

The American Journal of Human Genetics

Supplemental Data

## **Spell Checking Nature: Versatility of CRISPR/Cas9 for Developing Treatments for Inherited Disorders**

**Daria Wojtal, Dwi U. Kemaladewi, Zeenat Malam, Sarah Abdullah, Tatianna W.Y. Wong, Elzbieta Hyatt, Zahra Baghestani, Sergio Pereira, James Stavropoulos, Vincent Mouly, Kamel Mamchaoui, Francesco Muntoni, Thomas Voit, Hernan D. Gonorazky, James J. Dowling, Michael D. Wilson, Roberto Mendoza-Londono, Evgueni A. Ivakine, and Ronald D. Cohn**

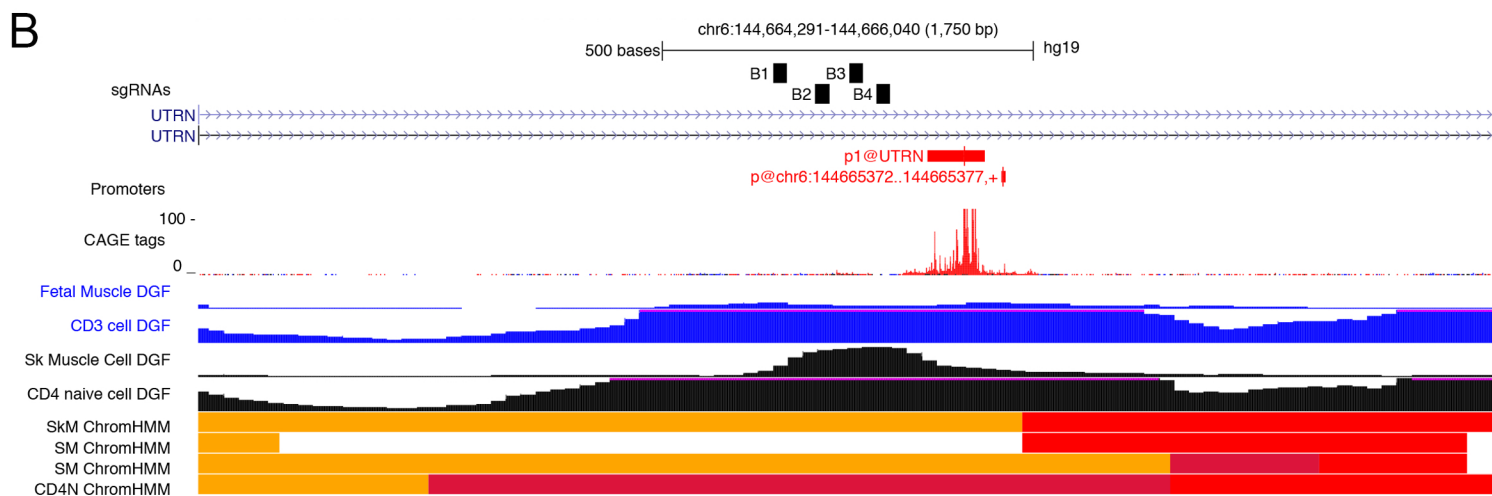
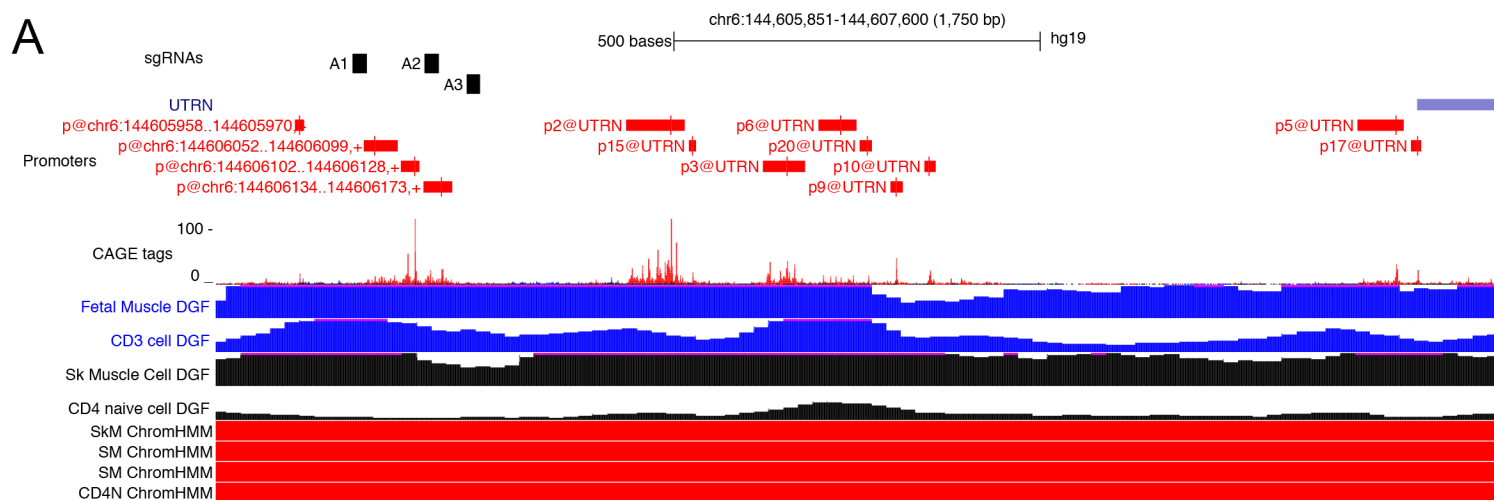


