
C2orf85	2	242814354	T>G	V>G

Supplementary Table 3: Frequencies of 127 somatic mutations characterized by the ultra-deep sequencing of regions R4 and R9 in patient 001.

Del = deletion, Ins = insertion, FS = frameshift. Regions containing more than 1% variant reads were called positive for the detected variant. The VHL mutation could not be analyzed by Next Generation Sequencing due to low coverage in the related region.

Gene	Chrom	Position	Variant	R4 Variant reads	R4 Total reads	R4 Percentage of mutant reads	R9 Variant reads	R9 Total reads	R9 Percentage of mutant reads
IL12RB2	1	67795338	T>C	145	466	31.12%	129	431	29.93%
BCAS2	1	115118221	C>T	87	293	29.69%	93	301	30.90%
IFI16	1	158984570	A>T	140	451	31.04%	175	697	25.11%
FCAMR	1	207134032	Del:C	131	482	27.18%	117	561	20.86%
PLB1	2	28801019	T>A	15	74	20.27%	26	87	29.89%
ALS2CR12	2	202216087	G>T	26	87	29.89%	21	92	22.83%
C2orf21	2	210658553	C>A	135	450	30.00%	118	390	30.26%
SGOL1	3	20215857	A>G	176	617	28.53%	185	632	29.27%
KLHL18	3	47385303	G>A	69	158	43.67%	66	172	38.37%
CLCN2	3	184072350	Del:C	30	140	21.43%	26	132	19.70%
WHSC1	4	1902962	G>T	69	224	30.80%	50	173	28.90%
ATXN1	6	16306994	G>A	113	412	27.43%	98	391	25.06%
DOPEY1	6	83848337	A>C	140	420	33.33%	106	461	22.99%
CCR6	6	167550439	T>A	178	537	33.15%	160	526	30.42%
INTS1	7	1512818	Del:T	30	100	30.00%	29	249	11.65%
PTPRZ1	7	121653122	A>G	251	796	31.53%	505	1121	45.05%
ZC3HC1	7	129666125	G>T	77	229	33.62%	119	263	45.25%
EXT1	8	118816994	C>G	122	380	32.11%	151	452	33.41%
RALGDS	9	135984240	T>C	24	62	38.71%	32	67	47.76%
MSRB2	10	23399206	Del:A	105	317	33.12%	51	289	17.65%
EIF4G2	11	10825745	Ins:A	124	504	24.60%	114	408	27.94%
ANO5	11	22294457	Del:T	170	549	30.97%	147	529	27.79%
C11orf68	11	65685051	C>T	43	132	32.58%	28	143	19.58%
MRPL51	12	6602286	Del:AA	62	171	36.26%	49	157	31.21%
KDM2B	12	121867936	G>A	143	413	34.62%	87	403	21.59%
TOX4	14	21966415	G>C	48	179	26.82%	57	149	38.26%
NUSAP1	15	41667966	Del:A	196	641	30.58%	122	489	24.95%
TCF12	15	57544686	G>A	48	187	25.67%	48	225	21.33%
ZC3H18	16	88694377	Del:T	117	364	32.14%	122	422	28.91%

DDX52	17	36002244	C>G	174	539	32.28%	145	536	27.05 %
ZNF519	18	14106244	T>C	142	416	34.13%	102	261	39.08 %
AKAP8	19	15471720	G>A	18	62	29.03%	10	39	25.64 %
CYP4F3	19	15769201	Del:CCCAAAG	68	263	25.86%	80	301	26.58 %
KIAA0355	19	34843646	C>A	47	116	40.52%	40	129	31.01 %
WDR62	19	36594487	T>C	24	84	28.57%	42	124	33.87 %
KLK4	19	51412045	C>T	30	123	24.39%	31	118	26.27 %
IGLON5	19	51827044	Del:C	33	98	33.67%	30	96	31.25 %
NLRP7	19	55451315	A>C	45	136	33.09%	25	104	24.04 %
MAGEB16	X	35820817	Del:TGATG	85	150	56.67%	93	176	52.84 %
SESN2	1	28599163	C>G	21	219	9.59%	48	140	34.29 %
CCBL2	1	89434486	G>C	46	339	13.57%	128	442	28.96 %
SETD2	3	47165518	Del:GA	73	403	18.11%	194	501	38.72 %
PLRG1	4	155467287	Ins:A	15	299	5.02%	107	330	32.42 %
CASP2	7	143001797	G>T	43	351	12.25%	190	397	47.86 %
SSNA1	9	140084290	Del:T	17	140	12.14%	74	149	49.66 %
TH	11	2185608	C>T	4	48	8.33%	21	70	30.00 %
PPFIA1	11	70224166	T>A	68	447	15.21%	116	374	31.02 %
CDKN1B	12	12871131	Del:G	7	74	9.46%	43	145	29.66 %
WSCD2	12	108603968	G>A	7	43	16.28%	19	102	18.63 %
ZNF780A	19	40581873	C>T	68	467	14.56%	160	511	31.31 %
PPP6R2	22	50878183	C>G	12	76	15.79%	20	67	29.85 %
MTOR	1	11174383	G>A	0	252	0.00%	73	171	42.69 %
UGT2A1	4	70460968	C>T	0	432	0.00%	73	394	18.53 %
ABHD11	7	73151579	C>T	1	147	0.68%	66	187	35.29 %
GALNT11	7	151805164	G>C	0	147	0.00%	35	169	20.71 %
RIMBP2	12	130927064	C>T	0	303	0.00%	72	351	20.51 %
PSMD7	16	74335551	G>A	0	402	0.00%	81	314	25.80 %
CENPN	16	81058350	C>G	0	292	0.00%	72	212	33.96 %
SOX9	17	70118864	C>A	0	73	0.00%	72	145	49.66 %
NPHS1	19	36322595	G>A	0	70	0.00%	10	34	29.41 %
RBFOX2	22	36157335	C>T	0	287	0.00%	97	326	29.75 %
KDM5C	X	53222717	Del:C	0	87	0.00%	47	79	59.49 %
KDM5C	X	53222723	C>G	0	87	0.00%	49	78	62.82 %
SATL1	X	84362594	C>G	0	221	0.00%	135	236	57.20 %
FLNA	X	153596345	Del:CCT	0	169	0.00%	84	198	42.42 %
ITGB3	17	45380104	G>T	23	186	12.37%	0	143	0.00%
LATS2	13	21562148	C>A	0	548	0.00%	0	387	0.00%
DIRAS3	1	68512683	C>T	0	398	0.00%	0	403	0.00%
NGEF	2	233791888	G>T	0	23	0.00%	19	63	30.16 %
ZNF493	19	21606235	T>G	0	323	0.00%	0	284	0.00%

SPATA21	1	16727227	G>A	105	378	27.78%	0	268	0.00%
DDX58	9	32487509	A>T	111	375	29.60%	1	338	0.30%
DAPK1	9	90317948	G>C	70	221	31.67%	0	190	0.00%
ALKBH8	11	107424632	T>A	78	271	28.78%	1	194	0.52%
KL	13	33635697	Del:C	91	337	27.00%	0	256	0.00%
ERCC5	13	103514868	C>A	114	332	34.34%	0	252	0.00%
DIO1	1	54371785	Del:AAC	0	162	0.00%	0	129	0.00%
PIAS3	1	145578718	C>G	0	322	0.00%	0	289	0.00%
MR1	1	181021644	A>T	0	176	0.00%	1	158	0.63%
C3orf20	3	14803014	G>C	0	280	0.00%	0	262	0.00%
SETD2	3	47161717	G>A	0	273	0.00%	1	317	0.32%
TNIK	3	170856133	Del:T	0	589	0.00%	0	692	0.00%
LIAS	4	39466770	Del:A	0	551	0.00%	0	506	0.00%
FBXO1	5	41934121	G>C	0	269	0.00%	0	289	0.00%
AKAP9	7	91690733	G>C	0	240	0.00%	0	404	0.00%
ITIH5	10	7621865	C>A	0	169	0.00%	0	158	0.00%
WDR24	16	735684	G>A	0	148	0.00%	0	188	0.00%
MYH8	17	10299657	T>G	0	210	0.00%	0	206	0.00%
TOM1	22	35723300	Del:G	0	157	0.00%	0	151	0.00%
SBF1	22	50900272	G>T	0	100	0.00%	0	106	0.00%
KDM5C	X	53228342	A>T	0	147	0.00%	0	149	0.00%
USP51	X	55514714	C>G	0	132	0.00%	0	156	0.00%
NAP1L3	X	92928275	G>A	0	95	0.00%	0	117	0.00%
ADAMTSL4	1	150532548	Del:AGGAC	0	331	0.00%	0	372	0.00%
DUSP12	1	161719926	T>G	0	13	0.00%	0	32	0.00%
SLC2A12	6	134328018	A>G	0	194	0.00%	1	191	0.52%
RAB27A	15	55497741	T>A	1	650	0.15%	0	634	0.00%
CIB2	15	78416057	Del:C	0	246	0.00%	0	183	0.00%
RPS8	1	45241793	G>T	0	60	0.00%	0	141	0.00%
FAM129B	9	130271358	C>A	0	16	0.00%	0	39	0.00%
PHF21B	22	45312250	G>T	0	26	0.00%	0	39	0.00%
HDAC6	X	48675014	C>A	0	35	0.00%	0	32	0.00%
MAP3K6	1	27688199	G>T	0	114	0.00%	0	78	0.00%
MAMLD1	X	149671727	T>G	0	22	0.00%	0	54	0.00%
DNMT3A	2	25523017	C>A	0	29	0.00%	1	52	1.92%
ZNF521	18	22805039	T>A	3	494	0.61%	0	345	0.00%
MMAB	12	109998874	G>T	0	95	0.00%	0	113	0.00%
DACH2	X	86071070	A>G	0	164	0.00%	0	180	0.00%
SLC2A1	1	43394624	DEL:catga<6>caaa	0	335	0.00%	38	301	12.62%
TM7SF4	8	105361447	G>A	0	407	0.00%	139	464	29.96%
ANKRD26	10	27326806	C>G	0	250	0.00%	61	245	24.90%
CD44	11	35231528	G>C	0	520	0.00%	124	554	22.38%
KRT4	12	53202570	C>T	1	389	0.26%	88	424	20.75%
KIAA1267	17	44116456	G>C	0	315	0.00%	30	175	17.14%

C3	19	6713210	G>C	0	149	0.00%	45	172	26.16 %
ADAMTS 10	19	8668659	C>T	0	118	0.00%	35	70	50.00 %
IFNAR1	21	34717554	G>A	0	694	0.00%	155	780	19.87 %
BCL11A	2	60688070	C>A	13	81	16.05%	1	176	0.57%
PLCL1	2	198948857	Del:A	101	558	18.10%	0	624	0.00%
SETD2	3	47059231	A>T	33	154	21.43%	0	155	0.00%
KIAA1524	3	108285373	Del:A	114	724	15.75%	0	849	0.00%
NRAP	10	115412717	T>A	64	341	18.77%	0	395	0.00%
LAMA3	18	21533020	T>C	95	509	18.66%	0	372	0.00%
CDH19	18	64235921	G>T	85	367	23.16%	1	343	0.29%
SUPT6H	17	27018057	Del:GGACAATTTCCCT	0	107	0.00%	0	93	0.00%
WDR7	18	54444053	Del:CT	0	1037	0.00%	0	1004	0.00%
C2orf85	2	242814354	T>G	3	42	7.14%	0	78	0.00%

Supplementary Table 4: Details of 119 somatic mutations identified in patient 002.

Del = deletion, Ins = insertion, FS = frameshift.

Gene	Chromosome	Position	Nucleotide Variant	Amino Acid Change
CCDC14	3	123680148	G>C	R>P
CLEC3B	3	45077050	C>A	C>X
DNHD1	11	6568627	G>C	W>S
FOXK2	17	80545003	G>T	Q>H
FUBP3	9	133488388	C>A	S>Y
GOLGA6D	15	75586748	G>A	G>S
KRT33B	17	39525884	A>C	N>T
MPHOSPH8	13	20245344	A>T	splice site
MYC	8	128752925	C>T	S>S
PACS1	11	66003390	G>A	Q>Q
PBRM1	3	52668751	Ins:A	FS
PCDH15	10	55782710	Del:G	FS
PCDHB4	5	140502541	A>G	T>A
POFUT2	21	46703346	C>T	P>L
POLG2	17	62481852	A>C	H>P
POLI	18	5182011	G>A	R>K

		3		
PPP1R1B	17	37791905	A>T	E>V
REG3G	2	79254935	C>A	G>G
SEC24A	5	134002515	C>G	P>A
SPNS1	16	28994190	A>G	I>V
SUN2	22	39134593	T>A	F>Y
THSD1	13	52971608	G>A	V>V
TNKS1BP1	11	57076542	G>A	E>K
TTN	2	179596625	G>A	E>E
VHL	3	10191513	Del:T	FS
ZEB1	10	31799718	T>G	F>C
ZNF14	19	19823124	Del:A	FS
SLFN5	17	33592101	Del:A	FS
ATXN3	14	92548802	A>T	K>I
ENPP5	6	46135891	G>A	G>R
RAI2	X	17819421	Del:AT	FS
SCNN1G	16	23223474	T>A	splice site
VARS	6	31760805	T>C	A>A
LCT	2	136594624	Ins:A	FS
VPS13A	9	79954536	C>T	L>L
PCLO	7	82763696	C>G	S>X
SETD2	3	47142997	Ins:CTTC	FS
AP001107.2	11	66114447	Del:G	FS
CCT3	1	156304539	Del:C	FS
NUP62	19	50412772	A>G	N>S
RNF8	6	37336689	T>A	F>I
SLC22A7	6	43266294	G>T	E>D
MYBPC1	12	102055018	C>T	Y>Y
FLG	1	152286668	C>A	H>N
AGER	6	3215094	C>A	P>H

		7		
ARHGEF7	13	111955383	C>T	R>C
DCLRE1A	10	115607083	A>C	T>P
KIAA1549	7	138593868	G>A	splice site
PTEN	10	89685315	Del:G	splice site
TTC15	2	3428416	G>T	V>L
POLG	15	898618111	G>A	R>H
CNTN2	1	205041237	G>A	A>T
FAM86A	16	5143514	C>A	R>R
TNR	1	175355422	C>T	T>M
ADAL	15	43632530	C>T	P>L
HSPA14	10	148918111	G>A	splice site
PVRL4	1	161059028	Del:AGC	Del:L
KDM5D	Y	21877290	G>T	V>L
SNORD55	1	45241736	G>A	R>H
TRIO	5	143595889	C>G	I>M
TNIK	3	171177819	G>T	E>X
WWC2	4	184201996	Del:AAG	Del:E
CAPN14	2	31424845	G>T	W>L
TNXB	6	32063513	Del:AC	FS
ENSG00000251322	22	51117068	C>T	A>V
BCL7A	12	122468662	C>A	P>H
SYNM	15	99672798	G>A	T>T
MLLT4	6	168352522	G>T	TT
NADSYN1	11	71174508	C>A	N>K
MERTK	2	112786314	C>T	P>L
GRIK5	19	42546817	C>T	L>L
TRIM71	3	32932868	C>G	I>M
PTCHD3	10	27702413	C>T	A>V
ZFHX4	8	77765426	C>T	A>V

TXLNB	6	139581546	Del:TC	FS
OBP2A	9	138439776	T>A	Y>N
OR10H3	19	15852408	T>G	I>S
PARP8	5	50090830	C>A	S>X
NAP1L5	4	89618484	Del:TCC	Del:E
LRRC31	3	169557957	G>A	R>Q
KIAA0146	8	48309021	C>G	S>C
GPR98	5	90073831	G>A	E>K
CPVL	7	29103731	A>C	E>D
DNAJC27	2	25190090	G>T	G>X
DPYSL4	10	134013888	C>A	T>T
LPHN2	1	82409438	C>G	P>A
ITGA6	2	173333974	G>A	G>E
PTEN	10	89725051	T>A	L>Q
COL2A1	12	48370920	G>T	G>X
FARP1	13	99045913	C>A	H>N
GRHL1	2	101391114	C>A	P>T
IQCA1	2	237253239	A>T	E>D
KIAA0892	19	19465166	G>T	E>D
MGA	15	42058932	G>A	L>3L
NPHP4	1	6038367	C>T	P>L
TRRAP	7	98581774	G>A	M>I
ZBTB46	20	62421407	C>T	P>L
SLC40A1	2	190440038	Del:C	FS
DCX	X	110653364	C>T	T>M
ADAM12	10	127737976	G>A	R>Q
C9orf102	9	98684586	G>C	splice site
HAAO	2	42996948	G>T	E>X
RBBP6	16	24580307	Ins:T	FS
SFXN5	2	7329879	Del:GCC	Del:A

		7		
SPTAN1	9	131395564	C>A	P>H
ENDOU	12	48110160	C>T	A>V
TRIP11	14	92465743	A>G	N>S
TOP1	20	39746848	G>A	R>H
TP53	17	7577539	C>T	R>W
SETD2	3	47158204	T>C	C>R
SFRS6	20	42088682	C>T	R>X
ST8SIA1	12	22440159	G>A	G>E
PGK1	X	77372833	G>C	E>Q
NOL11	17	65718738	T>A	I>I
GPR126	6	142725060	A>C	I>L
GPR149	3	154146593	C>A	S>Y
DNAH2	17	7681656	C>T	P>S
ATP11A	13	113481146	G>A	E>K
EMP1	12	13366418	G>A	W>X

Supplementary Figure Legends

Supplementary Figure 1: Regional validation of mutations identified by multi-region exome sequencing in patient 001

The figure shows 37 mutations validated using Sanger sequencing. Mutations have been characterised and validated, in most cases, from tumour regions R3, R4, R9, M2a and germline (blood/PBMC) DNA. Mutations are indicated with an arrow. Note that KDM5C_2 shows two distinct mutations. Five genes DIXDC1, HPS5, HMG20A, RLF and SSR3 did not validate after Sanger sequencing, results not shown.

