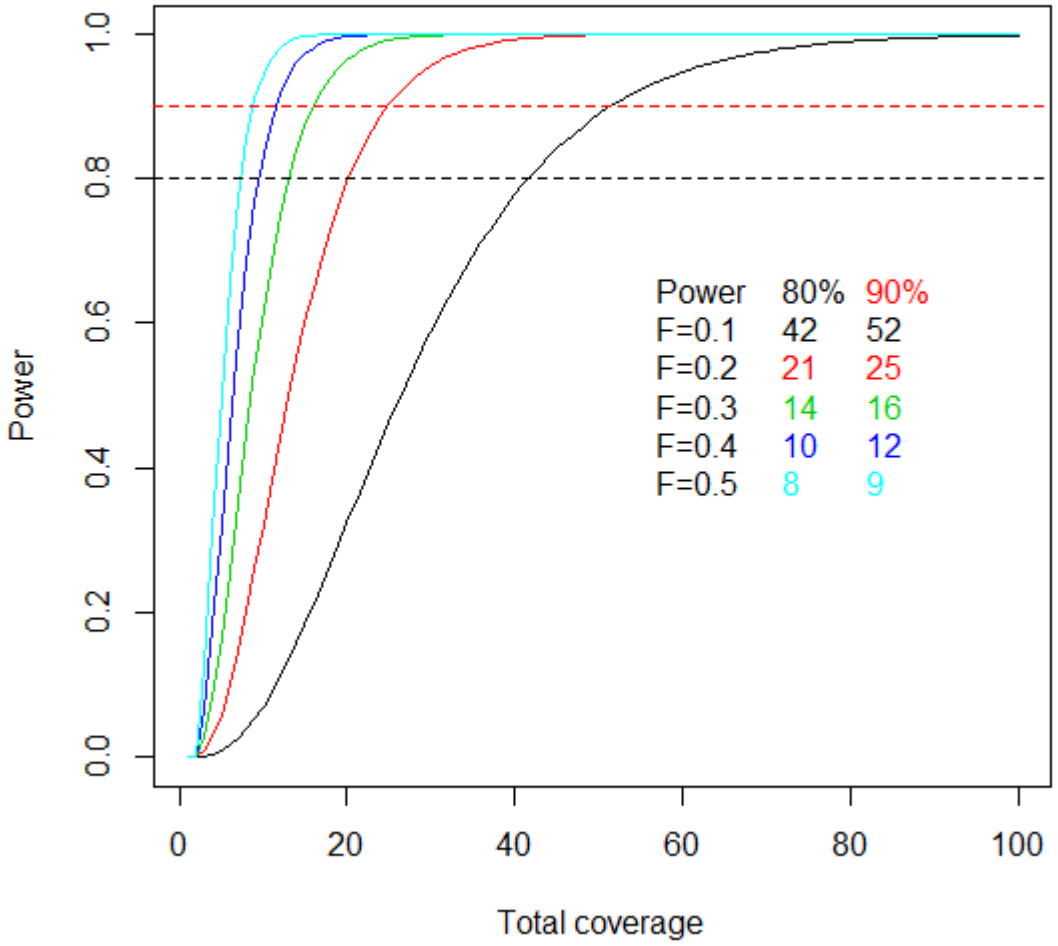
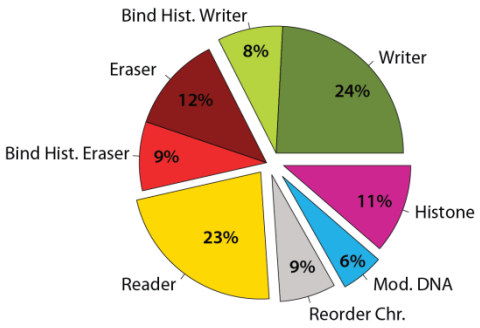
Supplementary Information: The landscape of somatic alterations in epigenetic genes across 1000 pediatric cancer genomes

