Supporting Information for

Landscape of Somatic Single Nucleotide and Copy Number Mutations in Uterine Serous Carcinoma

Siming Zhao[a], Murim Choi[a], John D. Overton[a], Stefania Bellone[b], Dana M. Roque[b], Emiliano Cocco[b], Federica Guzzo[b], Diana P. English[b], Joyce Varughese[b], Sara Gasparrini[b], Ileana Bortolomai[b], Natalia Buzac, Pei Hui[c], Maysa Abu-Khalaf[d], Antonella Ravaggi[e], Eliana Bignotti[e], Elisabetta Bandiera[e], Chiara Romani[e], Paola Todeschini[e], Renata Tassi[e], Laura Zanotti[e], Luisa Carrara[e], Sergio Pecorelli[e], Dan-Arin Silasi[b], Elena Ratner[b], Masoud Azodi[b], Peter E. Schwartz[b], Thomas J. Rutherford[b], Amy L. Stiegler[f], Shrikant Mane[a], Titus J. Boggon[f], Joseph Schlessinger[f], Richard P. Lifton[a,1] and Alessandro D. Santin[b]

[a]Department of Genetics, Howard Hughes Medical Institute, Yale University School of Medicine, New Haven, CT 06510; [b]Department of Obstetrics, Gynecology & Reproductive Sciences, [c]Department of Pathology, [d]Internal Medicine & Oncology, Yale University School of Medicine, New Haven, CT, 06510; [e] "Angelo Nocivelli" Institute of Molecular Medicine, Department of Obstetrics, & Gynecology, University of Brescia, Brescia 25123, Italy and [f]Department of Pharmacology, Yale University School of Medicine, New Haven, CT 06510.

[1]To whom correspondence should be addressed: richard.lifton@yale.edu

SI Materials and Methods

Patients and Specimens. The study protocol was approved by the Yale Human Investigation Committee. DNA was purified and libraries were prepared as described (1). DNA and RNA fractions were isolated from the tissue using an AllPrep DNA/RNA mini kit (Qiagen) per the manufacturer's procedure. Genomic DNA was prepared from venous blood, primary fibroblast cultures or frozen myometrium by standard procedures. Light microscopic evaluation was performed on a hematoxylin and eosin stained section of each frozen tumor specimen submitted to sequencing for assessment of percent tumor nuclei and percent necrosis in addition to other pathology annotations. We confirmed that each section contains a high purity of tumor epithelium (>60%) and minimal necrosis and used those sections for DNA extraction. Primary tumor cell lines purity was tested by morphology and by flow cytometry using pan-cytokeratin antibody staining (i.e., C-11-FITC, ab78478, Abcam Inc. Cambridge, MA). Only cell lines with tumor purity above 90% and less than 3 weeks of culture in vitro were used for sequencing.

Whole Exome Sequencing. Genomic DNA was captured on the NimbleGen 2.1M human exome array and subjected to 74 base paired-end reads on the Illumina HiSeq instrument as described (1). Sequence reads were mapped to the reference genome (hg18) using the ELAND program (1). Reads outside the targeted sequences were discarded and statistics on coverage were collected from the remaining reads using perl scripts. ELAND was also used for indel detection (1). For matched normal-tumor pairs, somatic mutations were called by comparing reference and non-reference reads from the matched pair by Fisher's exact test with tumor-specific thresholds determined from approximation of the null distribution (2). For unmatched tumors, SAMtools was used to call variant bases appended with quality scores. Among these, common variants that are listed in public databases were excluded and only rare variants were considered as potential somatic variants. Identified variants were annotated based on novelty, impact on the encoded protein, conservation, and expression using an automated pipeline.

qRT-PCR. RNA isolation from all primary USC cell lines was performed using TRIzol Reagent (Invitrogen, Carlsbad, CA, USA), according to the manufacturer's instructions. Quantitative PCR was carried out with a 7500 RealTime PCR System using the manufacturer's recommended protocol (Applied Biosystems, Foster City, CA, USA) to evaluate the expression of PIK3CA, PPP2R1A, FBXW7, HCFC1R1, CNTN4, LRRC7, MYO7B, MED12, TAF1, TTN, YIPF3, KRAS, CHD4, GRIN2A, ZEB2 and ATP1B4. The primers and probes were obtained from Applied Biosystems (i.e., FBXW7, Assay ID: Hs00217794_m1; CNTN4, Assay ID: Hs00330047_m1; HCFC1R1, Assay ID: Hs00372415_m1; TTN, Assay ID: Hs00399225_m1; TAF1, Assay ID: Hs00270322_m1; MED12, Assay ID: Hs00192801_m1; MYO7B, Assay ID: Hs00400099_m1; LRRC7, Assay ID: Hs00363532_m1; PIK3CA, Assay ID: Hs00180679_m1; PPP2R1A, Assay ID: Hs00204426_m1;YIPF3, Assay ID: Hs00209862_m1; KRAS, Assay ID: Hs00270666_m1; CHD4, Assay ID: Hs00172349_m1; GRIN2A, Assay ID: Hs00168219_m1; ZEB2, Assay ID: Hs00207691_m1; TP53, Assay ID: Hs01034249_m1ATP1B4, Assay ID: Hs00201320_m1). The comparative threshold cycle method was used to determine gene expression in each sample, relative to the value observed in non-malignant endometrial epithelial cell samples collected from similar age women, using glyceraldehyde-3-phosphate dehydrogenase (Assay ID Hs99999905_m1) RNA as an internal control.

1. Choi M et al. (2009). Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. *Proc Natl Acad Sci USA* 106:19096-19101.

2. Choi M et al. (2011) K+ channel mutations in adrenal aldosterone-producing adenomas and hereditary hypertension. *Science* 11;331:768-72.

3. Lee KR et al. (2003) Tumors of the Ovary and Peritoneum, In: World Health Organization Tumours of the Breast and Female Genital System, Lyon: IARC Press, 113-45.

