OMTM, Volume *26*

Supplemental information

PAM-altering SNP-based allele-specific CRISPR-Cas9 therapeutic strategies for Huntington's disease

Jun Wan Shin, Eun Pyo Hong, Seri S. Park, Doo Eun Choi, Sophia Zeng, Richard Z. Chen, and Jong-Min Lee

Table S1. Summary of polymorphic PAS in the analysis region.

224 PAS in the region (chr4:3056408-3116408; GRCh37/hg19) were polymorphic in 8,543 HD participants. For those 224 PAS, PAM-generating alleles (PGA) are grouped based on the allele and strand. Both reference and alternative alleles of 5 PAS generate the NGG PAM sites. In addition, two alternative alleles of one PAS generate PAM sites. Therefore, we identified 230 PGAs from 223 bi-allele SNPs and 1 multi-allele SNP in the region.

PAM-generating allele (PGA)	Frequency
Reference allele on the plus strand	59
Reference allele on the minus strand	73
Alternative allele on the plus strand	49
Alternative allele on the minus strand	39
Reference allele on the plus strand and alternative allele on the minus strand	1
Reference allele on the minus strand and alternative allele on the plus stand	4

Table S2. The mutant specificities of 230 PGAs from 224 polymorphic PAS.

Information of 230 PGAs from 224 PAS are shown. Genomic coordinate (BP, base pair) was based on GRCh37/hg19. PGA represents NGG PAM-generating allele. 'G' and 'C' alleles generate the NGG PAM sequence on the plus and minus strand, respectively. The frequency of PGA was based on 8,543 HD subjects. To evaluate the applicability of each PAS in the allele-specific CRISPR-Cas9, we calculated the proportions of HD subjects who carry the PAM site on the mutant *HTT*, normal *HTT*, both, or neither. The proportion of HD subjects who carry mutant-specific PAM sites for a given PAS represents the mutant specificity for that PAS.

Table S3. Predicted off-targets and experimental validation.

In order to evaluate the levels of off-target effects, we predicted potential off-targets using Cas-Offinder. Based on the PAM site generated by the corresponding PAS, a gRNA (20 nucleotide) was designed for each PAS.

A) Names and sequences of test gRNAs are shown in the second and third columns, respectively. L and R represent the left and right side of the transcription start site, respectively. Off-target prediction was based on 0, 1, 2, and 3 mismatches. $, on-target (i.e., *HTT)*. &, one on-target and one off-target. The gRNA with 0 predicted targets with 0 mismatch for L2 was due to the fact that the alternative allele at this location generates the PAM site. Sites with yellow highlight represent candidate targets, and were therefore further validated by MiSeq analysis (see below).

B) Subsequently, we treated a patient-derived iPSC and performed MiSeq analysis to experimentally validate the predicted off-targets of L4, R4, and R6. We focused on predicted off-targets that are located in the exons of protein-coding genes because off-targeting at intergenic regions or introns may have minimal impacts. Five, one, and one exonic off-targets were predicted for L4, R4, and R6, respectively; all predicted off-targets carry 3 mismatches. #, off-targeting potentially due to sequencing errors caused by repetitive sequences. Genomic location was based on GRCh38.

A. Predicted off-targets

PAS	gRNA	gRNA sequence	0 mismatch	1 mismatch	2 mismatches	3 mismatches
rs1313769	L1	GGCAGAGCTTGCAGTGAGCT	31	581	> 1000	> 1000
rs1313774	L2	GCATATAATCAAGAAATAAT	0	7	9	41
rs12506200	L3	CAGGCATGAGCCAGCATGCC	80	> 1000	> 1000	> 1000
rs2857935	L4	CCCGCTCCAGGCGTCGGCGG	1 $	0	0	13
rs28820097	R1	CAACAACTAAAAGCACAACA	1 $	0	7	70
rs7659144	R2	CCCATGGGCCATGTGGAAAT	1 $	0	1	12
rs7688390	R3	AGAATGGACATCATAAAGAT	1 $	0	0	26
rs16843804	R4	GTCGATGATCTCTTTAACCG	1 $	0	0	6
rs6828615	R5	TGGGCTCACGCCTGTAATCC	2 &	42	> 1000	> 1000
rs16843836	R6	GCTATGTTTATCCTGCAACC	1 $	0	0	5

B. Experimental validation of the predicted off-targets on exons

PAS	gRNA	Off-target location	# of mismatches	Type of sequence	Percentage
rs2857935	L4	chr20:21396423-21396445	3	Modified	0.9 #
				Unmodified	99.1
		chr20:51003962-51003984	3	Modified	0
				Unmodified	100
		chr19:1084312-1084334	3	Modified	0
				Unmodified	100
		chr7:7969448-7969470	3	Modified	0.4 #
				Unmodified	99.6
		chr10:101131872-101131894	3	Modified	4.3 #
				Unmodified	95.7
rs16843804	R4	chr10:77985187-77985209	3	Modified	0
				Unmodified	100
rs16843836	R6	chr1:39387262-39387284	3	Modified	0
				Unmodified	100

Table S4. Selective expression of normal *HTT* mRNA in the targeted clonal lines.