Supplementary Figure 2: Variant frequencies for R4 in ultra-deep sequencing data.

Each segment represents a mutation detected in R4 (radius = proportional to percentage of variant reads).

Supplementary Figure 3: Patient 001 phylogenetic tree based on synonymous and non-coding mutations.

We performed an exploratory analysis of heterogeneous synonymous and non-coding point mutations which are unlikely to be biased by selection pressures to reveal regional differences in the mutation spectrum and the overall mutational load of tumor 001. After clonal ordering, a phylogenetic tree was constructed which demonstrates similar ancestral relationships as the non-synonymous analysis (Figure 1C) with different branch lengths.

There was no significant difference in the mutations spectra in branch 1 vs. branch 2 with predominating G:C>A:T substitutions in 41% and 50%,

respectively, indicating that the mechanism of mutation generation may be identical in these main branches. An average of 2 synonymous mutations were detected per non-synonymous mutation in the main branch 1 and its sub-branches and an average of 1.5 synonymous mutations per non-synonymous mutation in the main branch 2 and sub-branches. Regions in branch 2 are tetraploid/aneuploid compared to diploid regions in branch 1. This may indicate that an increase in ploidy renders cancer cells more tolerant to non-synonymous mutations which are more likely to be detrimental to fitness. Larger series will be required to investigate whether polyploidy confers a selective advantage through such a buffer effect, limiting clonal extinction due to the accumulation of unfavorable mutations in genomically unstable cancer cells or those treated with DNA damaging agents.

Supplementary Figure 4: Somatic mutations by regions.

Different sample collection times are shown.

Supplementary Figure 5: Patient 001 SNP array data analysis.

A: Allelic imbalance (AI): red (metastatic sites) and blue (primary tumor sites) markers highlight chromosomal sections where AI was detected. **B**: Regional segmented probe intensities from the chromosome 3p from tumor 001. Sections with decreased intensity correspond to the sections of reported allelic imbalance and to the localizations of the VHL and SETD2 genes, strongly suggesting LOH in these sections. Dark blue, light blue and grey highlight SNPs present in sections of strong loss, loss and no copy number change, respectively.

Supplementary Figure 6: Immunohistochemical staining for H3K36me3 in patient 001.

Positive control: ccRCC sample with wild type (WT) SETD2. Black arrows: tumor cells, White arrows: stromal cells.

Supplementary Figure 7: Alignment of mTOR to PI3K-beta and the secondary structure for PI3K-beta (PDB code 2Y3A).

The mTOR repressor region (residues 2430-2450) is shown as a black bar.

Supplementary Figure 8: Regional validation of mutations identified by multi-region exome sequencing in patient 002

The figure shows 20 mutations validated using Sanger sequencing. Mutations have been characterised and validated, in most cases, from tumour regions R3, R4, R7, R9, M and germline (blood/PBMC) DNA. Mutations are indicated with an arrow. GRHL1 did not validate after Sanger sequencing and results are not shown.

Supplementary Figure 9: Somatic mutations in patient 002

Number of ubiquitous, shared and private mutations in samples analyzed with exome sequencing in patient 002.

Supplementary Figure 10: Ploidy profiles of patients 002-004

Ploidy profiles of each region in patients 002, 003 and 004. No sufficient data points were obtained to reliably estimate the ploidy of regions R1 and R8 from patient 002, region R1 from patient 003 and region R5 from patient 004.

Supplementary Figure 11: Allelic imbalance (AI) in patients 002, 003 and 004.

Blue markers highlight chromosomal sections where AI was detected.

Supplementary Figure 12: Immunohistochemical detection of Histone H3K36 methylation status in patient 001

Representative H3K36me3 staining of tumor cells (black arrows) from regions bearing different SETD2 mutations (R4 and M). Strong positive staining is seen in stromal cells (white arrows).

Supplementary Figure 13: Phospho-Akt staining for PTEN activity in patient 002.

Representative phospho-Akt (Ser473) staining of tumor regions with wild type (WT) PTEN (R4), PTEN carrying a missense (R7) or splice-site mutation (R10).

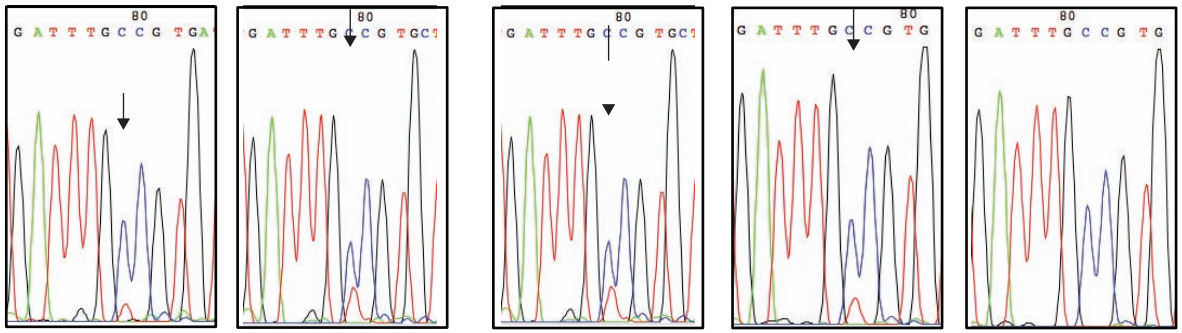
Supplementary Figure 14: Immunohistochemical detection of Histone H3K36 methylation status in patient 004

A: Representative H3K36me3 staining in tumour cells from regions with either wild type SETD2 (R4) or SETD2 containing a frame shift mutation (R6) in patient 004. **B:** Electropherograms showing wild type SETD2 (R4 and blood) and a frame shift mutation (R6) in patient 004. Investigation of the mutation was performed using Sanger sequencing.

References

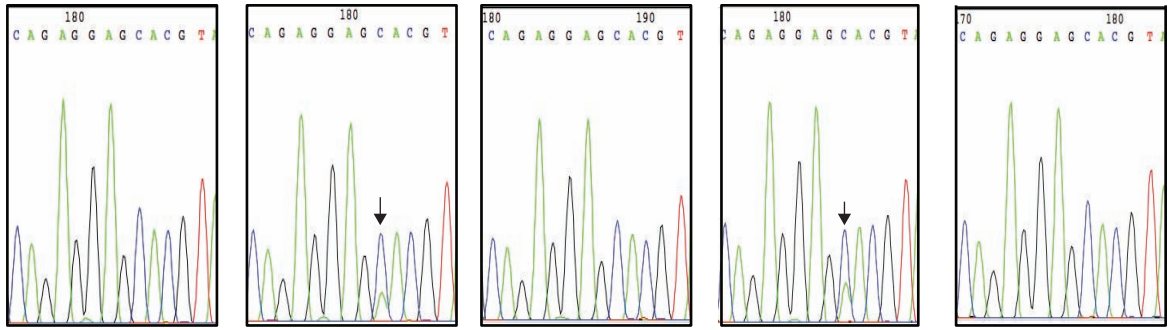
1. Agilent SureSelect protocol. http://www.chem.agilent.com/Library/usermanuals/Public/g3362-90001_sureselect_illuminahumanallexon_2.3.pdf. In.
2. Gnirke A, Melnikov A, Maguire J, et al. Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. Nature biotechnology 2009;27:182-9.
3. Tewhey R, Nakano M, Wang X, et al. Enrichment of sequencing targets from the human genome by solution hybridization. Genome Biol 2009;10:R116.
4. TruSeq Cluster Generation v5 protocol. https://icom.illumina.com/download/summary/m3k7sKMtGEOX1RjuSLQzjQ In.
5. Bentley DR. Whole-genome re-sequencing. Curr Opin Genet Dev 2006;16:545-52.
6. Bentley DR, Balasubramanian S, Swerdlow HP, et al. Accurate whole human genome sequencing using reversible terminator chemistry. Nature 2008;456:53-9.
7. TruSeq GAIIx SBS v5 reagent protocol. https://icom.illumina.com/download/summary/PRv_NCP-DEaVrFfTbSdQnw In.
8. Varela I, Tarpey P, Raine K, et al. Exome sequencing identifies frequent mutation of the SWI/SNF complex gene PBRM1 in renal carcinoma. Nature 2011;469:539-42.
9. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 2009;25:1754-60.
10. Li H, Handsaker B, Wysoker A, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics 2009;25:2078-9.
11. Goya R, Sun MG, Morin RD, et al. SNVMix: predicting single nucleotide variants from next-generation sequencing of tumors. Bioinformatics 2010;26:730-6.
12. dbSNP. http://www.ncbi.nlm.nih.gov/projects/SNP/. In.
13. R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing. http://www.R-project.org/ . In.
14. Morgan M, Anders S, Lawrence M, Aboyoun P, Pages H, Gentleman R. ShortRead: a bioconductor package for input, quality assessment and exploration of high-throughput sequence data. Bioinformatics 2009;25:2607-8.
15. Nakamura K, Oshima T, Morimoto T, et al. Sequence-specific error profile of Illumina sequencers. Nucleic Acids Res 2011;39:e90.
16. Merlo LM, Pepper JW, Reid BJ, Maley CC. Cancer as an evolutionary and ecological process. Nat Rev Cancer 2006;6:924-35.
17. Staaf J, Lindgren D, Vallon-Christersson J, et al. Segmentation-based detection of allelic imbalance and loss-of-heterozygosity in cancer cells using whole genome SNP arrays. Genome Biol 2008;9:R136.
18. Venkatraman ES, Olshen AB. A faster circular binary segmentation algorithm for the analysis of array CGH data. Bioinformatics 2007;23:657-63.
19. Irizarry RA, Hobbs B, Collin F, et al. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics (Oxford, England) 2003;4:249-64.
20. Brannon AR, Reddy A, Seiler M, et al. Molecular Stratification of Clear Cell Renal Cell Carcinoma by Consensus Clustering Reveals Distinct Subtypes and Survival Patterns. Genes Cancer 2010;1:152-63.

