KOnezumi documentation

Akihiro Kuno M.D., Ph.D.

Department of Anatomy and Embryology, University of Tsukuba

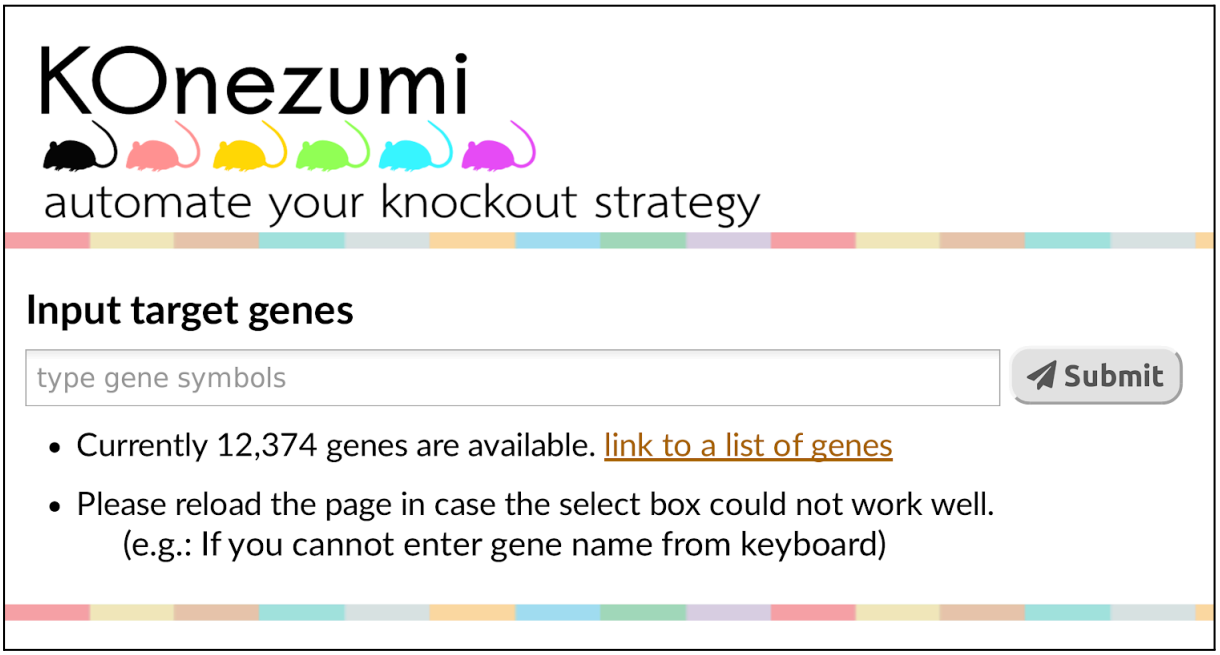
Contact: akuno [at] md.tsukuba.ac.jp

Last update: 2023-12-09

---

1. Input page

Availability of target genes



Currently, users can select from 12,374 genes. These genes have critical exons as defined by IMPC/IKMC as follows:

Critical exons are common to all transcript variants, and the deletion of these exons will disrupt more than fifty percent of the protein-coding sequence by frameshift mutation. (Skarnes et al. 2011)

Data of critical exons can be downloaded from FTP site at UCSC Genome Browser.

When you encounter an error with the selection box

You may experience input errors due to the response behavior of the selection box. If you cannot input gene names or remove them, please reload this page.

2. Overview of targeting gene

KOnezumi initially provides an overview of the target gene, including the locations of gRNAs and primers. The transcript variants of the target genes are based on the RefSeq database.