We performed MiSeq analysis of cDNA from total RNA using a primer set to quantify alleles at rs363099, which is heterozygous in the HD subjects with the most frequent diplotype. Mutant and normal *HTT* carry C and T allele at this location, respectively. Based on the allele at rs363099, we calculated the expression levels of mutant and normal *HTT* mRNA in each of the targeted clones. # of reads and percent values represent the number of MiSeq sequence reads and the relative proportion of corresponding allelic expression.

Cell	gRNA pair	Clonal line	Allele	MiSeq of cDNA	
				Number sequence reads	%
iPSC-A	L4-R4	1	Mutant	0	0
			Normal	69,732	100
		2	Mutant	0	0
			Normal	69,169	100
		3	Mutant	0	0
			Normal	66,331	100
		4	Mutant	0	0
			Normal	33,385	100
		5	Mutant	0	0
			Normal	75,267	100
	L4-R6	1	Mutant	0	0
			Normal	67,713	100
		2	Mutant	0	0
			Normal	80,816	100
		3	Mutant	0	0
			Normal	100,832	100
		4	Mutant	0	0
			Normal	62,525	100
		5	Mutant	0	0
			Normal	72,609	100
iPSC-B	L4-R4	1	Mutant	0	0
			Normal	58,789	100
		2	Mutant	0	0
			Normal	73,195	100
		3	Mutant	0	0
			Normal	47,992	100
		4	Mutant	0	0
			Normal	67,098	100
		5	Mutant	0	0
			Normal	61,895	100
	L4-R6	1	Mutant	0	0
			Normal	64,061	100
		2	Mutant	0	0
			Normal	74,280	100
		3	Mutant	0	0
			Normal	49,121	100
		4	Mutant	0	0
			Normal	52,415	100
		5	Mutant	0	0
			Normal	51,208	100

Table S5. Mutant specificity of TP-CRISPR strategies using two allele-specific gRNAs.

Since allele-specific TP-CRISPR strategies to prevent the expression of the mutant *HTT* require the use of two gRNAs simultaneously, we computed the levels of mutant specificity (i.e., the percentage of HD subjects who are eligible for a given mutant *HTT*-specific approach) of different gRNA combinations focusing on 10 candidate PAS. PAS on the rows and columns represent PAS upstream and downstream of the TSS, respectively. For example, the mutant specificity of L4-R4 combination was calculated by counting HD subjects who carry the NGG PAM site only on the mutant *HTT* at rs2857935 and rs16843804, revealing that 27% of HD subjects carry the PAM sites at both locations of the mutant *HTT.* Numbers in parentheses represent the mutant specificity of individual PAS in the same HD subjects.

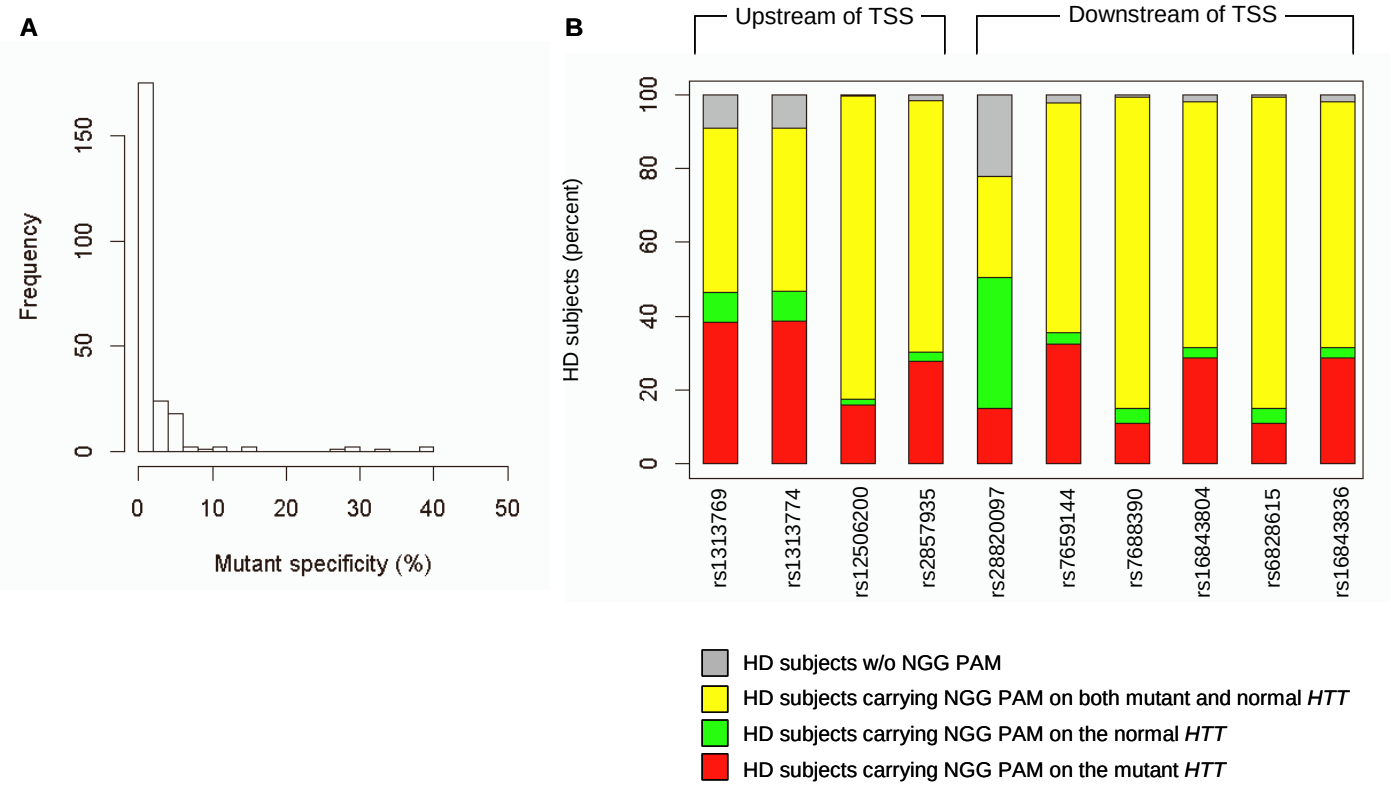
gRNA	R1	R2	R3	R4	R5	R6
PAS	rs28820097	rs7659144	rs7688390	rs16843804	rs6828615	rs16843836
(mutant specificity %)	(14.8)	(32.5)	(10.9)	(28.7)	(10.9)	(28.7)
L1	4.9	22.9	1.2	22.2	1.2	22.2
rs1313769						
(38.4)						
L2	4.9	22.9	1.2	22.2	1.2	22.2
rs1313774						
(38.5)						
L3	0.8	2.7	8.4	1.1	8.4	1.1
rs12506200						
(15.9)						
L4	0.1	26.9	0	27	0	27
rs2857935						
(27.6)						