Figure S1

**Figure S1. Zoomed in location of sgRNAs relative to *UTRN* transcription start sites (TSSs), DNase I hypersensitivity foot prints and chromatin state maps.**

Zoomed in view of Figure 1D. (A) sgRNAs targeting promoter A are plotted above experimentally determined TSSs. (B) sgRNAs targeting promoter B are plotted above experimentally determined TSSs. These have been obtained from over 300 primary tissues assay by FANTOM5. The maximum signal at each promoter region is shown below the TSSs. Digital DNase Footprinting (DGF) assays for Fetal Muscle and primary CD3 cells are shown in blue. DGF assays for skeletal muscle cells, skeletal muscle and naïve CD4 cells are shown in black. Chromatin State maps from the Roadmap Epigenomic Consortium are shown for skeletal muscle cells (SkM), skeletal muscle (SM) and naïve CD4 cells (CD4N). Red chromatin state indicates transcription start sites and yellow indicates enhancer states. The A guides all fall within muscle promoter regions. The B guides fall into an enhancer region immediately upstream of an annotated promoter region. In CD4 cells this region is considered an active promoter. At promoter B region there is a weak DGF footprint in muscle cells relative to CD4 cells. Data was plotted using the UCSC genome browser. Fantom5, DGF and chromatin state data was obtained via UCSC “Track Hubs”.