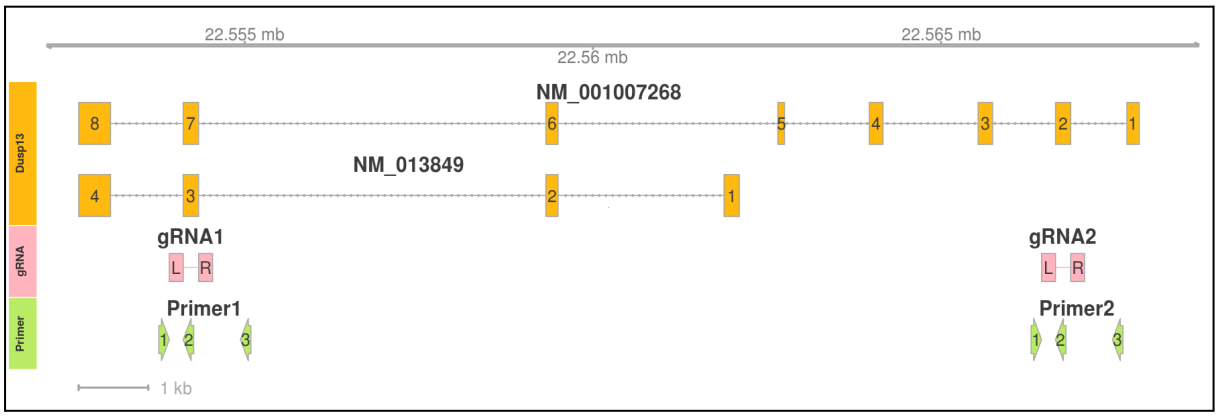
The protein-coding sequence (CDS) is colored in orange.

Untranslated regions (UTR) are not shown in this schema because several genes exhibit significant differences between their CDS and transcripts.

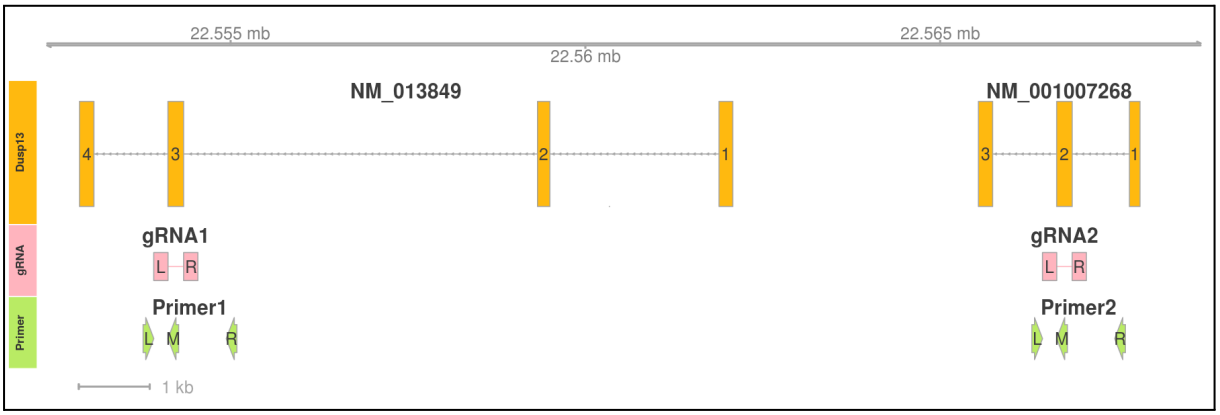
Example gene: Dusp13.

Dusp13 has two transcripts, NM_013849 and NM_001007268. At first glance, when viewing the "Schematic view of Dusp13 including UTR information", it appears that gRNA1 might be sufficient to disrupt this gene, as it targets common exons. However, the CDS information reveals that NM_013849 and NM_001007268 have completely different protein-coding exons. This indicates that both gRNA1 and gRNA2, or the complete deletion of its gene locus, are required to effectively disrupt Dusp13.

Schematic view of Dusp13 including UTR information

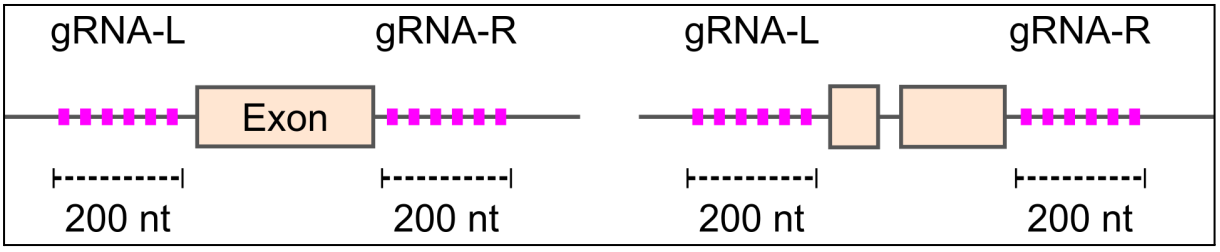


Schematic view of Dusp13 without UTR information



3. CRISPR gRNA candidates

KOnezumi generates candidate CRISPR gRNAs positioned within 200 base pairs (bp) of each flanking sequence of the target exons. These are designated as gRNA-L (left) and gRNA-R (right).