Table S6. Cumulative mutant specificities of TP-CRISPR strategies using one allele-specific and one non-allele-specific gRNAs.

In order to identify a set of gRNAs and target sites that can be applied to the maximum number of HD subjects for allele-specific TP-CRISPR targeting, we calculated cumulative mutant specificity by 10 iterations of identification of target population and re-calculation of the mutant specificity. Since simultaneous use of one allele-specific and one non-allele-specific gRNA may lead to allele-specific CRISPR editing, we calculated cumulative mutant specificity based on a single allele-specific gRNA approach. The third column shows cumulative mutant specificity. For example, three allele-specific TP-CRISPR strategies based on rs1313774, rs12506200, and rs568806386 (iteration 3) can be applied to at least ~60% of HD subjects. To identify alternative targets, linkage disequilibrium (LD) was calculated for the selected target SNP based on genotypes of 8,543 HD subjects; we took the top 5 PAS with the highest LD values (numbers below PAS).

Iteration	PAS	Cumulative mutant specificity (%)	Alternative target 1	Alternative target 2	Alternative target 3	Alternative target 4	Alternative target 5
			LD	LD	LD	LD	LD
1	rs1313774	38.44	rs1313769	rs2857935	rs16843804	rs16843836	rs7659144
			0.9974	0.3243	0.2868	0.2859	0.2412
2	rs12506200	50.59	rs13141939	rs13102260	rs7688390	rs6828615	rs35342954
			0.4149	0.4003	0.3043	0.3022	0.2875
3	rs568806386	59.73	rs28820097	rs762855	rs2285086	rs6446722	rs2024115
			0.0038	0.0025	0.002	0.002	0.0018
4	rs574984731	66.59	rs187490343	rs916170	rs28820097	rs1313774	rs1313769
			0.902	0.0063	0.0026	0.0009	0.0009
5	rs186788713	69.31	rs28820097	rs1313774	rs1313769	rs184373685	rs193177768
			0.002	0.001	0.001	0.0007	0.0007
6	rs61792502	70.21	rs61792500	rs61792503	rs61792505	rs61792472	rs56794194
			0.9983	0.9915	0.9865	0.9665	0.8475
7	rs551562237	71.07	rs73191179	rs28820097	rs2857935	rs16843836	rs16843804
			0.0009	0.0002	0.0002	0.0002	0.0002
8	rs7693317	71.75	rs7693317	rs10016755	rs28652828	rs28660254	rs80093929
			1	1	1	1	0.9788
9	rs13102260	72.17	rs13141939	rs35342954	rs35631490	rs13122415	rs13132932
			0.8983	0.6406	0.5061	0.5019	0.4939
10	rs7659144	72.49	rs55962025	rs2024115	rs16843804	rs16843836	rs2857935
			0.9341	0.8053	0.7662	0.7651	0.7591

Figure S1. Distributions of mutant specificities of polymorphic PAS.