SUPPLEMENTARY INFORMATION

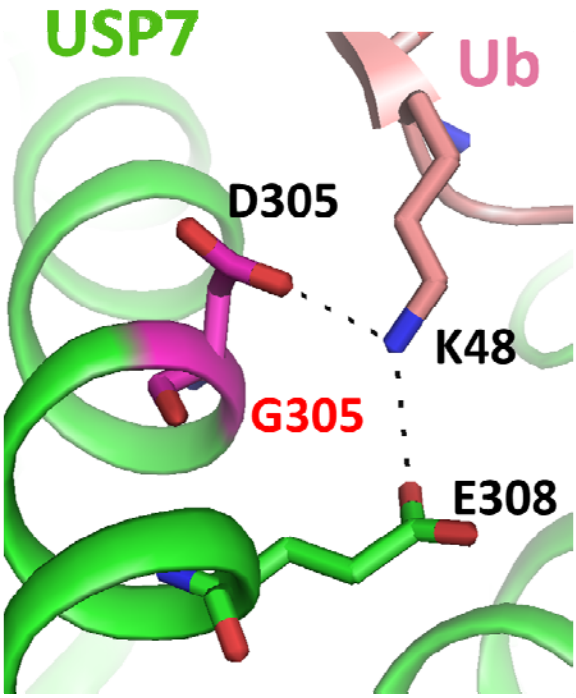
TABLE OF CONTENTS



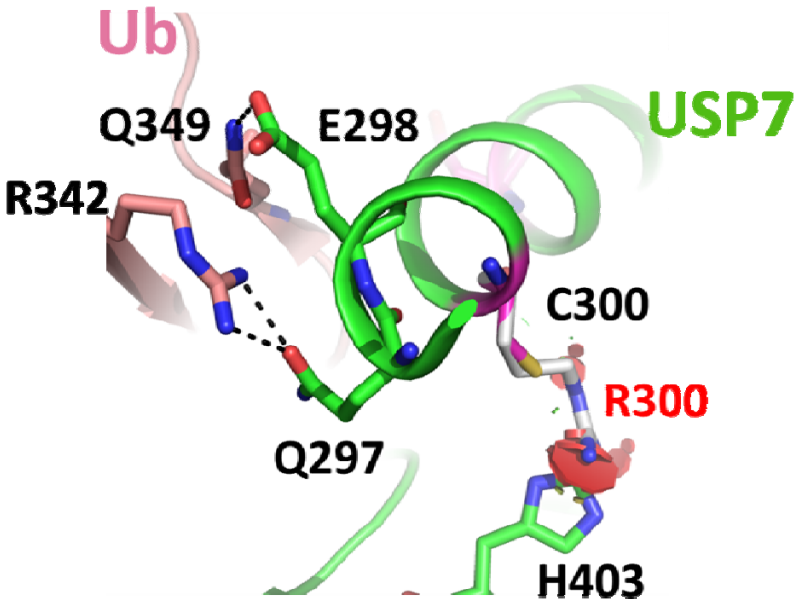
Supplementary Figure 1: Sequence coverage required for detecting mutations at various mutant allele frequency X-axis represents sequence coverage; y-axis represents the power for mutation detection based on binomial distribution of sampling 3 NGS reads that harbor a mutant at MAF 0.1 (black), 0.2 (red), 0.3 (green), 0.4 (blue) and 0.5 (cyan). The 2-by-2 table shows the coverage required for achieving 80% and 90% power for detecting mutations at different mutant allele fraction (F).