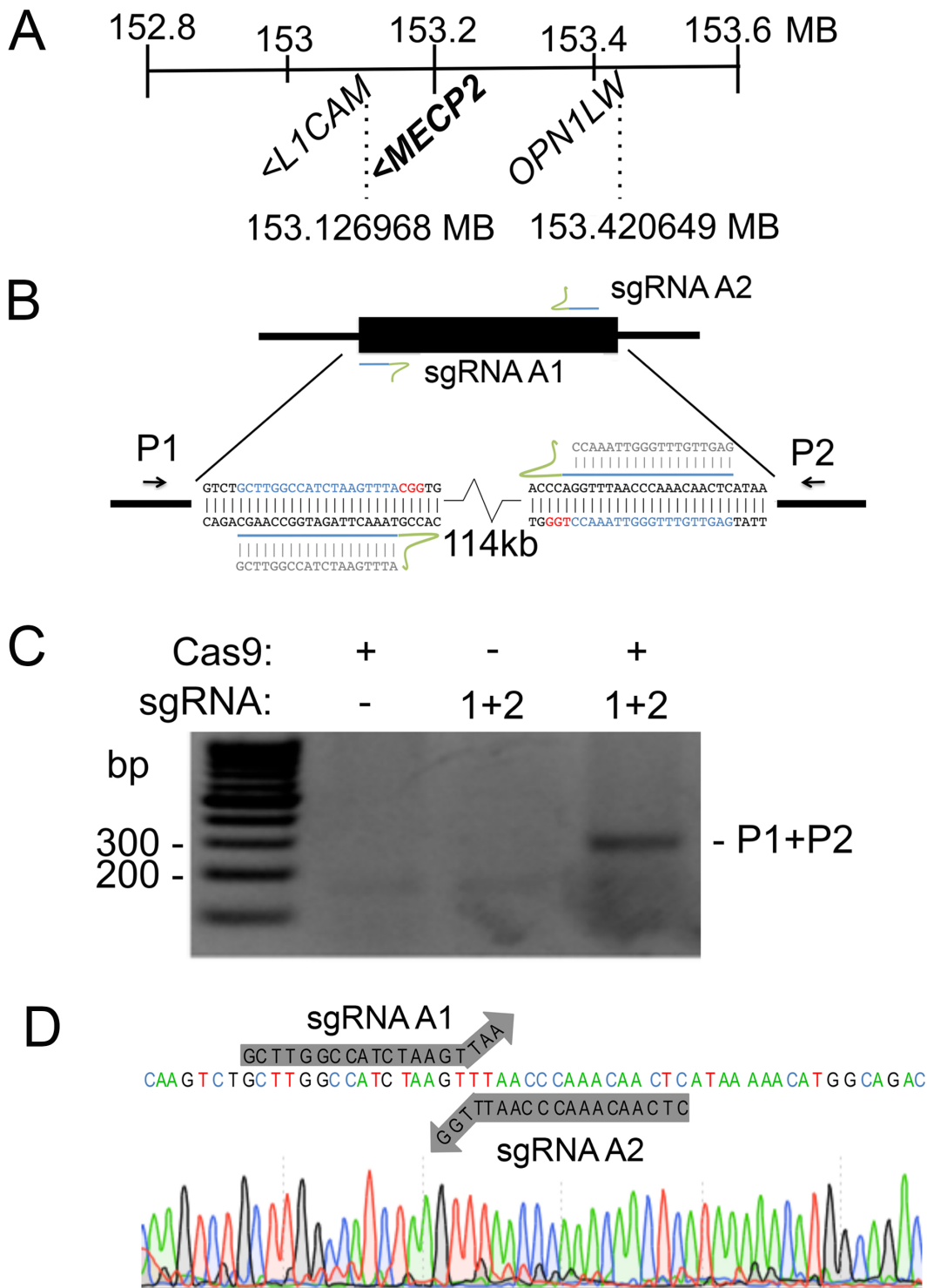


Figure S2

**Figure S2. Targeted deletion of a 114 kb fragment of DNA containing *MECP2* gene in WT genomic DNA.**

(A) Schematic depicting the q28 region (h19) of the X-chromosome which has been duplicated in this *MECP2* duplication syndrome-affected individual. The breakpoints of the duplication, as indicated by the dotted lines, fall within the *L1CAM* and *OPN1LW* genes and include intervening genes one of which is *MECP2*. (B) Relative position of *MECP2*: sgRNAs 1 and 2 to WT genomic DNA. (C) Presence of targeted deletion assessed using PCR with primers P1 and P2 in human primary fibroblasts nucleofected with corresponding CRISPR components. Intervening fragment between *MECP2*: sgRNAs 1 and 2 can only be amplified if 114 kb fragment is deleted. Band corresponding to a deletion is only detected in cells nucleofected with the 2 sgRNAs coupled with Cas9. (D) Sequencing read of deletion in human primary fibroblasts with annotated positions of the 2 sgRNAs used.