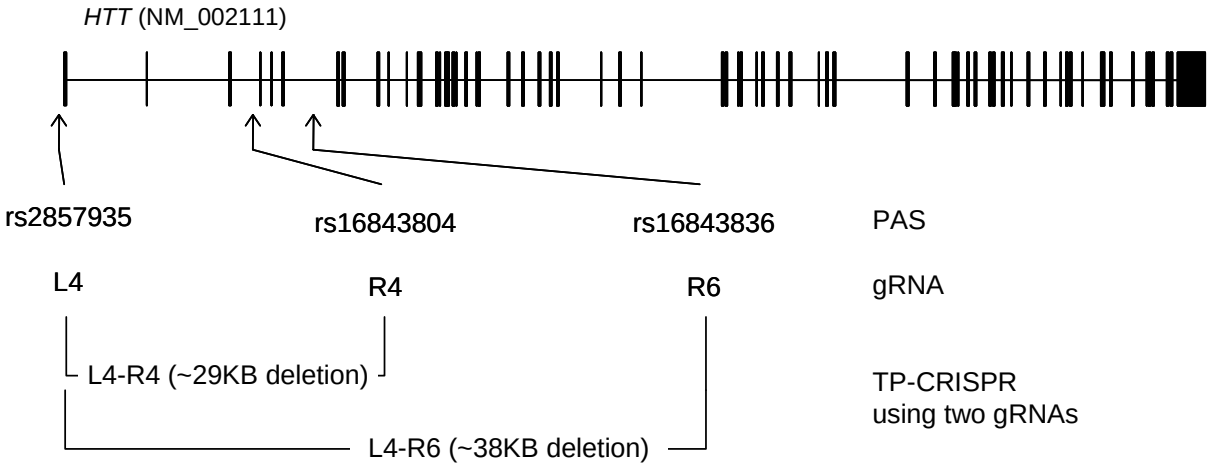


A) The levels of mutant specificity (i.e., percentage of HD subjects who carry the PAM site only on the mutant *HTT*) (X-axis) of 230 PGAs from 224 polymorphic PAS are summarized.

B) By applying the 10% mutant specificity threshold, we identified 10 PAS that show relatively high levels of mutant specificities. For each PAS, the percentages of HD subjects who carry the PAM site 1) only on the mutant *HTT* (red; mutant specificity), 2) only on the normal *HTT* (green), 3) on both mutant and normal *HTT* (yellow), and 4) on none of mutant and normal *HTT* (grey) are summarized.

Figure S2. Allele-specific TP-CRISPR strategies based on combinations of two PAS.

A



B

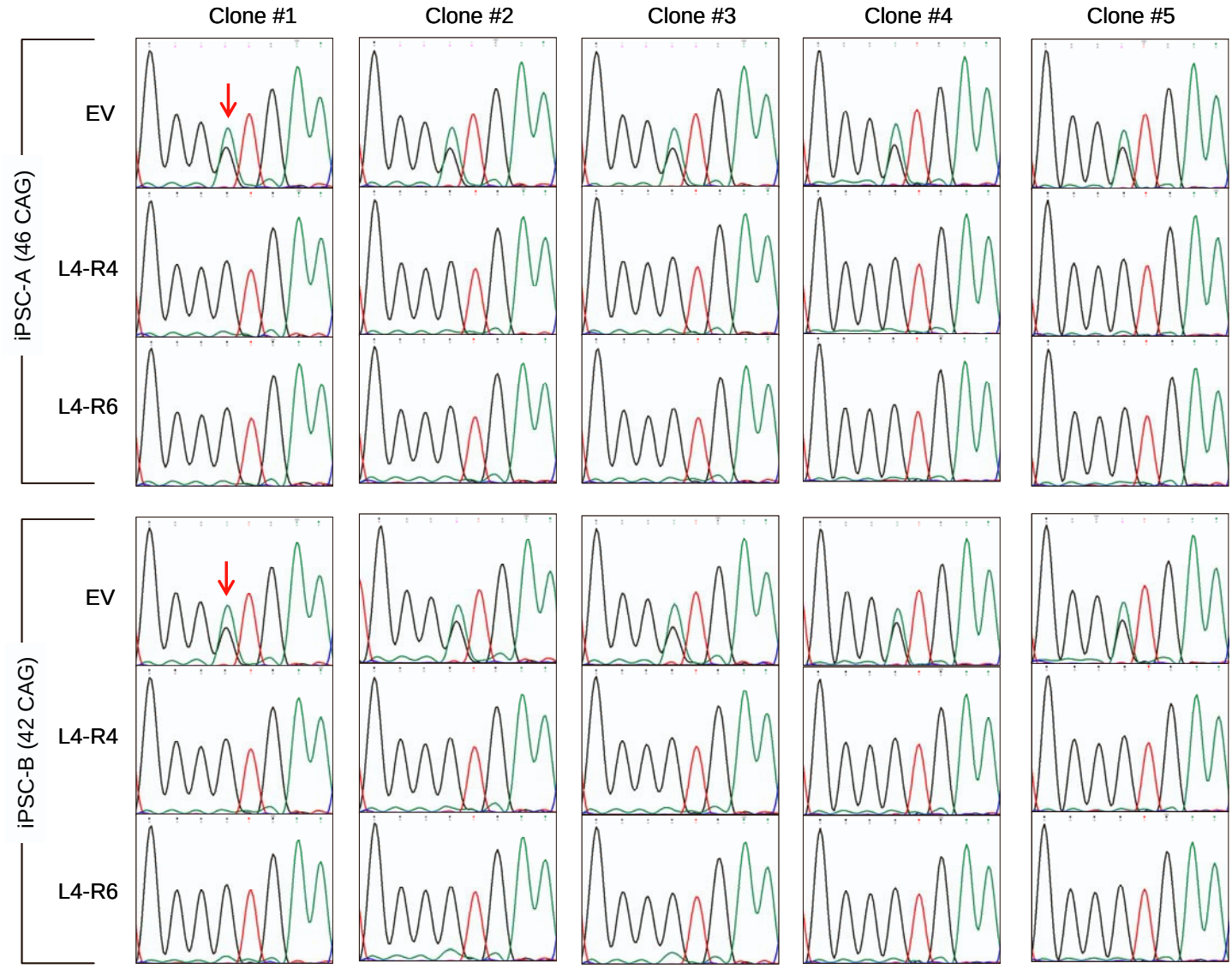
HD-derived iPSC	CRISPR-Cas9 treatment	Number of independent clones
iPSC-A (46 CAG)	EV	6
	TP-CRISPR : L4-R4	5
	TP-CRISPR : L4-R6	5
iPSC-B (42 CAG)	EV	6
	TP-CRISPR : L4-R4	5
	TP-CRISPR : L4-R6	5