Supplementary Figure 1



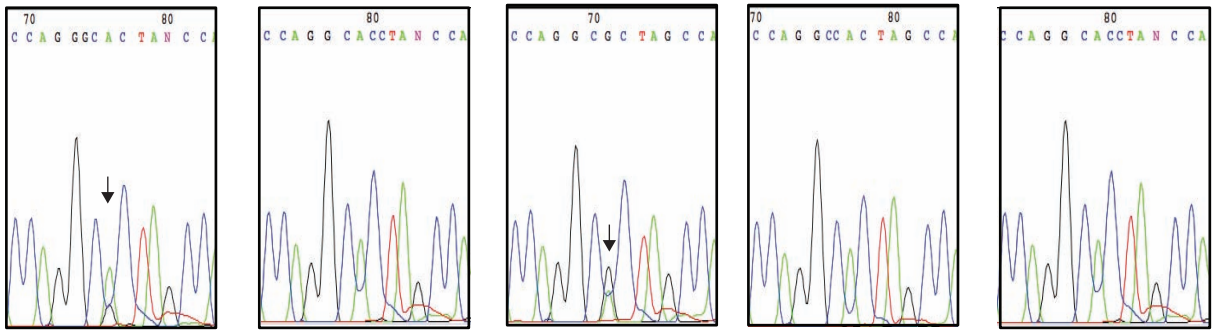
001/R3 001/R4 001/R9 001/M2a 001/Blood

[+] ATXN1



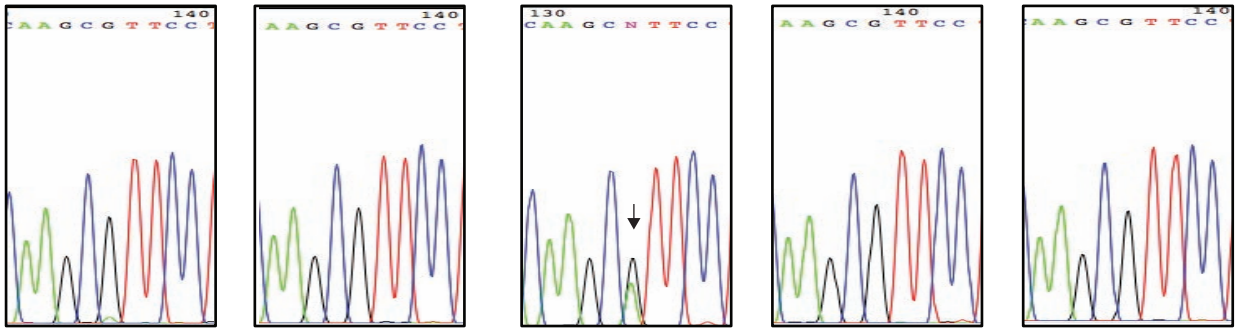
001/R3 001/R4 001/R9 001/M2a 001/Blood

[+] ERCC5



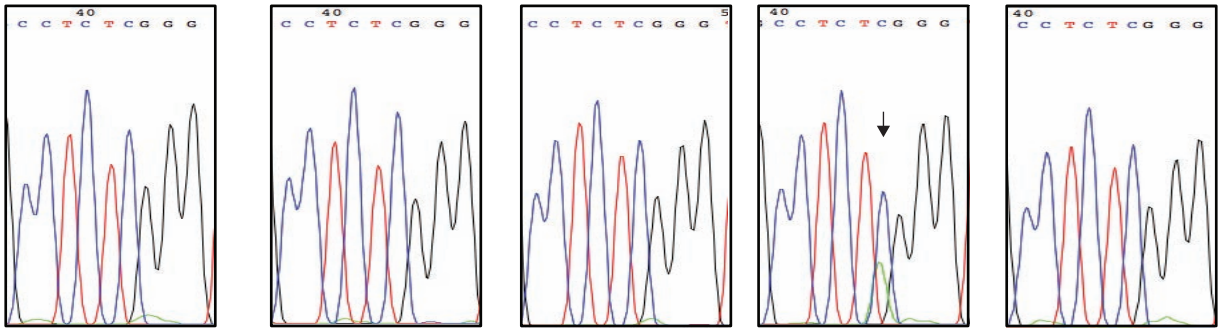
001/R3 001/R4 001/R9 001/M2a 001/Blood

[+] SATL1



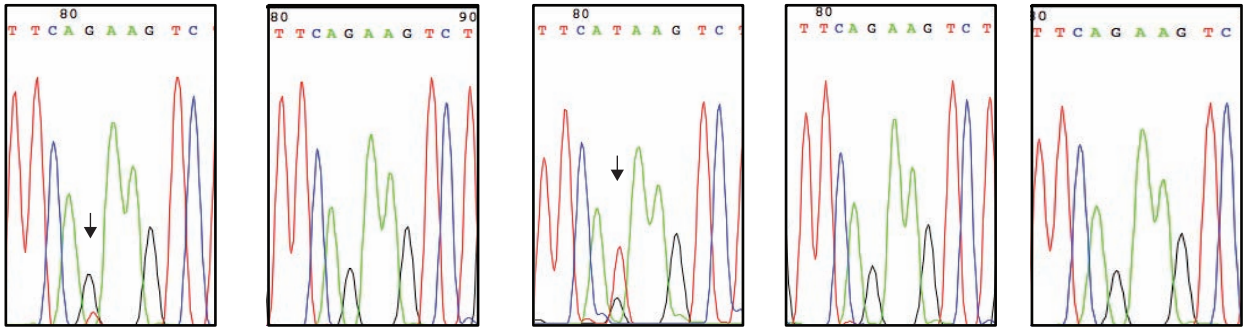
001/R3 001/R4 001/R9 001/M2a 001/Blood

[+] ADAMTS10



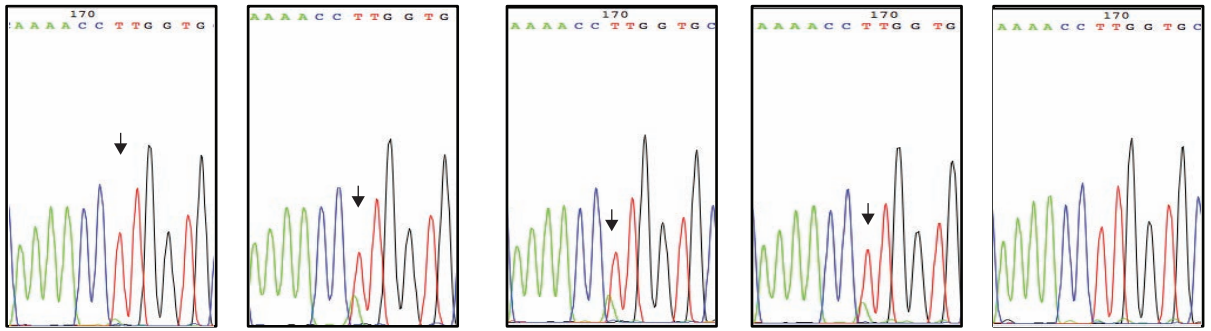
001/R3 001/R4 001/R9 001/M2a 001/Blood

[-] ITIH5



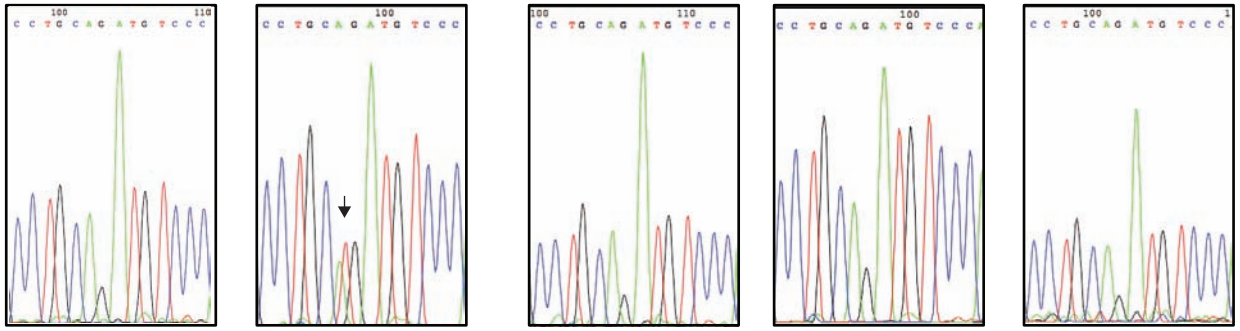
001/R3 001/R4 001/R9 001/M2a 001/Blood

[-] SOX9



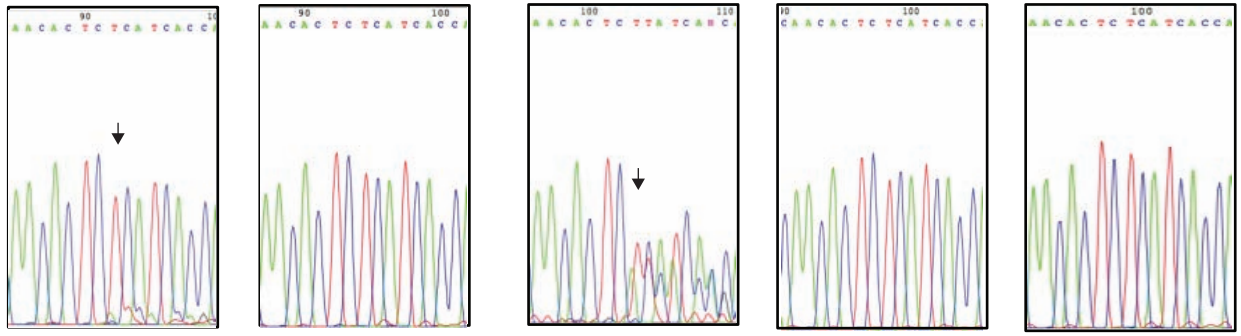
001/R3 001/R4 001/R9 001/M2a 001/Blood

CCR6



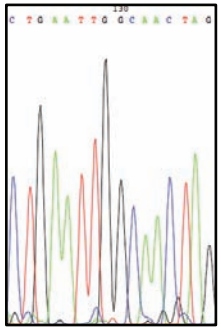
001/R3 001/R4 001/R9 001/M2a 001/Blood

[-] SETD2 (2)

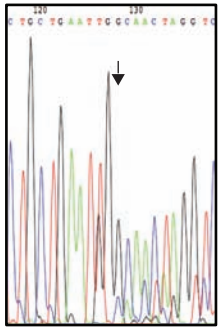


001/R3 001/R4 001/R9 001/M2a 001/Blood

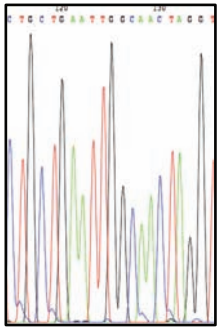
[+] SETD2 (3)