A

**MECP2:sgRNA A1**

Name	Result	Chr Position	Score	PCR Primer left	PCR Primer right	Amplicon Size
C-OFF 1	TTCTTGACACATCTAAGTTTATGG -- hit TGCTTGGCCATCTAAGTTTANRG -- query	Chr3:157313782-157313804	0.63	AGACCTAGGCTTCAGTTCACCTGT	CGTCTAATGCCACTGGCTGC	450
C-OFF 2	TGCTTTATCATCTAAGTTTATAG -- hit TGCTTGGCCATCTAAGTTTANRG -- query	Chr5:74363041-74363063	0.71	GTGGAGCCTGAAGATGTGACTG	GAGCCTTCCGCAGGTTGTAATC	454
C-OFF 3	TGCTTTATCATCTAAGTTTATGG -- hit TGCTTGGCCATCTAAGTTTANRG -- query	ChrX:94926307-94926329	0.71	GCTGAAATTTGGAGTGGCTATGGC	CACACATGATTATCGGGAGGAGG	524
C-OFF 4	TGCTTTATCATCTAAGTTTATGG -- hit TGCTTGGCCATCTAAGTTTANRG -- query	Chr1:216962432-216962454	0.71	GAGTTGCTTCTTACGGATGAGTGG	GCTGTGGTCTATTAAGTGTGCTGC	489
C-OFF 5	TGTTTGGAGATCTAAGTTTAGGG -- hit TGCTTGGCCATCTAAGTTTANRG -- query	Chr3:124377831-124377853	0.77	CACTGGCTTCTGCCTTCT	GCCCAGTCTGGGTTCTGATTG	450

**MECP2:sgRNA A2**

Name	Result	Chr Position	Score	PCR Primer left	PCR Primer right	Amplicon Size
C-OFF 6	TGATTTGTGTGTGTTAAACCAAG -- hit TGAGTTGTTTGGGTAAACCNRG -- query	ChrX:119997612-119997634	1.32	CCTAACTCATTCTATGAGGCCAGC	TAGGAGTGGTGAGAGAGGGAG	472
C-OFF 7	TGGTTTTGGGGTTAAACCCAG -- hit TGAGTTTTGGGTAAACCNRG -- query	Chr3:58164792-58164813	1.39	AGAAGGTGTGAGGATCCAGGTG	GTTTACAAGCACCCGTGAGG	450
C-OFF 8	AGAATTGTTTGGGGTAAACCAAG -- hit TGAGTTGTTTGGGTAAACCNRG -- query	Chr8:102985829-102985851	1.59	GGACTTGAACCTATCCTAGGTGC	CTGCTAAGGCATATCATAGGACGG	450
C-OFF 9	TGAATTGTTTGGGTAAACCCAGG -- hit TGAGTTGTTTGGGTAAACCNRG -- query	Chr22:42636319-42636340	1.98	TGACGTCACATTGCAACAAGGGG	TTCTGGGCCTCCACTGGCAA	450
C-OFF 10	TGAGTTTTAGAGTTAAACCCAGG -- hit TGAGTTTTGGGTAAACCNRG -- query	Chr3:141297296-141297317	2.04	GGTGCTGCCACTGCAACATG	CCTCTAGCAACCTGGTAAGGTA	450

**DMD:sgRNA 1**

Name	Result	Chr Position	Score	PCR Primer left	PCR Primer right	Amplicon Size
C-OFF 11	AATATTTATTAACACCCGAAG -- hit AATATTTCTTAAATACCCGNRG -- query	Chr5:98629458-98629479	2.97	GCCTACCTCAAATGAGAGCC	GCTGTCTATTTCCAGCACC	535
C-OFF 12	AATATCATCTCAAATCCCGTGG -- hit AATATCTTCTTAAATCCCGNRG -- query	Chr7:37522116-37522137	3.74	GGAGTGTCAGCCACAAATGCTTC	GGGTTGGGTTGTGTTTCATCAG	457