DNA, RNA, and protein analysis

A) Based on the transfection and MiSeq analysis, we identified 3 target PAS and gRNAs that showed high allele specificities and relatively good editing efficiencies. To further evaluate TP-CRISPR strategies using those 3 candidate PAS, we treated patient-derived iPSC lines with empty vector (EV) or two different combinations of gRNAs. The gRNA L4 was designed based on the PAM site generated by rs2857935, which showed 27.6% mutant specificity. This gRNA targets the left side of the transcription start site, and is therefore used for both combinations. The gRNAs R4 and R6 were designed based on PAM sites respectively generated by rs16843804 and rs16843836, which are located downstream of the TSS. L4-R4 and L4-R6 combinations were expected to excise approximately 29KB and 38KB from the mutant *HTT* in our representative HD iPSC lines carrying the most frequent diplotype.

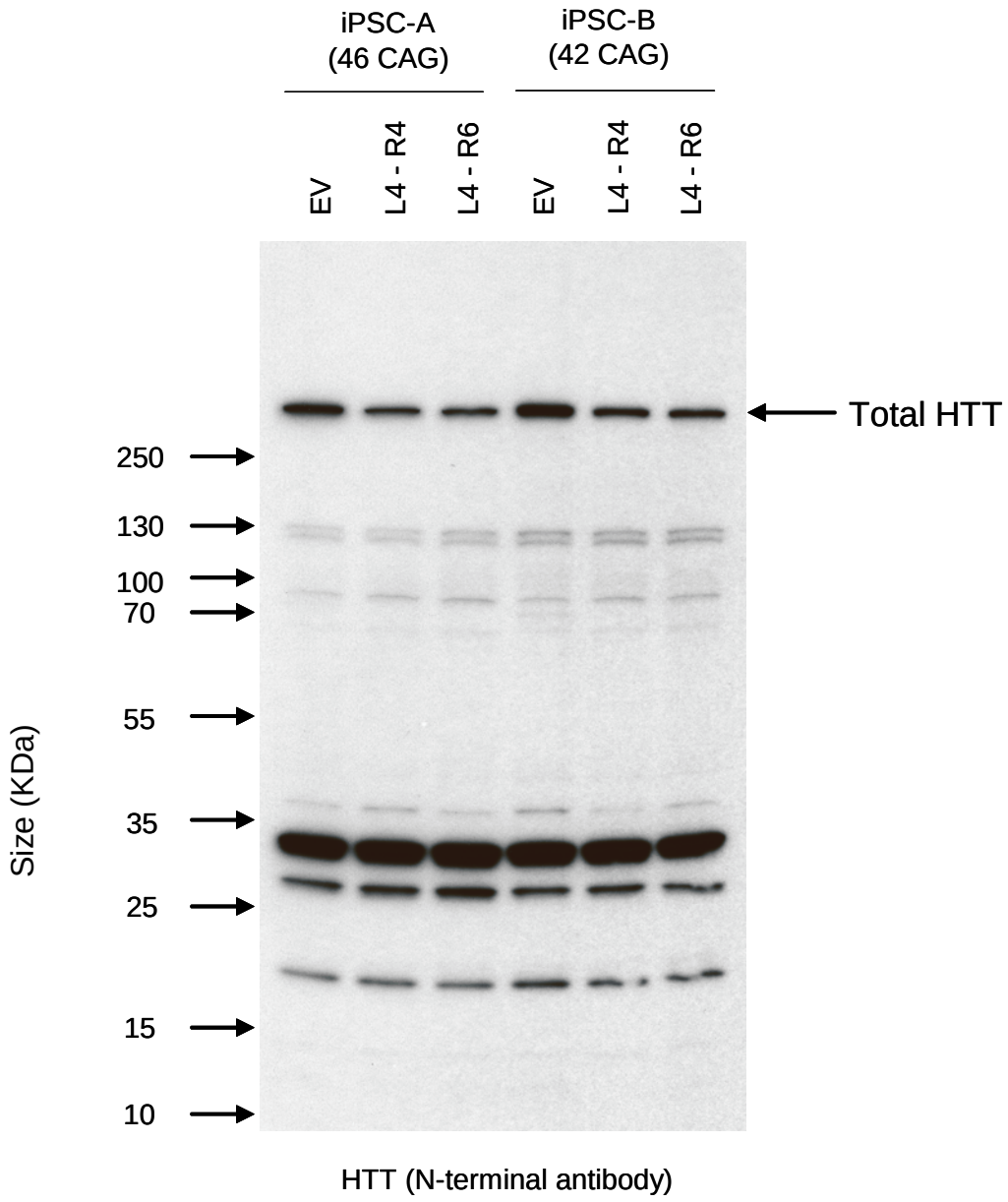
B) Two independent HD iPSC lines were treated with empty vector (EV), L4-R4 or L4-R6 gRNA combinations. Subsequently, we established 20 targeted and 12 empty vector treated clonal lines for subsequent characterization.

