eurofins | Genomics

Data Analysis Report

Project / Study: GEN190101_A

Project specification: CRISPR_NHEJ_Analysis_Example

Date: May 3, 2019

Table of Contents

1 Analysis of CRISPR/Cas9 gene editing events

The CRISPR/Cas9 system has been widely adopted for targeted genome editing in a wide variety of cells. The Cas9 nuclease can be targeted using a single-guide RNA (sgRNA) to create double-strand breaks (DSBs) at specific sites in the chromosomal DNA. Hereby, the presence of a protospacer adjacent motif (PAM) is mandatory to induce a DSB. The DSB is induced 3-4bp upstream of the PAM. Two major pathways are available that repair DSBs: the Homology-Directed Repair (HDR) pathway repairs DSBs in presence of a homologous donor sequence, and the the Non-Homologous End Joining (NHEJ) pathway that joins DNA ends without homology requirements. A repair by the NHEJ pathway is error-prone and often results in small indels (1-10bp) which can disrupt the targeted locus.

1.1 Project Details

Assay ID: Example

List of target regions:

- Target_1
- Target_2

Table 1: List of **edited samples**:

Sample name	Replicate name	Associated control sample
Edited_Sample_1	Edited_Sample_1	Control_Sample
Edited_Sample_2	Edited_Sample_2	Control_Sample

Table 2: List of **wildtype samples**:

Sample name	Replicate name
Control_Sample	Control_Sample

1.2 Results

1.2.1 Editing efficiency

Table 3: **Editing efficiency per sample and target region.** The listed numbers are the summed relative frequencies of identified variants.

Sample	Target_1	Target_2
Edited_Sample_1	76.8%	86.5%
Edited_Sample_2	76.5%	86.4%

1.2.2 NHEJ-induced Variants

1.2.2.1 Edited_Sample_1

Associated control: Control_Sample

Table 4: **Edited alleles frequency table for sample Edited_Sample_1 region Target_1.** *Aligned Sequence*: The observed read sequence, aligned to the reference. *Reference Sequence*: The reference sequence. *%Reads*: The fraction of reads with same sequence. *#Reads*: The number of reads with the same sequence. **Please note:** All alleles with variants that are close to the expected cut-site are listed here. However, only the 30 most frequent alleles are listed here. The full list of alleles is delivered as a separate file.

Target:	*Target_1*		
Aligned Sequence	**Reference Sequence**	**%Reads**	**#Reads**
AATTTTTTTTATA------GCCTTGTTCCGATTCAGTCAT	AATTTTTTTTATAGCCTTTGCCTTGTTCCGATTCAGTCAT	15.0%	833
AATTTTTTTTATAGCCT-----TTGTTCCGATTCAGTCAT	AATTTTTTTTATAGCCTTTGCCTTGTTCCGATTCAGTCAT	6.2%	345
AATTTTTTTTATAGCCTTT-CCTTGTTCCGATTCAGTCAT	AATTTTTTTTATAGCCTTTGCCTTGTTCCGATTCAGTCAT	3.8%	209
AATTTTTTTTATAGCCTTTGACCTTGTTCCGATTCAGTCA	AATTTTTTTTATAGCCTTTG-CCTTGTTCCGATTCAGTCA	2.8%	153
TTTTTTTTATAGCCTTTGGGCCTTGTTCCGATTCAGTCAT	TTTTTTTTATAGCCTTT--GCCTTGTTCCGATTCAGTCAT	2.4%	133
AATTTTTTTTATAGCCTTTG-CTTGTTCCGATTCAGTCAT	AATTTTTTTTATAGCCTTTGCCTTGTTCCGATTCAGTCAT	2.3%	127
AATTTTTTTTATAGCCTTTGCCCTTGTTCCGATTCAGTCA	AATTTTTTTTATAGCCTTTG-CCTTGTTCCGATTCAGTCA	1.8%	100
AATTTTTTTTATAGCCTTTGGACCTTGTTCCGATTCAGTC	AATTTTTTTTATAGCCTTTG--CCTTGTTCCGATTCAGTC	1.2%	69
AATTTTTTTTATAGCCTTTG--TTGTTCCGATTCAGTCAT	AATTTTTTTTATAGCCTTTGCCTTGTTCCGATTCAGTCAT	1.2%	66
AATTTTTTTTATAGCCTTTGTCCTTGTTCCGATTCAGTCA	AATTTTTTTTATAGCCTTTG-CCTTGTTCCGATTCAGTCA	1.0%	54
ATTTTTTTTATAGCCTTTGGCCTTGTTCCGATTCAGTCAT	ATTTTTTTTATAGCCTTT-GCCTTGTTCCGATTCAGTCAT	0.9%	49
----------	AATTTTTTTTATAGCCTTTGCCTTGTTCCGATTCAGTCAT	0.9%	48
TTTTTTTATAGCCTTTGAGGCCTTGTTCCGATTCAGTCAT	TTTTTTTATAGCCTTT---GCCTTGTTCCGATTCAGTCAT	0.7%	41
AATTTTTTTTATAGCCTTTGAACCTTGTTCCGATTCAGTC	AATTTTTTTTATAGCCTTTG--CCTTGTTCCGATTCAGTC	0.7%	39
AATTTTTTTTATAGCCTTTG---TGTTCCGATTCAGTCAT	AATTTTTTTTATAGCCTTTGCCTTGTTCCGATTCAGTCAT	0.7%	38
AATTTTTTTTATAGCCTTTGGGACCTTGTTCCGATTCAGT	AATTTTTTTTATAGCCTTTG---CCTTGTTCCGATTCAGT	0.6%	32
AATTTTTTTTATAGCCTT--CCTTGTTCCGATTCAGTCAT	AATTTTTTTTATAGCCTTTGCCTTGTTCCGATTCAGTCAT	0.6%	31
AATTTTTTTTAT----------	AATTTTTTTTATAGCCTTTGCCTTGTTCCGATTCAGTCAT	0.5%	30
TTTTTTTATAGCCTTTGGGGCCTTGTTCCGATTCAGTCAT	TTTTTTTATAGCCTTT---GCCTTGTTCCGATTCAGTCAT	0.5%	29
----------TGTTCCGATTCAGTCAT	AATTTTTTTTATAGCCTTTGCCTTGTTCCGATTCAGTCAT	0.5%	27
AATTTTTTTTATAGCCTTTGGAACCTTGTTCCGATTCAGT	AATTTTTTTTATAGCCTTTG---CCTTGTTCCGATTCAGT	0.5%	26
AATTTTTTTTATAGCCT---CCTTGTTCCGATTCAGTCAT	AATTTTTTTTATAGCCTTTGCCTTGTTCCGATTCAGTCAT	0.5%	26
AATTTTTTTTATAGCCTTTG--------CGATTCAGTCAT	AATTTTTTTTATAGCCTTTGCCTTGTTCCGATTCAGTCAT	0.5%	25
TTTTTTTTATAGCCTTTGAGCCTTGTTCCGATTCAGTCAT	TTTTTTTTATAGCCTTT--GCCTTGTTCCGATTCAGTCAT	0.4%	23
----------CCTTGTTCCGATTCAGTCAT	AATTTTTTTTATAGCCTTTGCCTTGTTCCGATTCAGTCAT	0.4%	21
AATTTTTTTTATAGCCTTTGTCCCTTGTTCCGATTCAGTC	AATTTTTTTTATAGCCTTTG--CCTTGTTCCGATTCAGTC	0.4%	20
AATTTTTTTTATAGC-----CCTTGTTCCGATTCAGTCAT	AATTTTTTTTATAGCCTTTGCCTTGTTCCGATTCAGTCAT	0.4%	20
AATTTTTTTTATAGCCTTTGGCCCTTGTTCCGATTCAGTC	AATTTTTTTTATAGCCTTTG--CCTTGTTCCGATTCAGTC	0.3%	19
AATTTTTTTTATAGCCTTTGACCCTTGTTCCGATTCAGTC	AATTTTTTTTATAGCCTTTG--CCTTGTTCCGATTCAGTC	0.3%	18
AATTTTTTTTATAGCCTTTGCCCCTTGTTCCGATTCAGTC	AATTTTTTTTATAGCCTTTG--CCTTGTTCCGATTCAGTC	0.3%	18
...	...	...	...
SUM		**76.8%**	**4256**