KOnezumi generates rational CRISPR gRNAs, prioritizing those with higher sequence specificity and predicted cutting efficiency. It is particularly recommended to use

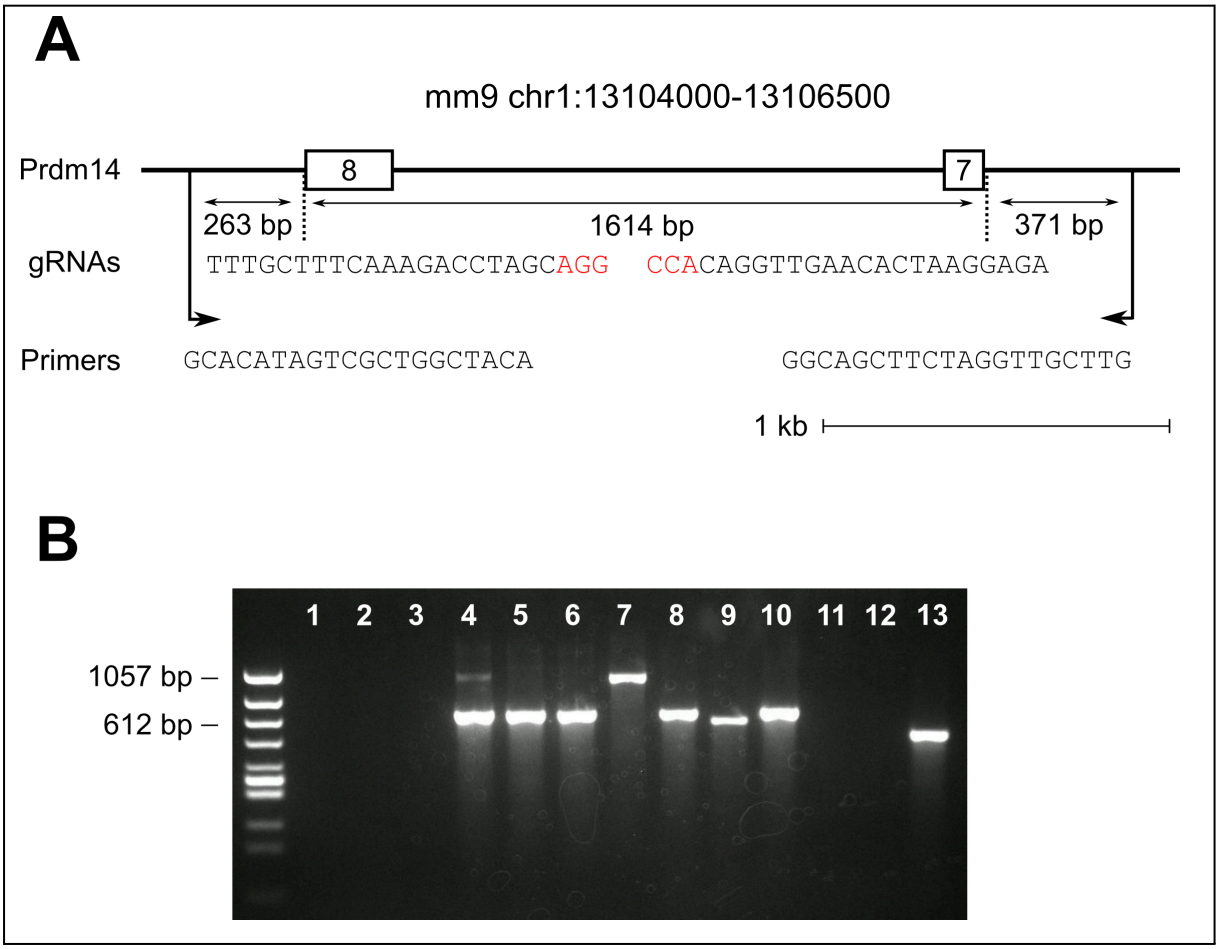
gRNAs highlighted in bold. These gRNAs boast **a MIT specificity score greater than 50 and a Doench/Fusi cutting efficiency score above 55**.