Figure S3. The lack of re-integration of the excised region in the targeted clonal lines by allele-specific TP-CRISPR.



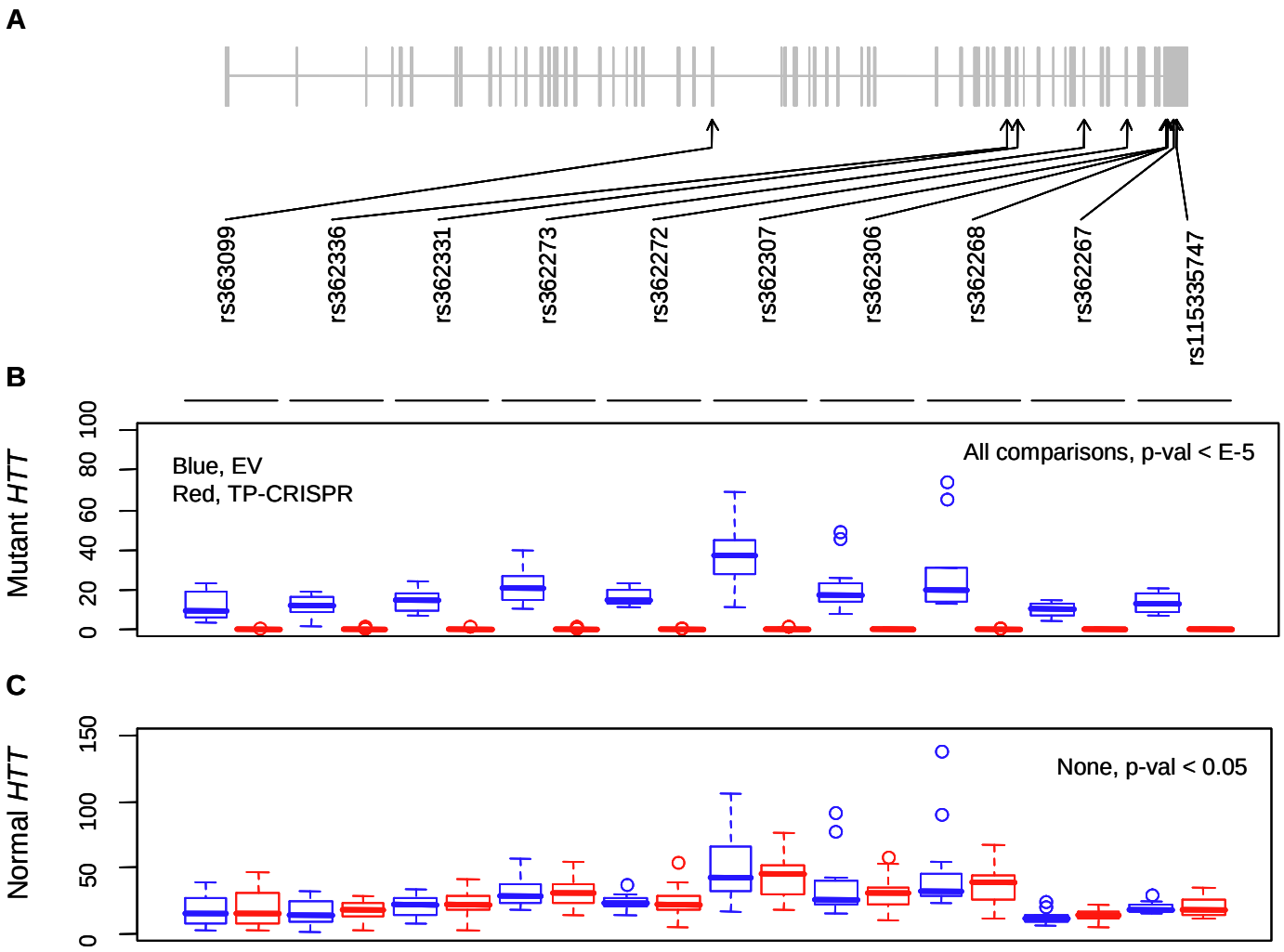
To determine whether the excised DNA region got integrated into the genome elsewhere, we performed DNA PCR and Sanger sequencing analysis focusing on determining the genotype at rs2285086 (red arrows), which is located in the middle of the excised region. Heterozygous genotype shown in the empty vector-treated clones (EV) indicates the presence of mutant and normal *HTT*. The loss of heterozygosity and signals for 'G' allele at rs2285086 (red arrows) in the targeted clonal lines refute the integration of excised mutant DNA in the genome. All 20 independent targeted clones from iPSC-A (A) and iPSC-B (B) confirmed the loss of heterozygosity, actually meaning hemizygosity.