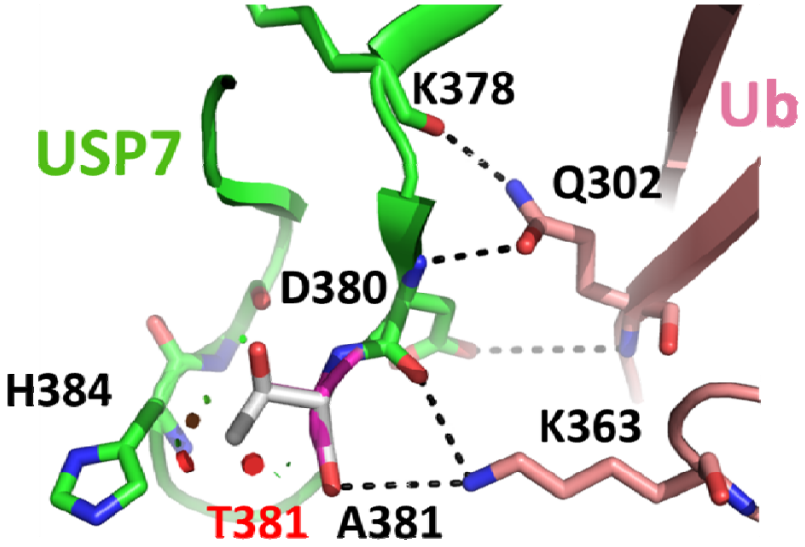
Supplementary Figure 2: Distribution of somatic SNV/Indels identified within each class from the dominant clone The classes of epigenetic proteins are colored according to Figure 1a. Where Eraser (dark red) and Bind Hist. Eraser (red) is one slice, Writer (dark green) and Bind Hist. Writer (green) are a second slice and Reader (yellow), Reorder Chr. (gray), Mod. DNA (blue) and Histone (pink) are another.



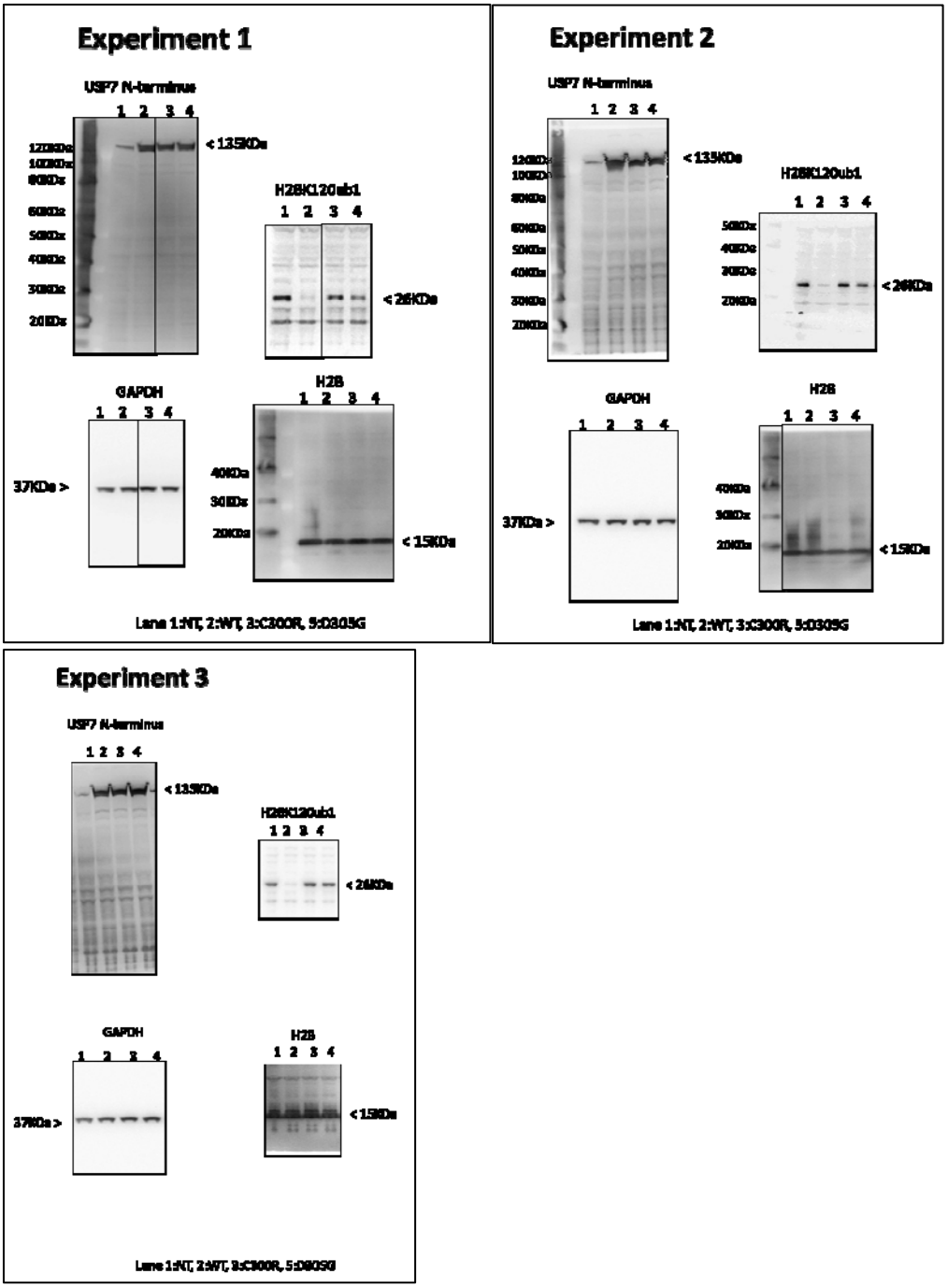
Supplementary Figure 3: Modeled D305G missense mutation in USP7 Modeled missense mutations on USP7 (rendered in green). The side chain of Aspartic acid 305 (D305: illustrated as a purple stick) is involved in a hydrogen bond/electrostatic interaction with that of Lysine 48 (K48) of ubiquitin (labeled Ub: peach cartoon). The Lys side chain also interacts in a bifurcated manner with that of Glutamic acid 308 (E308). Mutation of D305 to Glycine would eliminate this interaction and disrupt interactions of USP7 with ubiquitin. Additional residues are represented as sticks and discussed in the text.