Table S1 Clinical features of uterine serous carcinoma patients

PairID/ SampleID	Sample type	Status	Age	Ethinicity[1]	Stage[2]	Histology[3]	Chemo therapy[4]
ARK-15N	Cell line	Matched	67	E	IIIC	PURE	YES
ARK-17	Cell line	Matched	59	E	IIIA	PURE	NO
ARK-18	Cell line	Matched	62	A	IIB	PURE	NO
ARK-19	Cell line	Matched	65	E	IA	PURE	YES
ARK1	Cell line	Matched	62	A	IVA	PURE	NO
ARK11	Cell line	Matched	80	A	IIIC	MIXED	NO
ARK13N	Cell line	Matched	67	E	IVB	MIXED	YES
ARK2	Cell line	Matched	63	A	IVB	PURE	NO
ARK6	Cell line	Matched	48	E	IB	MIXED	NO
ARK7	Cell line	Matched	75	E	IIC	PURE	NO
ARK8	Cell line	Matched	88	E	IIIA	PURE	NO
ARK9	Cell line	Matched	73	A	IIIC	MIXED	NO
FF-1	Fresh frozen tissue	Matched	67	E	IIIC	PURE	NO
FF-10	Fresh frozen tissue	Matched	69	E	IVA	PURE	NO
FF-11	Fresh frozen tissue	Matched	58	A	IA	MIXED	NO
FF-12	Fresh frozen tissue	Matched	80	E	IIB	MIXED	NO
FF-13	Fresh frozen tissue	Matched	75	E	IIC	PURE	NO
FF-14	Fresh frozen tissue	Matched	59	E	REC*	PURE	YES
FF-16	Fresh frozen tissue	Matched	54	E	IIIC	MIXED	NO
FF-17	Fresh frozen tissue	Matched	73	A	IVB	PURE	NO
FF-19	Fresh frozen tissue	Matched	74	E	IVB	PURE	NO
FF-2	Fresh frozen tissue	Matched	36	A	IB	PURE	NO
FF-3	Fresh frozen tissue	Matched	78	E	IC	PURE	NO
FF-35	Fresh frozen tissue	Matched	63	A	IVB	PURE	NO
FF-4	Fresh frozen tissue	Matched	64	E	IB	PURE	NO
FF-40	Fresh frozen tissue	Matched	57	E	IB	PURE	NO
FF-41	Fresh frozen tissue	Matched	65	E	IA	PURE	NO
FF-42	Fresh frozen tissue	Matched	71	H	IA	PURE	YES
FF-43	Fresh frozen tissue	Matched	67	E	IIIC	PURE	NO
FF-5	Fresh frozen tissue	Matched	66	A	IB	MIXED	NO
FF-6	Fresh frozen tissue	Matched	63	A	IA	PURE	NO
FF-7	Fresh frozen tissue	Matched	49	A	IVA	PURE	NO
FF-8	Fresh frozen tissue	Matched	66	A	IVB	PURE	NO
FF-9	Fresh frozen tissue	Matched	56	E	IIIC	MIXED	NO
USC-FF-18	Fresh frozen tissue	Unmatched	72	E	IIIC	PURE	NO
USC-FF-20	Fresh frozen tissue	Unmatched	65	E	IIC	PURE	NO
USC-FF-21	Fresh frozen tissue	Unmatched	70	A	IIIA	PURE	NO
USC-FF-22	Fresh frozen tissue	Unmatched	73	A	IIB	PURE	NO
USC-FF-23	Fresh frozen tissue	Unmatched	60	E	IB	MIXED	NO
USC-FF-24	Fresh frozen tissue	Unmatched	60	E	IIIA	MIXED	NO

USC-FF-25	Fresh frozen tissue	Unmatched	76	A	IVB	PURE	NO
USC-FF-26	Fresh frozen tissue	Unmatched	76	A	IIIC	PURE	NO
USC-FF-27	Fresh frozen tissue	Unmatched	57	E	IIC	PURE	NO
USC-FF-28	Fresh frozen tissue	Unmatched	71	E	IIIA	PURE	NO
USC-FF-29	Fresh frozen tissue	Unmatched	70	A	IB	PURE	NO
USC-FF-30	Fresh frozen tissue	Unmatched	77	E	IIIA	MIXED	NO
USC-FF-31	Fresh frozen tissue	Unmatched	70	E	IIA	PURE	NO
USC-FF-32	Fresh frozen tissue	Unmatched	75	E	IIIC	PURE	YES
USC-FF-33	Fresh frozen tissue	Unmatched	66	A	IIIC	PURE	YES
USC-FF-34	Fresh frozen tissue	Unmatched	62	E	IA	PURE	NO
USC-FF-36	Fresh frozen tissue	Unmatched	55	A	IIB	PURE	NO
USC-FF-37	Fresh frozen tissue	Unmatched	74	H	IIIC	PURE	YES
USC-FF-38	Fresh frozen tissue	Unmatched	76	E	IIIC	PURE	NO
USC-FF-39	Fresh frozen tissue	Unmatched	81	A	IVB	PURE	NO
USPC-ARK-20	Cell line	Unmatched	42	E	IIB	PURE	NO
USPC-ARK10	Cell line	Unmatched	79	E	IVB	PURE	NO
USPC-ARK4	Cell line	Unmatched	82	E	IVB	PURE	NO

[1] Race information comes from PCA analysis. A, African; E, European; H, Hispanic.
[2] The staging used in the table is the 1988 FIGO (International Federation of Gynaecology and Obstetrics) staging system.
*REC denotes that recurrent tumor was sequenced.
[3] For uniformity with the current literature we have used the World Health Organization guidelines for epithelial tumors (i.e., USC that contain <10% of a second malignant component are considered 'pure' USC) (3).
[4] Chemotherapy yes means that the tumor biopsy we have sequenced was collected after the patient had received chemotherapy (i.e, carboplatin and paclitaxel).

Table S2 Exome run quality summary for all samples

Sample origin	Matched		Unmatched
Number	**34**	**34**	**23**
Status	**Normal**	**Tumor**	**Tumor**
Lanes used	1/3	2/3	2/3
Single end / Paired ends	PE	PE	PE
Read length	74 bp	74 bp	74 bp
# of reads per lane (M)	89	188	175
Median coverage (X)	84	165	144
Mean coverage (X)	100	199	170
% on genome	92.33%	92.21%	92.60%
% on target	73.43%	68.98%	62.76%
% of bases covered at least 4x	97.16%	97.93%	97.78%
% of bases covered at least 8x	95.63%	97.06%	96.89%
% of bases covered at least 20x	90.33%	94.66%	94.20%
Mean error rate	0.42%	0.49%	0.47%
% of PCR duplicate	4.95%	9.98%	7.63%

Table S3 Mismatch repair gene mutations and POLE mutations in hypermutated samples