Guide sequences are linked to CRISPRdirect (Naito et al. 2015). Users can verify the actual number of off-target sites by visiting the website.

ID	chr	start	end	+/-	Guide sequence	PAM	MIT	Doench/Fusi	Moreno-Mateos
gRNA-L	chr18	12081723	12081746	+	AGCACCTCTAGCAACCCCCA	AGG	78	86%(62)	59%(49)
gRNA-L	chr18	12081728	12081751	+	CTCTAGCAACCCCCAAGGAG	AGG	70	84%(61)	52%(46)
gRNA-L	chr18	12081752	12081775	+	AAATGATCACCCCACCACAT	GGG	72	97%(69)	72%(56)
gRNA-L	chr18	12081761	12081784	-	ATAAAGTGACCCATGTGGTG	GGG	72	97%(69)	71%(55)
gRNA-L	chr18	12081762	12081785	-	GATAAAGTGACCCATGTGGT	GGG	63	94%(66)	39%(40)
gRNA-L	chr18	12081763	12081786	-	TGATAAAGTGACCCATGTGG	TGG	67	95%(67)	79%(60)
gRNA-L	chr18	12081766	12081789	-	CACTGATAAAGTGACCCATG	TGG	73	100%(82)	74%(57)
gRNA-R	chr18	12083480	12083490	+	AGCCTTGATGCAGGTAACTG	TGG	73	81%(60)	25%(33)
gRNA-R	chr18	12083480	12083492	-	AGCCACAGTTACCTGCATCA	AGG	70	75%(58)	9%(21)
gRNA-R	chr18	12083480	12083497	+	ATGCAGGTAACTGTGGCTCG	AGG	80	69%(56)	80%(61)
gRNA-R	chr18	12083480	12083499	+	GCAGGTAACTGTGGCTCGAG	GGG	72	78%(59)	95%(77)
gRNA-R	chr18	12083517	12083540	+	GGGAGTTGCCTTGCTTTACA	GGG	65	75%(58)	88%(67)
gRNA-R	chr18	12083525	12083548	-	CCCATCTGCCCTGTAAAGCA	AGG	65	78%(59)	31%(36)
gRNA-R	chr18	12083599	12083622	+	GGGCATATCTCAGGCCACAC	AGG	63	86%(62)	89%(68)

The efficiency of removing target exon(s) by two gRNAs

The efficiency of excising target exon(s) using two gRNAs is of critical importance. To validate this, we introduced two gRNAs along with the Cas9 protein into fertilized eggs via electroporation and subsequently performed genotyping PCR on the blastocysts. The results showed that 8 out of 13 blastocysts (61.5%) exhibited the deletion band. Notably, lane No.7 displayed a longer deletion band, leading us to consider **7 out of 13 blastocysts (53.9%)** as practically viable for generating knockout mice.