Supplementary Figure 4: Modeled C300R missense mutation in USP7 Modeled missense mutations on USP7 (rendered in green). The side chain of Cysteine 300 (C300; illustrated as a purple stick) is buried and participates in hydrophobic interactions to stabilize an ubiquitin binding helix on one side of the ubiquitin binding pocket. Mutation of C300 to a large polar Arginine (R300: illustrated as a white stick) would physically perturb the local environment through van der Waal clashes. This would alter the conformation of residues Glutamic acid 298 (E298) and Glutamine 297 (Q297) that selectively interact with ubiquitin (labeled Ub: peach cartoon) and would weaken USP7/ubiquitin interactions. Additional residues are represented as sticks and discussed in the text. Clashes between the mutation and surrounding residues are represented by colored disks (red indicates severe clash).



Supplementary Figure 5: Modeled A381R missense mutation in USP7 Modeled missense mutations on USP7 (rendered in green). The mutation of Alanine 381 (A381: illustrated as a purple stick) to Threonine (T381: illustrated as a white stick) is identified in a single sample of hypodiploid leukemia. A381 is located within an ordered loop and its backbone carbonyl hydrogen bonds to the side chain amine group of Lysine 363 (K363) of a bound ubiquitin (labeled Ub: peach cartoon). The methyl side chain of A381 is directed away from the bound ubiquitin and towards residue Histidine 384 (H384) within the loop. The A381T mutation is a minor amino acid substitution and does not appear to drastically clash with any surrounding residues. As the backbone carbonyl is the important interaction with ubiquitin, a mutation of A381T may alter the structure of the loop and could at most slightly alter ubiquitin binding. Additional residues are represented as sticks and discussed in the text. Clashes between the mutation and surrounding residues are represented by colored disks (red indicates severe clash).