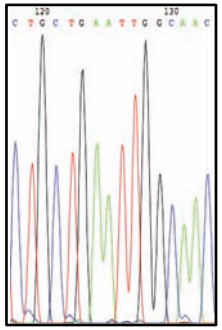
001/R3



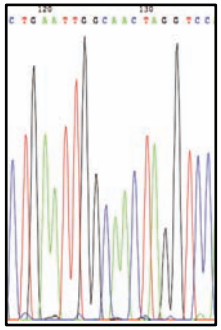
001/R4



001/R9

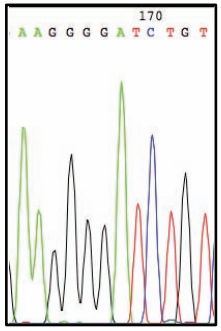


001/M2a

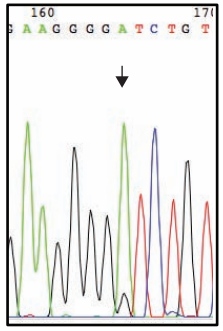


001/Blood

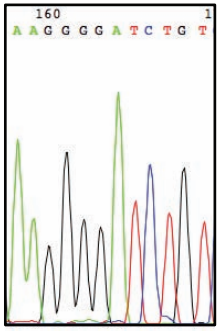
PLCL1



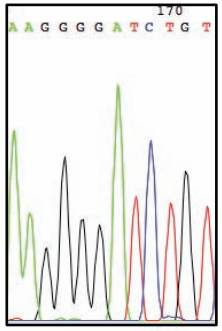
001/R3



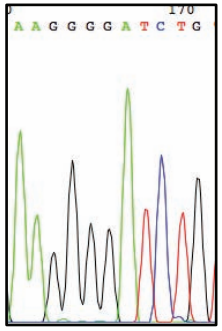
001/R4



001/R9

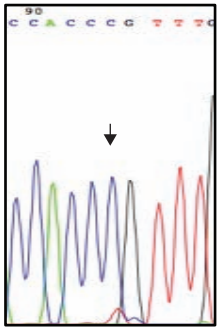


001/M2a

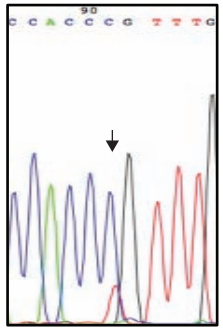


001/Blood

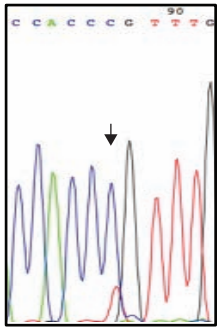
LAMA3



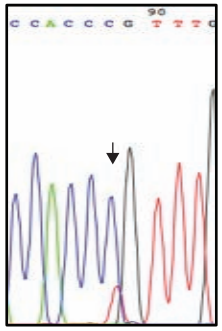
001/R3



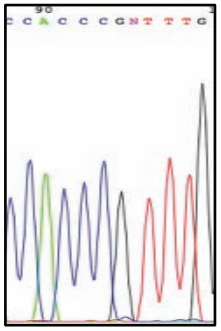
001/R4



001/R9

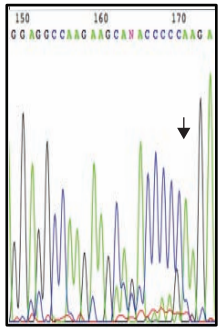


001/M2a

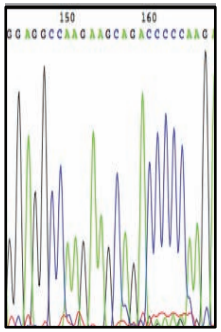


Blood

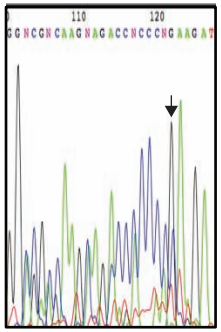
KDM2B



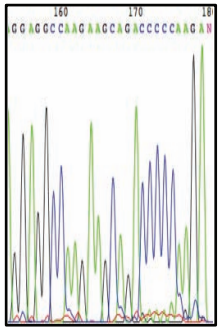
001/R3



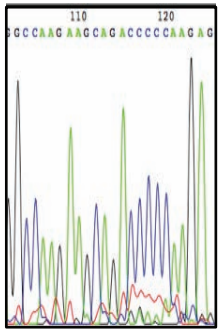
001/R4



001/R9

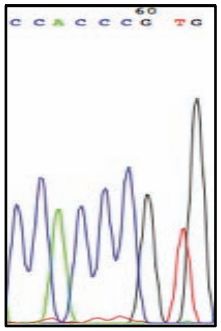


001/M2b

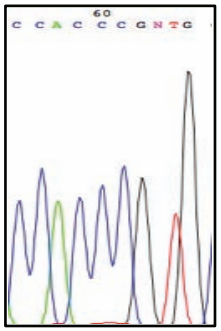


001/Blood

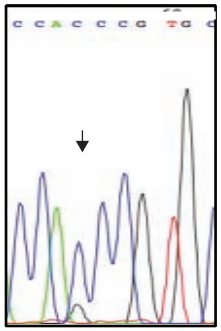
FLNA



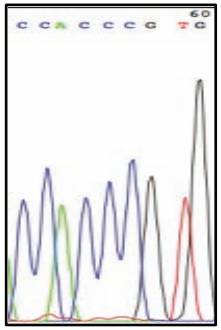
001/R3



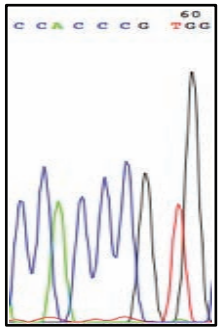
001/R4



001/R9

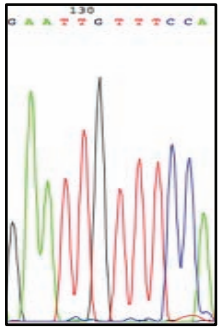


001/M2a

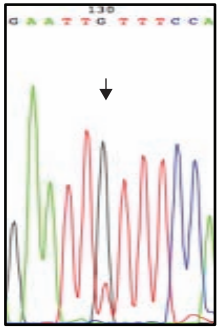


001/Blood

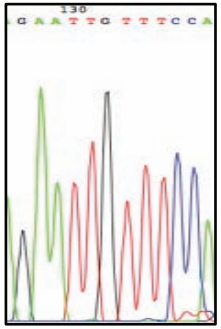
KIAA1267



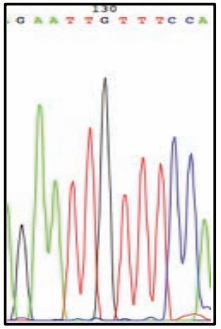
001/R3



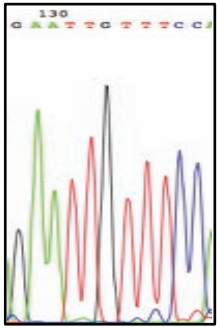
001/R4



001/R9

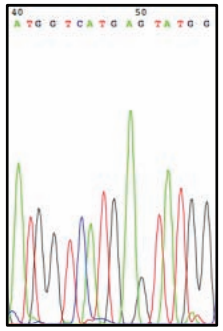


001/M2a



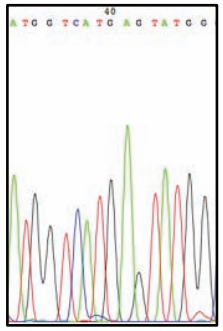
001/Blood

[-] CDH19

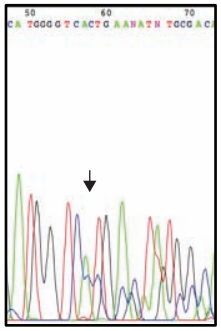


001/R3

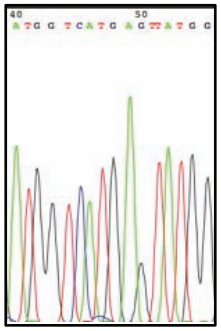
SLC2A1



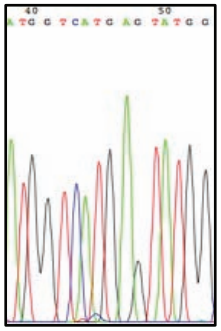
001/R4



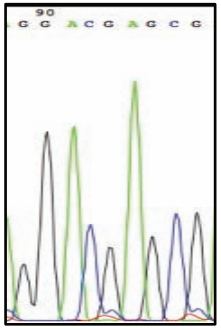
001/R9



001/M2a

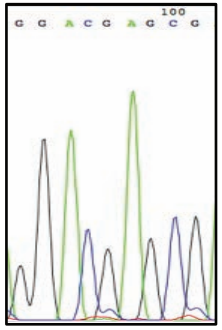


001/Blood

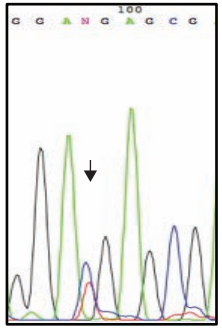


001/R3

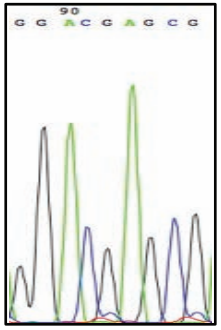
TM7SF4



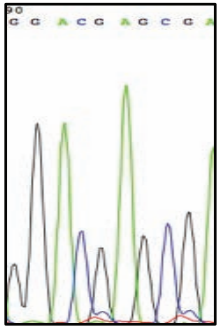
001/R4



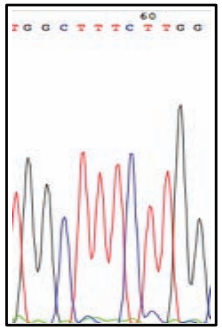
001/R9



001/M2a

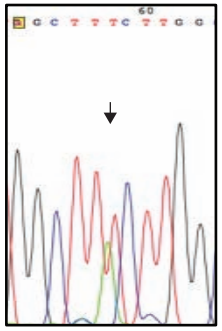


001/Blood

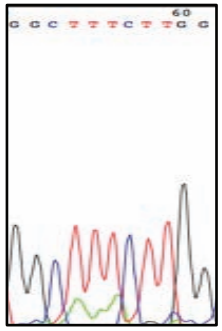


0001/R3

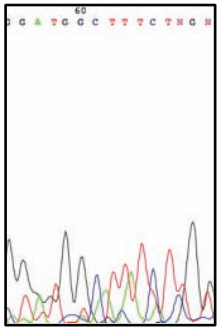
NRAP



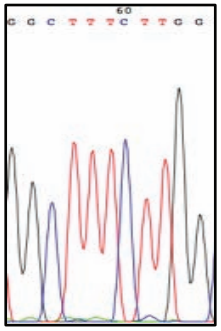
001/R4



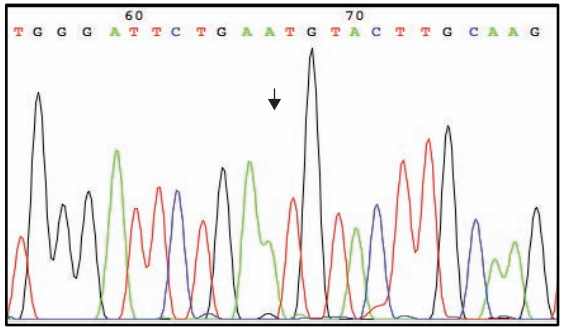
001/R9



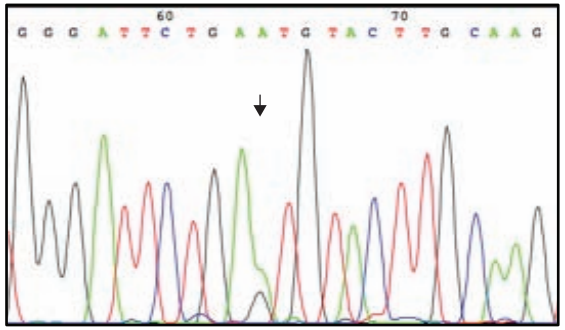
001/M2a



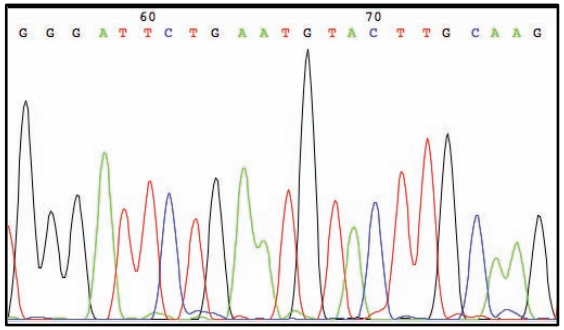
001/Blood



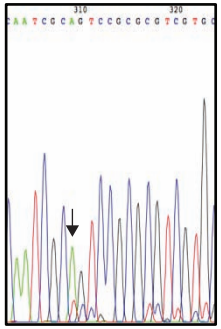
001/R1
SGOL1



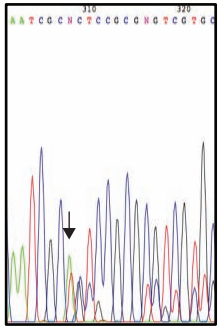
001/M2a



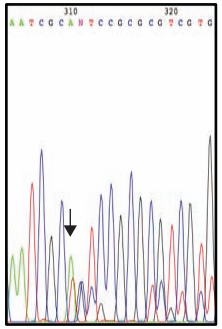
001/Blood



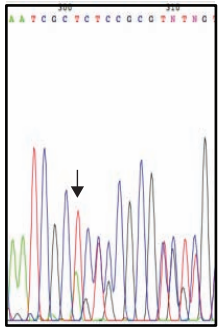
001/R1
VHL



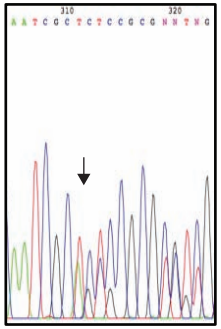
001/R2



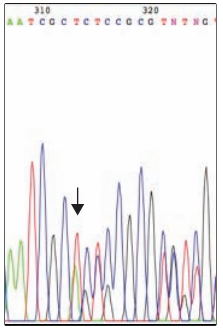
001/R3



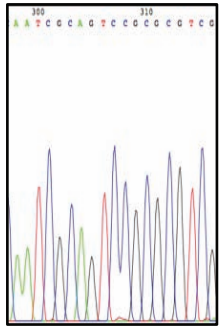
001/R4



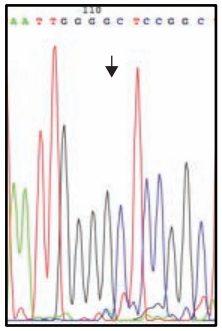
001/R9



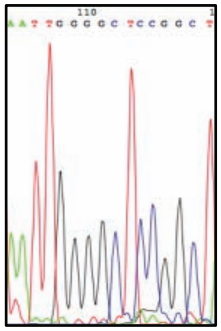
001/M2a



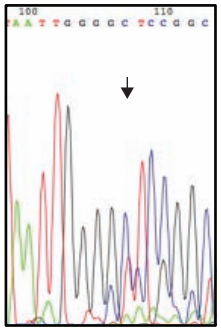
001/Blood



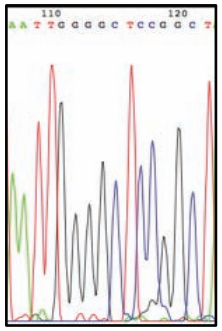
001/R3
CDKN1B



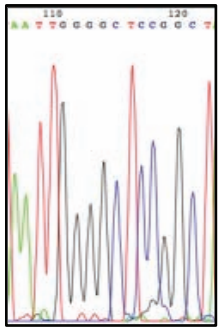
001/R4



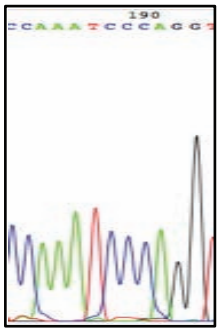
001/R9



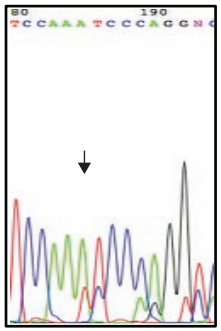
001/M2a



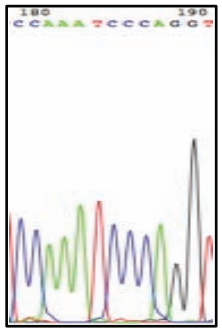
001/Blood



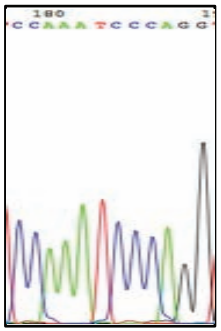
001/R3



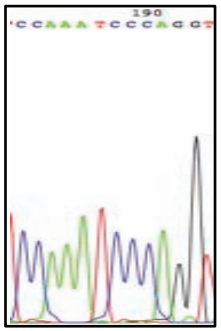
001/R4



001/R9

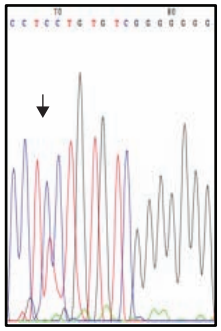


001/M2a

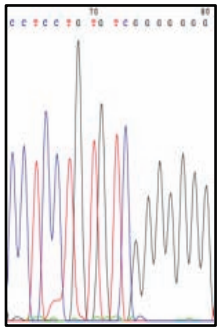


001/Blood

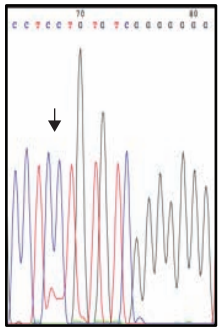
KIA1524



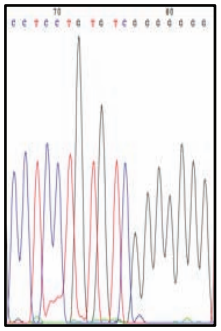
001/R3



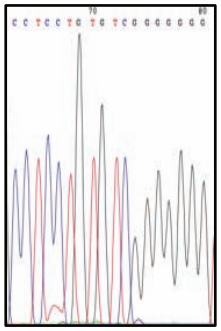
001/R4



001/R9

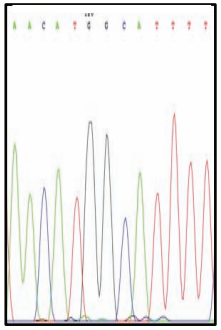


001/M2a

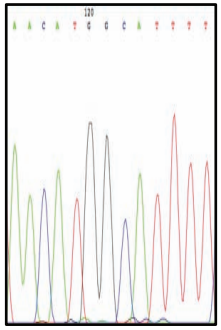


001/Blood

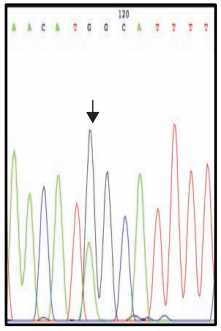
NPHS1



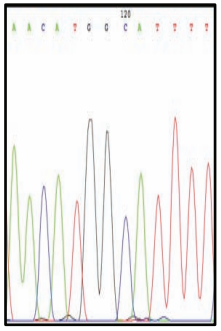
001/R2



001/R3

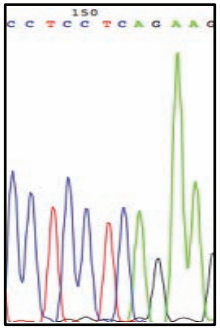


001/M2a

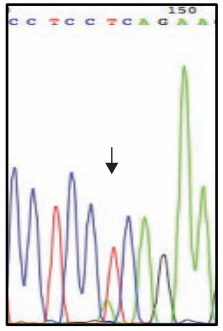


001/Blood

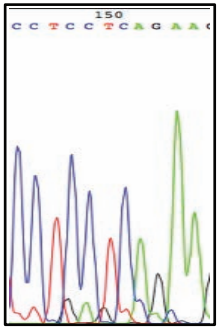
SETD2_1



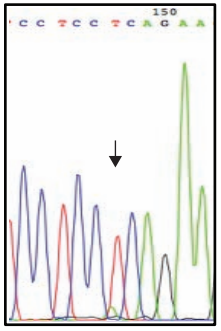
001/R3



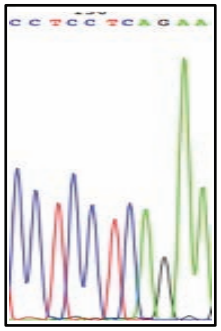
001/R4



001/R9

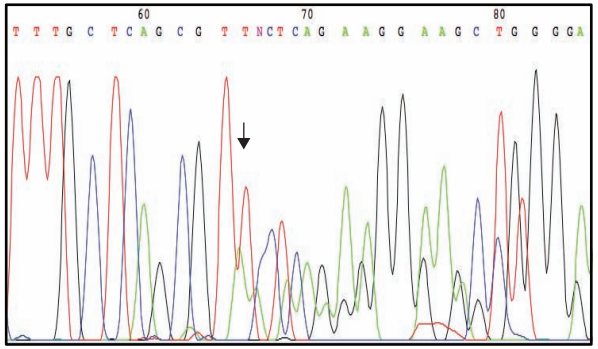


001/M2a



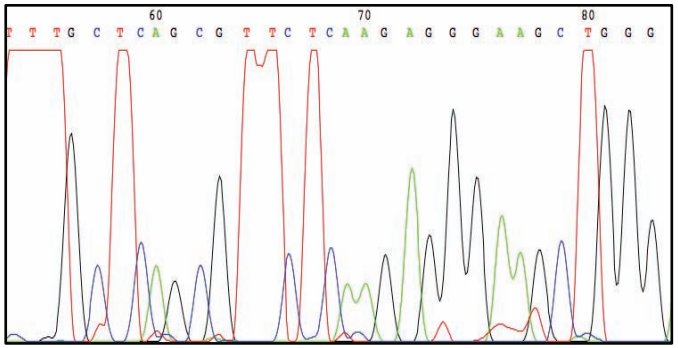
001/Blood

[-] ALKBH8

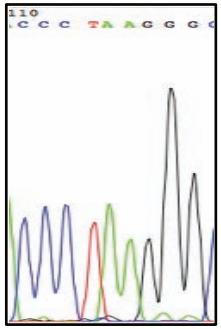


001/R9

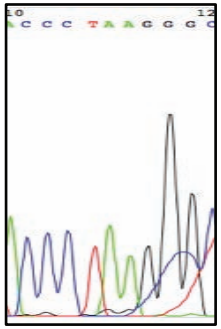
SSNA1



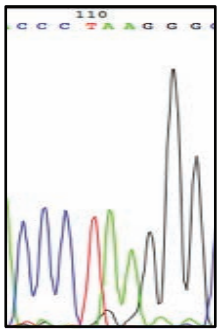
Blood



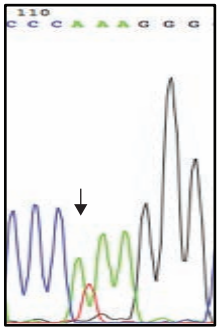
001/R3



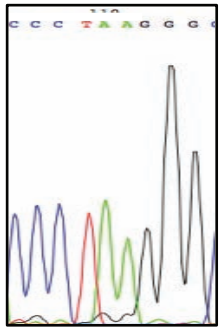
001/R4



001/R9

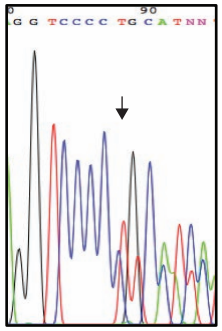


001/M2a

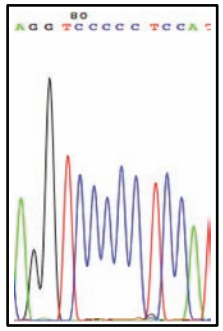


001/Blood

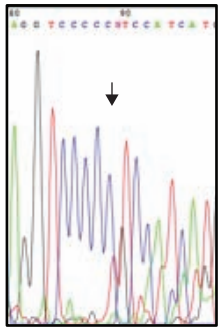
KDM5C (1) splice



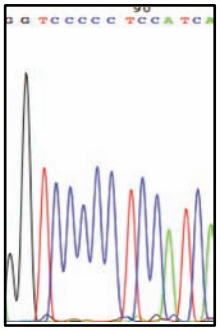
001/R3



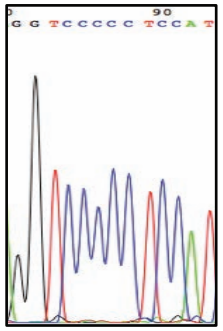
001/R4



001/R9

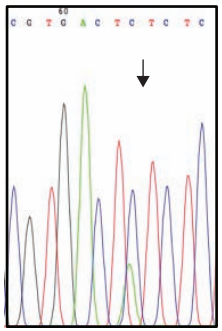


001/M2a

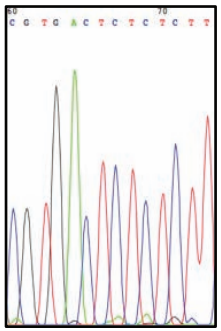


001/Blood

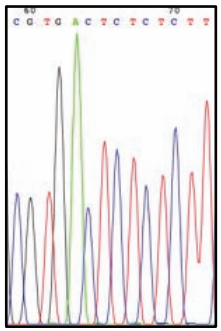
[-] KDM5C (2)