Sample	Matched/ unmatched	Gene	Status	AA change	AA location/ protein length	P-value/QS*	LOH**
ARK6T	Matched	MLH1	Missense	E89D	89/756	1.23E-29	no
		MLH1	Missense	P705S	705/756	3.05E-18	no
		MLH3	Missense	A1246T	1246/1453	7.30E-16	no
		MSH6	Nonsense	E1322X	1322/1360	8.68E-11	no
		POLE	Missense	C1642Y	1642/2286	4.74E-06	no
		POLE	Missense	A1967V	1967/2286	2.98E-65	no
		POLE	Missense	G2076V	2076/2286	9.83E-12	no
		POLE	Missense	L2207I	2207/2286	1.11E-11	no
		POLE	Missense	D368Y	368/2286	3.19E-25	no
		POLE	Missense	H67N	67/2286	7.44E-25	no
		POLE	Missense	A832T	832/2286	1.28E-07	no
FF-4T	Matched	MLH3	Nonsense	E1288X	1288/1453	1.72E-09	no
		MSH2	Nonsense	E580X	580/934	4.34E-07	no
		MSH3	Missense	A396T	396/1137	3.88E-07	no
		PMS1	Missense	R93C	93/932	2.68E-10	no
		PMS2	Missense	L266I	266/862	1.24E-07	no
		POLE	Missense	M295R	295/2286	2.32E-08	no
FF-9T	Matched	MLH3	Missense	P1178H	1178/1453	2.66E-15	no
		MSH2	Nonsense	R680X	680/934	5.71E-23	no
		MSH2	Missense	E749A	749/934	1.65E-16	no
		MSH3	Missense	Y1011H	1011/1137	2.22E-24	no
		MSH6	Missense	A1055T	1055/1360	1.78E-06	no
		PMS1	Missense	L252R	252/932	6.13E-15	no
		POLE	Missense	T1052M	1052/2286	1.08E-05	no
		POLE	Missense	V1452A	1452/2286	9.36E-17	no
		POLE	Missense	V411L	411/2286	7.41E-11	no
		POLE	Missense	R742C	742/2286	6.04E-23	no
		POLE	Missense	R77C	77/2286	4.76E-22	no
FF-40T	Matched	MSH2	Missense	D91Y	91/934	3.45E-05	no
		POLE	Missense	F1099S	1099/2286	8.74E-05	no
		POLE	Missense	F1672L	1672/2286	3.55E-23	no
		POLE	Missense	V411L	411/2286	5.98E-15	no
		POLE	Missense	Y470H	470/2286	2.21E-32	no
		POLE	Missense	S928I	928/2286	6.23E-29	no
FF-24	Unmatched	MLH1	Missense	L585F	585/756	228	no
		MSH3	Missense	M953I	953/1137	228	no
		MSH6	Missense	M1326T	1326/1360	228	no
		POLE	Missense	P1164S	1164/2286	140	no

Underlined mutations highlights damaging mutations and recurrent mutations
*Fisher p-value, showing likelihood of the somatic calls being real. Most of cases, $p = 10^{-4}$ is the cutoff value. For calls from unmatched samples, quality scores from Samtools are given since Fisher p-values are not available for these cases
**LOH, loss of heterozygosity

Table S4 Significantly duplicated intervals

Chr	Start	End	Size	Frequency	Gene number : gene list*	Cancer gene**
chr3	173500000	195000000	21500000	15	**128**	ETV5, PIK3CA, LPP, SOX2, BCL6, EIF4A2
chr1	223500000	224500000	1000000	14	**15**:DNAH14, EPHX1, LBR, H3F3A, LIN9, H3F3AP4, ENAH, LEFTY2, SRP9, LEFTY1, ACBD3, TMEM63A, C1orf55, PYCR2, MIXL1	H3F3A
chr17	35000000	35500000	500000	11	**21**:NEUROD2, TCAP, PGAP3, PPP1R1B, ORMDL3, ERBB2, GSDMA, GSDMB, PSMD3, MIR4728, CSF3, MED24, PNMT, GRB7, THRA, STARD3, SNORD124, LRRC3C, MIEN1, IKZF3, ZPBP2	ERBB2
chr19	34500000	35500000	1000000	11	**7**:PLEKHF1, VSTM2B, CCNE1, POP4, LOC284395, C19orf12, URI1	CCNE1
chr8	55000000	142000000	87000000	11	**373**	NCOA2, CHCHD7, EXT1, TCEA1, NDRG1, MYC, PLAG1, COX6C, HEY1
chr20	29500000	30000000	500000	10	**16**:BCL2L1, TPX2, PDRG1, REM1, MYLK2, FOXS1, ID1, COX4I2, TTLL9, PSIMCT-1, DUSP15, DEFB124, DEFB123, HM13, MIR3193, LINC00028	
chr11	75500000	76000000	500000	9	**4**:UVRAG, PRKRIR, WNT11, C11orf30	
chr12	23500000	24000000	500000	9	**1**:SOX5	
chr14	21500000	22000000	500000	9	**0**	
chr13	111000000	113000000	2000000	8	**13**:MCF2L-AS1, PROZ, C13orf35, F7, PCID2, MCF2L, TUBGCP3, CUL4A, SPACA7, LAMP1, ATP11A, F10, SOX1	
chr18	12500000	23000000	10500000	8	**57**	SS18, ZNF521
chr16	29500000	30000000	500000	7	**24**:DOC2A, ASPHD1, LOC440356, PRRT2, CDIPT, QPRT, SLC7A5P1, PPP4C, SPN, MVP, FAM57B, ZG16, ALDOA, INO80E, SEZ6L2, TAOK2, KCTD13, MAZ, KIF22, C16orf92, C16orf53, TMEM219, C16orf54, HIRIP3	
chr6	11000000	11500000	500000	7	**6**:SYCP2L, ERVFRD-1, NEDD9, C6orf228, ELOVL2, LOC100506409	

* Gene names listed if gene number less than 50. Otherwise, gene names can be accessed using the coordinates provided in the table via portals such as the UCSC genome browser (http://genome.ucsc.edu/cgi-bin/hgGateway)
** Cancer gene: COSMIC cancer census gene

