Supplementary Materials

Functional Interrogation of Lynch Syndrome Associated *MSH2* Missense Variants via CRISPR-Cas9 Gene Editing in Human Embryonic Stem Cells

Abhijit Rath[1], Akriti Mishra[2], Victoria Duque Ferreira[3], Chaoran Hu[4, 5], Gregory Omerza[6], Kevin Kelly[6], Andrew Hesse[6], Honey V. Reddi[6],James P. Grady[5], Christopher D. Heinen[1]

[1] Center for Molecular Oncology and Institute for Systems Genomics, UConn Health, Farmington, CT 06030-3101, USA
[2] Department of Molecular and Cell Biology, University of Connecticut, Storrs, CT 06269, USA
[3] University of St. Joseph, West Hartford, CT 06117, USA
[4] Department of Statistics, University of Connecticut, Storrs, CT 06269, USA
[5] Connecticut Institute for Clinical and Translational Science, UConn Health, Farmington, CT 06030, USA
[6] Clinical Genomics Laboratory, The Jackson Laboratory for Genomic Medicine, Farmington, CT 06032, USA

Corresponding Author: Christopher D. Heinen, 263 Farmington Avenue, Farmington, CT 06030-3101, Phone: 860-679-8859, Email: cheinen@uchc.edu

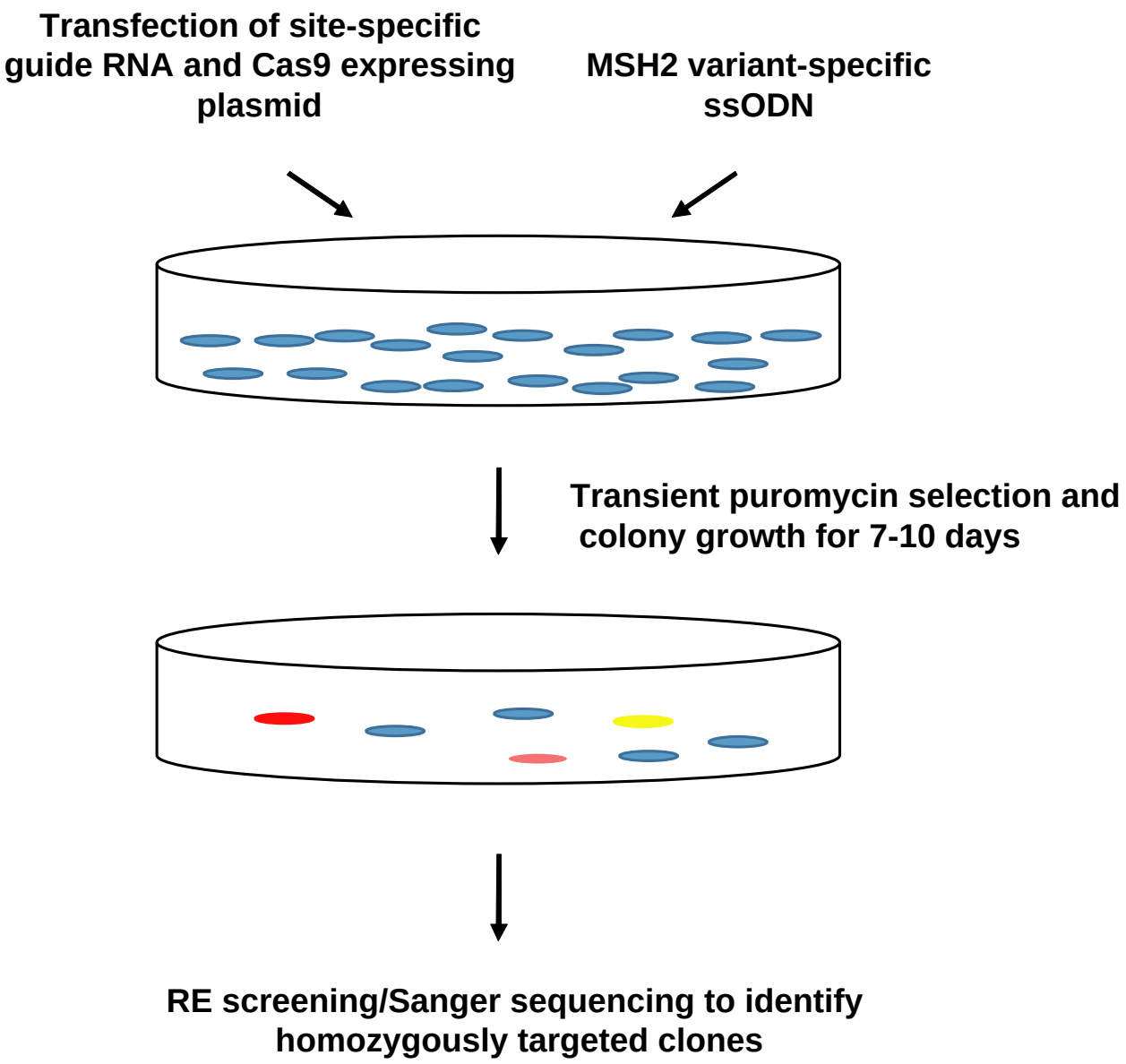


Figure S1. Outline of CRISPR-Cas9 gene editing process. Schematic showing the general framework of the CRISPR-Cas9 gene targeting process to generate clonal homozygous *MSH2* variant lines in H1 hESCs. H1 WT cells are indicated in blue. All other colors demonstrate the clonal heterogeneity in the cell population obtained after targeting.