Figure S4. Mutant *HTT*-specific TP-CRISPR does not produce exon 1 huntingtin protein.



To determine whether excision of DNA (involving TSS and expanded CAG repeat) from the mutant *HTT* resulted in the production of exon 1 huntingtin protein, we analyzed whole cell lysate from EV-treated and targeted clonal lines by immunoblot analysis. We ran gels for 1 hr to retain small molecular weight proteins. Transferred membranes were probed using N-terminal antibody (N17). The size of exon1 fragment protein with 46 and 42 glutamines is approximately 12KDa.

Figure S5. ASE analysis of RNAseq data.



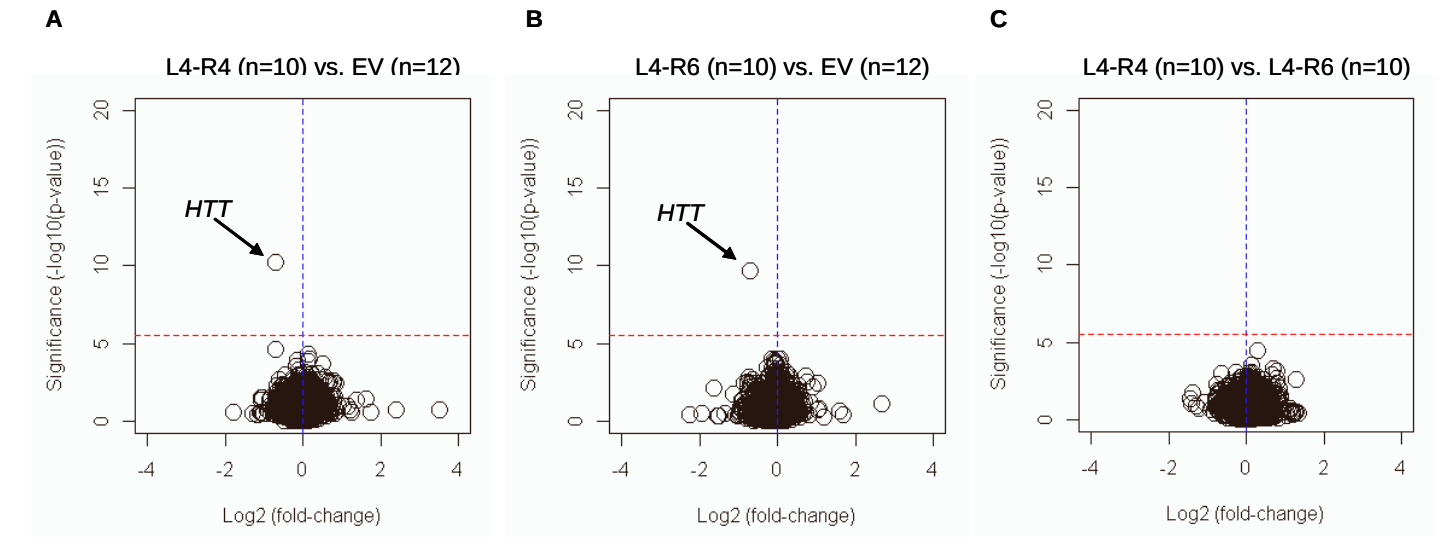
For allele-specific expression (ASE) analysis to determine the allele specificity of TP-CRISPR strategies, we counted alleles of 10 heterozygous exonic SNPs in our RNAseq data.

A) Locations of the 10 heterozygous exonic SNPs are shown relative to the RefSeq NM_002111.

B) Counts of alleles for the 10 heterozygous exonic SNPs on the mutant *HTT* in the EV-treated clones (blue) and TP-CRISPR-targeted clones (red; L4-R4 and L4-R6 combinations) are summarized. All pairs showed nominal p-value < 1E0-5 (Bonferroni corrected p-value < 0.05). N=12 for each blue box; N=20 for each red box.

B) The same analysis approach was applied to alleles of 10 heterozygous exonic SNPs that are on the normal *HTT*. Blue and red boxes represent alleles of normal *HTT* in the EV-treated and TP-CRISPR targeted clonal lines, respectively. None of the exonic SNP sites showed nominal p-value < 0.05. N=12 for each blue box; N=20 for each red box.