Table S5 Significantly deleted intervals

Chr	Start	End	Size	Frequency	Gene number : gene list *	Cancer gene**
chr22	49000000	49500000	500000	18	**24**:LMF2, CHKB-CPT1B, ODF3B, SBF1, MIOX, MAPK11, MAPK12, CHKB, PLXNB2, SYCE3, FAM116B, ADM2, CPT1B, PPP6R2, KLHDC7B, SHANK3, MAPK8IP2, NCAPH2, ARSA, TYMP, LOC100144603, HDAC10, TUBGCP6, SCO2	
chr19	1500000	2000000	500000	17	**17**:REXO1, SCAMP4, KLF16, TCF3, ONECUT3, UQCR11, MIR1909, ADAT3, ATP8B3, MKNK2, CSNK1G2-AS1, LOC100288123, MBD3, CSNK1G2, MEX3D, FAM108A1, BTBD2	TCF3
chr16	85500000	88000000	2500000	15	**35**:TRAPPC2L, LOC100287036, CDH15, ZCCHC14, ZFPM1, JPH3, KLHDC4, ANKRD11, MGC23284, PABPN1L, CDT1, CBFA2T3, CTU2, C16orf95, LOC400558, ZC3H18, MAP1LC3B, LINC00304, CA5A, SLC7A5, MIR4722, BANP, IL17C, PIEZO1, ACSF3, GALNS, APRT, CYBA, MVD, SNAI3, RNF166, SLC22A31, ZNF469, FBXO31, ZNF778	CBFA2T3
chr4	99000000	179000000	80000000	14	**310**	FBXW7, TET2, IL2, RAP1GDS1
chr17	28000000	32000000	4000000	13	**63**	TAF15
chr8	2000000	3500000	1500000	13	**2**:MYOM2, CSMD1	
chr15	73000000	85000000	12000000	12	**135**	
chr3	53000000	55500000	2500000	12	**15**:CACNA2D3, SELK, ESRG, ACTR8, CACNA1D, PRKCD, SFMBT1, TKT, LRTM1, DCP1A, CHDH, RFT1, WNT5A, MIR1303, IL17RB	
chr9	94500000	96000000	1500000	11	**19**:MIRLET7F1, MIRLET7D, FGD3, ANKRD19P, NINJ1, FAM120A, ZNF484, PTPDC1, LOC158257, WNK2, BARX1, C9orf89, C9orf129, MIR4291, MIRLET7A1, FAM120AOS, PHF2, BICD2, SUSD3	
chr11	0	500000	500000	10	**19**:LOC100133161, ODF3, RIC8A, NLRP6, ANO9, IFITM5, IFITM2, PTDSS2, IFITM1, PSMD13, SCGB1C1, B4GALNT4, PKP3, ATHL1, SIGIRR, IFITM3, BET1L, SIRT3, RNH1	
chr13	28500000	31000000	2500000	10	**15**:C13orf33, TEX26, USPL1, UBL3, LOC440131, LOC728437, LINC00426, KATNAL1, SLC7A1, TEX26-AS1, MTUS2, ALOX5AP, B3GALTL, HMGB1, HSPH1	
chr19	50000000	54500000	4500000	9	**186**	ERCC2

* Gene names listed if gene number less than 50. Otherwise, gene names can be accessed using the coordinates provided in the table via portals such as the UCSC genome browser (http://genome.ucsc.edu/cgi-bin/hgGateway)
** Cancer gene: COSMIC cancer census gene

Table S6 Comparison with high-grade ovarian serous carcinoma (HGS-OvCa)

	USC(n=52)		HGS-OvCa(n=316)	
	Affected sample number	**Percentage**	**Affected sample number**	**Percentage**
PIK3CA*	**12**	23.08%	**5**	1.58%
CHD4*	**10**	19.23%	**8**	2.53%
FBXW7*	**9**	17.31%	**3**	0.95%
PPP2R1A*	**8**	15.38%	**4**	1.27%
TAF1*	**7**	13.46%	**5**	1.58%
KRAS*	**3**	5.77%	**2**	0.63%
PTEN*	**3**	5.77%	**0**	0.00%
HCFC1R1*	**2**	3.85%	**0**	0.00%
CDKN1A*	**2**	3.85%	**1**	0.32%
CTDSPL*	**2**	3.85%	**0**	0.00%
YIPF3*	**2**	3.85%	**0**	0.00%
SPOP*	**2**	3.85%	**1**	0.32%
FAM132A*	**2**	3.85%	**0**	0.00%
TP53*#	**31**	59.62%	**303**	95.89%
BRAC1#	**0**	0.00%	**11**	3.48%
CSMD3#	**0**	0.00%	**19**	6.01%
NF1#	**1**	3.33%	**13**	4.11%
CDK12#	**0**	0.00%	**9**	2.85%
FAT3#	**2**	6.67%	**19**	6.01%
GABRA6#	**0**	0.00%	**7**	1.90%
BRCA2#	**0**	0.00%	**10**	3.16%
RB1#	**0**	0.00%	**6**	1.90%

. *Significantly mutated genes in USC

#Significantly mutated genes in HGS-OvCa

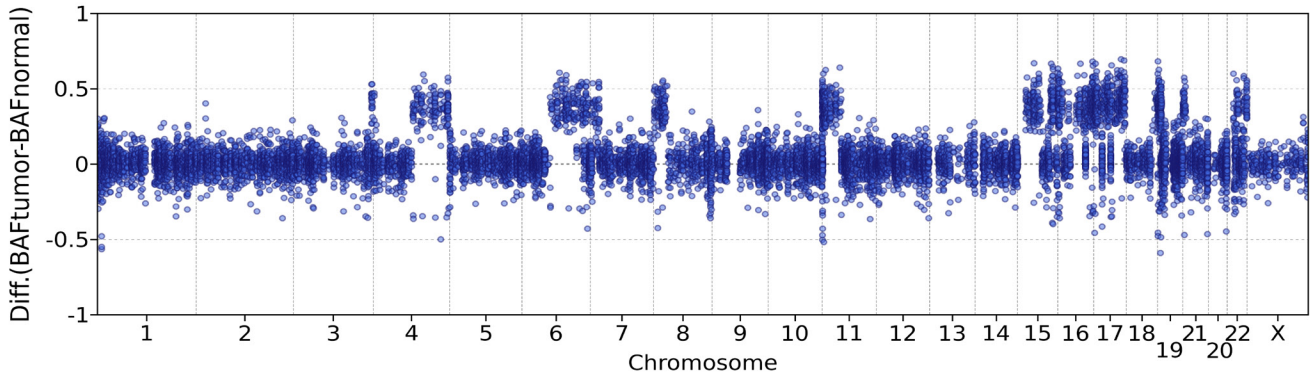

Fig. S1. Loss of heterozygosity calling from FF-12 tumor-normal pair.