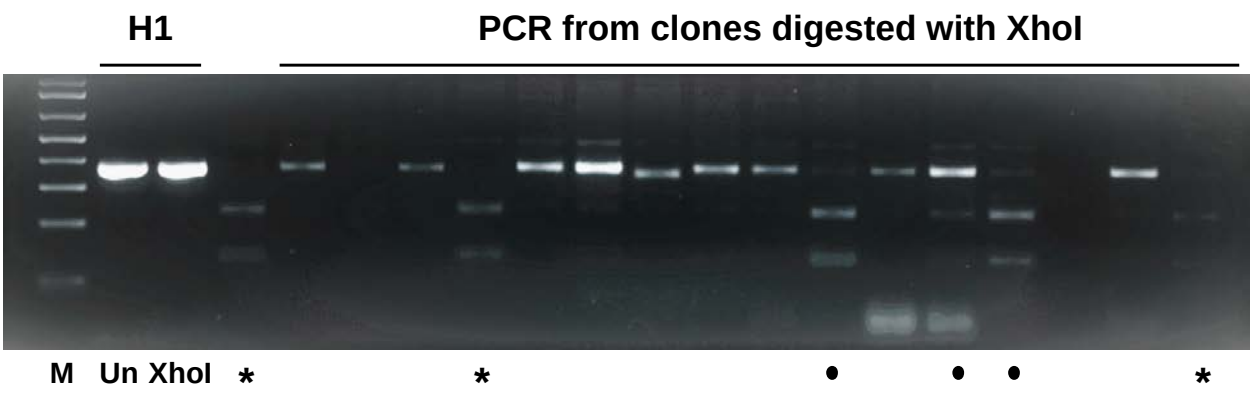
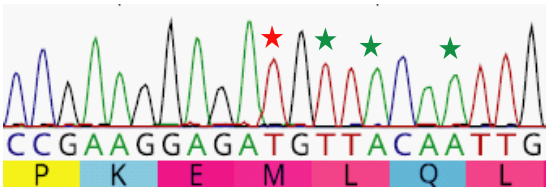


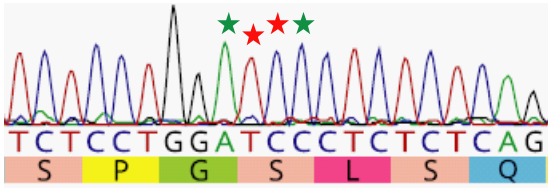
Figure S2. Sample restriction enzyme screen of PCR amplicons from p.C697R targeting experiment. A representative restriction enzyme (RE) digestion screening procedure is shown. To facilitate RE screening for MSH2 p.C697R, the restriction site for XhoI was introduced via silent mutations into the targeting ssODN along with the codon-altering mutation. Individual clones were selected following targeting and genomic DNA prepared for PCR of the target region. XhoI digestion of the PCR fragment generated a 224 and 137 bp product indicating biallelic homozygous (*) insertion of the mutation. The presence of some heterozygous clones (•) resulted in the appearance of both the digested and undigested products (361 bp). Homozygous clones were subsequently expanded and sequenced using Sanger sequencing to definitively check for homozygous insertion of the desired codon altering mutation.