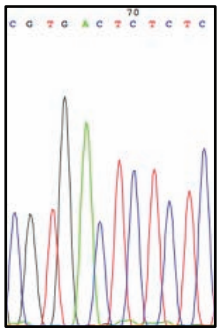
001/R3



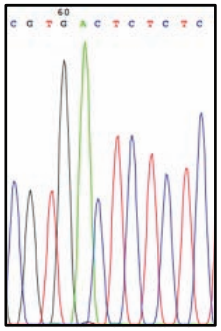
001/R4



001/R9

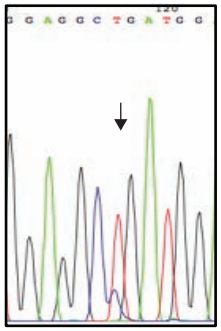


001/M2a

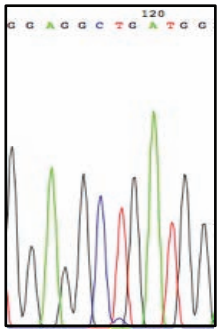


001/Blood

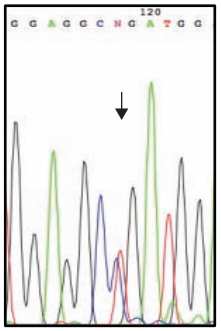
[-] LATS2



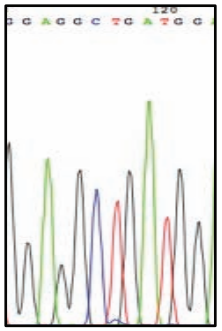
R3



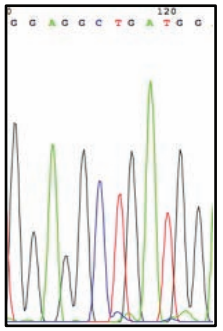
R4



R9

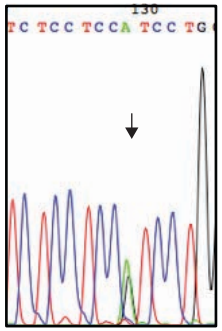


M2a

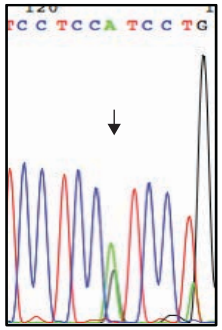


Blood

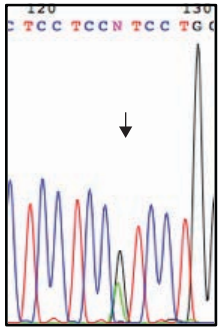
MTOR



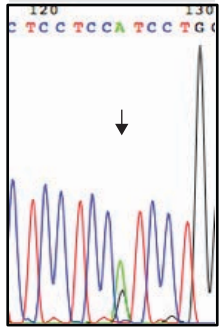
001/R3



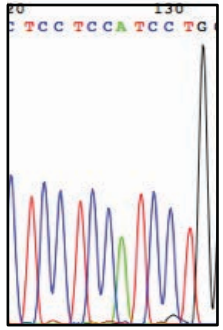
001/R4



001/R9

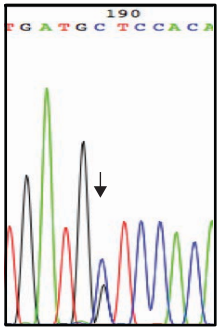


001/M2a

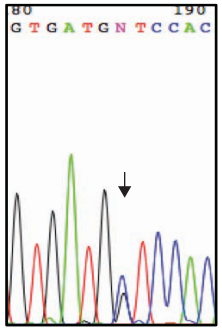


001/Blood

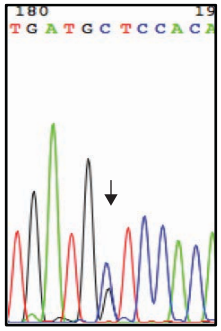
RALGDS



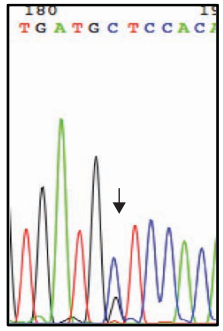
001/R3



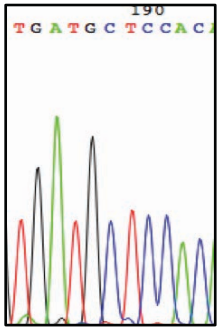
001/R4



001/R9

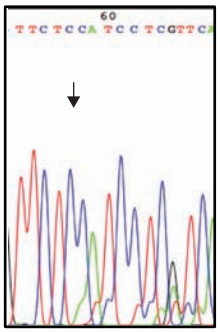


001/M2a

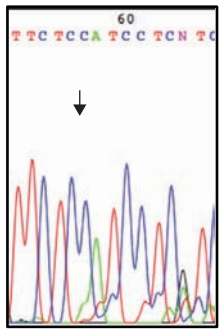


001/Blood

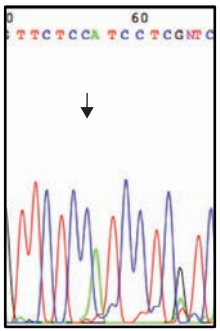
DDX52



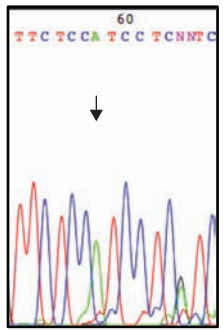
001/R3



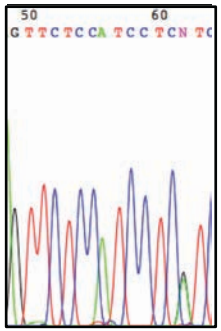
001/R4



001/R9

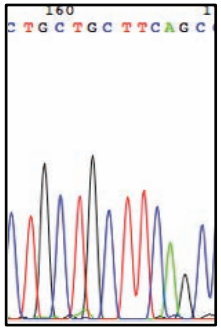


001/M2a

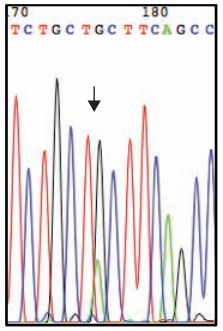


001/Blood

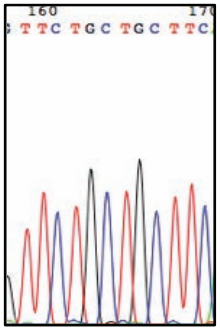
IGLON5



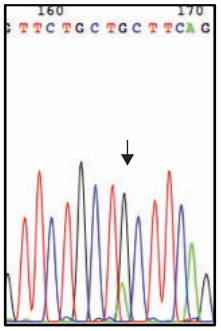
001/R3
SPATA21



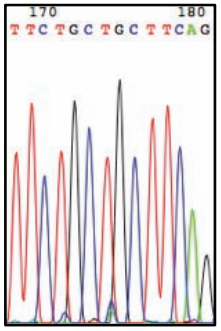
001/R4



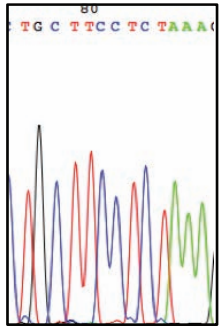
001/R9



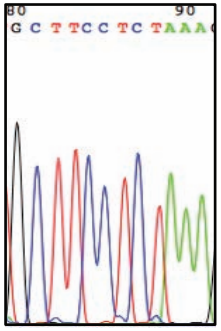
001/M2a



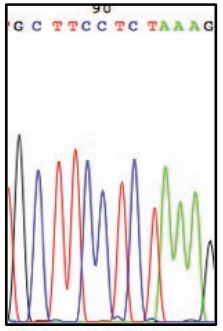
001/Blood



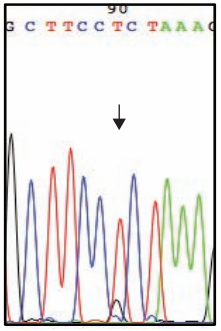
001/R3
MYH8



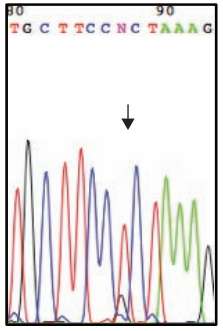
001/R4



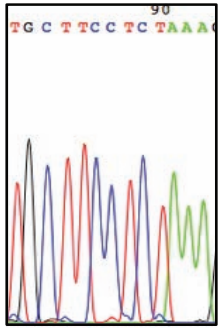
001/R9



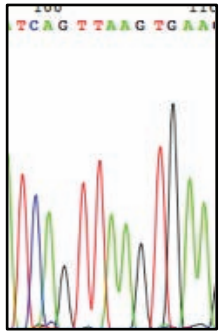
001/M1



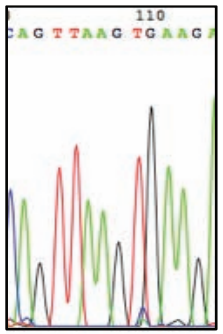
001/M2a



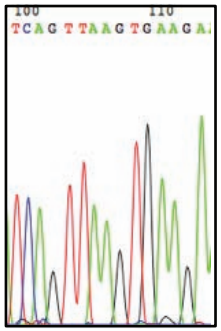
001/Blood



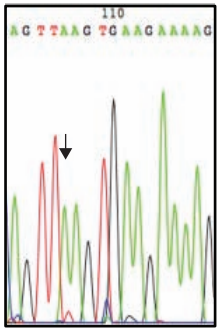
001/R3
RAB27A



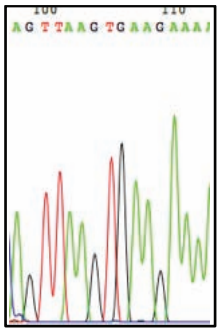
001/R4



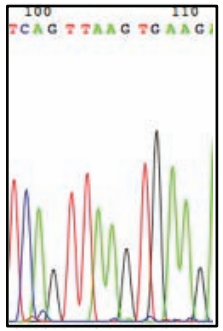
001/R9



001/M1

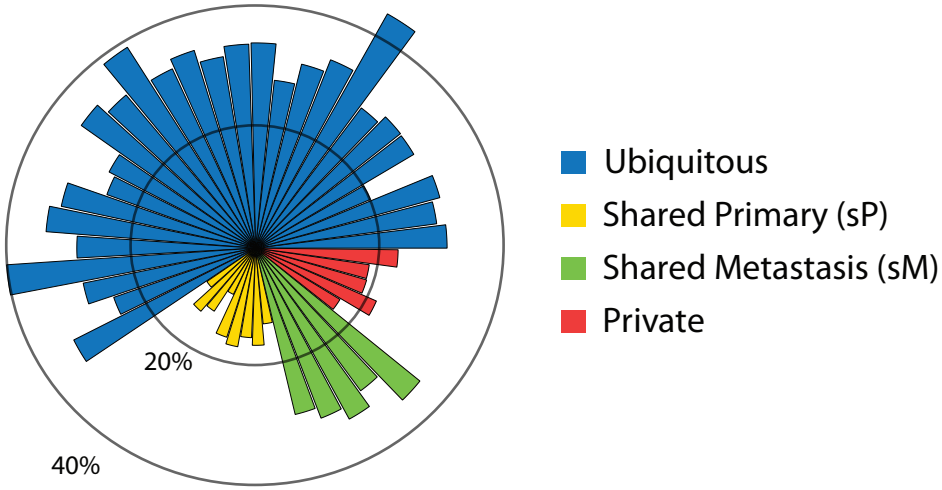