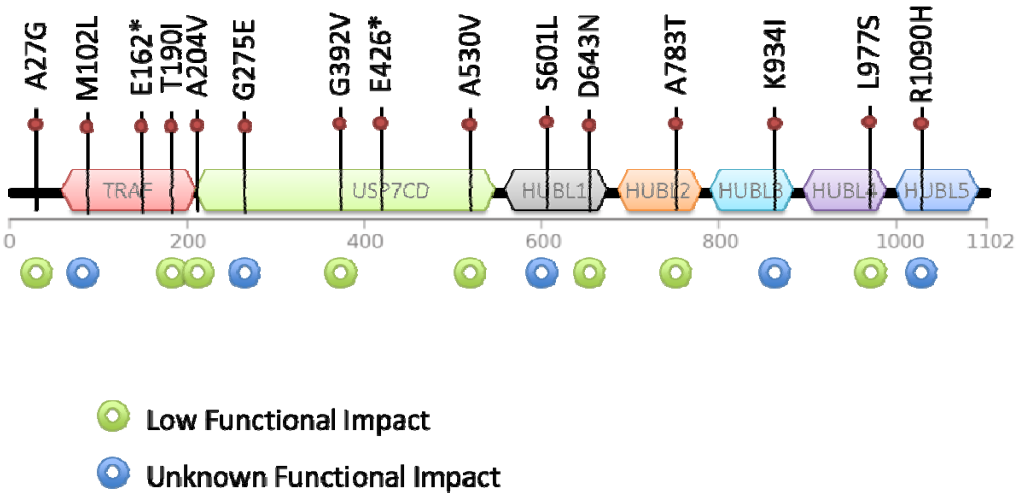


Supplementary Figure 6: Western blot gels for 3 replicates of wild-type and mutant USP7 assay 293T cells were transfected with wild-type USP7 and mutants (C300R and D305G). Total protein of 293T cells was extracted at 72 hours post transfection. Protein levels of USP7, total H2B and H2B ubiquityl Lys120 were detected by western blot with indicated antibodies. GAPDH was adopted as loading control on the same gel. Briefly, the blots were incubated in the 1:1000 diluted primary antibodies overnight at 4 degree C and followed by incubating in 1:5000

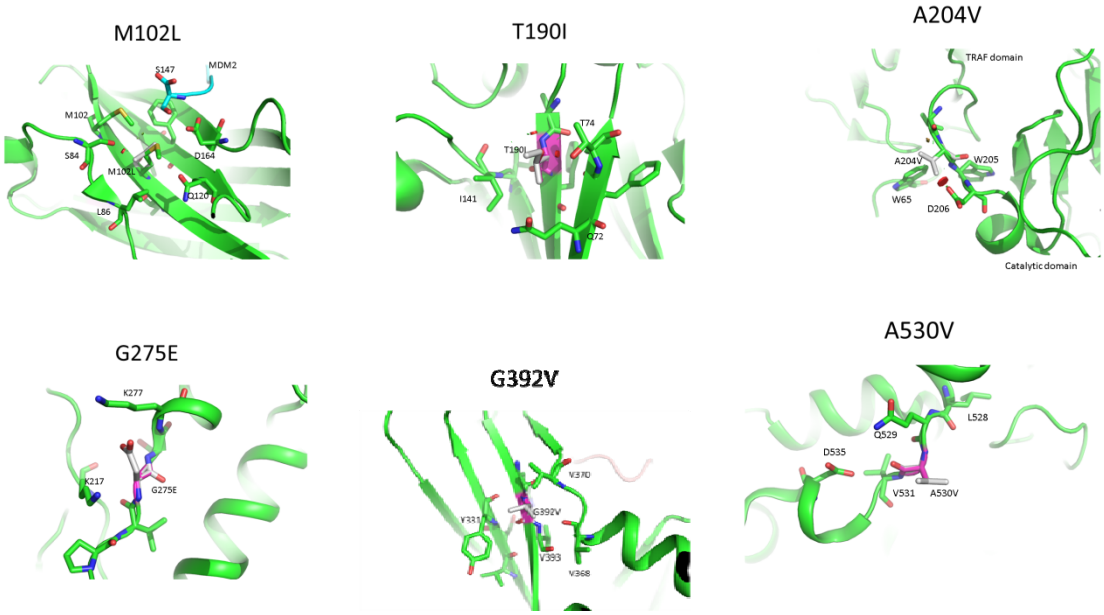
diluted secondary antibodies. The protein bands were detected by SuperSignal West Femto Maximum Sensitivity. The size of protein was determined by MagicMark XP western Protein Standard (Life Technologies, Grand Island, NY). Consistent results were observed for three replicates (Experiment 1, 2, 3).