Class 1

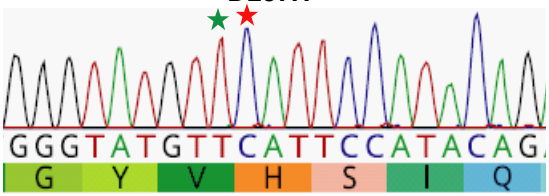
T8M



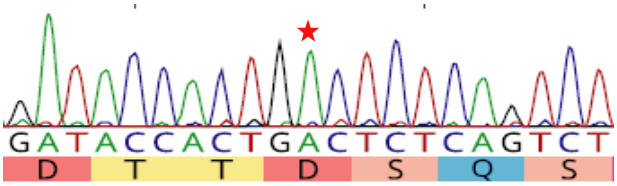
N127S



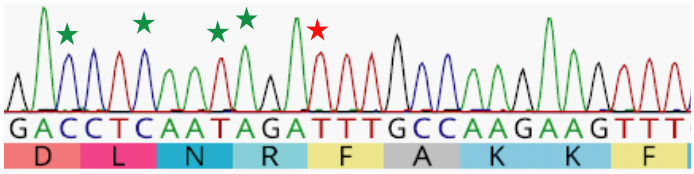
D167H



G322D

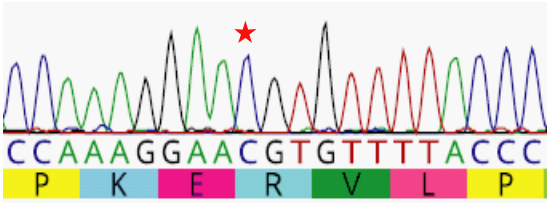


L390F

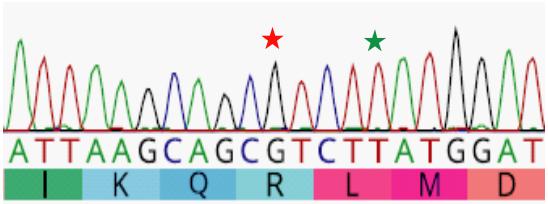


Class 5

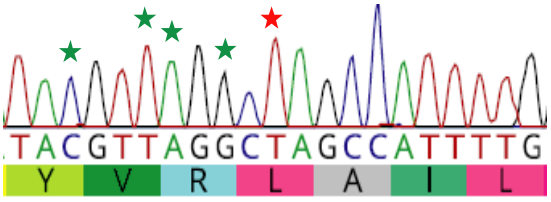
C199R



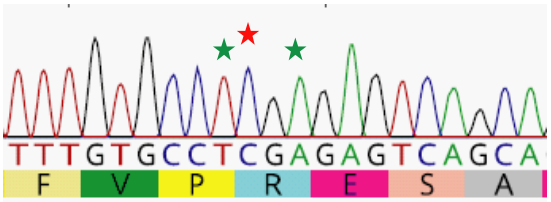
P349R



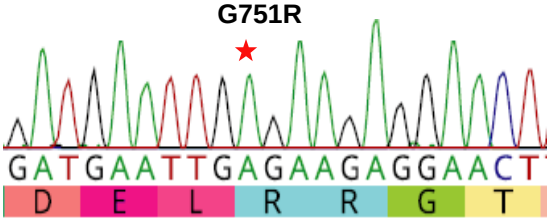
P622L



C697R



G751R



Class 3

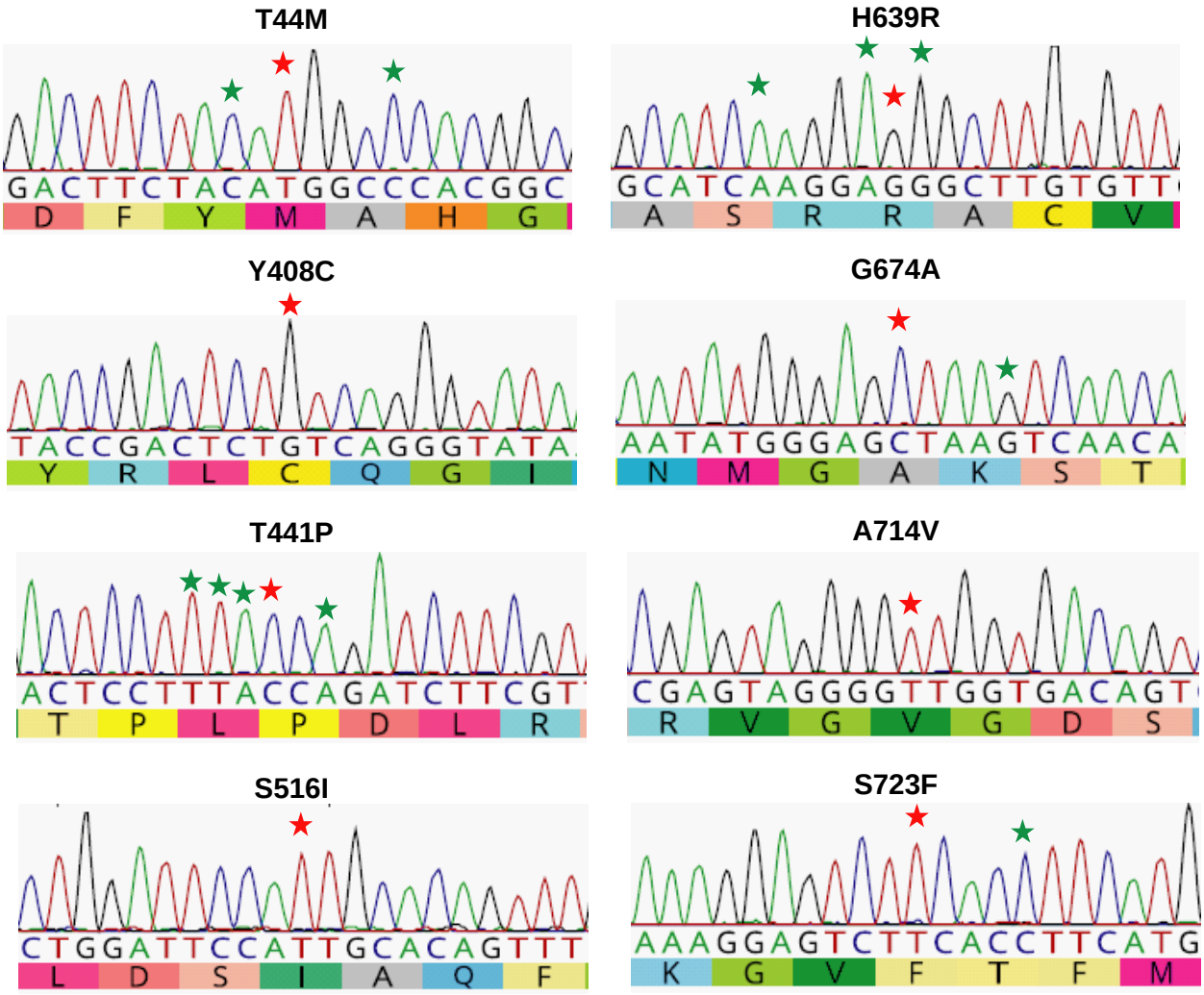


Figure S3. Sanger sequencing results of the targeted region in the homozygous *MSH2* variant expressing cell lines. Images display DNA chromatograms for each generated cell line showing homozygous insertion of the codon-altering mutations (red stars) and silent mutations (green stars).