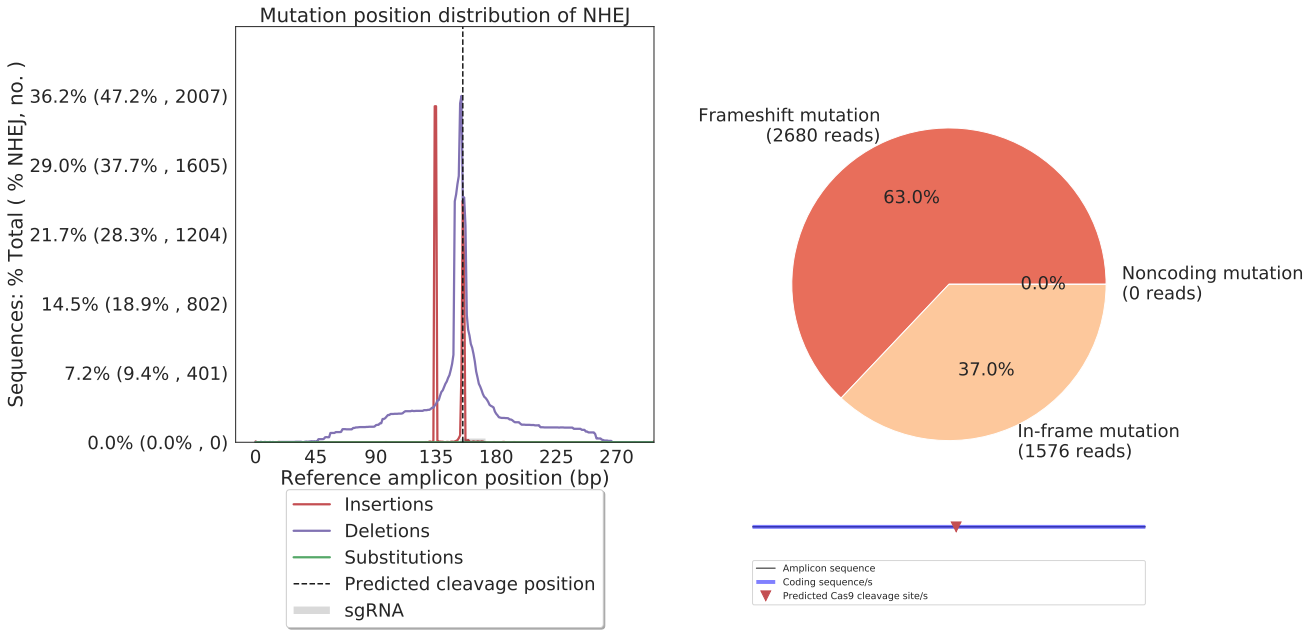

Figure 1: **Mutation rate and frameshift classification plots for sample Edited_Sample_1 region Target_1.** *Left:* The cumulative fraction of mutations at each position of the target region is shown.*Right:* The proportion of frameshift and in-frame mutations, as well as of mutations outside of coding region(s) is shown.

Figure 2: **Alleles around cutsite for sample Edited_Sample_1 region Target_1.** Please note that only alleles with a frequency of at least 0.2% and at most 50 alleles are shown.

Table 5: **Edited alleles frequency table for sample Edited_Sample_1 region Target_2.** *Aligned Sequence*: The observed read sequence, aligned to the reference. *Reference Sequence*: The reference sequence. *%Reads*: The fraction of reads with same sequence. *#Reads*: The number of reads with the same sequence. **Please note:** All alleles with variants that are close to the expected cut-site are listed here. However, only the 30 most frequent alleles are listed here. The full list of alleles is delivered as a separate file.

Target:	*Target_2*		
Aligned Sequence	**Reference Sequence**	**%Reads**	**#Reads**
AATTTTTTTTATA------GCCTTGTTCCGATTCAGTCAT	AATTTTTTTTATAGCCTTTGCCTTGTTCCGATTCAGTCAT	28.0%	3159
AATTTTTTTTATAGCCT-----TTGTTCCGATTCAGTCAT	AATTTTTTTTATAGCCTTTGCCTTGTTCCGATTCAGTCAT	11.0%	1243
AATTTTTTTTATAGCCTTT-CCTTGTTCCGATTCAGTCAT	AATTTTTTTTATAGCCTTTGCCTTGTTCCGATTCAGTCAT	2.4%	272
AATTTTTTTTATAGCCTTTG-CTTGTTCCGATTCAGTCAT	AATTTTTTTTATAGCCTTTGCCTTGTTCCGATTCAGTCAT	1.6%	185
TTTTTTTTATAGCCTTTGGGCCTTGTTCCGATTCAGTCAT	TTTTTTTTATAGCCTTT--GCCTTGTTCCGATTCAGTCAT	1.4%	163
AATTTTTTTTATAGCCTTTGACCTTGTTCCGATTCAGTCA	AATTTTTTTTATAGCCTTTG-CCTTGTTCCGATTCAGTCA	1.3%	149
AATTTTTTTTATAGCCTTTGCCCTTGTTCCGATTCAGTCA	AATTTTTTTTATAGCCTTTG-CCTTGTTCCGATTCAGTCA	1.3%	142
----------	AATTTTTTTTATAGCCTTTGCCTTGTTCCGATTCAGTCAT	1.1%	129
AATTTTTTTTATAGCCTTTG--TTGTTCCGATTCAGTCAT	AATTTTTTTTATAGCCTTTGCCTTGTTCCGATTCAGTCAT	0.9%	107
AATTTTTTTTATAGCCTTTGGACCTTGTTCCGATTCAGTC	AATTTTTTTTATAGCCTTTG--CCTTGTTCCGATTCAGTC	0.9%	106
AATTTTTTTTATAG----------	AATTTTTTTTATAGCCTTTGCCTTGTTCCGATTCAGTCAT	0.9%	104
AATTTTTTTTATAG----------CCGATTCAGTCAT	AATTTTTTTTATAGCCTTTGCCTTGTTCCGATTCAGTCAT	0.9%	98
ATTTTTTTTATAGCCTTTGGCCTTGTTCCGATTCAGTCAT	ATTTTTTTTATAGCCTTT-GCCTTGTTCCGATTCAGTCAT	0.8%	88
AATTTTTTTTATAGCCTTTGAACCTTGTTCCGATTCAGTC	AATTTTTTTTATAGCCTTTG--CCTTGTTCCGATTCAGTC	0.7%	82
----------CTTGTTCCGATTCAGTCAT	AATTTTTTTTATAGCCTTTGCCTTGTTCCGATTCAGTCAT	0.7%	75
----------TGTTCCGATTCAGTCAT	AATTTTTTTTATAGCCTTTGCCTTGTTCCGATTCAGTCAT	0.6%	70
AATTTTTTTTATAGCCTT--CCTTGTTCCGATTCAGTCAT	AATTTTTTTTATAGCCTTTGCCTTGTTCCGATTCAGTCAT	0.6%	67
AATTTTTTTTATAGCCT---CCTTGTTCCGATTCAGTCAT	AATTTTTTTTATAGCCTTTGCCTTGTTCCGATTCAGTCAT	0.5%	56
AATTTTTTTTATAGCCTTTG--------CGATTCAGTCAT	AATTTTTTTTATAGCCTTTGCCTTGTTCCGATTCAGTCAT	0.5%	54
----------TTGTTCCGATTCAGTCAT	AATTTTTTTTATAGCCTTTGCCTTGTTCCGATTCAGTCAT	0.5%	52
AATTTTTTTTATAGCCTTTG----------	AATTTTTTTTATAGCCTTTGCCTTGTTCCGATTCAGTCAT	0.5%	51
AA-TTTTTTTATA------GCCTTGTTCCGATTCAGTCAT	AATTTTTTTTATAGCCTTTGCCTTGTTCCGATTCAGTCAT	0.4%	46
----------GCCTTGTTCCGATTCAGTCAT	AATTTTTTTTATAGCCTTTGCCTTGTTCCGATTCAGTCAT	0.4%	46
AATTTTTTTTATAGCCTTTGTCCTTGTTCCGATTCAGTCA	AATTTTTTTTATAGCCTTTG-CCTTGTTCCGATTCAGTCA	0.4%	46
----------TTCAGTCAT	AATTTTTTTTATAGCCTTTGCCTTGTTCCGATTCAGTCAT	0.4%	45
AATTTTTTTTATAGCCTTTGGGACCTTGTTCCGATTCAGT	AATTTTTTTTATAGCCTTTG---CCTTGTTCCGATTCAGT	0.4%	45
----------GATTCAGTCAT	AATTTTTTTTATAGCCTTTGCCTTGTTCCGATTCAGTCAT	0.4%	43
TTTTTTTATAGCCTTTGAGGCCTTGTTCCGATTCAGTCAT	TTTTTTTATAGCCTTT---GCCTTGTTCCGATTCAGTCAT	0.4%	43
----------CCGATTCAGTCAT	AATTTTTTTTATAGCCTTTGCCTTGTTCCGATTCAGTCAT	0.4%	41
TTTTTTTATAGCCTTTGGGGCCTTGTTCCGATTCAGTCAT	TTTTTTTATAGCCTTT---GCCTTGTTCCGATTCAGTCAT	0.4%	41
...	...	...	...
SUM		**86.5%**	**9749**