B

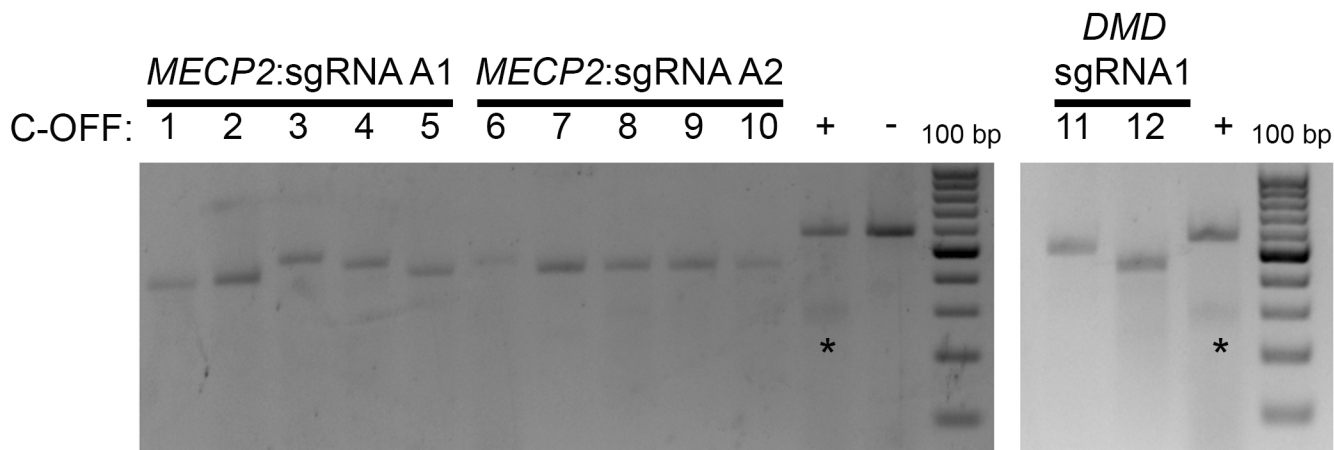


Figure S3

**Figure S3. Analysis of top 12 off-target hits predicted by the COSMID Tool.**

(A) Table identifying the sequence, score, position and primers used to amplify each off-target loci for *MECP2*: sgRNAs 1 and 2 as well as *DMD*: sgRNA 1. (B) Indel detection for predicted off-target sites shows no significant indel activity at all of the loci.

**Table S1: sgRNAs sequence identities and specificity scores**

Disorder	sgRNA ID	Strand	Sequence	PAM	Specificity Score CRISPR MIT Design Tool
DMD	A1	-	GCGCCCGTCAATCAGCGC	CGG	Not applicable
DMD	A2	+	GATCAGCCCCACTACGTTCC	CGG	Not applicable
DMD	A3	-	GAGAGCGCCGAGGGGAGC	CGG	Not applicable
DMD	B1	+	GTCATAGGAACATGAATAG	AGG	Not applicable
DMD	B2	-	GTCAGTAAAACTCCTTAGGC	AGG	Not applicable
DMD	B3	+	GTTTAGAGGAGGTGGGGTT	AGG	Not applicable
DMD	B4	-	GCTTTATTTTTCCCATGAG	AGG	Not applicable
DMD - Duplication	<i>DMD</i> : sgRNA 1	-	AATATCTTCTTAATACCCG	AGG	75.25
MECP2- Duplication	<i>MECP2</i> : sgRNA A1	+	GCTTGGCCATCTAAGTTTA	CGG	74.43
MECP2- Duplication	<i>MECP2</i> : sgRNA A2	-	GAGTTGTTTGGGTTAAACC	TGG	71.85
Achondroplasia	<i>FGFR3</i> : sgRNA 1	+	GCATCCTCAGCTACAGGGT	GGG	58
Achondroplasia	<i>FGFR3</i> : sgRNA 2	+	GCAGGCATCCTCAGCTAC	AGG/GGG	63