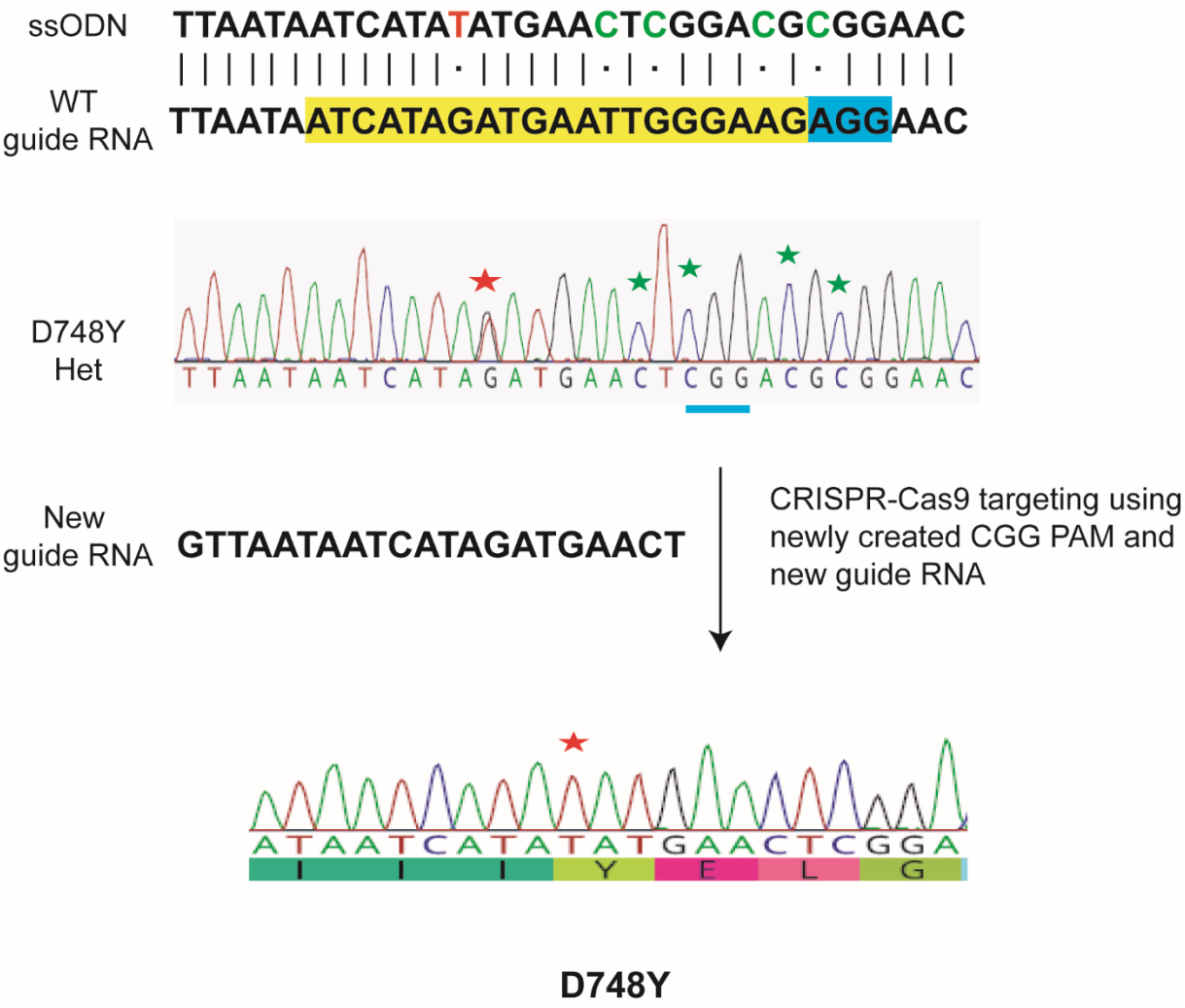


Figure S4. Generation of p.D748Y homozygous cell line from a p.D748Y heterozygous line. The schematic shows the WT genomic DNA (gDNA) sequence and the ssODN originally used to generate p.D748Y using the "AGG" PAM (light blue) and guide RNA sequence (yellow) in our first attempt. The solid vertical bars (I) represent matching ssODN sequence with the WT gDNA, whereas the dots (**.**) represent mismatches. The codon-altering mutation is shown in red and the silent mutations are shown in green. Initial targeting resulted in p.D748Y heterozygous (D748Y Het) line with incorporation of all the silent mutations in homozygous fashion essentially creating a new guide RNA binding site. The use of "GGG" PAM in the WT sequence was not possible