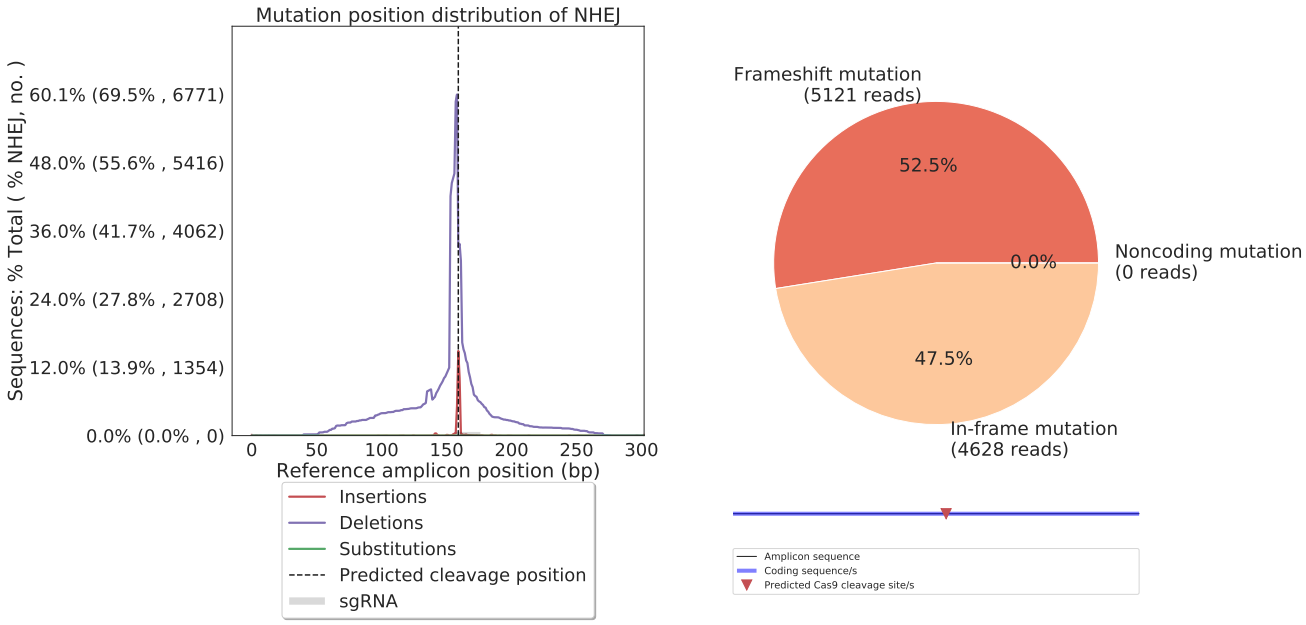

Figure 3: **Mutation rate and frameshift classification plots for sample Edited_Sample_1 region Target_2.** *Left:* The cumulative fraction of mutations at each position of the target region is shown.*Right:* The proportion of frameshift and in-frame mutations, as well as of mutations outside of coding region(s) is shown.

Figure 4: **Alleles around cutsite for sample Edited_Sample_1 region Target_2.** Please note that only alleles with a frequency of at least 0.2% and at most 50 alleles are shown.

1.2.2.2 Edited_Sample_2

Associated control: Control_Sample

Table 6: **Edited alleles frequency table for sample Edited_Sample_2 region Target_1.** *Aligned Sequence*: The observed read sequence, aligned to the reference. *Reference Sequence*: The reference sequence. *%Reads*: The fraction of reads with same sequence. *#Reads*: The number of reads with the same sequence. **Please note:** All alleles with variants that are close to the expected cut-site are listed here. However, only the 30 most frequent alleles are listed here. The full list of alleles is delivered as a separate file.