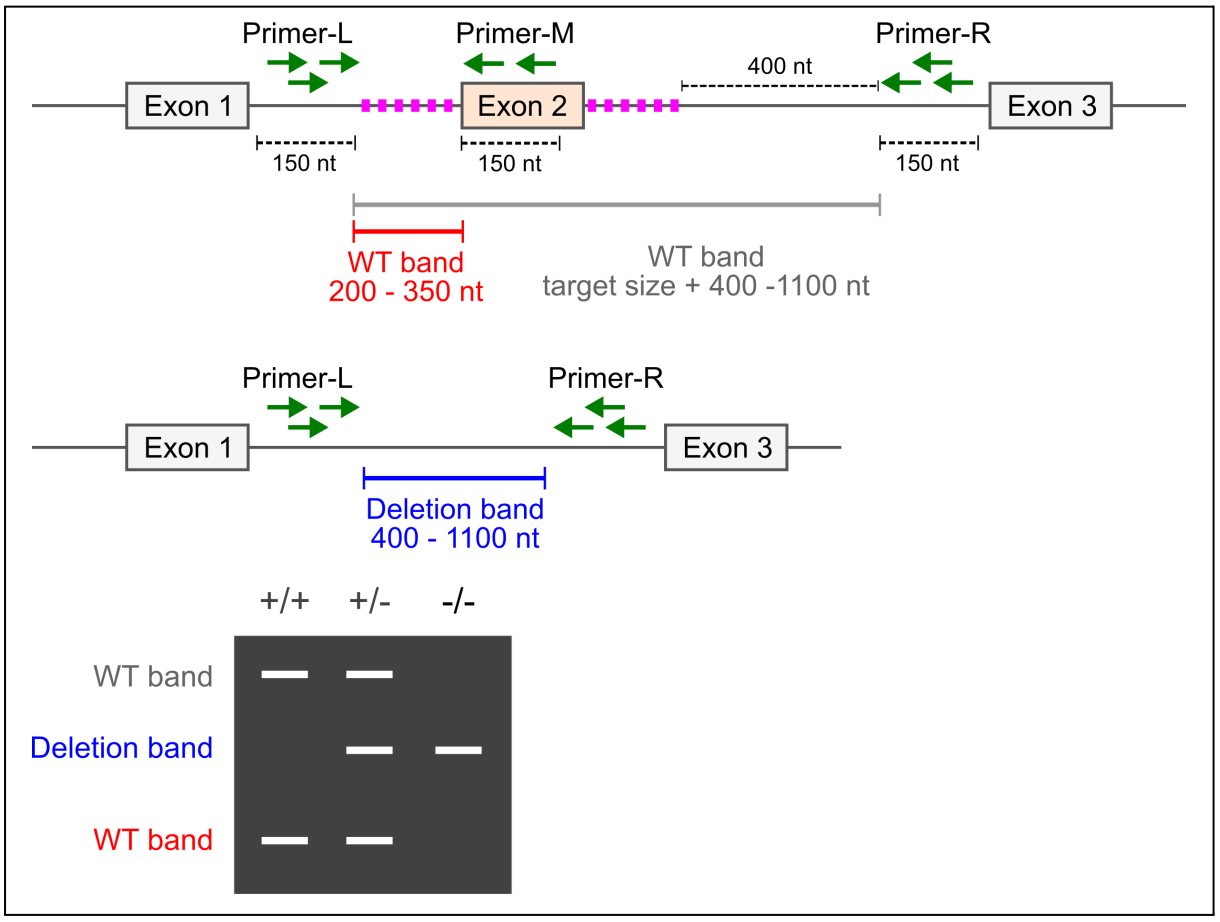


(A) Schema of location of gRNAs and primers. The boxes and numbers at the *Prdm14* gene locus mean targeting coding exons and exon numbers, respectively. Dot lines denote the location of gRNAs. Red colour nucleotides indicate PAM. The deletion length will be 1614 bp. Arrow lines indicate the location of PCR primers. The product size of the deletion band will be 634 bp (263 bp + 371 bp). **(B)** Genotyping of the blastocysts. 8 samples had the deletion bands (approximately 600 bp to 1100 bp).

4. Genotyping PCR primers

Genotyping PCR primers are positioned in the left genomic region of the gRNA-L cutting site (Primer-L; left), within the region targeted for deletion (Primer-M; middle), and in the right genomic region of the gRNA-R cutting site (Primer-R; right).

Gel electrophoresis will reveal a larger deletion band using Primer-L and Primer-R compared to the wild-type band produced by Primer-L and Primer-M.



Another wild-type band by Primer-L and Primer-R will be displayed when the size of target exons are very small (some hundreds of nucleotides). But the band will be larger than the deletion band so users can easily recognize the deletion band.