during the initial round of targeting because of the presence of an identical sequence at another genomic locus. The p.D748Y Het line was subsequently re-targeted using the new guide RNA and PAM sequence (CGG, underlined with blue bar) displayed here to generate p.D748Y expressing cells. The silent mutations in the final p.D748Y line are not shown for clarity.

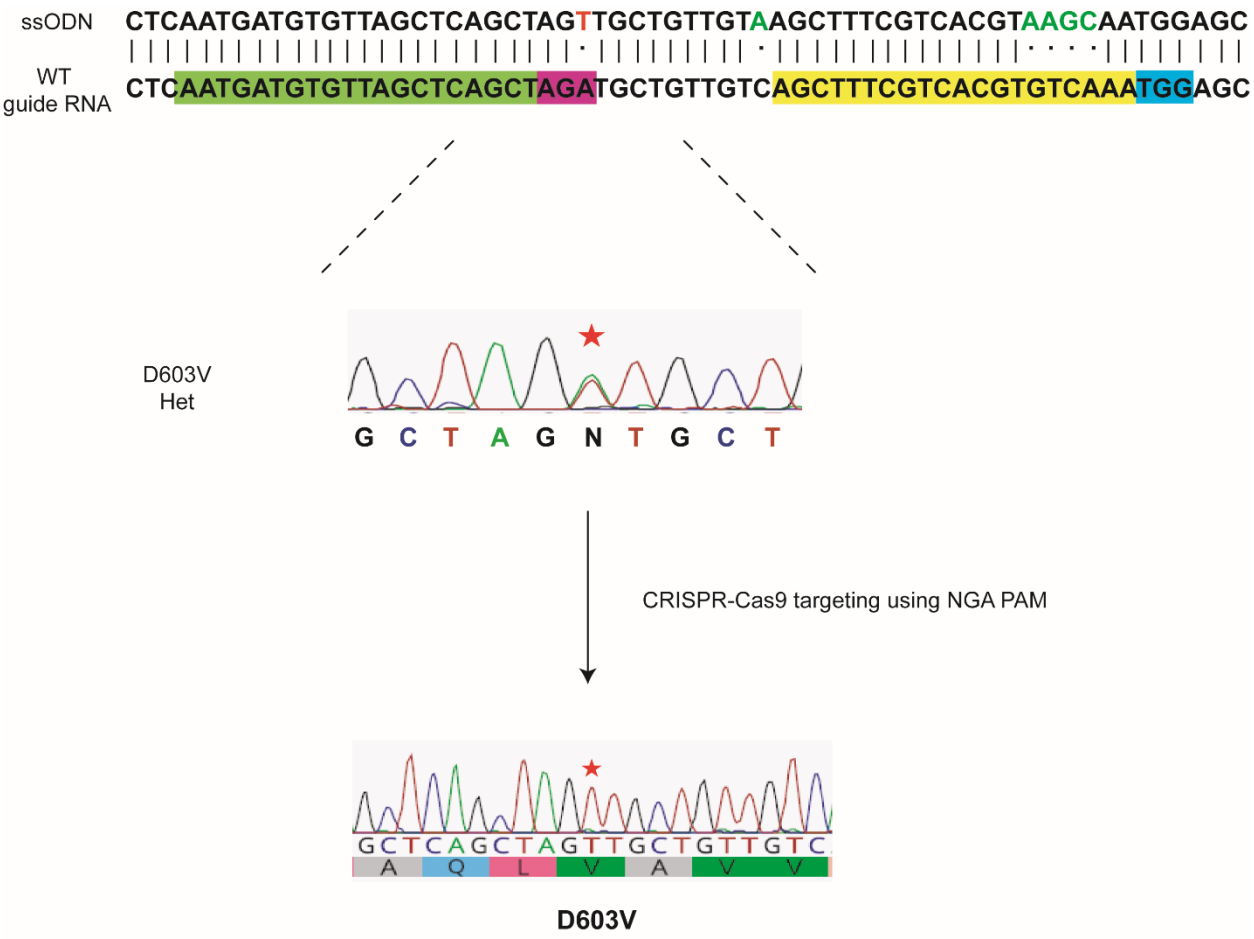


Figure S5. Generation of p.D603V cell line. Upon failure to generate p.D603V using “TGG” PAM (light blue) and the corresponding guide RNA (yellow), we switched to “AGA” PAM (magenta) and the appropriate guide RNA (green) for our second attempt. A p.D603V heterozygous (D603V Het) clone was obtained which was subsequently re-targeted to generate p.D603V-expressing cells.