Figure S6. DGE analysis of RNAseq data.



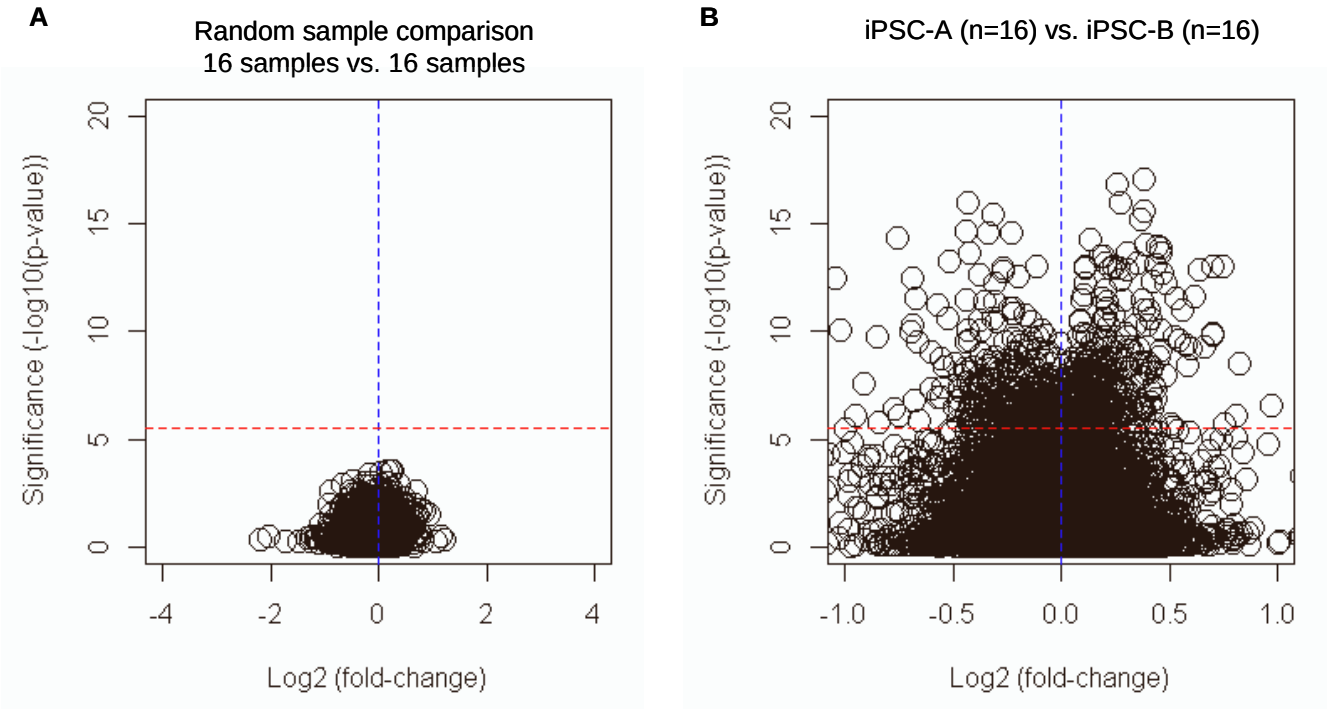
DGE analysis of RNAseq data was performed to identify significantly altered genes by TP-CRISPR strategies.

A) DGE analysis was performed by comparing 10 clones for L4-R4 combination to 12 EV-treated controls. A red horizontal and a blue vertical line represent Bonferroni-corrected significance and zero fold-change, respectively. No genes were significantly altered in TP-CRISPR targeted clonal lines except *HTT* (black arrow).

B) Similarly, DGE analysis was performed by comparing 10 clones for L4-R6 combination to 12 EV-treated clonal lines.

C) To determine whether L4-R4 combination produced different transcriptome changes compared to L4-R6 combination, we performed DGE analysis by comparing 10 clonal lines by L4-R combination to 10 clonal lines by L4-R6 combination, revealing zero significant genes by Bonferroni correction (red horizontal line).

Figure S7. Power and sensitivity of our RNAseq analysis.



The shape of the volcano plot was atypical in our main DGE analysis (i.e., TP-CRISPR vs. EV), potentially due to the fact that only one gene is significantly altered.

A) To understand the reason behind the atypical shape of the volcano plot in our targeted clones vs. EV comparison, we generated two groups of randomly assigned samples and performed DGE analysis. Since samples are randomly mixed in two groups, we did not expect to see any significantly altered genes. The shapes of volcano plots for true sample comparison and random sample comparison were similar except for *HTT*. A red horizontal and a blue vertical line represent Bonferroni-corrected significance and zero fold change, respectively.

B) To judge whether our RNAseq study had sufficient power and sensitivity, we performed DGE analysis by comparing all 16 clonal lines from iPSC-A with 16 clonal lines from iPSC-B. Due to the different genetic background of two lines, many genes are predicted to be significantly different. The volcano plot focused on log2(fold-change) -1 to 1 (X-axis) to highlight the levels of significance. By Bonferroni corrected p-value, 664 genes (above the red dotted line) were significantly different between two lines.