To design PCR primers, we used Primer3 version 2.4.0 (Untergasser et al. 2012; Koressaar and Remm 2007) with default parameters. Then, we utilized Bowtie version 1.2.2 (Langmead et al. 2009) to select highly specific PCR primers. We mapped each primer candidate to mm9 genome by Bowtie allowing 0, 1, or 2 mismatches or gaps in the alignment. Next, we counted the numbers of mapped sequences to score their sequence specificity. It is recommended to use primers with higher scores. Especially using primers of Score 3 will avoid nonspecific amplification by PCR.

Guide sequences are linked to GGGenome. Users can check actual numbers of

off-target sites using the website.

ID	Score	Penalty	Primer sequence	TM	GC-percent	Self-any	Self-end	Hairpin-TH	End-stability
Primer-L	3	0.042581	TCCCCAGCACTCAGGACATA	59.957	55.000	0.00	0.00	0.00	2.2900
Primer-L	3	0.323451	GGTAACCAAGGGCTTCCAGG	60.323	60.000	0.00	0.00	35.09	4.4500
Primer-L	3	0.332972	AACCAAGGGCTTCCAGGATC	59.667	55.000	0.00	0.00	33.73	3.3600
Primer-L	3	0.394899	CAAGGGCTTCCAGGATCCAC	60.395	60.000	15.82	0.00	0.00	4.0200
Primer-L	3	0.404040	CTGGTAACCAAGGGCTTCCA	59.596	55.000	0.00	0.00	30.80	3.5300
Primer-M	3	0.106967	GCCTCCCTCCATCCAACTTC	60.107	60.000	0.00	0.00	0.00	3.0100
Primer-M	3	0.178990	AGTTCCACACCTTCAAGGCC	60.179	55.000	2.20	0.00	39.63	5.1900
Primer-M	3	0.182537	CACCCAGTTCCACACCTTCA	59.817	55.000	0.00	0.00	0.00	3.0200
Primer-M	3	0.255033	AACCTTACCTTCCCATCGGC	59.745	55.000	0.00	0.00	0.00	5.5400
Primer-M	3	0.260230	TTCCACACCTTCAAGGCCAA	59.740	50.000	5.01	0.00	39.63	4.5200
Primer-R	0	0.034059	GGTGTGTGTGGAGACCAGAG	59.966	60.000	0.00	0.00	32.34	3.3500
Primer-R	0	0.034059	GTGTGTGTGGAGACCAGAGG	59.966	60.000	0.00	0.00	32.34	3.6900
Primer-R	0	0.250038	CAGGTGTGTGTGGAGACCAG	60.250	60.000	0.00	0.00	32.34	4.0000
Primer-R	2	1.783071	TGTACAGGTGTGTGTGGAGA	58.217	50.000	0.00	0.00	37.52	3.7100
Primer-R	3	1.411246	TGTACAGGTGTGTGTGGAGAC	59.589	52.381	0.00	0.00	37.52	3.3600