Heterozygous SNP locus from normal sample of FF-12 was extracted and the difference of B allele frequency (BAF) change in tumor sample is plotted along the genome. Regions with obvious shift in BAF change were called as LOH by manual curation. For example, chr4, 6, 7, 8, 11, 15, 16, 17, 19, 20, 22 all have LOH regions. Purity was estimated by averaged absolute BAF change in LOH regions multiplied by 2.

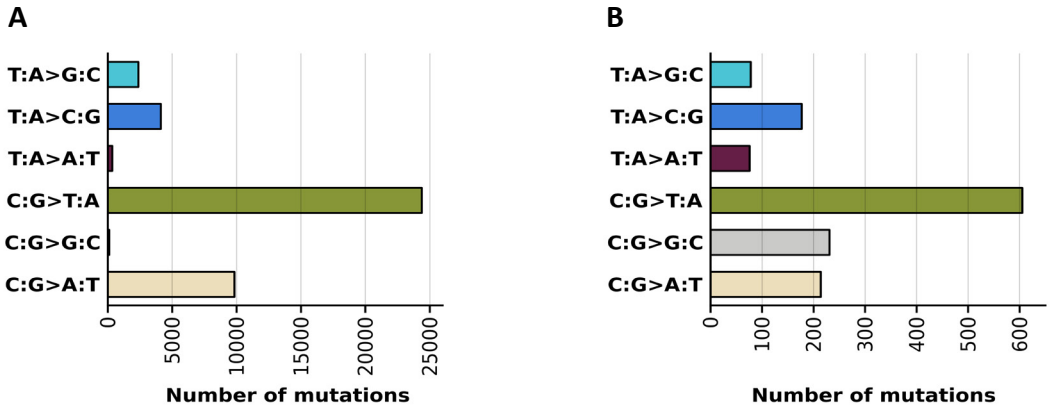

Fig. S2. Mutation spectrum in USC.

The numbers of base substitution in each of the 6 classes are shown. A. Plot for 4 tumors with hypermutator phenotype. B. Plot for 30 moderately mutated tumors.

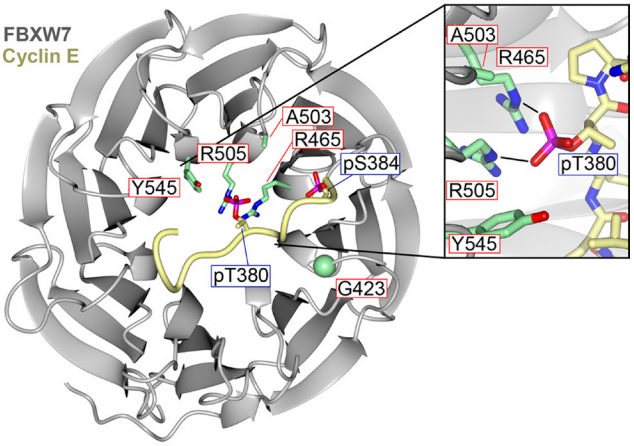

Fig. S3. Mapping of USC mutations onto the crystal structure of FBXW7 in complex with its high-affinity recognition motif in cyclin E.

FBXW7 is shown in grey and the doubly-phosphorylated cyclin E peptide in yellow, with phosphorylated residues labeled (PDB ID: 2OVQ). Residues of FBXW7 found mutated in USC are shown in green and labeled. Inset shows close up of the interaction with black lines indicating hydrogen-bonds.

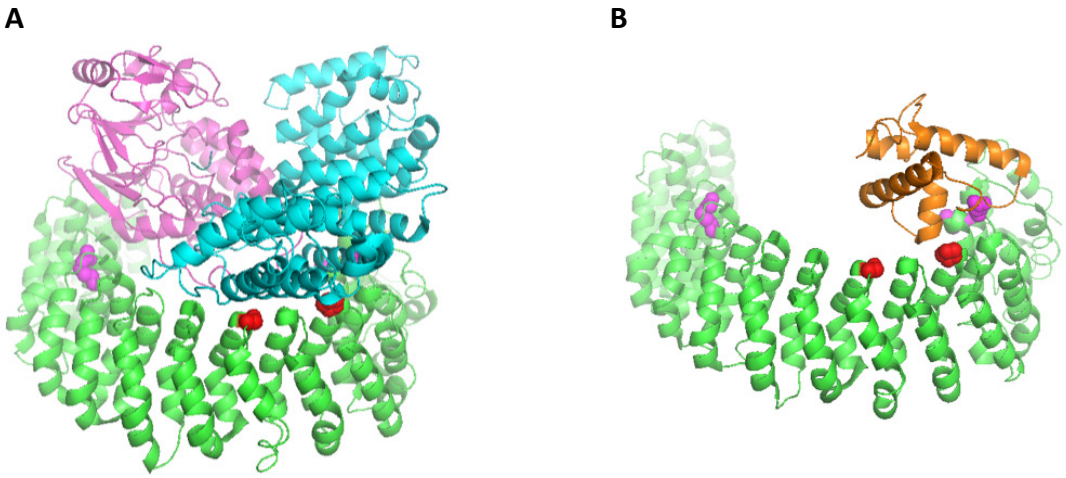

Fig. S4. Somatic mutations in PPP2R1A may affect its interaction with other proteins.
A. PP2A Holo-enzyme P179 and S256 make up part of the A-B interface. Green, subunit A; blue, subunit B; light purple, subunit C. Recurrently mutated amino acids P179 and S256 are labeled in red, singletons are labeled in purple. **B**. **SV40 and subunit A** SV40 virus binds regulatory subunit A with an overlapping site to the A-B interaction. Green, subunit A; orange, SV40 ; red, P179 (recurrently mutated) and S256 (recurrently mutated); purple, positions for singleton mutations.