Target:	*Target_1*		
Aligned Sequence	**Reference Sequence**	**%Reads**	**#Reads**
AATTTTTTTTATA------GCCTTGTTCCGATTCAGTCAT	AATTTTTTTTATAGCCTTTGCCTTGTTCCGATTCAGTCAT	14.7%	792
AATTTTTTTTATAGCCT-----TTGTTCCGATTCAGTCAT	AATTTTTTTTATAGCCTTTGCCTTGTTCCGATTCAGTCAT	6.3%	340
AATTTTTTTTATAGCCTTT-CCTTGTTCCGATTCAGTCAT	AATTTTTTTTATAGCCTTTGCCTTGTTCCGATTCAGTCAT	4.0%	217
TTTTTTTTATAGCCTTTGGGCCTTGTTCCGATTCAGTCAT	TTTTTTTTATAGCCTTT--GCCTTGTTCCGATTCAGTCAT	2.6%	140
AATTTTTTTTATAGCCTTTGACCTTGTTCCGATTCAGTCA	AATTTTTTTTATAGCCTTTG-CCTTGTTCCGATTCAGTCA	2.3%	126
ATTTTTTTTATAGCCTTTGGCCTTGTTCCGATTCAGTCAT	ATTTTTTTTATAGCCTTT-GCCTTGTTCCGATTCAGTCAT	2.0%	109
AATTTTTTTTATAGCCTTTGCCCTTGTTCCGATTCAGTCA	AATTTTTTTTATAGCCTTTG-CCTTGTTCCGATTCAGTCA	1.7%	94
AATTTTTTTTATAGCCTTTG-CTTGTTCCGATTCAGTCAT	AATTTTTTTTATAGCCTTTGCCTTGTTCCGATTCAGTCAT	1.7%	90
AATTTTTTTTATAGCCTTTGGACCTTGTTCCGATTCAGTC	AATTTTTTTTATAGCCTTTG--CCTTGTTCCGATTCAGTC	1.5%	79
AATTTTTTTTATAGCCTTTGAACCTTGTTCCGATTCAGTC	AATTTTTTTTATAGCCTTTG--CCTTGTTCCGATTCAGTC	1.5%	78
AATTTTTTTTATAGCCTTTG--TTGTTCCGATTCAGTCAT	AATTTTTTTTATAGCCTTTGCCTTGTTCCGATTCAGTCAT	1.4%	76
AATTTTTTTTATAGCCTT--CCTTGTTCCGATTCAGTCAT	AATTTTTTTTATAGCCTTTGCCTTGTTCCGATTCAGTCAT	1.0%	55
------------------GCCTTGTTCCGATTCAGTCAT	AATTTTTTTTATAGCCTTTGCCTTGTTCCGATTCAGTCAT	0.9%	48
TTTTTTTATAGCCTTTGGGGCCTTGTTCCGATTCAGTCAT	TTTTTTTATAGCCTTT---GCCTTGTTCCGATTCAGTCAT	0.8%	44
TTTTTTTTATAGCCTTTGAGCCTTGTTCCGATTCAGTCAT	TTTTTTTTATAGCCTTT--GCCTTGTTCCGATTCAGTCAT	0.8%	43
AATTTTTTTTATAGCCTTTGGGACCTTGTTCCGATTCAGT	AATTTTTTTTATAGCCTTTG---CCTTGTTCCGATTCAGT	0.6%	34
AATTTTTTTTATAGCCTTTGTCCTTGTTCCGATTCAGTCA	AATTTTTTTTATAGCCTTTG-CCTTGTTCCGATTCAGTCA	0.6%	34
----------GCCTTGTTCCGATTCAGTCAT	AATTTTTTTTATAGCCTTTGCCTTGTTCCGATTCAGTCAT	0.6%	33
AATTTTTTTTATAGCCTTTG---TGTTCCGATTCAGTCAT	AATTTTTTTTATAGCCTTTGCCTTGTTCCGATTCAGTCAT	0.6%	31
AATTTTTTTTATAGCCT---CCTTGTTCCGATTCAGTCAT	AATTTTTTTTATAGCCTTTGCCTTGTTCCGATTCAGTCAT	0.6%	31
TTTTTTTATAGCCTTTGAGGCCTTGTTCCGATTCAGTCAT	TTTTTTTATAGCCTTT---GCCTTGTTCCGATTCAGTCAT	0.5%	27
AATTTTTTTTATAGCCTTTG--------CGATTCAGTCAT	AATTTTTTTTATAGCCTTTGCCTTGTTCCGATTCAGTCAT	0.4%	24
AATTTTTTTTATAGCCTTTGAAACCTTGTTCCGATTCAGT	AATTTTTTTTATAGCCTTTG---CCTTGTTCCGATTCAGT	0.4%	22
AATTTTTTTTATAG---------------------------	AATTTTTTTTATAGCCTTTGCCTTGTTCCGATTCAGTCAT	0.4%	22
AA----TTTTATA------GCCTTGTTCCGATTCAGTCAT	AATTTTTTTTATAGCCTTTGCCTTGTTCCGATTCAGTCAT	0.3%	18
AATTTTTTTTATAGCCTTTGGCCCTTGTTCCGATTCAGTC	AATTTTTTTTATAGCCTTTG--CCTTGTTCCGATTCAGTC	0.3%	17
AATTTTTTTTATAGCCTTTGGGGCCCTTGTTCCGATTCAG	AATTTTTTTTATAGCCTTTG----CCTTGTTCCGATTCAG	0.3%	17
AATTTTTTTTATAGCCTTTG--------------------	AATTTTTTTTATAGCCTTTGCCTTGTTCCGATTCAGTCAT	0.3%	17
AATTTTTTTTATAGCCTTTGTTCCTTGTTCCGATTCAGTC	AATTTTTTTTATAGCCTTTG--CCTTGTTCCGATTCAGTC	0.3%	16
AATTTTTTTTATAG-------------CCGATTCAGTCAT	AATTTTTTTTATAGCCTTTGCCTTGTTCCGATTCAGTCAT	0.3%	16
...	...	...	...
SUM		**76.5%**	**4110**

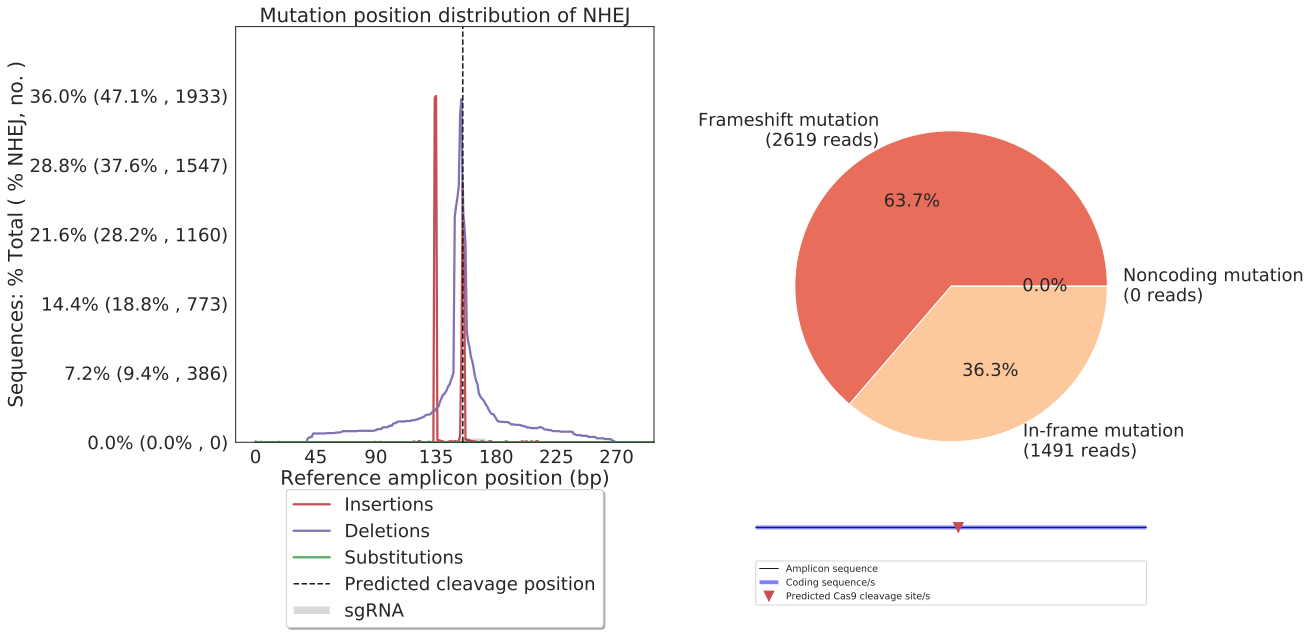

Figure 5: **Mutation rate and frameshift classification plots for sample Edited_Sample_2 region Target_1.** *Left:* The cumulative fraction of mutations at each position of the target region is shown.*Right:* The proportion of frameshift and in-frame mutations, as well as of mutations outside of coding region(s) is shown.