Table S2: Off target analysis performed based on CRISPR Design Tools algorithm

MECP2: sgRNA A1

Targets	Sequence	CRISPR Design Tools Score <sup>1</sup>	Cosmid Score <sup>2</sup>	Locus (Gene ID) <sup>3</sup>	Indels (%)
ON	GCTTGGCCATCTAAGTTTACGG	100	0	chrX:+153273069	N/D
OFF 1	AGGTGGCCATCTAAGTTTATAG	1.7	1.21	chr8:+119018322	N/D
OFF 2	AAGTTGGCCCTCTAAGTTTAAAG	1.3	1.28	chr18:-70219106	N/D
OFF 3	TGAATGTTTCATCTAAGTTTAAAG	0.9	Not identified	chr12:-55085547	N/D
OFF 4	TTCTTGAGCGTCTAAGTTTAAAG	0.9	Not identified	chr11:-125813383	N/D
OFF 5	TGCTAGCCCTCTAAGTTTAAAG	0.9	Not identified	chr9:-73214989	N/D
OFF 6	TGCTTGACCTCTAAGTTTAAAG	0.8	1.43	chrX:-139329098	N/D
OFF 7	CATTTTGCATCTAAGTTTATGG	0.8	Not identified	chr5:-53752882	0.86
OFF 8	TACACTGCCATCTAAGTTTAAAG	0.8	Not identified	chr17:+35634370	N/D
OFF 9	TTCTTGGAAAGTCTAAGTTTAAAG	0.8	Not identified	chr6:-91201638	N/D
OFF 10	TGATGTGCTTCTAAGTTTAAAG	0.8	Not identified	chrX:-43157660	N/D
OFF 11	AACATGGCCATCTAAGTTTGGG	0.7	Not identified	chr4:-160123587	N/D
OFF 12	TCGTTGGACATTTAAGTTTAAAG	0.7	Not identified	chr12:+66851859	N/D
OFF 13	AGCTGGCCCTCTAAGTTTAAAG	0.7	Not identified	chr7:-16277369	N/D
OFF 14	TTCTTGGACTTCTAAGTTTGTGG	0.6	Not identified	chr7:-129847577	N/D
OFF 15	TCATGGACATCCAAAGTTTAAAG	0.6	Not identified	chr1:+72535577	N/D
OFF 16	TGTTTGGCCCTCTAATTTAAAG	0.5	2.95	chr14:-45589728	N/D
OFF 17	TGTTTGGCCCTCTAATTTAAAG	0.5	Not identified	chr2:+155525161	N/D
OFF 18	TTCTAGCCATCTAAGTTTCAAG	0.5	Not identified	chr3:-6566451	N/D
OFF 19	TGCAAGCCATCTAAGTTTAAAG	0.5	Not identified	chr5:-59096266	N/D
OFF 20	TTCTTGCCCTCTAAGTTTCTGG	0.4	Not identified	chr17:-75376272	N/D