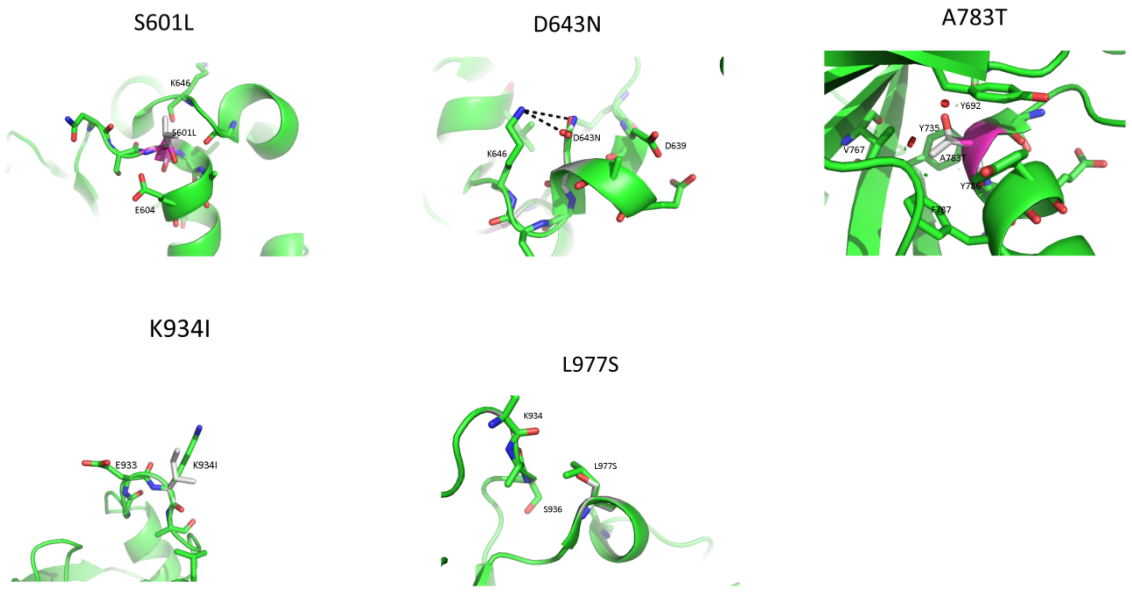
Supplementary Information: The landscape of somatic alterations in epigenetic genes across 1000 pediatric cancer genomes

a

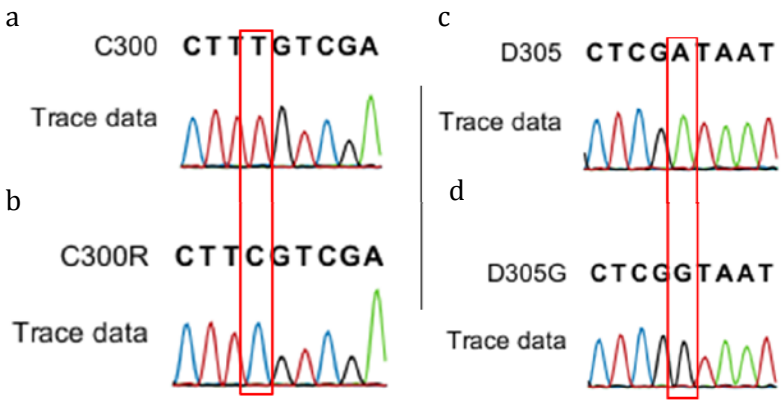


b





Supplementary Figure 7: Analysis of somatic non-synonymous mutations in USP7 found in adult primary tumors Confirmed somatic sequence mutations in the COSMIC database (a) Distribution of 15 confirmed non-synonymous mutations in USP7 in six adult cancers from a total of 885 sequenced primary tumors. Predicted functional impact was assessed based on mutation location and USP7 function. (b) Using the available crystal structural information, 11 out of 13 missense mutations could be modeled. Eight of the mutations (A27G, T190I, A204V, G392V, A530V, D643N, A783T, and L977S) appear benign and probable passenger events based on their location in the structure, the surrounding environment and extent of the amino acid change. The additional five mutations (M102L, G275E, S601L, K934I and R1090H) are located within regions of the structure that might affect binding of protein partners or catalytic activity but the functional impact is not obvious. We conclude that these previously described lesions in USP7 play passenger roles. Further details are provided in Supplementary table 9. USP7 protein (rendered in green) contains the wild-type residue (green/magenta - stick) and modeled mutant residue (white - stick). Surrounding residues (green - sticks), bound peptides (teal - cartoon), and interactions (black dots) are highlighted.