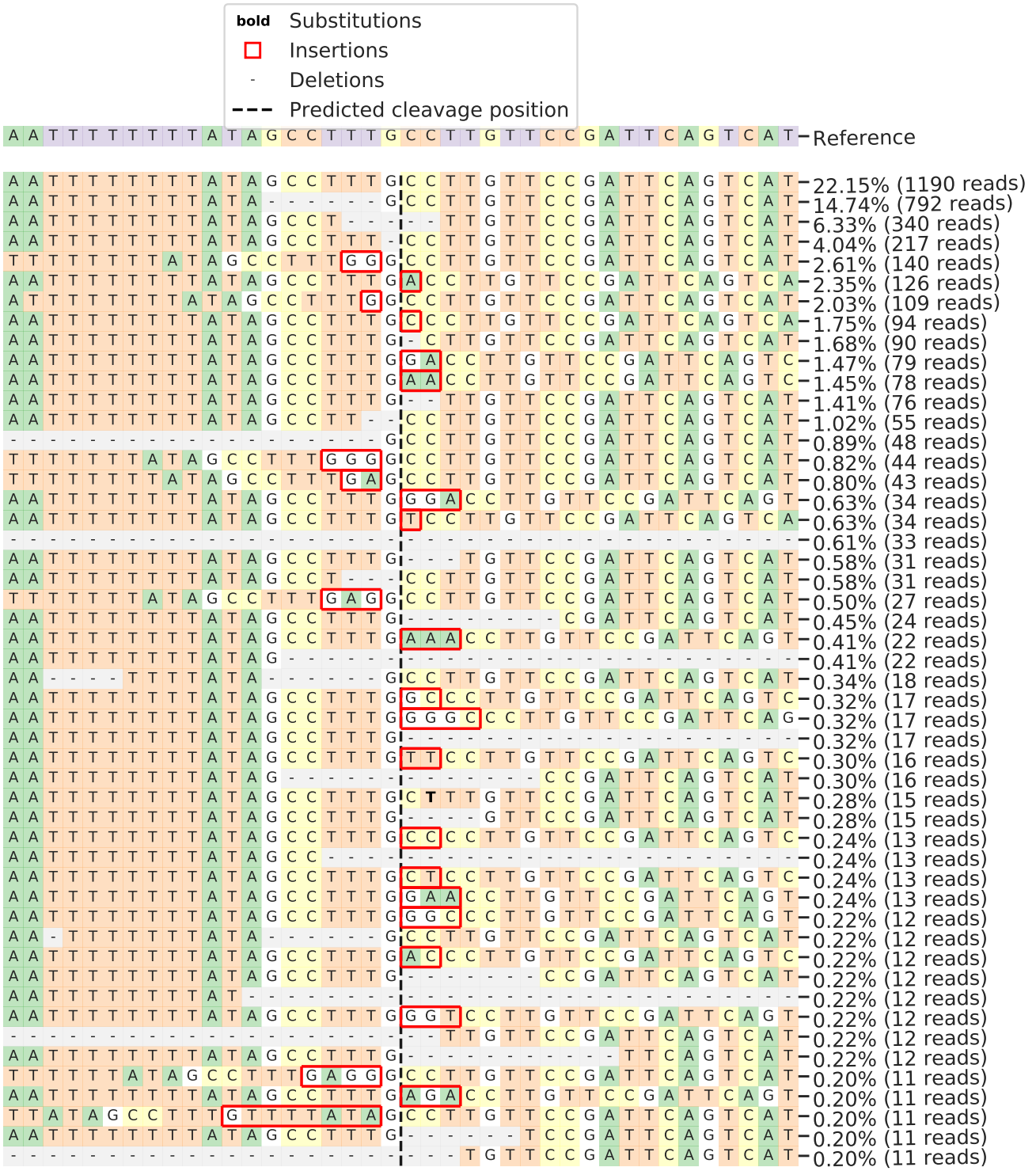

Figure 6: **Alleles around cutsite for sample Edited_Sample_2 region Target_1.** Please note that only alleles with a frequency of at least 0.2% and at most 50 alleles are shown.

Table 7: **Edited alleles frequency table for sample Edited_Sample_2 region Target_2.** *Aligned Sequence*: The observed read sequence, aligned to the reference. *Reference Sequence*: The reference sequence. *%Reads*: The fraction of reads with same sequence. *#Reads*: The number of reads with the same sequence. **Please note:** All alleles with variants that are close to the expected cut-site are listed here. However, only the 30 most frequent alleles are listed here. The full list of alleles is delivered as a separate file.

Target:	*Target_2*		
Aligned Sequence	**Reference Sequence**	**%Reads**	**#Reads**
AATTTTTTTTATA------GCCTTGTTCCGATTCAGTCAT	AATTTTTTTTATAGCCTTTGCCTTGTTCCGATTCAGTCAT	28.8%	3185
AATTTTTTTTATAGCCT-----TTGTTCCGATTCAGTCAT	AATTTTTTTTATAGCCTTTGCCTTGTTCCGATTCAGTCAT	10.5%	1158
AATTTTTTTTATAGCCTTT-CCTTGTTCCGATTCAGTCAT	AATTTTTTTTATAGCCTTTGCCTTGTTCCGATTCAGTCAT	2.4%	270
TTTTTTTTATAGCCTTTGGGCCTTGTTCCGATTCAGTCAT	TTTTTTTTATAGCCTTT--GCCTTGTTCCGATTCAGTCAT	1.6%	177
AATTTTTTTTATAGCCTTTGACCTTGTTCCGATTCAGTCA	AATTTTTTTTATAGCCTTTG-CCTTGTTCCGATTCAGTCA	1.4%	155
AATTTTTTTTATAGCCTTTGCCCTTGTTCCGATTCAGTCA	AATTTTTTTTATAGCCTTTG-CCTTGTTCCGATTCAGTCA	1.3%	139
AATTTTTTTTATAGCCTTTG-CTTGTTCCGATTCAGTCAT	AATTTTTTTTATAGCCTTTGCCTTGTTCCGATTCAGTCAT	1.2%	131
----------	AATTTTTTTTATAGCCTTTGCCTTGTTCCGATTCAGTCAT	1.2%	128
----------GCCTTGTTCCGATTCAGTCAT	AATTTTTTTTATAGCCTTTGCCTTGTTCCGATTCAGTCAT	1.0%	112
AATTTTTTTTATAGCCTTTG--TTGTTCCGATTCAGTCAT	AATTTTTTTTATAGCCTTTGCCTTGTTCCGATTCAGTCAT	0.9%	102
AATTTTTTTTATAGCCTTTGAACCTTGTTCCGATTCAGTC	AATTTTTTTTATAGCCTTTG--CCTTGTTCCGATTCAGTC	0.8%	93
AATTTTTTTTATAGCCTTTGGACCTTGTTCCGATTCAGTC	AATTTTTTTTATAGCCTTTG--CCTTGTTCCGATTCAGTC	0.8%	93
AATTTTTTTTATAG----------CCGATTCAGTCAT	AATTTTTTTTATAGCCTTTGCCTTGTTCCGATTCAGTCAT	0.8%	90
ATTTTTTTTATAGCCTTTGGCCTTGTTCCGATTCAGTCAT	ATTTTTTTTATAGCCTTT-GCCTTGTTCCGATTCAGTCAT	0.7%	82
AATTTTTTTTATAG----------	AATTTTTTTTATAGCCTTTGCCTTGTTCCGATTCAGTCAT	0.6%	71
----------GATTCAGTCAT	AATTTTTTTTATAGCCTTTGCCTTGTTCCGATTCAGTCAT	0.6%	68
----------CTTGTTCCGATTCAGTCAT	AATTTTTTTTATAGCCTTTGCCTTGTTCCGATTCAGTCAT	0.6%	67
AATTTTTTTTATAGCCTT--CCTTGTTCCGATTCAGTCAT	AATTTTTTTTATAGCCTTTGCCTTGTTCCGATTCAGTCAT	0.6%	62
----------TTGTTCCGATTCAGTCAT	AATTTTTTTTATAGCCTTTGCCTTGTTCCGATTCAGTCAT	0.5%	60
AATTTTTTTTATAGCCTTTG--------CGATTCAGTCAT	AATTTTTTTTATAGCCTTTGCCTTGTTCCGATTCAGTCAT	0.5%	55
TTTTTTTATAGCCTTTGGGGCCTTGTTCCGATTCAGTCAT	TTTTTTTATAGCCTTT---GCCTTGTTCCGATTCAGTCAT	0.5%	54
AATTTTTTTTATAGCC----------	AATTTTTTTTATAGCCTTTGCCTTGTTCCGATTCAGTCAT	0.4%	48
TTTTTTTTATAGCCTTTGAGCCTTGTTCCGATTCAGTCAT	TTTTTTTTATAGCCTTT--GCCTTGTTCCGATTCAGTCAT	0.4%	47
----------TTCAGTCAT	AATTTTTTTTATAGCCTTTGCCTTGTTCCGATTCAGTCAT	0.4%	45
----------CAT	AATTTTTTTTATAGCCTTTGCCTTGTTCCGATTCAGTCAT	0.4%	43
AATTTTTTTTATAGCCT---CCTTGTTCCGATTCAGTCAT	AATTTTTTTTATAGCCTTTGCCTTGTTCCGATTCAGTCAT	0.4%	39
AATTTTTTTTATAGCCTTTGGGACCTTGTTCCGATTCAGT	AATTTTTTTTATAGCCTTTG---CCTTGTTCCGATTCAGT	0.3%	36
----------TGTTCCGATTCAGTCAT	AATTTTTTTTATAGCCTTTGCCTTGTTCCGATTCAGTCAT	0.3%	36
----------AGTCAT	AATTTTTTTTATAGCCTTTGCCTTGTTCCGATTCAGTCAT	0.3%	36
AATTTTTTTTATAGCCTTTG---TGTTCCGATTCAGTCAT	AATTTTTTTTATAGCCTTTGCCTTGTTCCGATTCAGTCAT	0.3%	36
...	...	...	...
SUM		**86.4%**	**9547**