5. Nucleotide sequences of target exons

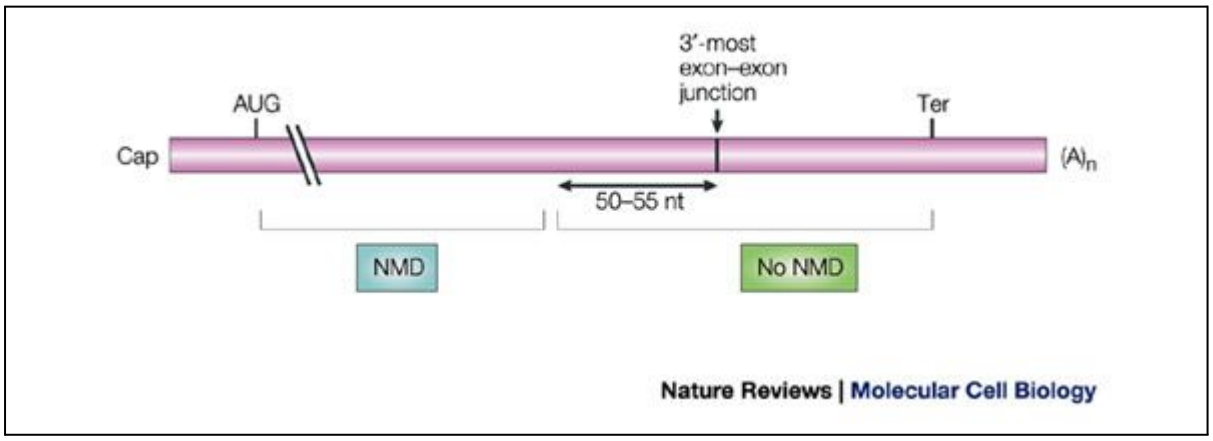
KOnezumi outputs nucleotide sequences of target exons. The list of sequences of the target exons can be applicable to confirm removal by DNA sequencing.

```
>Cables1|NM_001146287|gRNA target exon sequence|chr18:12081868-12081965
GGACCTGAAGTTGGATGGAGGGAGGCAGTCAGCAGGGGCCATGAGCCTGAAGGAGATCATTGGCCTTGAAGGTGTGGAAC
TGGGTGCCGATGGGAAG

>Cables1|NM_001146287|gRNA target exon sequence|chr18:12083323-12083480
ACAGTGTCTTATACCCAGTTTCTACTACCCACAAATGCCTTTGGGAACCGTAGAAATACCATAGACTCCACCGCCTCCTT
CTCCCAGTTTCGGAGCCTGAGCCACCGAAGCCTCTCCATGGGCCGGGCTGGCAGCACCCAGGGGAGCCTTGATGCAG

>Cables1|NM_022021|gRNA target exon sequence|chr18:12081868-12081965
GGACCTGAAGTTGGATGGAGGGAGGCAGTCAGCAGGGGCCATGAGCCTGAAGGAGATCATTGGCCTTGAAGGTGTGGAAC
TGGGTGCCGATGGGAAG

>Cables1|NM_022021|gRNA target exon sequence|chr18:12083323-12083480
ACAGTGTCTTATACCCAGTTTCTACTACCCACAAATGCCTTTGGGAACCGTAGAAATACCATAGACTCCACCGCCTCCTT
CTCCCAGTTTCGGAGCCTGAGCCACCGAAGCCTCTCCATGGGCCGGGCTGGCAGCACCCAGGGGAGCCTTGATGCAG
```

6. Protein-coding sequences (CDSs) of deletion transcripts

The position of a Premature Termination Codon (PTC) induced by frameshift mutation serves as a crucial indicator for determining whether aberrant transcripts will undergo degradation. If the PTC is situated over 50-55 nucleotides upstream of the final

exon-exon junction, these transcripts are subject to degradation through Nonsense-Mediated mRNA Decay (NMD) (Maquat, 2004).



When transcripts with deletions acquire a Premature Termination Codon (PTC), KOnezumi identifies the Coding DNA Sequences (CDSs) of these deletion products, highlighting the locations of frameshift nucleotides, the PTC, and the last exon-exon junction with colored annotations. The distance of the PTC from the last exon junction is displayed in the first row of the FASTA format output.