Cell Lines	ClinVar ID	Mutation in cDNA	Genomic DNA sequence for sgRNA
Class 1			
T8M	90964	c.23C>T	GGCCGAAGGAGACGCTGCAGT
ssODN	GCGCACGAAGCCGACCTCGGCCGCGCTCTCCAATTGTAACATCTCCTTCGGCTGCACCGCCATGTCGAAACCTCCTCACCTCCTGGTTGA (antisense)		
N127S	36577	c.381_382delinsTC	GTCAAACTGAGAGAGATTGCC
ssODN	ATTTTTAAAATTTTATTTTTACTTAGGCTTCTCCTGGATCCCTCTCTCAGTTTGAAGACATTCTCTTTGGTAACAATGATATGTCAGCTTCCATTGGTG (sense)		
D167H	91112	c.499G>C	GACAGGTTGGAGTTGGGTATG
ssODN	ACACAGTCCTAGTTTCCTCTGTATGGAATGAACATACCCAACTCCAACCTGTCTCTGGCCATCAACTGCGGACATTTTAACACCCACAAC (antisense)		
G322D	1762	c.965G>A	GGTTCTGTTGAAGATACCAC
ssODN	TATTTTTTGTTTACTAGGGTTCTGTTGAAGATACCACTGACTCTCAGTCTCTGGCTGCCTTGCTGAATAAGTGTAAAACC (sense)		
L390F	41641	c.1168C>T	GTTGGCAAGTCGGTTAAGATCT
ssODN	ACTTTACAAGAAGATTTACTTCGTCGATTCCCAGACCTCAATAGATTTGCCAAGAAGTTTCAAAGACAAGCAGCAAACTTACAAGATTGT (sense)		
Class 3			
T44M	90627	c.131C>T	GGCGACTTCTATACGGCGCA
ssODN	CCGGGCGGCCAGCAGCGCGTCCTCGCCGTGGGCCATGTAGAAGTCGCCCCGGTCGAAAAGGCGCACTGTGGTGGTCGGCTTCTCCGGCAT (antisense)		
Y408C	90574	c.1223A>G	GAGATTGTTACCGACTCTATC
ssODN	CTGTATAACATTAGGTAGTTGATTTATACCCTGACAGAGTCGGTAACAATCTTGTAAGTTTGCTGCTTGTCTTTGAAACTTCTTGGCAAG (antisense)		
T441P	90628	c.1321A>C	GTCAGAACGAAGATCAGTAAG
ssODN	TATTGTTGGCAGTTTTTGTGACTCCTTTACCAGATCTTCGTTCTGACTTCTCCAAGTTTCAGGAAATGATAGAAACAACTTTAGATATGG (sense)		
S516I	90694	c.1547G>T	GTAATATCCAAACTGTGCAC
ssODN	GGACCCTGGCAAACAGATTAAACTGGATTCCATTGCACAGTTTGGATATTACTTTCGTGTAACCTGTAAGGAAGAAAAAGTCCTTCGTAA (sense)		
D603V	480909	c. 1808A>T	GAATGATGTGTTAGCTCAGCT
ssODN	TTGACACGTGAGCAAAGCTGACAACAGCAACTAGCTGAGCTAACACATCATTGAGTGTCTGCATTGGTTCTACATAGCCTGTATAAAAAT (antisense)		
H639R	90818	c.1916A>G	GAACTTCAACACAAGCATGCC
ssODN	ACAAGGAAGAATTATATTAAAAGCATCAAGGAGGGCTTGTGTTGAAGTTCAAGATGAAATTGCATTTATTCCTAATGACGTATACTTTGA (sense)		
G674A	NA	c.2021G>C	GTTTGTAGGCCCCAATATGGG
ssODN	CCCAGTTTGTCGAATATATGTTGACTTAGCTCCCATATTGGGGCCTACAAAACAAATTATATCAGAAAGCAAGATTTTAAACTTCTTTCT (antisense)		
A714V	90907	c.2141C>T	GATCTTAGCCCGAGTAGGGGC
ssODN	ACGTGGAGACTCCTTTCAATTGACTGTCACCAACCCCTACTCGGGCTAAGATGCAGTCCACAATGGACACTTCTGCTGACTCACATGGCA (antisense)		
S723F	90913	c.2168C>T	GCAACATTTCAGCCATGAACG
ssODN	GGGCTGGTGACAGTCAATTGAAAGGAGTCTTCACCTTCATGGCTGAAATGTTGGAAACTGCTTCTATCCTCAGGTAAGTGCATCTCCTAGTC (sense)		
D748Y	90941	c.2242G>T	GTTTGTAGGCCCCAATATGGG
ssODN	AACCCAAATCCATCGTAGGTAGAAGTTCCGCGTCCGAGTTCATATATGATTATTAATGAATCTTTGGTTGCAGACCTGAAGCACATAATT (antisense)		
D748Y[a]	90941	c.2242G>T	GTTAATAATCATAGATGAACT
ssODN	GGGAAATTTCATGTAATTATGTGCTTCAGGTCTGCAACCAAAGATTCATTAATAATCATATATGAGCTCGGACGCGGAACTTCTACCTACGATGGATTTGGGTTAGCATG (sense)		
Class 5			
C199R	91146	c.595T>C	GCCGGGTAAAACACATTCCTT
ssODN	ATCTTGAGGCTCTCCTCATCCAGATTGGACCAAAGGAACGTGTTTTACCCGGAGGAGAGACTGCTGGAGACATGGGGAAACTGAGACAGG (sense)		
P349R	90513	c.1046C>G	GATTCTGTTCTTATCCATGAG
ssODN	AAAGACTTGTTAACCAGTGGATTAAGCAGCGTCTTATGGATAAGAACAGAATAGAGGAGAGGTATGTTATTAGTTTATACTTTCGTTAGT (sense)		