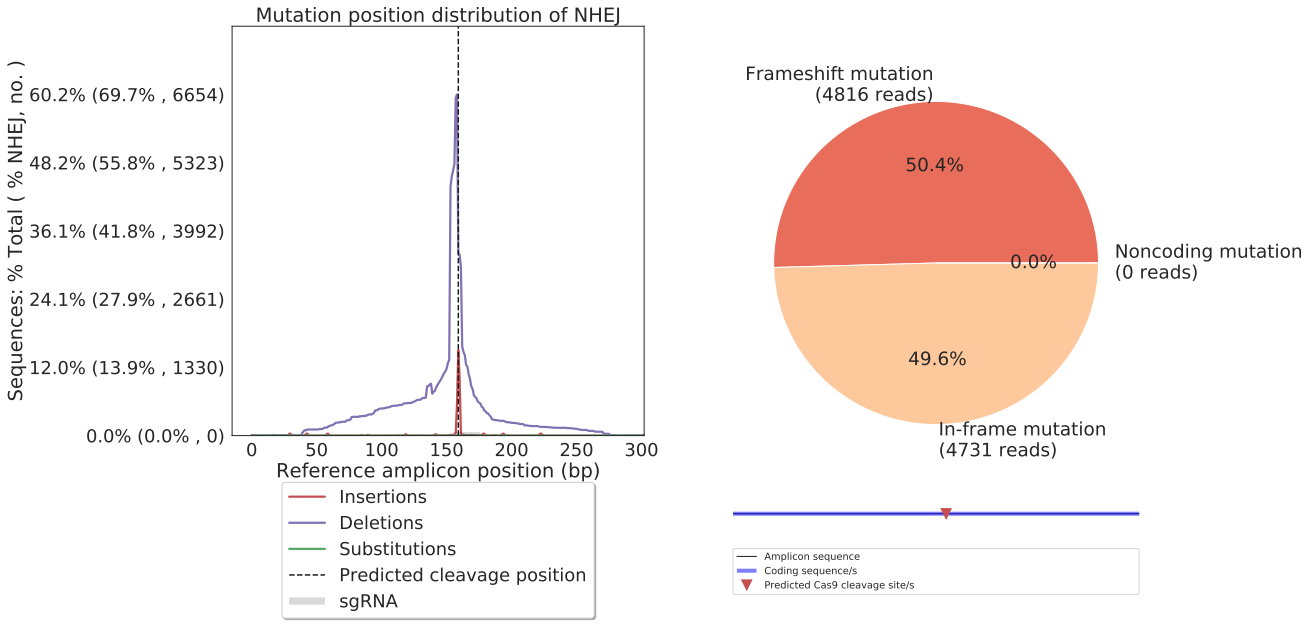

Figure 7: **Mutation rate and frameshift classification plots for sample Edited_Sample_2 region Target_2.** *Left:* The cumulative fraction of mutations at each position of the target region is shown.*Right:* The proportion of frameshift and in-frame mutations, as well as of mutations outside of coding region(s) is shown.

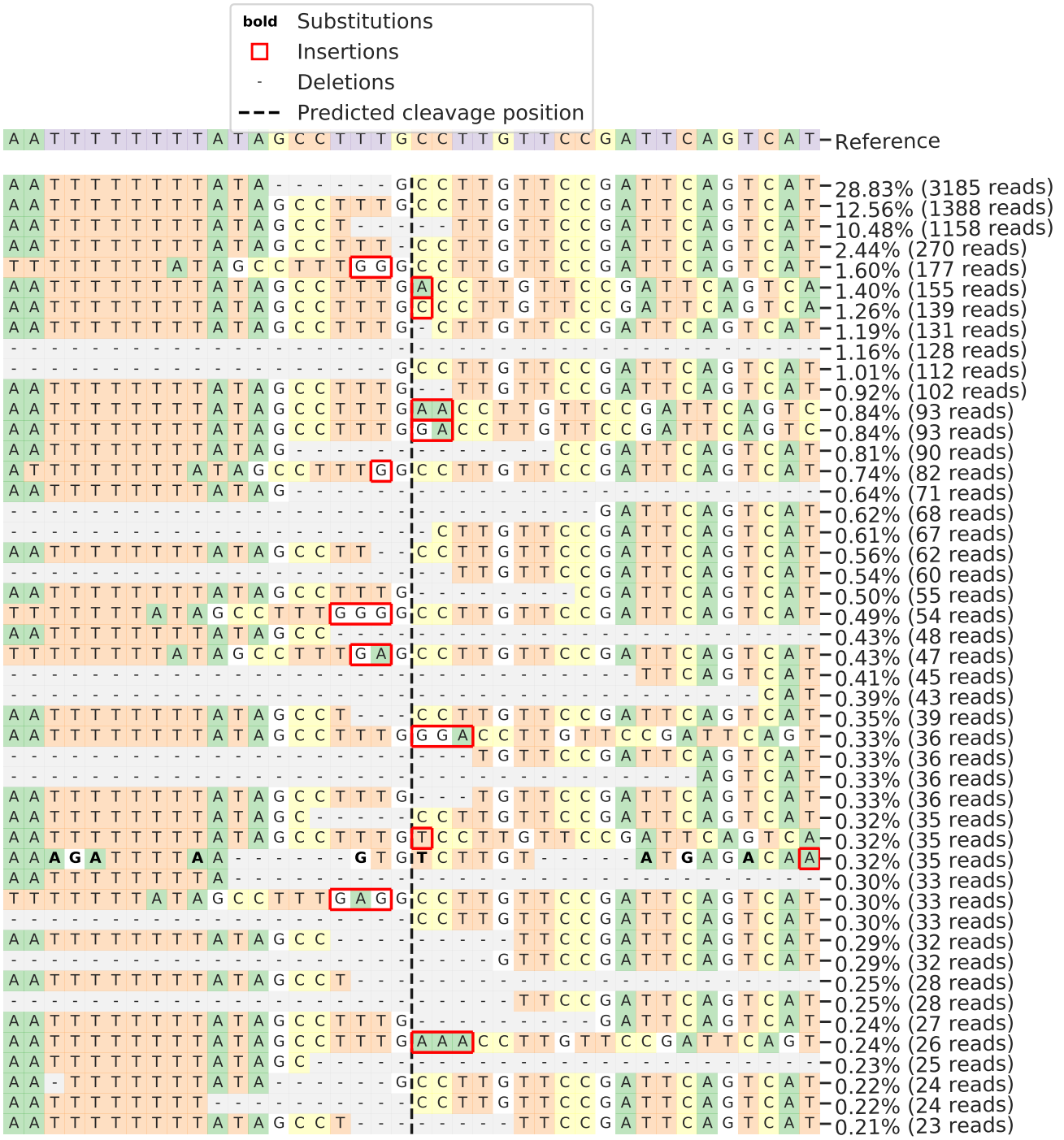

Figure 8: **Alleles around cutsite for sample Edited_Sample_2 region Target_2.** Please note that only alleles with a frequency of at least 0.2% and at most 50 alleles are shown.

1.2.2.3 Control_Sample

Associated control: none

Table 8: **Edited alleles frequency table for sample Control_Sample region Target_1.** *Aligned Sequence*: The observed read sequence, aligned to the reference. *Reference Sequence*: The reference sequence. *%Reads*: The fraction of reads with same sequence. *#Reads*: The number of reads with the same sequence. **Please note:** All alleles with variants that are close to the expected cut-site are listed here. However, only the 30 most frequent alleles are listed here. The full list of alleles is delivered as a separate file.