Utilizing this data, users can predict whether the deleted transcripts will be translated into proteins or targeted for degradation via Nonsense-Mediated mRNA Decay (NMD).

```
>Cables1|NM_001146287|gRNA-deleted CDS|PTC is 412 nt upstream of the last exon junction
ATGGCGGCGGCGACCGCCACCGCGGGCACGGCCGCCTGCAGCAGCAGCAGCAGCAGTCGCGGCGGCAGCACGGACGCCGC
GGCCACCAGCGGGGTGCAGCCGCCGCCGCCTCCTCCGGCCACCGCGCCGCCGGAACCTCTTCGTAAGCCGCGCATGGACC
CGCGGCGCCGCCAGGCTGCCCTGTCCTTCCTCACCAACATTTCTCTGGACGGACGGCCTCCGCTGCAGGACCACGAGTGG
GGCGGCGGCGAGGAGGGCGGCGGGACCAAGCCCGGAGCAAGGGCTCGGCTCAGCCTGCTGGCCGCTGGCTGCAACGCGTT
CTCGGCACCGGGCACCGCGGCCGCGCCGTGGACCGCCGGCTCGGGTTCCAGCCCTTGCCCGCTGCCGCCGTCTCTGGTCC
CCCGGGTGTTGGGAGAGCCATCTCAGCCGCCCCGTTCAGCTCCAGCTGTCACCGGTGCTCAGCTGCAGTTGCCTGACGGG
CCGGGAGGCGCCGGACAGGAGGAGCTGGAGGAGGACGATGCCTTCACCAACGTGCAGGTGCCATCGGCCAGCTTTTTGGG
CTCCGGGACCCCCGGGAGTACGAGCGGCAGTCGGGGCCGCCTCAACTCGTTCACTCAGGGAATCCTGCCCATCGCCTTCT
CCAGGCAGAACTCGCAGAACTACTGCGCCTTGGAGCAGTCAGGCCAGGGCGGCAGCACCAGCGCCTTGGAGCAGCTGCAG
AGGTCCAGACGTCGCCTCATCTCCCAGAGATCATCTTTGGAGACCCTGGAAGATATTGAAGAGAATGCCCCTCTCCGGAG
ATGTCGAACACTCTCAGGTTCTCCCAGACCAAAGAATTTCAAGAAAATTCATTTTATCAAGAATATGCGACAGCACGATA
CCAAGAATGGCAGAATAGTCCTTATCAGTGGCAGAAGATCCTTCTGTAGTATATTTTCCGTGCTGCCGTATCGAGACAGT
ACCCAAGTCGGGAAGTGACCTGGGAGACTTTATGGACTATGACCCAAACCTCCTGGATGACCCCCAGTGGCCTTGCGGCA
AGCACAAGCGAGTTCTGACCTTCCCTTCATACATGACCACAGTGATTGACTACGTGAAGCCCTCAGATCTCAAGAAGGAC
ATGAATGAAACCTTCAAGGAGAAGTTTCCTCACATTAAATTAACACTCAGCAAAATCAGGAGCCTGAAAAGAGAGATGCG
GAAGCTTGCCCAGGAGGACTGTGGCTTTGAGGAGCCCACGGTGGCCATGGCCTTTGTCTACTTTGAGAAGCTGGCCCTCA
GGGGGAAGCTGAACAAGCAGAACCGGAAGCTCTGTGCGGGAGCCTGTGTGCTGCTAGCGGCCAAGGTCGGAAGTGACCTC
AGAAAGCACGAGGTCAAGCACCTGATTGACAAACTGGAAGAGAAGTTCCGACTGAACAGACGAGAACTGATTGCCTTTGA
ATTCCCAGTGTTAGTGGCCTTGGAGTTTGCACTTCACCTGCCAGAGCACGAAGTCATGCCCCACTATAGACGCCTGATCC
AGAGCTCCTAG
```

References

Koressaar, Triinu, and Maido Remm. 2007. “Enhancements and Modifications of Primer Design Program Primer3.” *Bioinformatics* 23 (10): 1289–91.

Langmead, Ben, Cole Trapnell, Mihai Pop, and Steven L. Salzberg. 2009. “Ultrafast and Memory-Efficient Alignment of Short DNA Sequences to the Human Genome.” *Genome Biology* 10 (3): R25.

Maquat, Lynne E. 2004. “Nonsense-Mediated mRNA Decay: Splicing, Translation and mRNP Dynamics.” *Nature Reviews. Molecular Cell Biology* 5 (2): 89–99.

Naito, Yuki, Kimihiro Hino, Hidemasa Bono, and Kumiko Ui-Tei. 2015. “CRISPRdirect: Software for Designing CRISPR/Cas Guide RNA with Reduced off-Target Sites.” *Bioinformatics* 31 (7): 1120–23.

Skarnes, William C., Barry Rosen, Anthony P. West, Manousos Koutsourakis, Wendy Bushell, Vivek Iyer, Alejandro O. Mujica, et al. 2011. “A Conditional Knockout Resource for the Genome-Wide Study of Mouse Gene Function.” *Nature* 474 (7351): 337–42.

Untergasser, Andreas, Ioana Cutcutache, Triinu Koressaar, Jian Ye, Brant C. Faircloth, Maido Remm, and Steven G. Rozen. 2012. “Primer3--New Capabilities and Interfaces.” *Nucleic Acids Research* 40 (15): e115.