P622L	1753	c.1865C>T	**G**AAAATGGCTGGTCGTACATA
ssODN	TTGCTCACGTGTCAAATGGAGCACCTGTTCCATACGTTAGGCTAGCCATTTTGGAGAAAGGACAAGGAAGAATTATATTAAAAGCATCCA (sense)		
C697R	90882	c.2089T>C	GACACTTCTGCTGACTCACA
ssODN	TCGGGCTAAGATGCAGTCCACAATGGACACTTCTGCTGACTC**T**C**G**AGGCACAAAACACCCAATTTGGGCCATGAGTACTATCACCCCAGT (sense)		
G751R	90943	c.2251G>A	**G**TTAATAATCATAGATGAATT
ssODN	CAAATCCATCGTAGGTAGAAGTTCCTCTTC**T**CAATTCATCTATGATTATTAATGAATCTTTGGTTGCAGACCTGAAGCACATAATTACAT (antisense)		
		Revertant Line	
I516S		c.1547T>G	GTAATATCCAAACTGTGCAA
ssODN	GGACCCTGGCAAACAGATTAAACTGGATTCCA**G**TGCACAGTTTGGATATTACTTTCGTGTAACCTGTAAGGAAGAAAAAGTCCTTCGTAA (sense)		

Supp. Table S1 Sequence Details for Single Guide RNAs and ssODNs Used for Gene Editing

[a] D748Y was created by re-targeting original clone using newly created PAM site inserted during initial unsuccessful targeting event

G – for single guide RNA (sgRNA) not beginning with Guanine, an extra G was added to facilitate efficient transcription from U6 promoter

Red Base – codon altering variant

Green Base – silent mutations

NA – Not Applicable

Cell Line	Forward Primer	Reverse Primer
Class 1		
T8M	AGTAGCTAAAGTCACCAGCGTGC	CATGTACTTGATCACCCCCTGGG
N127S	AGCAGGAGAATCGATTGAACCC	TGTCTCTGGCCATCAACTGC
D167H	CTTAGGCTTCTCCTGGCAATC	CTTTCCTAGGCCTGGAATCTCC
G322D	CCTCTGTTTTTCATGGCGTAGTAAG	GCCTCTCATGGATAAGAACAGAATAGAG
L390F	CCCAGCAGATTCAAGCTTTT	GGACAGCACATTGCCAAGTA
Class 3		
T44M	AGAAGCCGACCACCACAGT	GTCGTGCCAGGCATTTTAAT
Y408C	CCCAGCAGATTCAAGCTTTT	GGACAGCACATTGCCAAGTA
T441P	CCTTTTGGATCAAATGATGCTT	CCACTGTCCACAAAGGTGCT
S516I	GGTTTACCCAGAAAGCAGCTTTC	CGACTTGCAAACCTGTTGGTA
D603V	ATGGGTTTTGAATTCCCAAATGG	CACAAAGCCCAAAAACCAGGTT
H639R	TGTAAATTAGGAAATGGGTTTTGAA	ACAAAACGTTACCCCCACAA
G674A	CTAGGCCACAGTCAAATTACAGG	TTCCAACATTTCAGCCATGAACG
A714V	TATGTCAGTGTAAACCTACGCG	TATCTTCAAGGGACTAGGAGATGC
S723F	CAGTGTACAGTTTAGGACTAACAATCC	CAGATGTTTTTACATGAGAATCTGCAA
D748Y	GCCCTTGCCCATTTTTCTAT	TTGGCCAAGGCAGTAAGTTC
Class 5		
C199R	TCAGTTTGAAGACATTCTCTTTGG	TCACTAGACTCAATTTGCTTACCTG
P349R	TGTTCCTCTGTTTTTCATGGCG	AGTGGTATAATCATGTGGGTAACT
P622L	CAGGCTATGTAGAACCAATGCAG	CCAGTAATGATGTGGAACATCTG
C697R	AGGCTGTGGTTCTGCCTTTA	GGAGATGCACTTACCTGAGGA
G751R	GCCCTTGCCCATTTTTCTAT	TTGGCCAAGGCAGTAAGTTC
MSH2-KO	AGTAGCTAAAGTCACCAGCGTGC	CATGTACTTGATCACCCCCTGGG