Target:	*Target_1*		
Aligned Sequence	**Reference Sequence**	**%Reads**	**#Reads**
SUM		**0.0%**	**0**

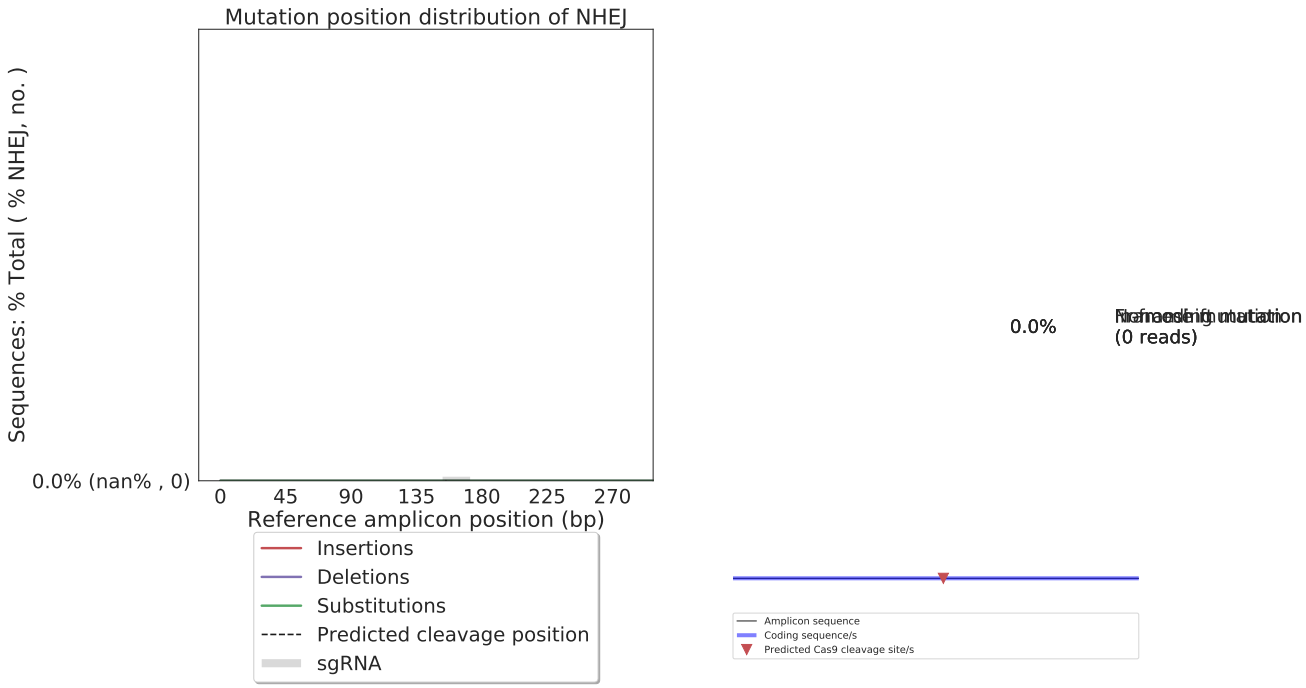

Figure 9: **Mutation rate and frameshift classification plots for sample Control_Sample region Target_1.** *Left:* The cumulative fraction of mutations at each position of the target region is shown.*Right:* The proportion of frameshift and in-frame mutations, as well as of mutations outside of coding region(s) is shown.

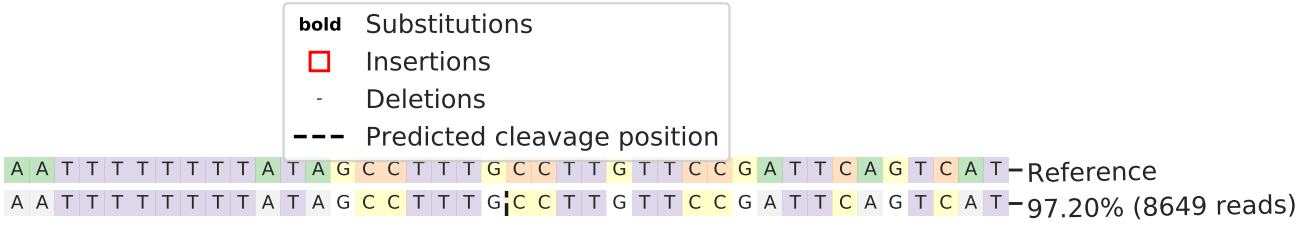

Figure 10: **Alleles around cutsite for sample Control_Sample region Target_1.** Please note that only alleles with a frequency of at least 0.2% and at most 50 alleles are shown.

Table 9: **Edited alleles frequency table for sample Control_Sample region Target_2.** *Aligned Sequence*: The observed read sequence, aligned to the reference. *Reference Sequence*: The reference sequence. *%Reads*: The fraction of reads with same sequence. *#Reads*: The number of reads with the same sequence. **Please note:** All alleles with variants that are close to the expected cut-site are listed here. However, only the 30 most frequent alleles are listed here. The full list of alleles is delivered as a separate file.

Target:	*Target_2*		
Aligned Sequence	**Reference Sequence**	**%Reads**	**#Reads**
TTTTTT----GCCGTCGTTGCCTTGTTCCGATTCAGTCAT	TTTTTTTATAGC---CTTTGCCTTGTTCCGATTCAGTCAT	0.0%	1
SUM		**0.0%**	**1**

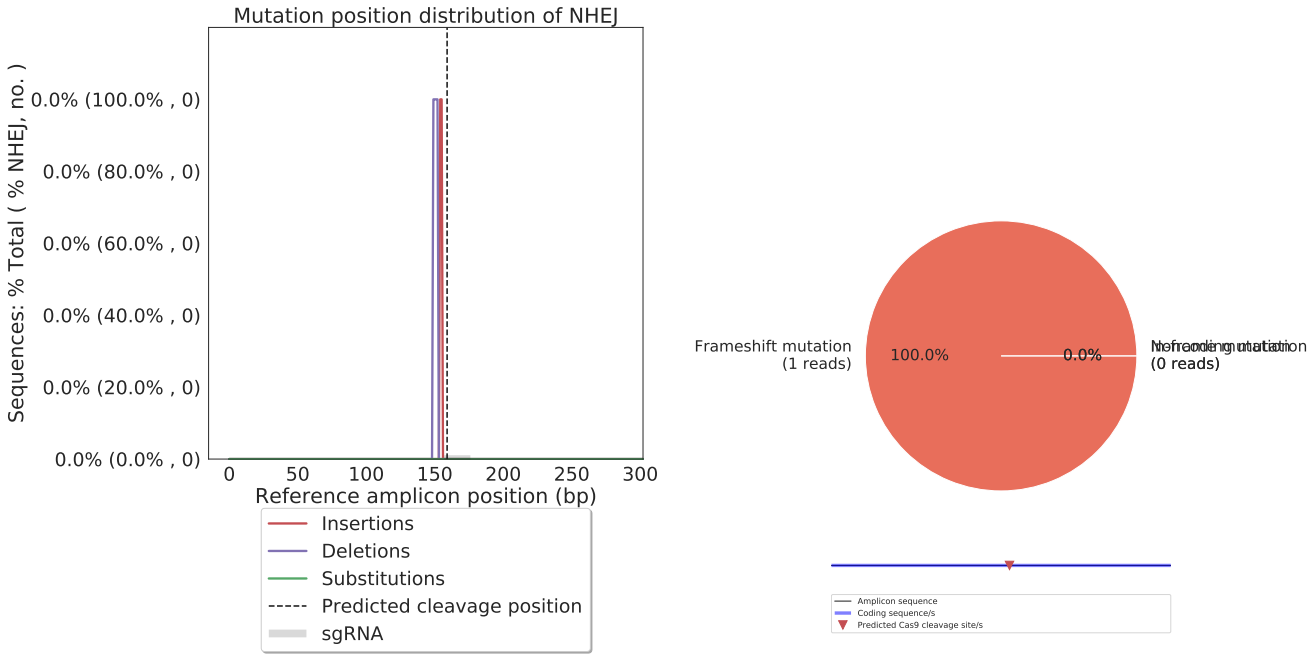

Figure 11: **Mutation rate and frameshift classification plots for sample Control_Sample region Target_2.** *Left:* The cumulative fraction of mutations at each position of the target region is shown.*Right:* The proportion of frameshift and in-frame mutations, as well as of mutations outside of coding region(s) is shown.