Supplementary Figure 8: Sequencing traces of wild-type and mutant *USP7* plasmids Traces of wild-type plasmid (a and c) and mutant USP7 plasmids (b and d) used in transfection experiments are centered at codon C300 (TGT) and D305 (GAT). The rectangles highlight the wild-type and mutant alleles in these two codons.

Supplementary Table 1: The 15 genes with the highest mutation frequency in PCGP cohorts

Gene	Complex	#Total mutations	Disease	AA Change[+]	#Mutated tumors	*p-value
ASXL1	PRC2	5	HGG	R693* (n=2)	2	2.65E-02
			AML	R693*	1	
			EPD	R693*	1	
			NBL	D943N	1	
ASXL2	PRC2	7	CBF	T740fs, S524fs, R443_K444fs, T740fs, A1276_V1277fs, N590fs	6	5.13E-10
			TALL	Q1017*	1	
ATRX		16	HGG	L2112_F2113fs, D1791H, T1610R, D2144G, R188*	5	6.42E-05
			NBL	L407F, A1690D, E555*, R2188Q, K425_E426fs	5	1.83E-04
			ACT	R2164S, Q811*, E2253*	3	5.25E-04
			OS	Y266*, D1383fs, R1803C	3	1.20E-03
CREBBP	CREB	8	HYPO	K389_M395>K, R1169C, P1241_E20splice, P1241_E20exon	4	7.65E-06
			MB	R483*, Q541*	2	2.56E-02
			NBL	E1421*	1	
EZH2	PRC2	8	TALL	G5R, F723fs, V442D	3	4.45E-04
			CBF	T126_V127fs, T53fs (SAME SAMPLE)	1	
			HYPO	N675K	1	
			INF	Y646C	1	
			EWS	A682G	1	
H3F3A		35	HGG	K27M (N=30), G35R (N=2)	32	4.12E-89
			TALL	K28R, K37R	2	3.38E-04
			LGG	K27M	1	
KDM6A	COMPASS-C	15	MB	G298fs, G1242V, G359fs, S54_E2splice, V1336_E27splice_region, L989*	6	1.06E-08
			TALL	V1113fs (N=2), S537_G540fs, Q542K, R1111fs, W287*, S1114*	5	9.53E-06
			CBF	R1213*, E19_splice	2	3.32E-03
			ETV	G137V	1	
			MLL	G1242D	1	
MLL2	MLL, ASCOM	7	TALL	R5154Q	1	
			CBF	E1581_E19exon	1	
			HGG	R2687*, R2830* (SAME SAMPLE)	1	
			HYPO	V4642I	1	
			MB	V4194fs	1	
			OS	P2197L	1	
NSD2		5	E2A	E1099K (N=2)	2	6.30E-05
			ETV	E1099K, D1125N	2	1.99E-03
			INF	E1099K	1	
PHF6	NuRD	22	TALL	L244_E8splice, R274Q, M125I, M81_E4splice, R319*, C215fs, S246fs, F263fs, F263V, R116*, R129*, C297Y, C242fs, S246Y, Y103*, G213fs, E139*, R129*, H239fs, K323_E9splice	19	9.73E-40
			AML	C242Y, V307fs	2	NO_BMR
			E2A	R129*	1	
SETD2		6	ERG	(T152fs, C1754R [SAME SAMPLE]), S1572_E6splice	2	1.56E-03
			TALL	V1190M	1	
			AML	S1572fs	1	
			CBF	I1883fs	1	
SMARCA4	BAF/PBAF	7	MB	T910M (N=2), G1232C, P946L, H884P	5	9.82E-07
			NBL	G271W	1	
			TALL	E821K	1	
			HGG	R513Q	1	
SMC1A		5	CBF	*insR131, R92G, G1131R, R196C	4	1.16E-06
			EPD	T182_F184del	1	
USP7	DUU	6	TALL	D380fs, T177fs, D305G, R340fs, V203fs	5	3.01E-06
			HYPO	A381T	1	
ZMYM3	NuRD	5	MB	R647*, Q1199E, L1111fs	3	4.13E-04
			HGG	C968fs	1	
			RHB	C743*	1	