Supp. Table S2 Primers Used to Amplify Target Region for Sequence Verification

Cell Line	Forward Primer	Reverse Primer	Expected Amplicon size (bp)
C199R	GAATCTGCAGAGTGTTGTGC	CAGCACTATTCATCTGCTCTCC	551
D603V	AATATGAAGAAGCCCAGGATGCC	CAGCCATGAACGTGGAGACTCC	477
C697R	TGTTCCATATGTACGACCAGCC	GTTGGTATCTGATTGGCCAAGGC	540

Supp. Table S3 Primers Used to Amplify Target Region for cDNA amplification

Cell Line	No. of homozygous clones obtained	No. of clones screened	Homozygous targeting efficiency (%)	Distance of codon-altering mutations from PAM	Distance of silent mutations from PAM
Class 1					
T8M	5	24	21[d]	9	2, 5, 7
N127S	2	24	8[e]	4, 5	3, 6
D167H	1	40	2.5[d]	PAM	-[f]
G322D	1	7	14[d]	PAM	-
L390F	4	32	13[d]	14	4, 7, 10, 11
Class 3					
T44M	3	32	9[e]	7	3, 9
Y408C	2	16	13[d]	3	-
T441P	5	53	9[e]	4	1,5
S516I	2	16	13[d]	1	-
D603V[a]	0	45	0[e]	31	3, 4, 5, 6, 21
D603V[b]	0[c]	94	0[d]	PAM	-
H639R	4	23	17[e]	4	PAM, 3, 5
G674A	1	22	5[e]	PAM	4 (3' of PAM)
A714V	3	11	27[d]	1	-
S723F	2	35	6[e]	PAM	2
D748Y	0[c]	23	0[e]	14	PAM, 2, 6, 8
Class 5					
C199R	1	16	6[d]	7	-
P349R	3	32	9[d]	PAM	3
P622L	2	24	8[d]	11	3, 6, 9
C697R	5	48	10[e]	1	PAM, 3
G751R	3	48	6[d]	PAM	-

Supp. Table S4 Targeting Efficiency and Relative Position of Target Codon to PAM Site

[a] Initial attempt to target via an NGG PAM site failed to produce any targeted clones

[b] Targeting via an NGA PAM site

[c] Heterozygous clones were obtained that were re-targeted

[d] Determined by Sanger sequencing

[e] Determined by restriction enzyme digestion screening

[f] A single nucleotide silent mutation was introduced within the PAM

Cell Line	I	II	III	IV	V
Class 1					
T8M	✓	✓	✓	✓	ND
N127S	✓	✓	✓	✓	✓
D167H	✓	✓	ND	✓	✓
G322D	✓	✓	✓	✓	✓
L390F	✓	✓	✓	✓	✓
Class 3					
T44M	✓	✓	✓	✓	✓
Y408C	✓	✓	✓	✓	✓
T441P	✓	✓	✓	✓	✓
S516I	✓	✓	✓	✓	ND
D603V	✓	✓	✓	✓	✓
H639R	✓	✓	✓	✓	✓
G674A	✓	✓	✓	✓	✓
A714V	✓	✓	✓	✓	✓
S723F	✓	✓	✓	✓	✓
D748Y	✓	✓	✓	✓	✓
Class 5					
C199R	✓	✓	✓	✓	✓
P349R	✓	✓	✓	✓	✓
P622L	✓	✓	✓	✓	✓
C697R	✓	✓	ND	✓	✓
G751R	✓	✓	✓	✓	✓

Supp. Table S5 Examination of Top Five Candidate Loci for Off-target Cleavage by CRISPR-Cas9

✓ - Sequence at candidate off-target site matches expected H1 WT sequence

ND - Sites could not be amplified due to presence of repetitive sequences

Cell Line	BAT-26[a]	NR-21[a]	BAT-25[a]	MONO-27[a]	NR-24[a]	Penta D[a]	Penta E[a]
H1 (WT)	0/1	0/1	0/1	0/1	0/1	0/1	0/1
MSH2-KO	1/1	1/1	1/1	1/1	1/1	0/1	1/1
H639R	0/12	0/12	1/12	0/12	0/12	0/12	0/12

Supp. Table S6 Microsatellite Instability with Expanded Five Mononucleotide Marker Panel

[a] Number of clones with altered alleles compared to WT sequence