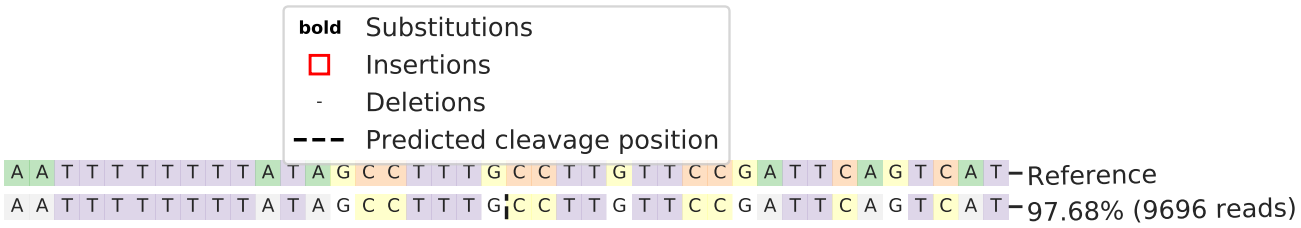

Figure 12: **Alleles around cutsite for sample Control_Sample region Target_2.** Please note that only alleles with a frequency of at least 0.2% and at most 50 alleles are shown.

1.2.3 FASTQ Read Statistics

No	Replicate	Read Pairs	Yield (Kbp)	%Q30	Mean Q
1	Control_Sample	20,000	12,000	88.62	35.20
2	Edited_Sample_1	20,000	12,000	84.13	34.11
3	Edited_Sample_2	20,000	12,000	83.03	33.82
	Total/Average	**60,000**	**36,000**	**85.26**	**34.38**

Table 10: FASTQ processing results.

Remarks:

- All reads are passed filter, i.e. reads have passed the default Illumina filter procedure (chastity filter).
- "Yield (Kbp)": number of bases called in kilobases.
- "%Q30": represents the percentage of bases with a quality score of at least 30 (inferred base call accuracy of 99.9%). The Q-score is a prediction of the probability of a wrong base call.

1.2.4 Preprocessing and Mapping

Table 11: **Preprocessing and mapping results.** *Replicate*: Replicate name, see table 1 or 2. *Raw reads*: The number of unmodified (raw) reads that went into the analysis. *Merged reads*: The number of reads remaining after the preprocessing step, and percent of raw reads. *On-target reads*: The numer of reads that were mapped to any target region, and percent of preprocessed reads.

No	Replicate	Raw reads	Merged reads	On-target reads
1	Control_Sample	40,000	38,692 (96.7%)	37,656 (99.2%)
2	Edited_Sample_1	40,000	37,782 (94.5%)	35,702 (97.7%)
3	Edited_Sample_2	40,000	37,908 (94.8%)	35,282 (96.9%)

1.2.5 Coverage Statistics

Table 12: **Coverage statistics for each target region.** *Replicate*: Replicate name, see table 1 or 2. *Donwsampled reads*: Number of remaining reads after downsampling, and percent of on-target reads. *Coverage median*: Median coverage of the target after downsampling. *Percent covered*: Proportion of target region that is covered by reads.

Target: *Target_1*

No	Replicate	Downsampled reads	Coverage median	Percent covered
1	Control_Sample	8,900 (100.0%)	8,898	100.0%
2	Edited_Sample_1	5,788 (100.0%)	5,481	100.0%
3	Edited_Sample_2	5,667 (100.0%)	5,303	100.0%

Target: *Target_2*

No	Replicate	Downsampled reads	Coverage median	Percent covered
1	Control_Sample	9,928 (100.0%)	9,926	100.0%
2	Edited_Sample_1	12,063 (100.0%)	10,923	100.0%
3	Edited_Sample_2	11,974 (100.0%)	10,561	100.0%

1.3 Methods

The Eurofins Genomics CRISPR NHEJ analysis pipeline performs the following operations:

1. Clipping of reads using Trimmomatic (v 0.36). Adapter sequences are removed from read sequences. If less than 40 bp of read sequence remains, the read is removed. If a read is removed, its mate is removed as well.
2. Paired-end reads were merged (assembled) using the software FLASH2 (v 2.2.00) to obtain a single, longer read that covers the full target region. A minimal overlap of 10bp and a maximum of 15% mismatch rate in the overlap region were required.
3. Mapping of preprocessed reads using BWA MEM (v 0.7.15). Reads are mapped to the reference sequences with default alignment parameters.
4. Aligned reads are filtered and downsampled. Reads aligning outside of the expected target regions are removed. Reads with ambiguous alignments (mapping quality of 1) are removed. If more than 20000 reads were aligned to a target region, the reads aligning to this target region are randomly downsampled to roughly 20000.
5. Identification and quantification of sequence alleles using CRISPResso (v 1.0.13). Only mutations within a 5bp window around the expectd cutsite are considered as editing mutations.

1.4 Deliverables

Table 14: List of delivered files, formats, and recommended programs to access the data.

File	Format	Program To Open File
*_Report.pdf	PDF	PDF reader
STEP.metrics.csv	CSV	Spreadsheet Editor
REPLICATE_[1|2].fastq.gz	FASTQ	none
REPLICATE.extendedFrags.fastq.gz	FASTQ	none
REPLICATE.bam	BAM	IGV, Tablet
REPLICATE.bam.bai	BAI	none

Table 15: Short descriptions of file contents.

File	Description
*_Report.pdf	This report.
REPLICATE_[1|2].fastq.gz	Contains the unprocessed raw read data in FASTQ format. These files were the starting point of the analysis.
REPLICATE.extendedFrags.fastq.gz	Contains the assembled reads in FASTQ format. These files were used for mapping.
REPLICATE.bam	Contains preprocessed, mapped reads in BAM format. These files were downsampled and then used for variant calling.
REPLICATE.bam.bai	The index file associated with REPLICATE.bam
Alleles_frequency_table.txt	Lists observed allele sequences aligned to reference sequence and associated quantifications.
Alleles_frequency_table_around_cut_site_[ACGT].txt	Same as above, but restricted to a 40bp window around the expected cut site.
efficiency.table.csv	This file summarizes the editing efficiency per sample and target region (sum of alternative allele frequencies).
alignment.metrics.csv	This file contains various alignment metrics.
coverage.metrics.csv	This file contains various coverage metrics.
merging.metrics.csv	This file contains various coverage metrics.

Table 16: Descriptions of file formats.

Format	Description
FASTQ	A text-based format for storing both a biological sequence and its corresponding quality scores.
BAM	Compressed binary version of the Sequence Alignment/Mapping (SAM) format, a compact and index-able representation of nucleotide sequence alignments.
CSV	Comma separated table style text file. It can be imported into spreadsheet editors like MS OFFICE Excel.
TSV	Tab separated table style text file. It can be imported into spreadsheet editors like MS OFFICE Excel.
TXT	Text file of arbitrary style. It can be opened by any text editor. We recommend to use Notepad++.

Eurofins Genomics' products, services and applications reach the best quality and safety levels. They are carried out under strict QM and QA systems and comply with the following standards:

ISO 17025	Accredited analytical excellence	GLP	The gold standard to conduct non-clinical safety studies
ISO 13485	Oligonucleotides according to medical devices standard	GCP	Pharmacogenomic services for clinical studies
		cGMP	Products and testing according to pharma and biotech requirements

Eurofins Genomics Europe Sequencing GmbH • Jakob-Stadler-Platz 7 • 78467 Constance • Germany