001/M2a

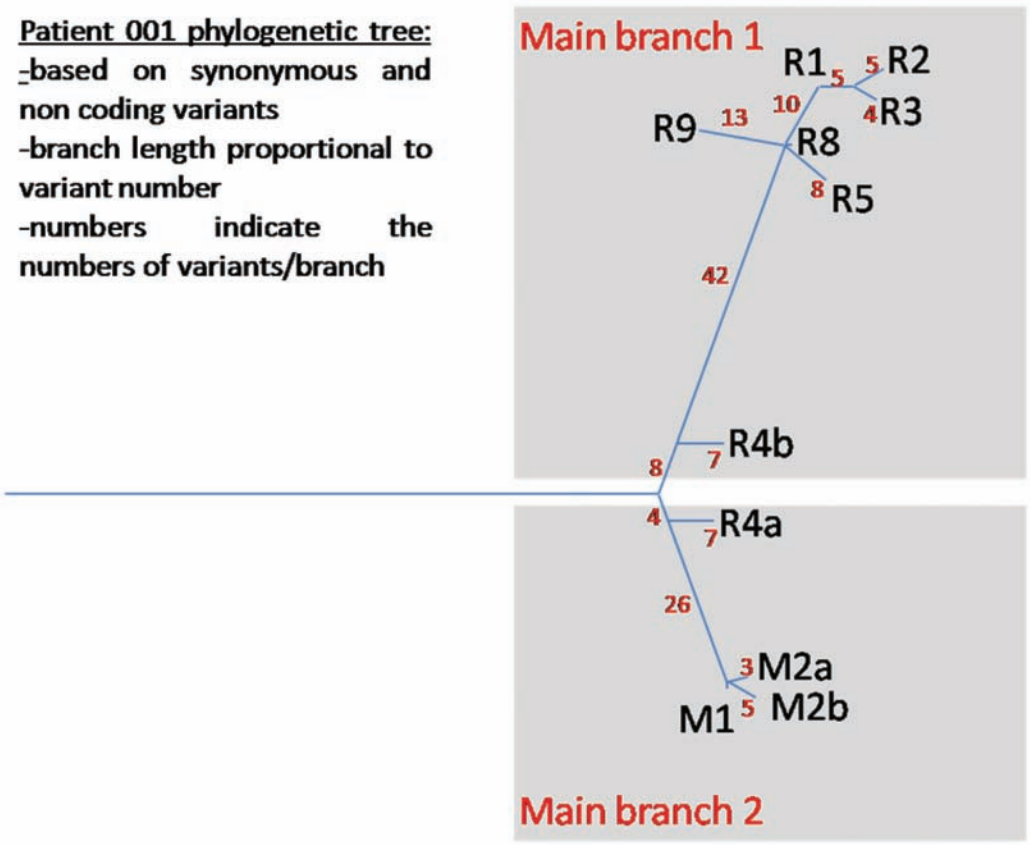


001/Blood

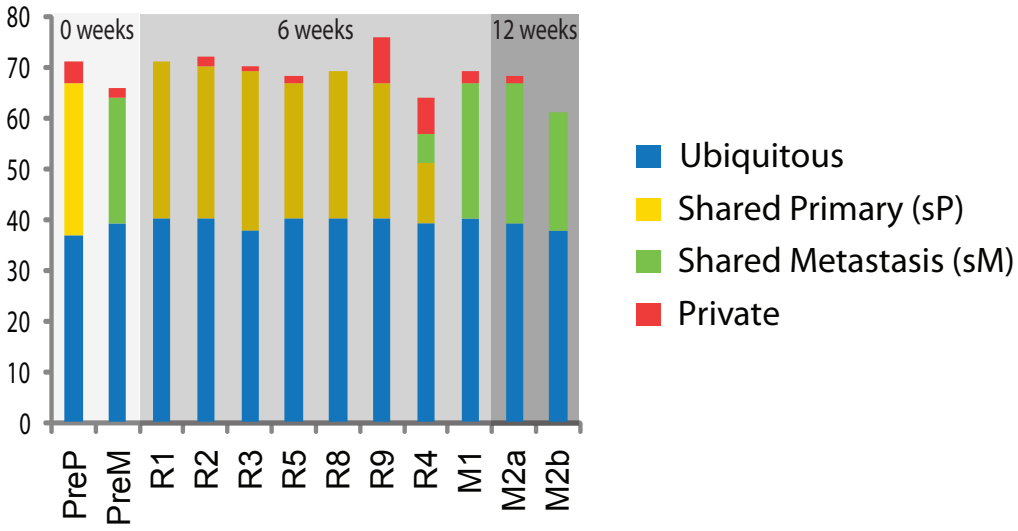
Supplementary Figure 2



Supplementary Figure 3

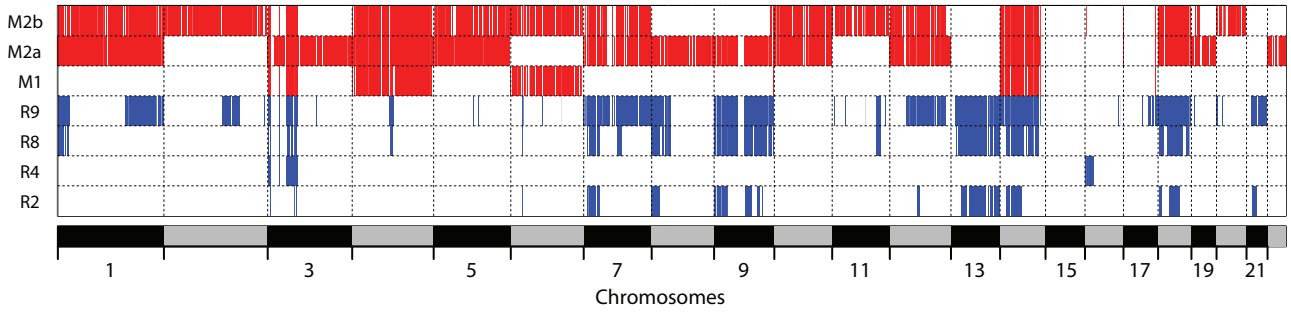


Supplementary Figure 4

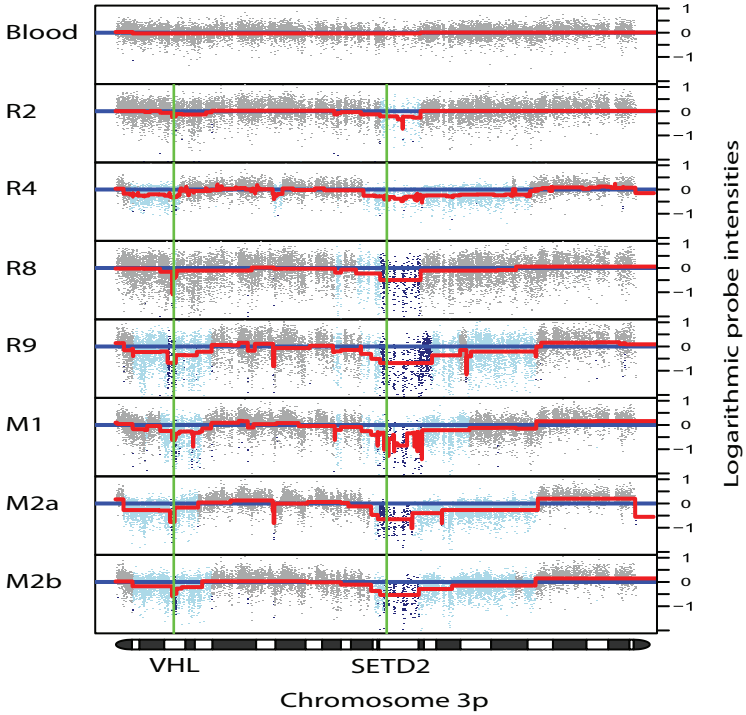


Supplementary Figure 5

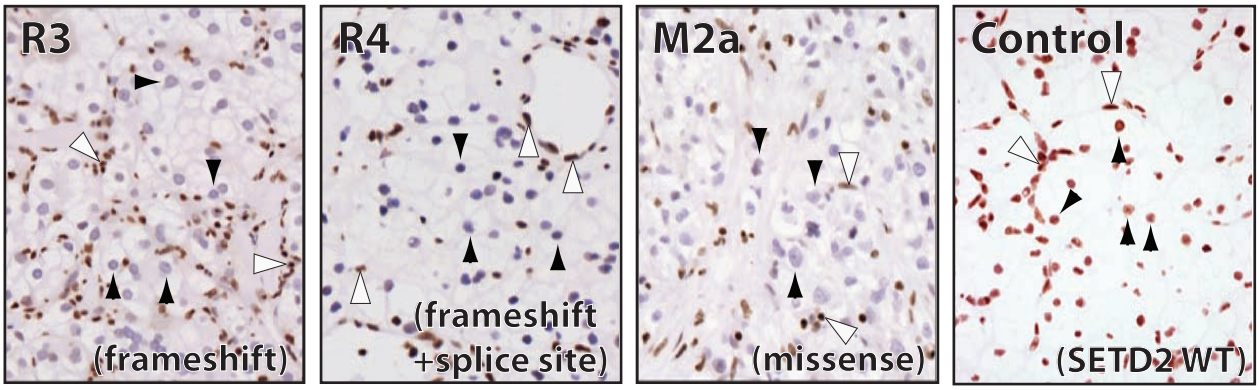
A



B

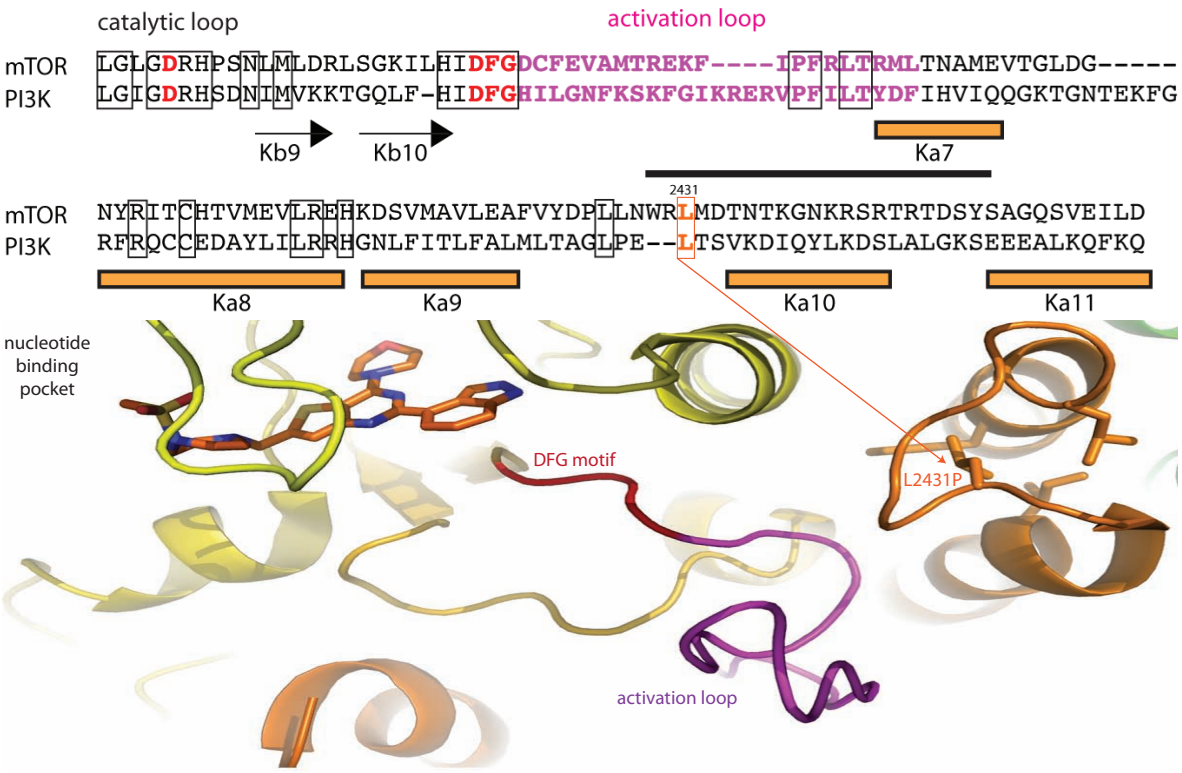


Supplementary Figure 6

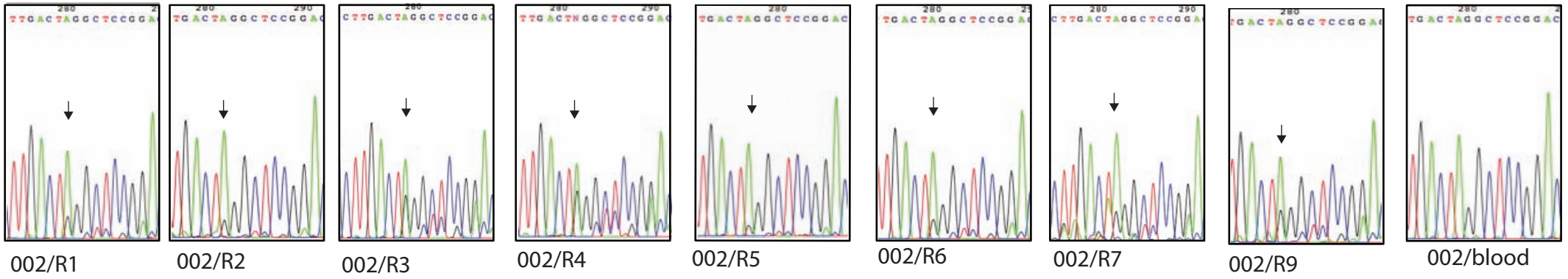


H3K36me3

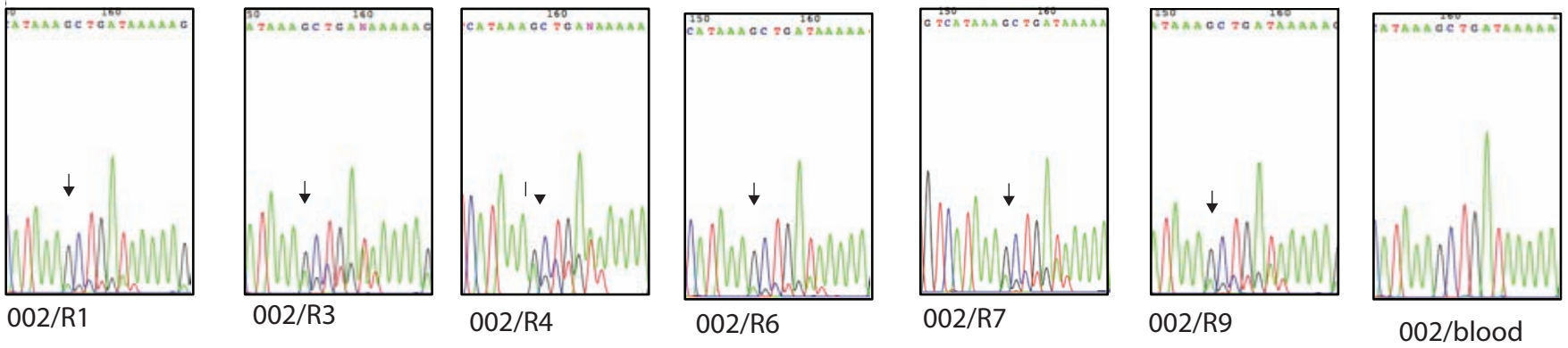
Supplementary Figure 7



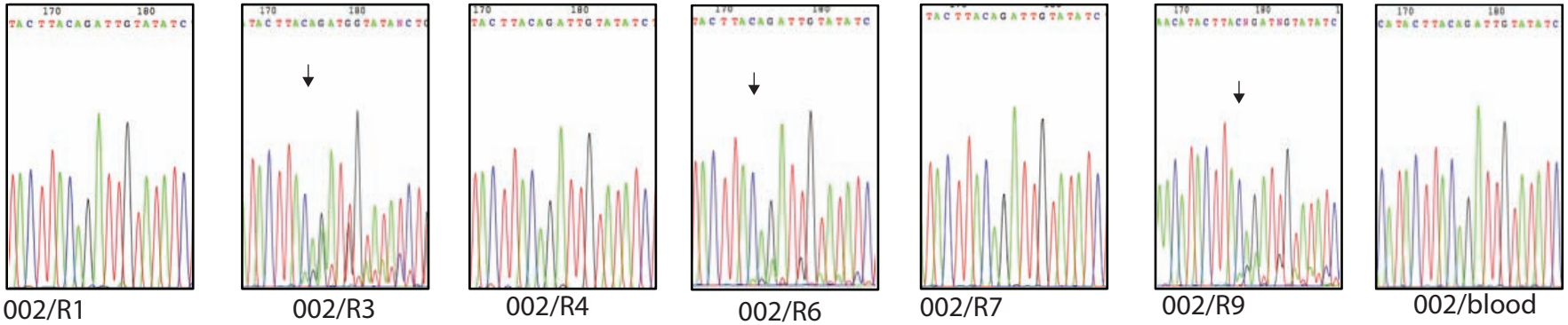
Supplementary Figure 8



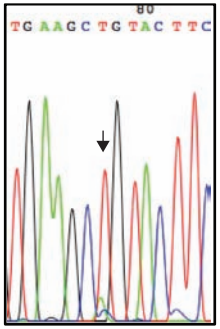
VHL



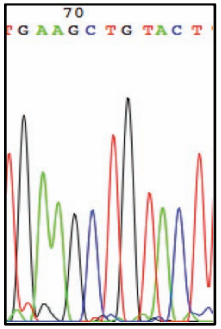
PBRM1



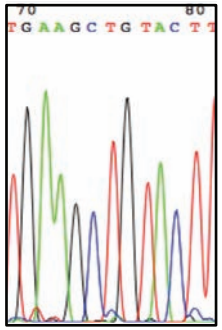
PTEN



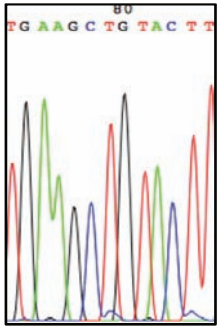
PTEN 002/R7



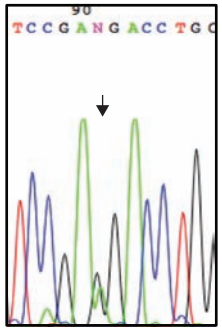
002/R2



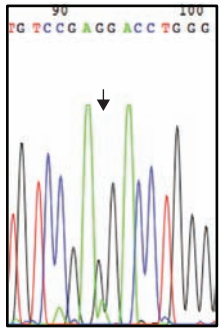
002/R6



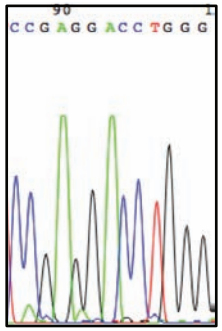
002/blood



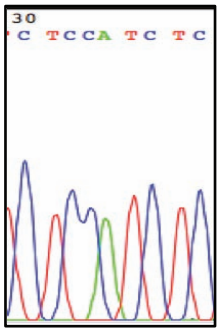
002/R7
MYC



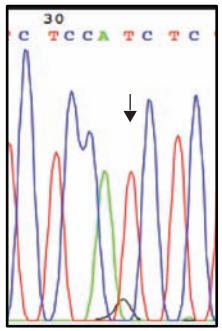
002/R9



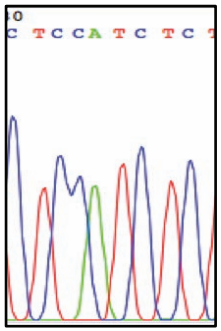
002/blood



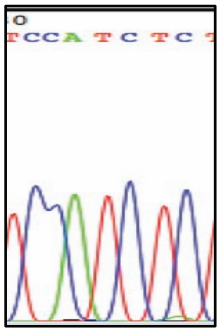
002/R3
OR10H3



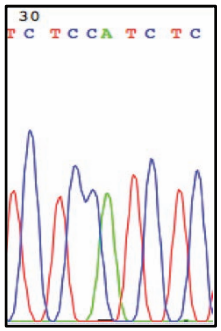
002/R4



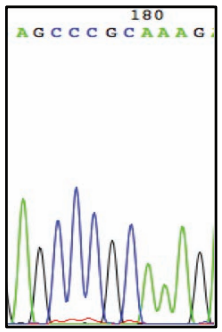
002/R5



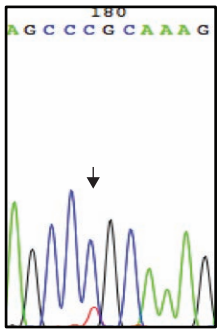
002/M



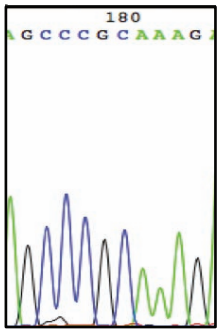
002/blood



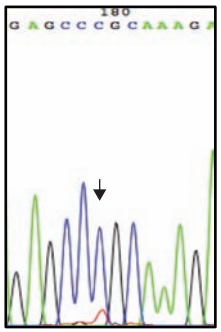
002/R1



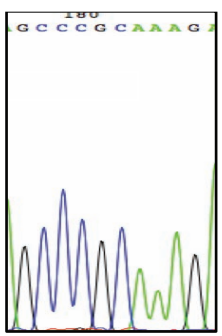
002/R3



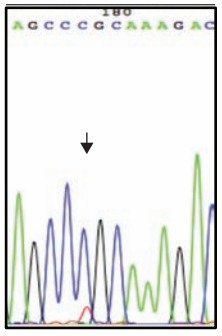
002/R4



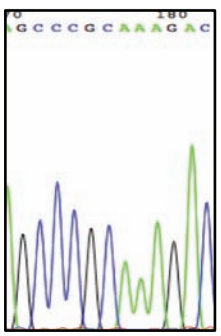
002/R6



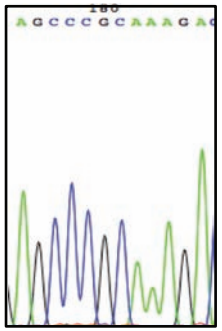
002/R7



002/R9

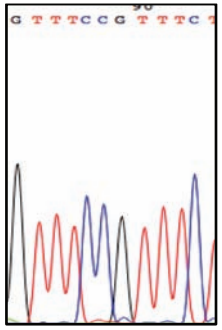


002/M

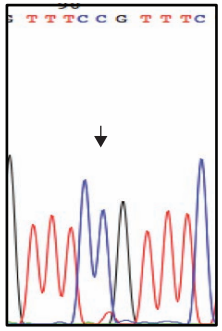


002/Blood

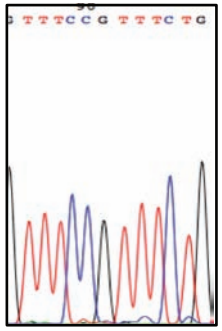
ARHGEF7



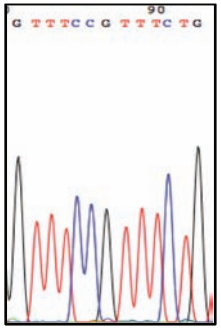
002/R1



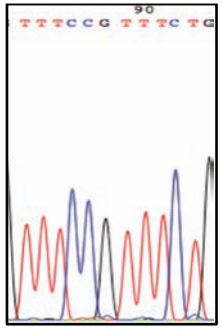
002/R3



002/R4

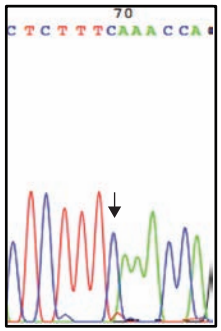


002/M

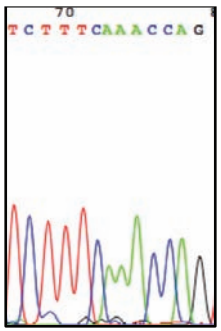


002/Blood

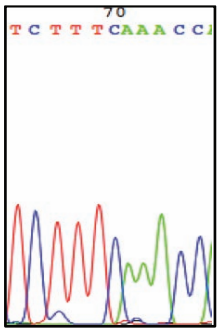
SYNM