MECP2: sgRNA A2

Targets	Sequence	CRISPR Design Tools Score <sup>1</sup>	Cosmid Score <sup>2</sup>	Locus (Gene ID) <sup>3</sup>	Indels (%)
ON	GAGTGTGTTGGGTTAAACCTGG	100	0	chrX:-153386812	N/D
OFF 1	TGTGTTGTTTGGGTGAAACCGAG	2.7	2.05	chr5:-20399210	N/D
OFF 2	TGAGTTTTTGGGTTAAGCCTGG	1.2	4.23	chr7:+32206618	N/D
OFF 3	ACTGTTCTTTGGGTTAAACCAAG	0.9	Not identified	chr3:-33704075	N/D
OFF 4	AGTTTTGTTTGGTTAAACCTAG	0.8	Not identified	chr16:-20201186	N/D
OFF 5	TGAGCTATTTGGTTAAACCAAG	0.7	1.52	chr22:+26856195	N/D
OFF 6	GCAGATGTTTGGATTAACCAAG	0.6	Not identified	chr8:-65798479	0.87
OFF 7	GAAGATGTTTGGGTTAAACCTAG	0.5	Not identified	chrX:-116193921	N/D
OFF 8	TGAATTCCTTTGGTTAAACCCAG	0.5	Not identified	chr18:-73217777	N/D
OFF 9	TGAGGTCTCAGGGTTAAACCAAG	0.5	Not identified	chr5:-2440574	N/D
OFF 10	TCATTTCTTTGGGTTAAACATAG	0.5	Not identified	chr8:-22631876	N/D
OFF 11	TGAGATGATGGGTTAAAGCTGG	0.5	Not identified	chr17:+74157644 (NM_052916)	N/D
OFF 12	TTATTGGTTTGGTTAAACCAAG	0.4	Not identified	chr5:+115222903	N/D
OFF 13	AGGGTTTTTGGCTAAACCTAG	0.4	Not identified	chr6:+151273474	N/D
OFF 14	TGGATTGTTTTGGTTAAACCTGG	0.4	Not identified	chr2:-55249220	N/D
OFF 15	TGACTAGTTGGGTTAAACATGG	0.4	Not identified	chr20:+55793327	N/D
OFF 16	TGAGTTGAAAGGGTTAAACACAG	0.4	Not identified	chr1:+180894617	N/D
OFF 17	TTTGTCTTTGGGTTAAAGCCAG	0.3	Not identified	chr18:+4272341	N/D
OFF 18	TGAGCCATATGGGTTAAACCAAG	0.3	Not identified	chr5:-58907249	1.00
OFF 19	TTAGTCATTTCCGTTAAACCAAG	0.3	Not identified	chr10:+15514525	N/D
OFF 20	AGAGGTGTGGGTTAAAGCAGG	0.3	Not identified	chr4:+11194960	N/D

DMD: sgRNA 1

Targets	Sequence	CRISPR Design Tools Score <sup>1</sup>	Cosmid Score <sup>2</sup>	Locus (Gene ID) <sup>3</sup>	Indels (%)
ON	ATATCTTCTTAAATACCCGAAGG	100	0	chrX:+32461612	N/D
OFF 1	AGTGTCTCTTAAATACCTGCAG	1.1	5.3	chr2:+225344609	N/D
OFF 2	AAAACCTTCAGAAATACCCGGAG	0.7	Not identified	chr6:+19965267	N/D
OFF 3	AATCTCTTCTCAATACCCCTGG	0.7	6.97	chr18:+23119458	N/D
OFF 4	AAGAGCTGCTTAAATACCCCTGAG	0.7	Not identified	chr11:+46836502	2.53
OFF 5	CATATCTCTTAAATAGCCCTGG	0.6	9.12	chr8:+67600685	N/D
OFF 6	ATTAGCATCTTAAATACCCGAAG	0.5	Not identified	chr7:+130907163	N/D
OFF 7	GATATACCTGAAATACCCGTAG	0.5	Not identified	chrX:-7222056	N/D
OFF 8	ATTTCTTATGAAATACCCGAAG	0.5	Not identified	chr14:-91038783	N/D
OFF 9	TAAATCCTCTTAAATACCCTAAG	0.5	6.89	chr1:-19924480	N/D
OFF 10	GAAATCTCATAAATACCCAGGAG	0.5	6.16	chr9:-34930580	N/D
OFF 11	AATTACTTCATAAATACCCGTAGG	0.5	Not identified	chr5:-170343578 (NM_022897)	N/D
OFF 12	AATTTCAACTTAAATACCCCTGG	0.5	Not identified	chr13:-109558634	1.26
OFF 13	AATTCATCTTAAATACCCCTAAG	0.5	Not identified	chr7:-10351725	N/D
OFF 14	CATCTTTCTTAAATACCCCAAGG	0.4	Not identified	chr2:+208081568	N/D
OFF 15	AGTTTCTTGTAAATACCCCAAGG	0.4	Not identified	chr6:+8264834	N/D
OFF 16	ATTCTCTTTTAAATACCCACAG	0.4	Not identified	chr5:+146530145	N/D
OFF 17	AATTTCTCTTAAATACCCCAAG	0.4	Not identified	chr19:-22631201	N/D
OFF 18	AATATTTCTTCAATACCCCTGG	0.4	7.01	chr5:-146140359	N/D
OFF 19	AATATTTCTTCAATACCCCTAAG	0.4	7.01	chr1:-185743949	N/D
OFF 20	AATACCTCATAAAGTACCCGAAG	0.4	1.99	chr13:+23073576	N/D