#Only mutations with mutant allele frequency >=0.3 are included, *raw p-values were calculated only for diseases with recurrently mutated genes (see methods for procedure in calculating p-value), AA Change+: blue text indicates loss-of-function (based on a majority rule of the function predictors), red text indicates known gain-of-function and black text is unknown.

Supplementary Table 2: Predicted functional impact of USP7 mutations in COSMIC

SNV	Domain	PDB	Tissue	Structural Environment	Predicted Functional Impact based on Structural Model.	MA	P2	S
E162*	-	-	L. Intest.	-	-			
E426*	-	-	Neuro	-	-			
A27G	-	-	Ovary	N-terminal.	The location and amino acid change indicate low functional impact.	n	b	t
M102L	TRAF	2FOP	Ovary	Buried hydrophobic at protein-protein interface.	Residue has been shown to be perturbed during peptide binding. Given the change is not drastic (hydrophobic to hydrophobic) unfavorable and limited functional data, functional impact would be unknown. (PMID:15808506)	m	b	t
T190I	TRAF	2FOP	L. Intest.	Solvent exposed, opposite face to protein-protein binding surface.	Loss of some stability due to change in hydrophobicity but location and surrounding residues indicates a low functional impact.	n	pod	t
A204V	loop	2FIZ	L. Intest.	Solvent exposed loop	Exposed hydrophobic residue small impact on stability but low functional impact.	l	b	t
G275E	CD	1NBF	CNS	Solvent exposed, adjacent to catalytic center	May impact switching loop and catalytic residue orientation. Potential for Impact but requires movement.	l	prd	t
G392V	CD	1NBF	Breast	Partial solvent exposed, Adjacent to Ub binding site	Gly side chain hydrogen is directed away from Ub and pointed towards pocket of hydrophobic amino acids. Val would likely be directed in the pocket as well. Low functional impact	m	prd	d
A530V	CD	1NBF	L. Intest.	Solvent exposed, base of catalytic domain	Loss of some stability in the region from exposed hydrophobic. Far from binding and active site, low functional impact.	n	b	t
S601L	HUBL1	2YLM	Breast	Solvent exposed, caps amino end of helix	Exposed hydrophobic residue is potentially destabilizing. Functional impact unknown.	n	b	t
D643N	HUBL1	2YLM	L. Intest.	Solvent exposed, side chain interaction with K646	Similar sized amino acid and slight loss of negative charge but appears to be low functional impact.	m	b	t
A783T	HUBL2	2YLM	L. Intest.	Buried in hydrophobic pocket.	Small amino acid change but little side chain clash, indicate a potential fit and low functional impact.	l	prd	t
K934I	HUBL3	2YLM	Ovary	Solvent exposed loop	Potential impact from increase in hydrophobicity and would likely cause structural movements to the loop. Functional impact unknown.	n	b	t
L977S	HUBL4	2YLM	Skin	Partial solvent exposed	Location and type of interactions indicates a low functional impact.	n	b	t
R1090H	HUBL5	2YLM	L. Intest.	Adjacent to activation peptide residues S1089. Y1091, Y1093, I1100 (PMID: 21981925).	Low impact, both residues are available for donating a hydrogen bond. Location of mutation however, would increase potential for functional impact (PMID: 21981925).	l	prd	t

Predicted results from Mutation assessor (MA), polyphen2 (P2) and Sift (S) have been abbreviated. Where MA: low=L, neutral=n, medium=m; P2: benign=b, probably damaging=prd, possibly damaging=pod; and S: tolerated=t, damaging=d.