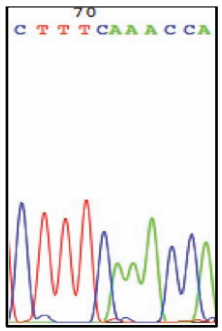
002/R7



002/R2

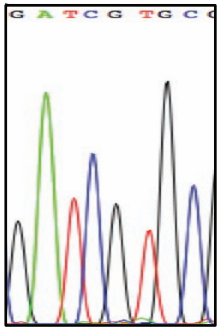


002/R6

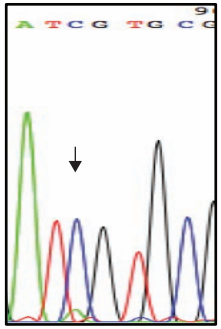


002/blood

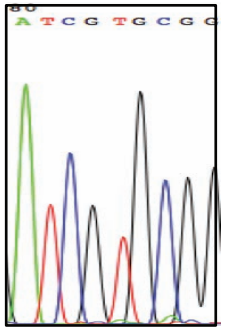
MGA



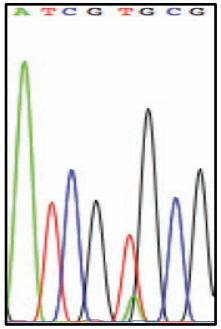
002/R2
GRIK5



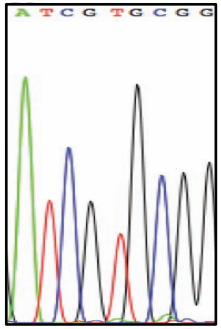
002/R3



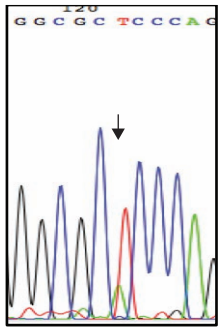
002/R4



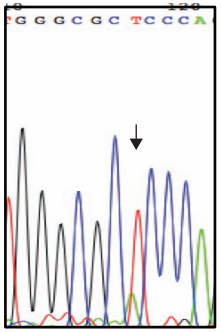
002/M



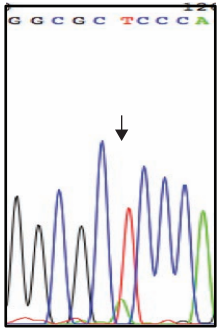
002/Blood



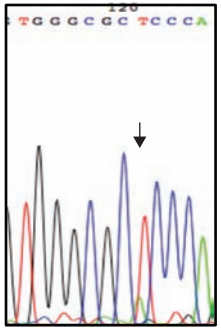
002/R3
PPP1RIB



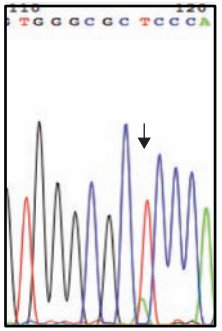
002/R4



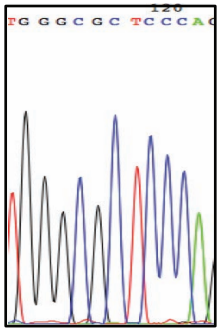
002/R7



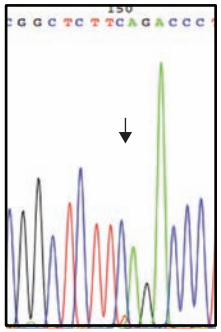
002/R9



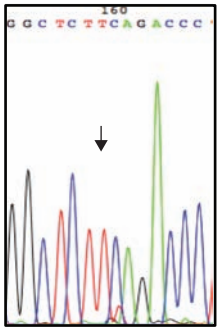
002/M



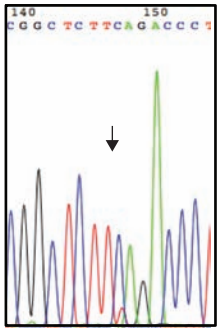
002/Blood



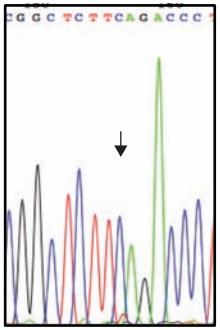
002/R3
TNKS1BP1



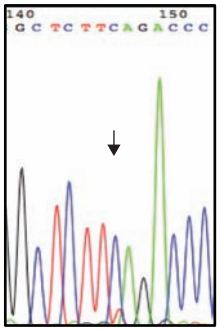
002/R4



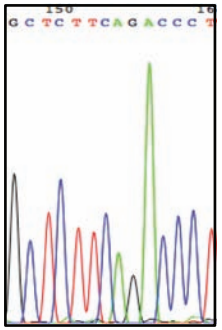
002/R7



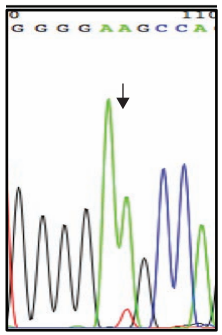
002/R9



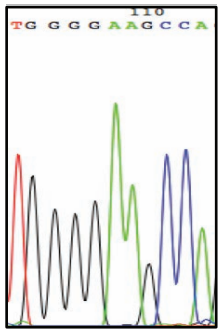
002/M



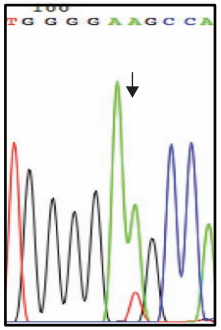
002/Blood



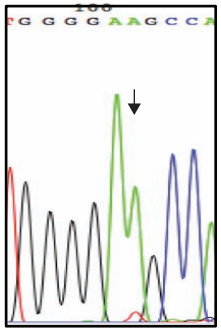
002/R3



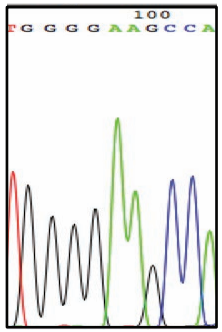
002/R4



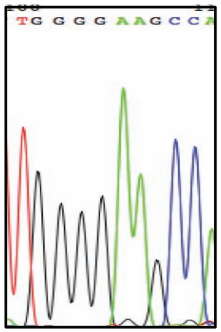
002/R7



002/R9

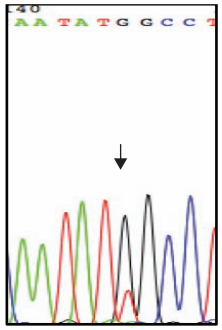


002/M

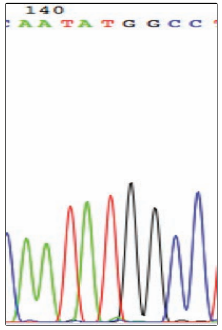


002/Blood

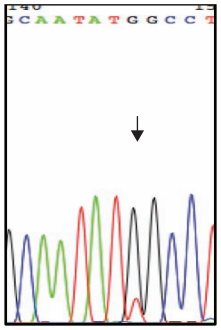
RNF8



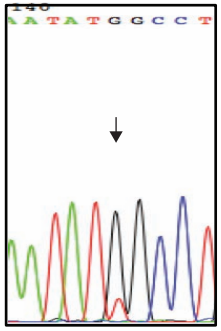
002/R3



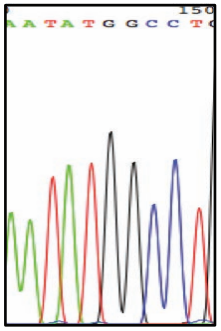
002/R4



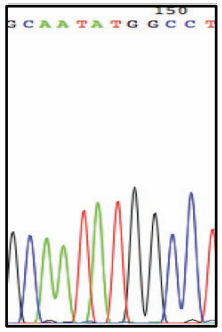
002/R7



002/R9

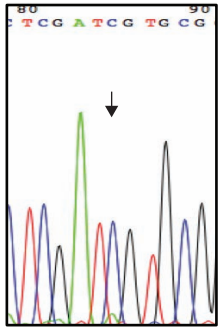


002/M

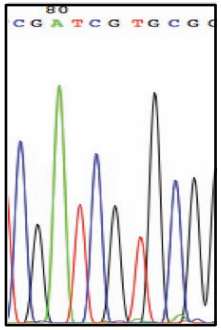


002/Blood

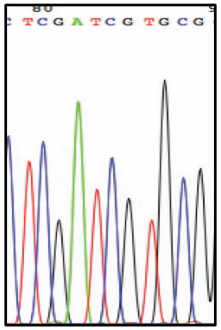
FLG



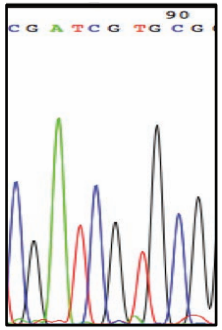
002/R3



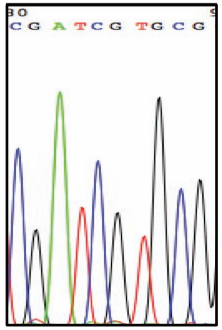
002/R4



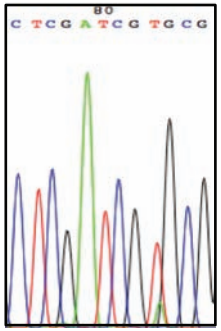
002/R7



002/R9

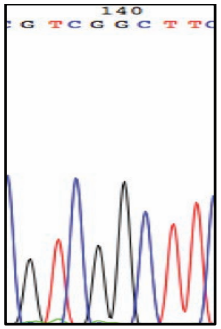


002/M

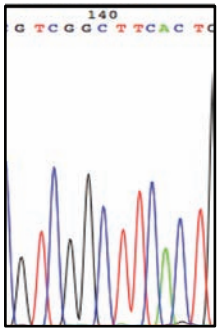


002/Blood

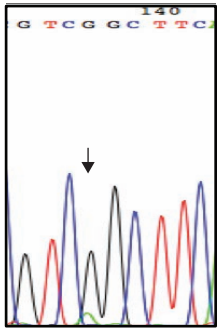
MLLT4



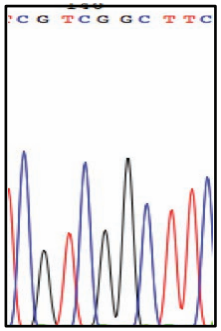
002/R2