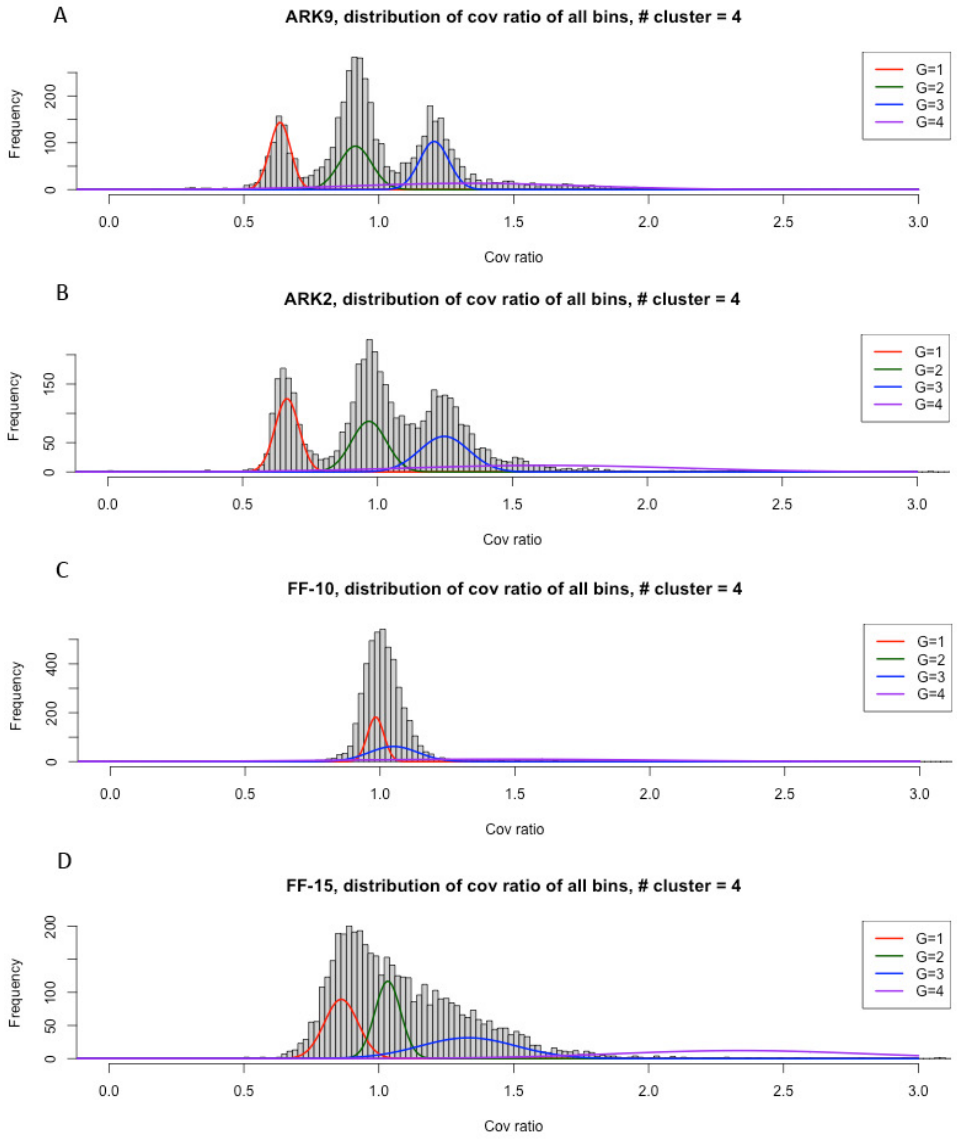

Fig. S5. Patterns of distribution of coverage depth ratio in different USC samples.
The distribution of coverage depth ratio (cov ratio) was plotted for each of the 34 matched normal tumor pairs. Unsupervised clustering was performed to identify potential clusters with coverage ratio deviated from 1 caused by copy number variation (red, green, blue and purple curves). 4 examples representing typical distribution patterns in these samples are shown. Only samples with cov ratio patterns similar to A, B and C were used for further CNV analysis (n=25) while samples like D were discarded in CNV analysis due to noise.