<sup>1</sup> CRISPR Design Tool: <http://crispr.mit.edu/>

<sup>2</sup> COSMID: <https://crispr.bme.gatech.edu/>

<sup>3</sup> aligned against hg19

N/D: Not determined

**Table S3: Targeted elimination of a dominant-negative, gain-of-function allele.**

	Cas9	<i>FGFR3</i> : sgRNA 1 (Allele-specific)
Sequences	%	%
CgGGGTGG	47.97	42.481
CaGGGTGG	36.74	26.514
CaG-GTGGG	0.537	10.851
CaGG-TCGG	0.333	2.364
CAGGTGGC	0.365	1.952
CCAGGGTG	0.95	1.51
CGGGTGGC	0.769	1.151
CTGGGTGG	1.196	0.893
CATGGGTG	0.649	0.87
CATGGGTC	0.61	0.577
CATGGTGG	0.145	0.565
CCAGGTGG	0.011	0.546
CGGGTCGG	0.534	0.465
CCAGGGTC	0.508	0.441
CCGGGGTG	0.574	0.407
CAGGGGTG	0.383	0.392
ACAGGGTG	0.352	0.349
CATGGTCG	0.107	0.322
ACAGGTGG	0.102	0.276
CGGTGGGC	0.022	0.26
GGGTGGGC	0.45	0.236
AGGGTGGG	0.261	0.232
CCGGGTGG	0.157	0.227
CGGGGTGCG	0.441	0.219
ACAGGGTC	0.285	0.218
CCATGGGT	0.174	0.212
CGGTGGCT	0.026	0.163
CCAGGTGCG	0.013	0.16
CGGGGGTG	0.183	0.155
CTGGGTGCG	0.161	0.153
AGGGTGGC	0.11	0.143
CAGTGGGC	2E-04	0.133
ACAGGTGCG	0.057	0.132
AGGGTCGG	0.156	0.123
GGGGTGGG	0.379	0.122
TGGGGTGG	0.26	0.119
CAGGGCGG	0.178	0.112
AGGTGGGC	0.011	0.106
ACCAGGGT	0.104	0.105
CTCGGGGT	0.389	0.085
CTCGGGTG	0.175	0.049
GGGTGGGT	0.128	0.06
CTGGGGTC	0.122	0.052

	Cas9	<i>FGFR3</i> : sgRNA 2 (PAM-specific)
Sequences	%	%
AGCTACgG	38.667	49.113
AGCTACaG	54.883	32.805
A - - TACaGGG	0.001	9.602
ATACGGAG	0.00008	0.645
AGCTACGA	0.692	0.605
AGTACAGG	0.029	0.488
ACTACAGG	0.065	0.419
AGCTCACG	0.535	0.374
AGTACGGA	0.003	0.369
GAGCTACG	0.264	0.257
AGCTCACA	0.639	0.251
ACTACGGA	0.008	0.228
GCTACGGG	0.133	0.22
GCTACAGG	0.244	0.205
ATACGAGG	0.00001	0.182
AGTACGGT	0.004	0.16
GAGCTACA	0.333	0.159
AGTACCG	0.133	0.158
ATACGGTG	0.00008	0.158
AGCTGACG	0.111	0.156
AGCTGCGG	0.103	0.152
AGCTACCA	0.217	0.132
AGCCTACG	0.095	0.112
ACTACGGT	0.009	0.11
GGCTACGG	0.13	0.103
AGCTATGG	0.042	0.102
GGCTACAG	0.177	0.098
AGCTGCAG	0.151	0.081
AGCCACAG	0.135	0.073
AGCCTACA	0.118	0.061