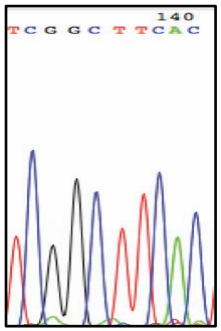
002/R4



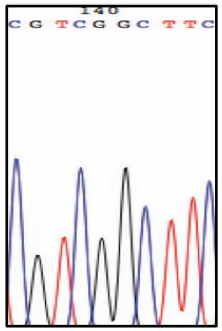
002/R7



002/R9

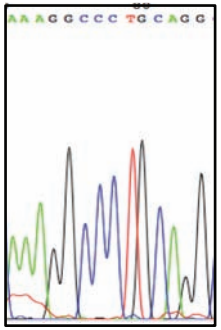


002/M

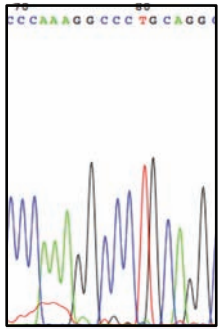


002/Blood

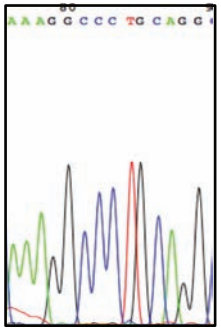
NPHP4



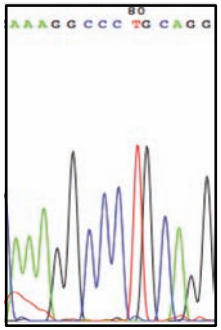
002/R2



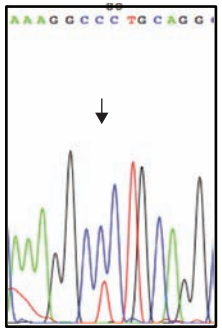
002/R4



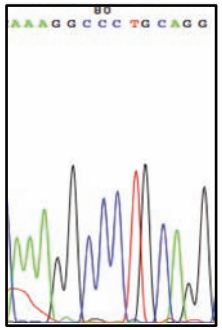
002/R7



002/R9

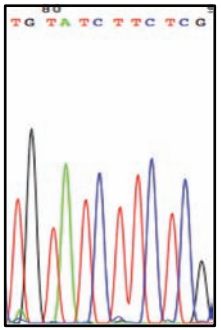


002/M

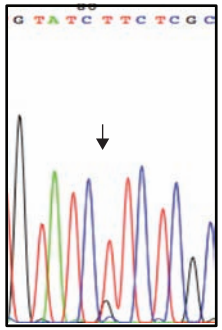


002/Blood

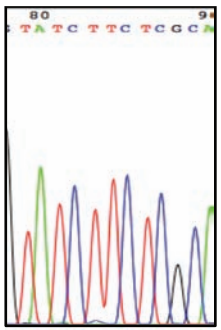
DNAH2



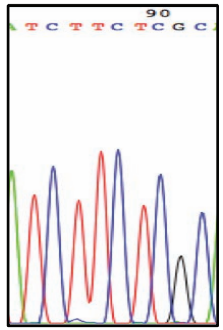
002/R2



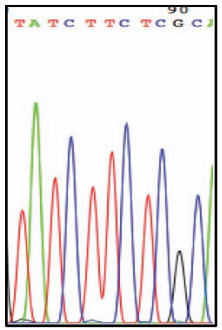
002/R4



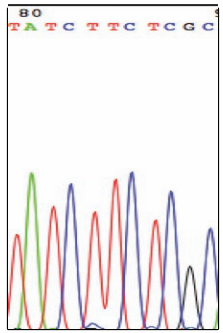
002/R7



002/R9

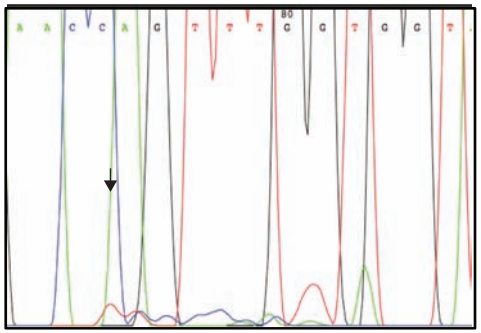


002/M

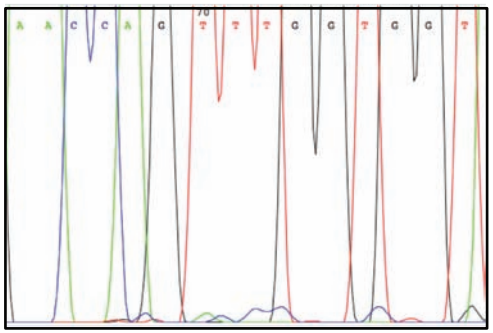


002/Blood

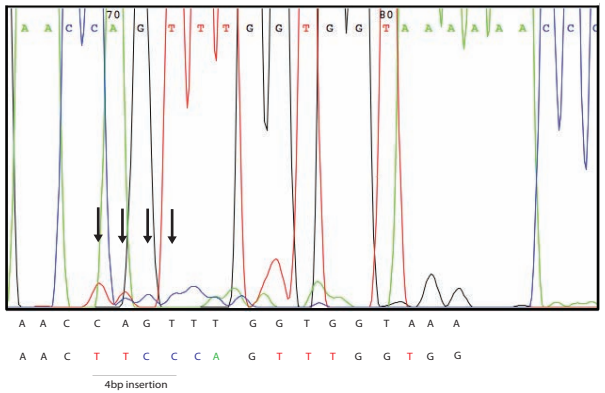
CPVL



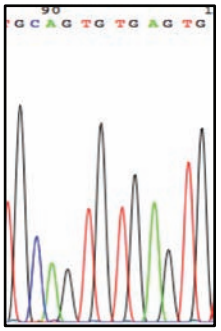
002/R9



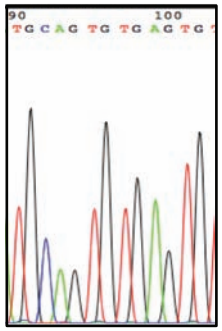
002/Blood



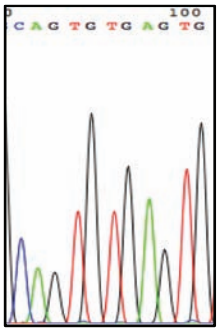
SETD2_1



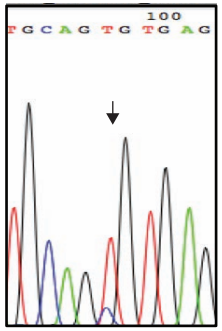
002/R4



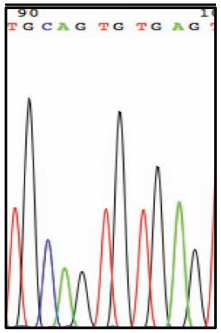
002/R7



002/R9

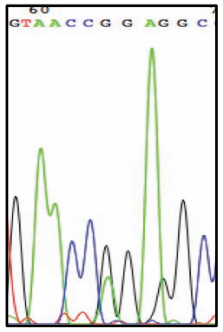


002/M

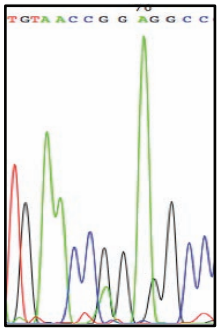


002/Blood

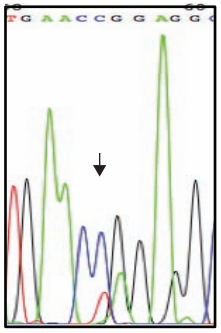
SETD2_2



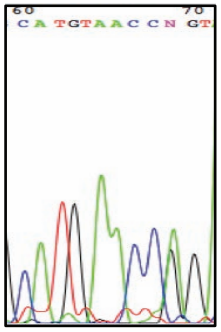
002/R4



002/R9



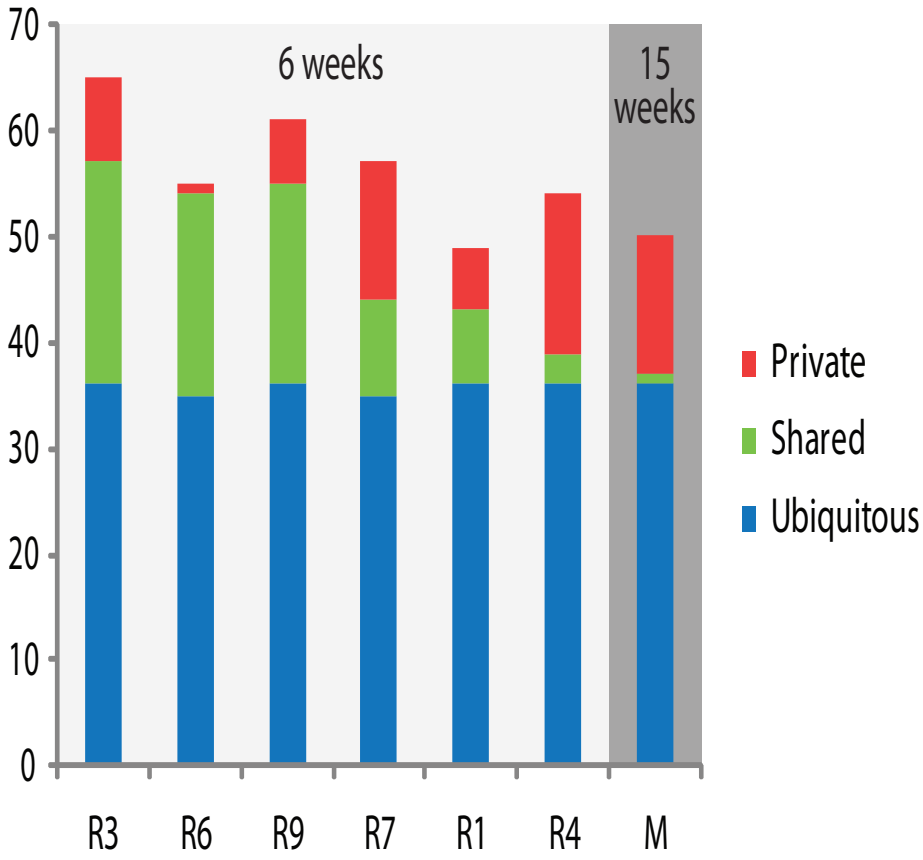
002/M



002/Blood

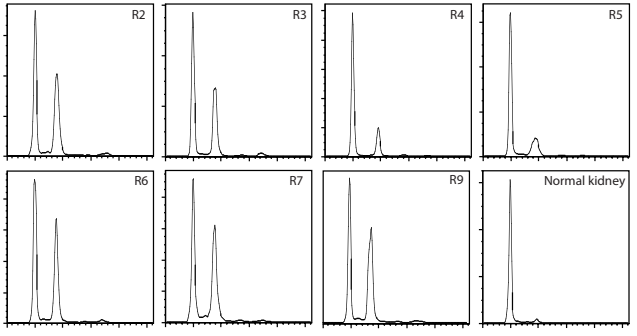
TP53

Supplementary Figure 9

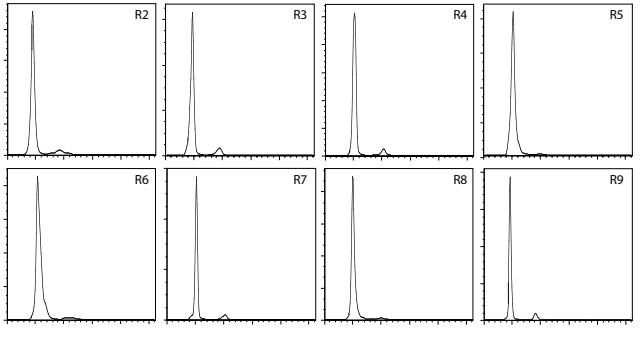


Supplementary Figure 10

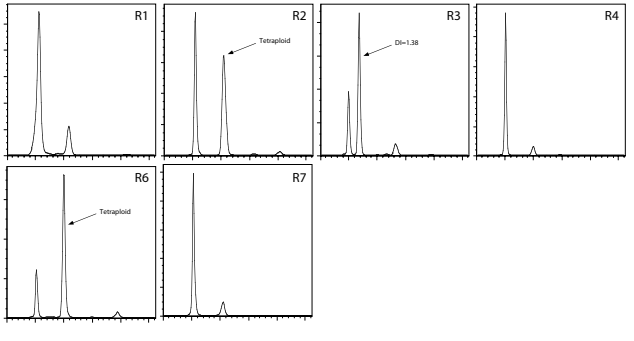
Patient 002



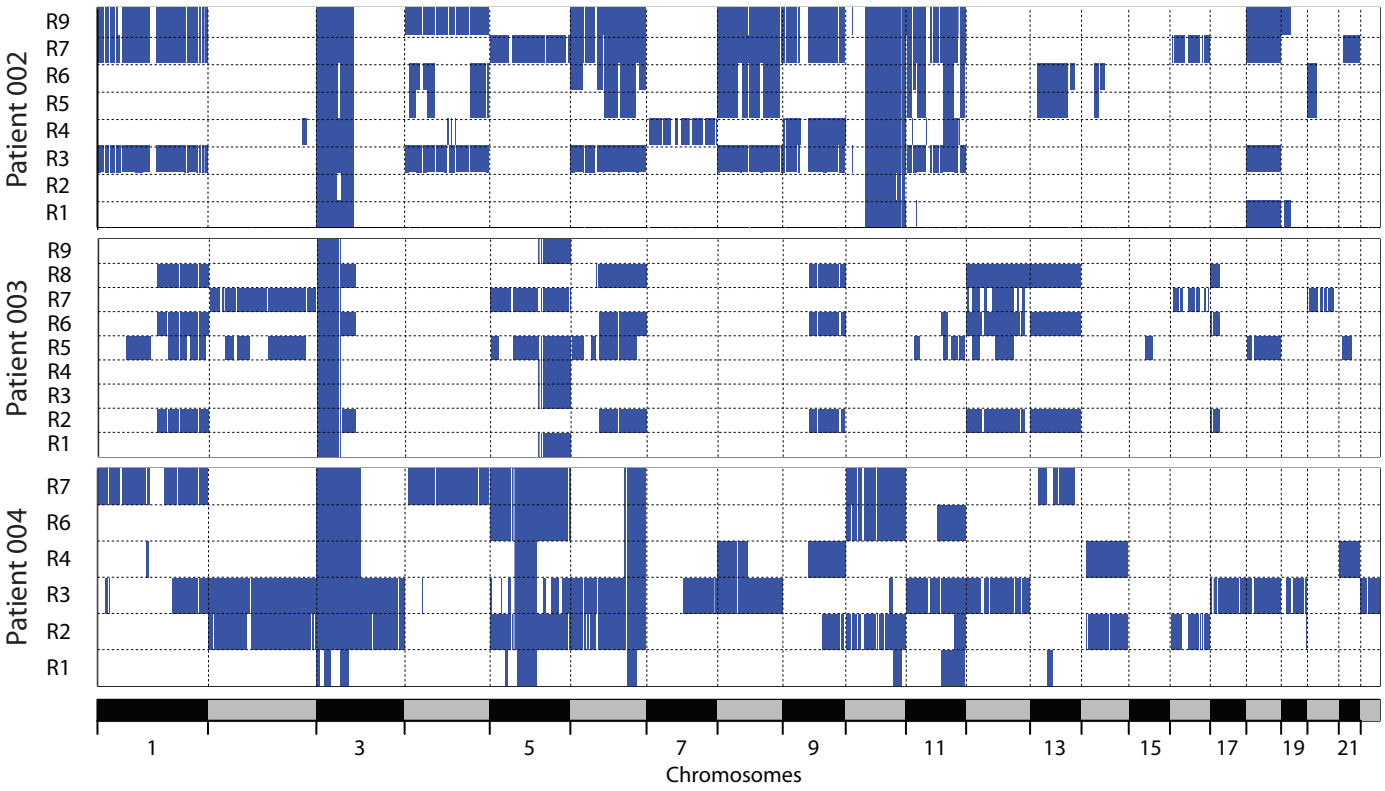
Patient 003



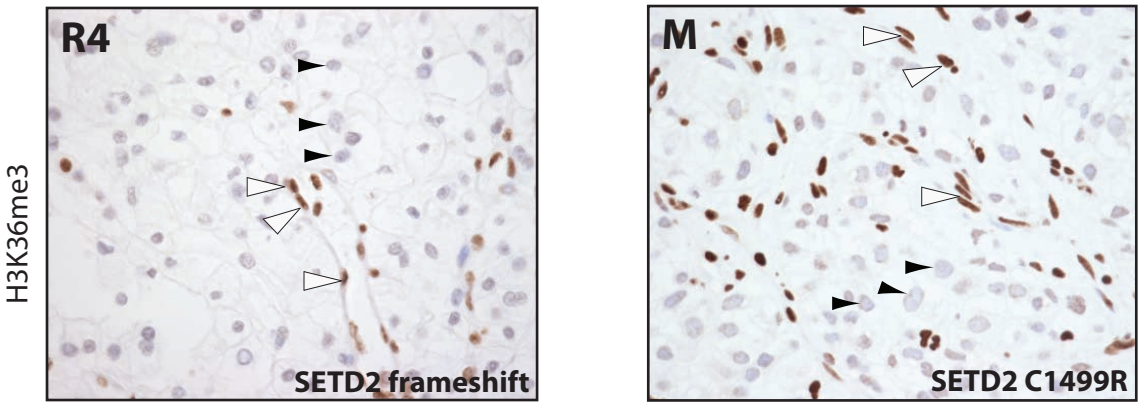
Patient 004



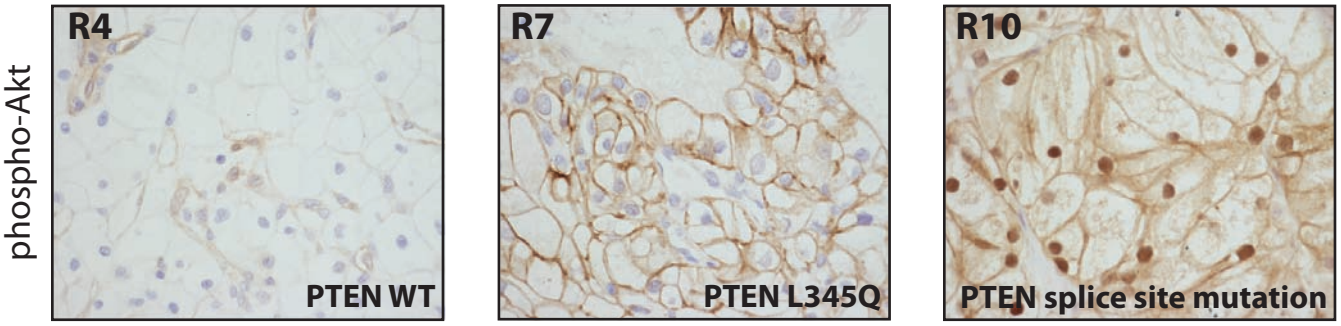
Supplementary Figure 11



Supplementary Figure 12



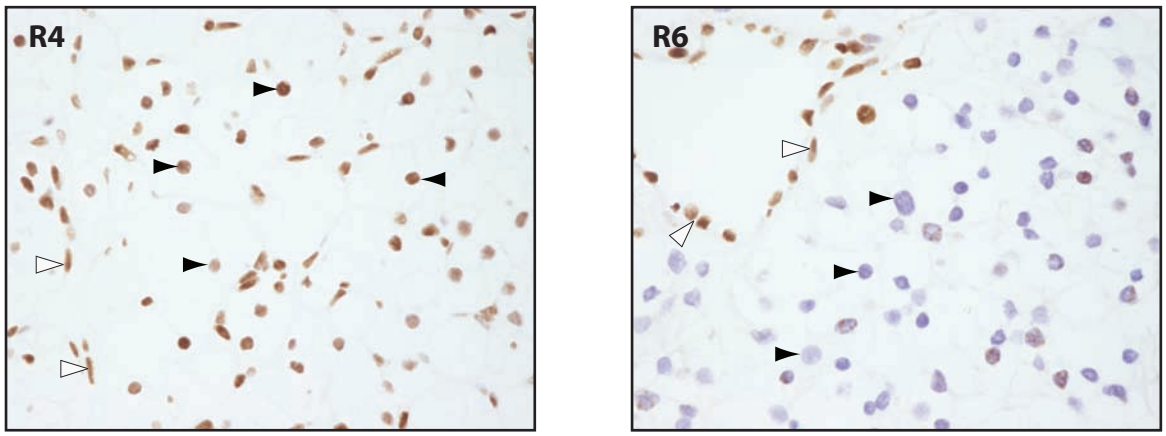
Supplementary Figure 13



Supplementary Figure 14

Patient 004 H3K36me3

A



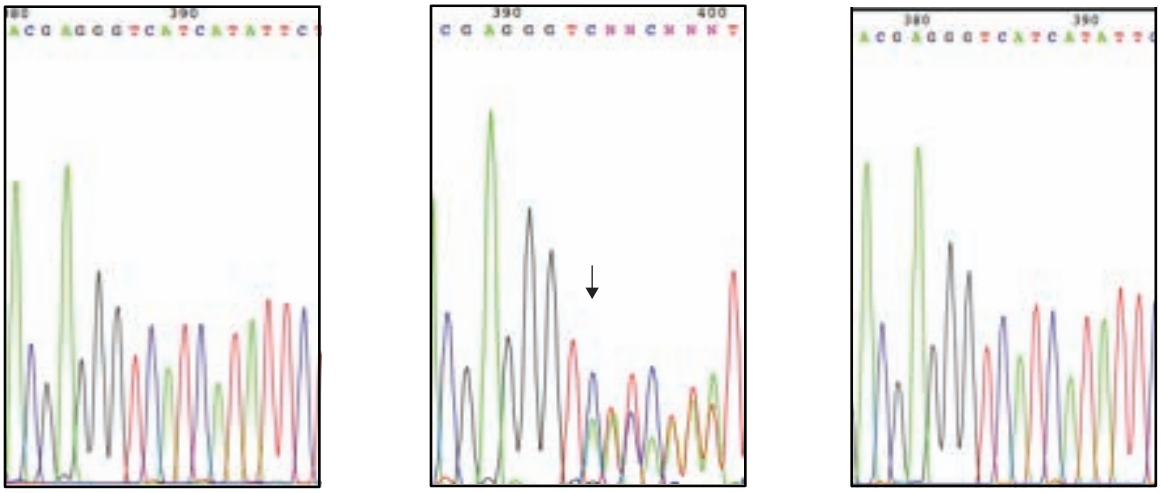
SETD2 WT

SETD2 Chr3:47164498 del C

B

SETD2 WT

SETD2 Chr3:47164498 Del:C



R4

R6

Blood