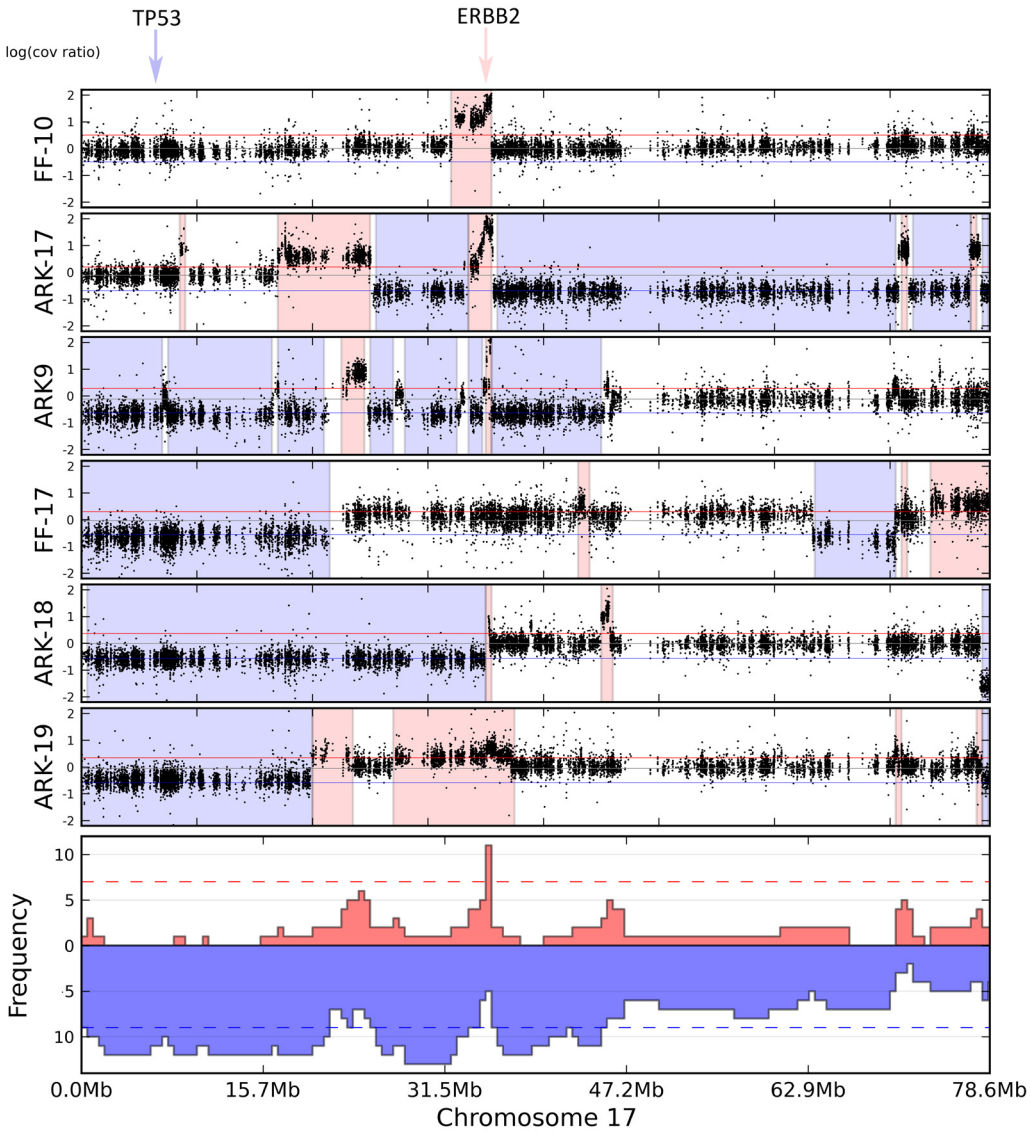

Fig. S6. Focal amplification on chromosome 17 (ERBB2) and chr17p deletion (TP53).
Log of coverage depth ratio is plotted across chromosome 17. 6 samples (FF-10, ARK-17,ARK9, FF-17, ARK-18 and ARK-19) are shown in parallel with the bottom subplot showing CNV frequency for all samples analyzed (n=25). The positions of ERBB2 and TP53 are marked. Horizontal red line, CNV duplication cluster center; horizontal blue line, CNV deletion cluster center; horizontal grey line, copy neutral cluster center; horizontal red and blue dotted line in bottom frequency plot, CNV gain and loss genome wide significance level.

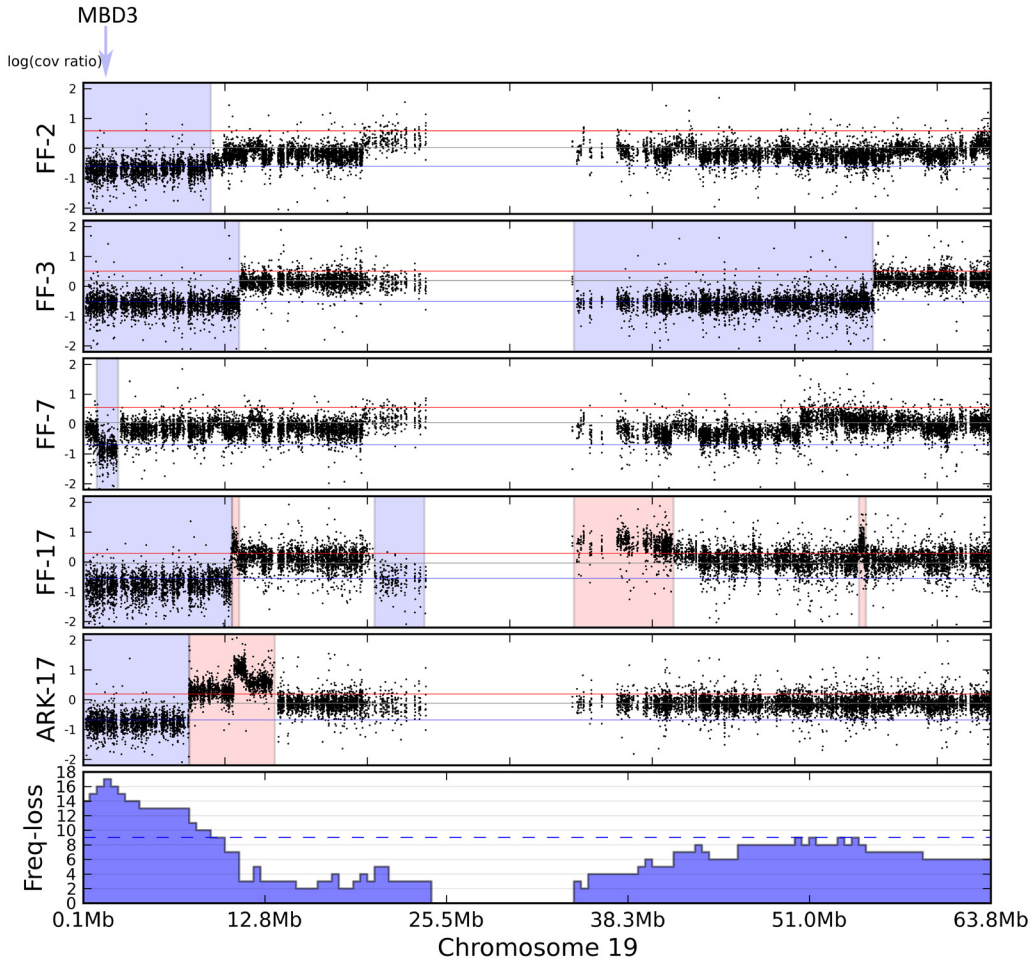

Fig. S7. Focal deletion on chromosome 19p (MBD3).
See Fig. S6 for legend.

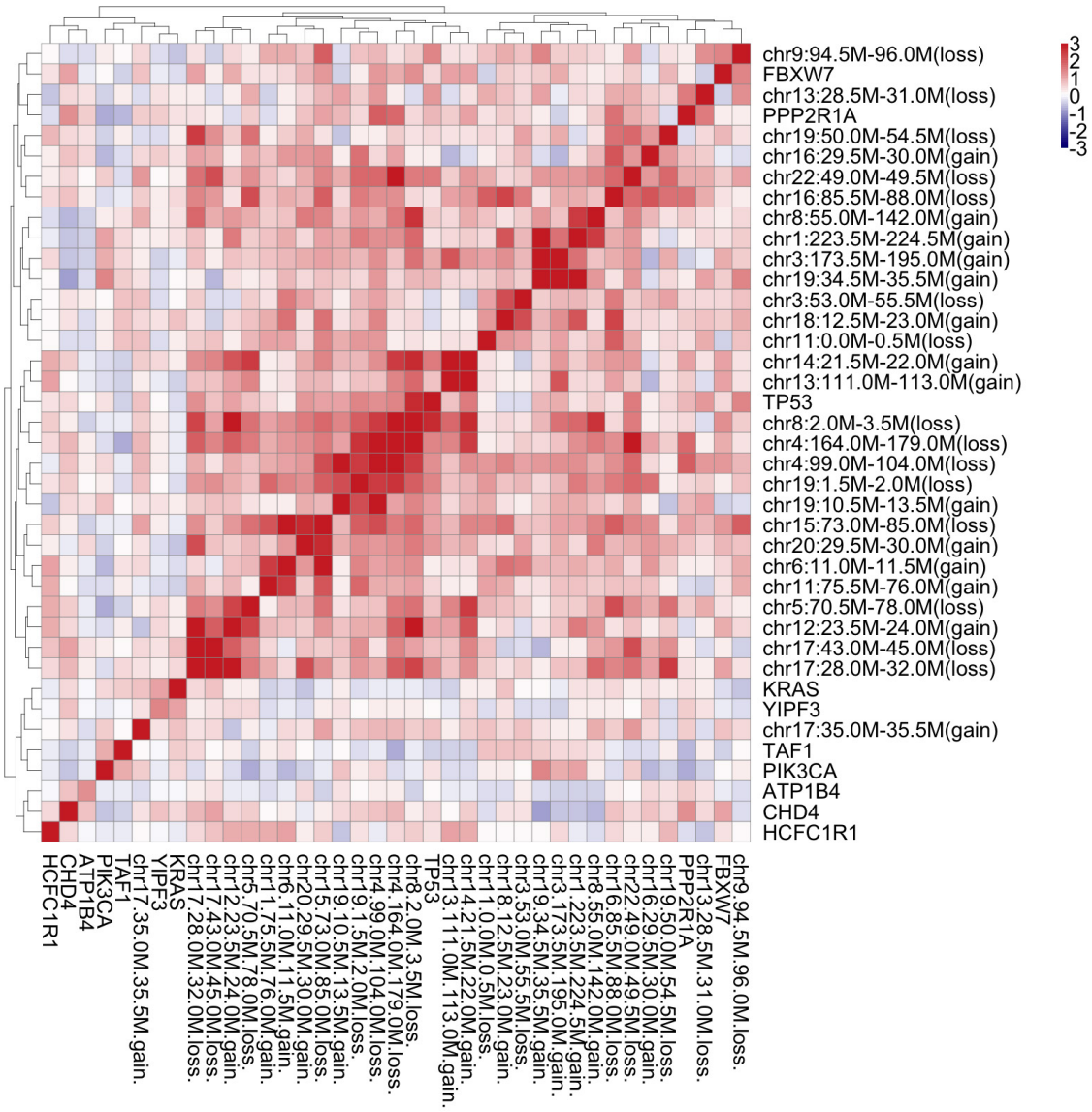

Fig. S8. Pairwise correlation/exclusion study for significantly mutated genes and significant CNVs.

Heat map depicting patterns of correlation and exclusion in USC is shown. Significantly mutated genes from 30 matched samples and significant CNV events from 25 matched samples were included in this study. Pairwise correlation or exclusion p value was calculated based on permutation. Red, correlation; blue, exclusion.

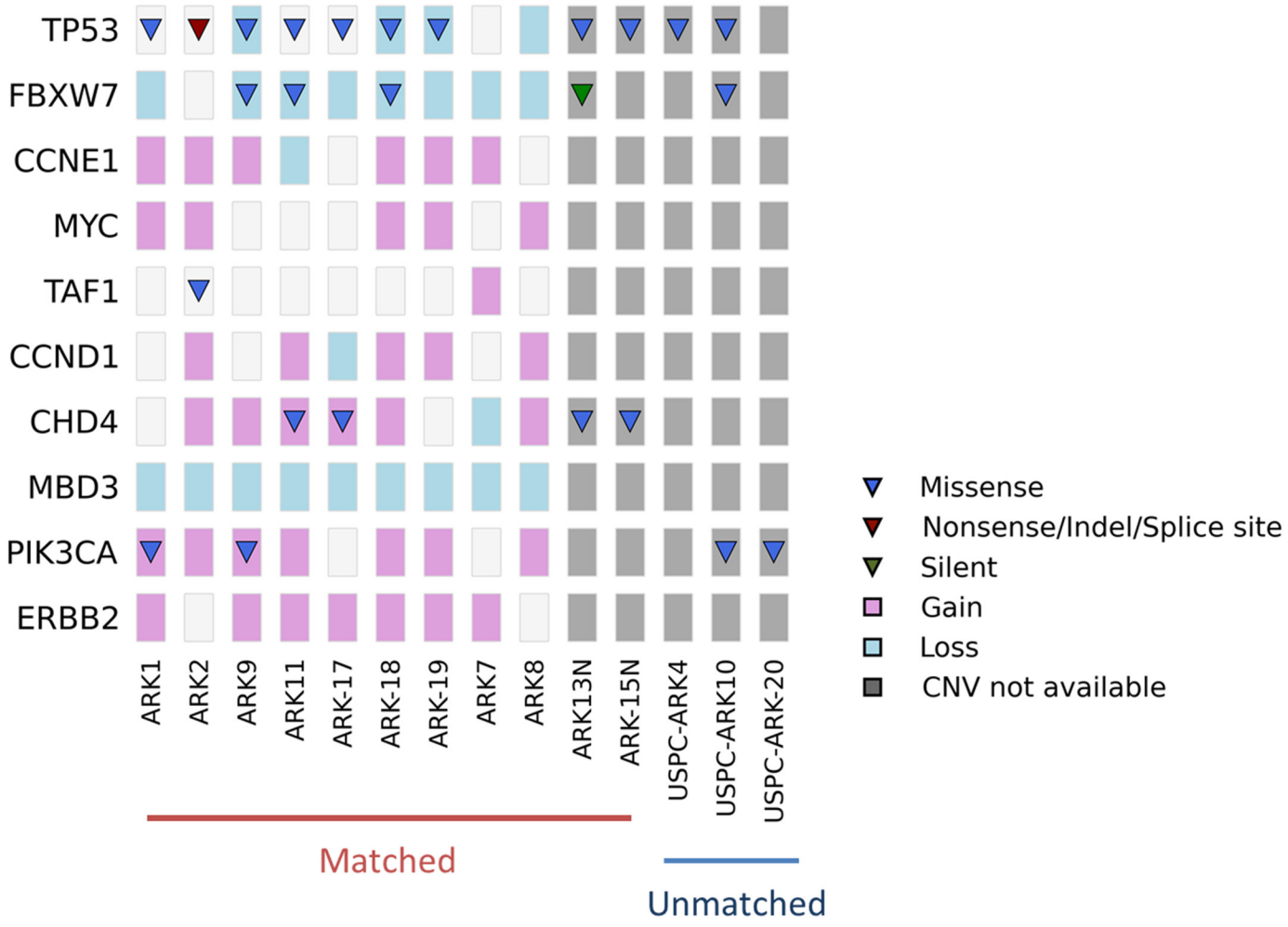

Fig. S9. Mutation profile in 14 USC cell lines.
The distribution of SNVs and CNVs in 14 USC cell lines is shown. CNV calls were not made in cell lines without matched normal DNA for comparison. One additional cell line (not shown) had the hypermutator phenotype.