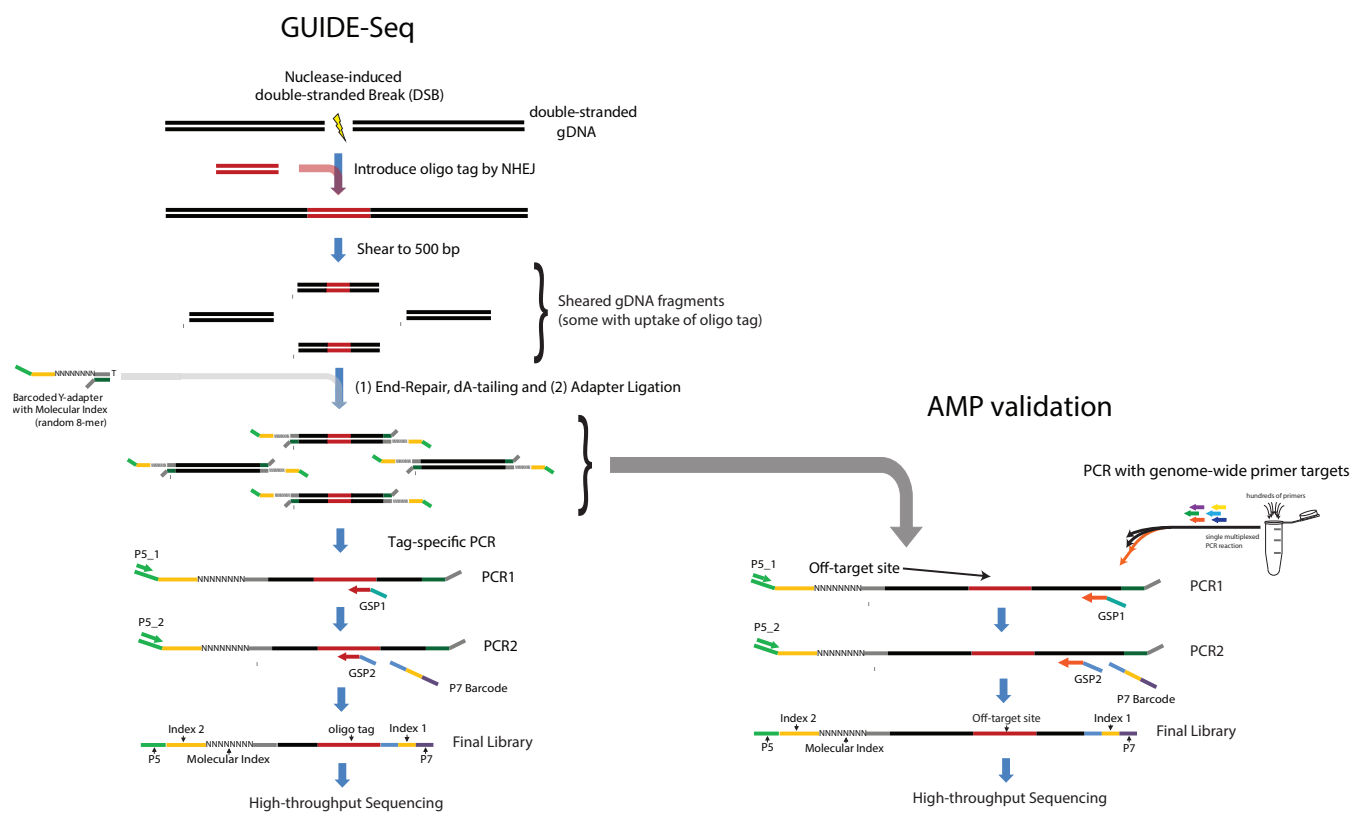
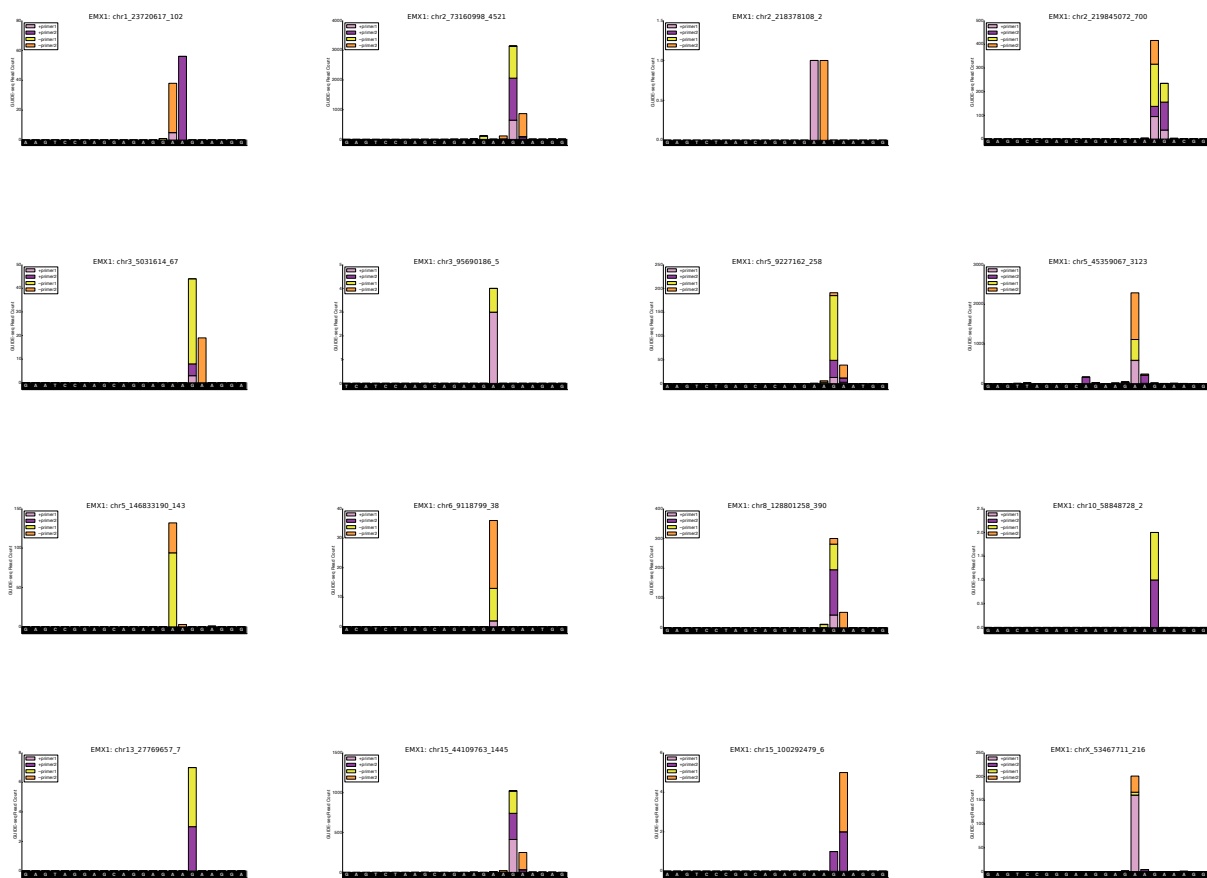


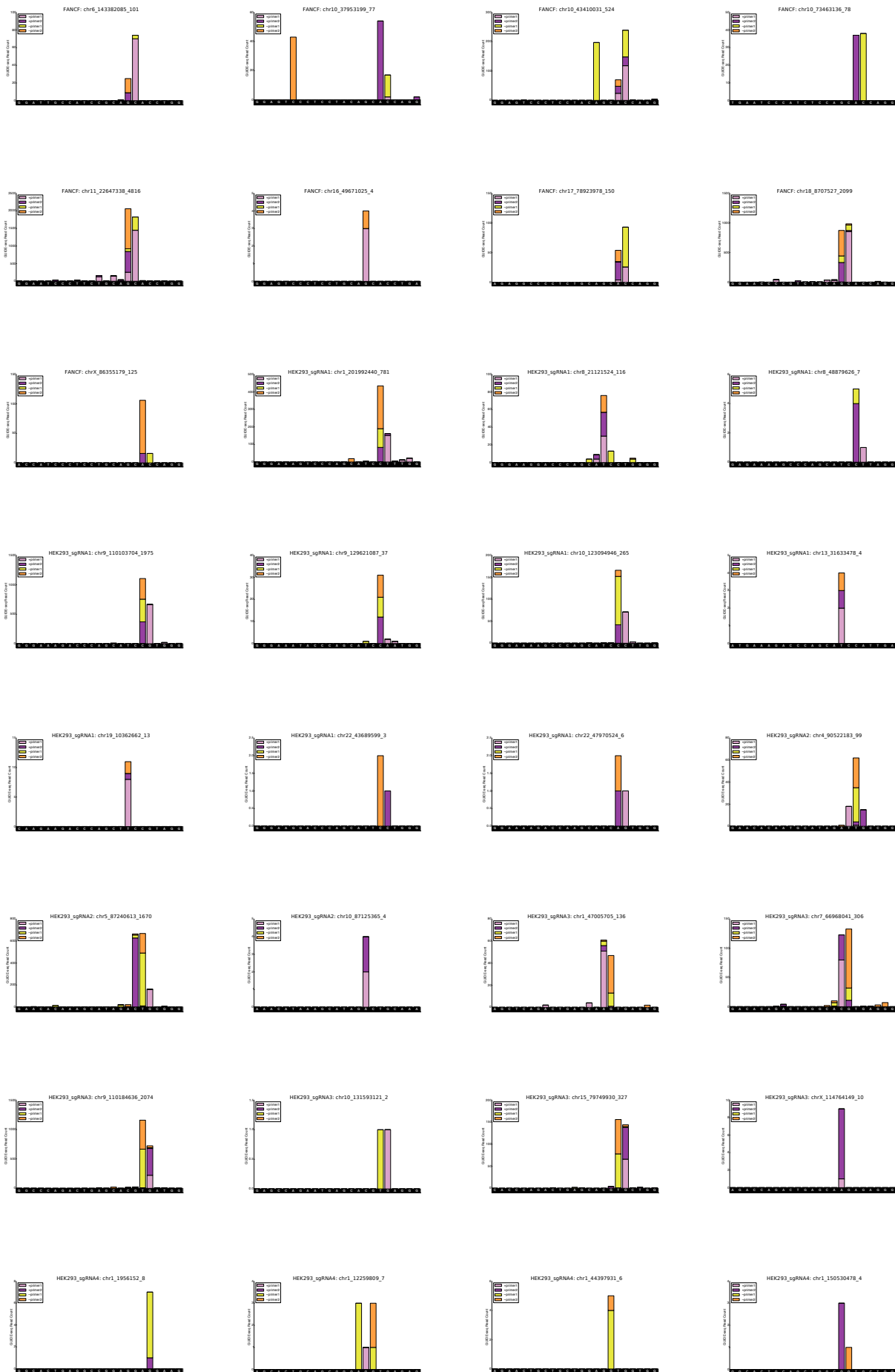
Supplementary Fig. 1 Detailed schematic overview of GUIDE-Seq and AMP-based sequencing for validation of dsODN insertions and indel mutations

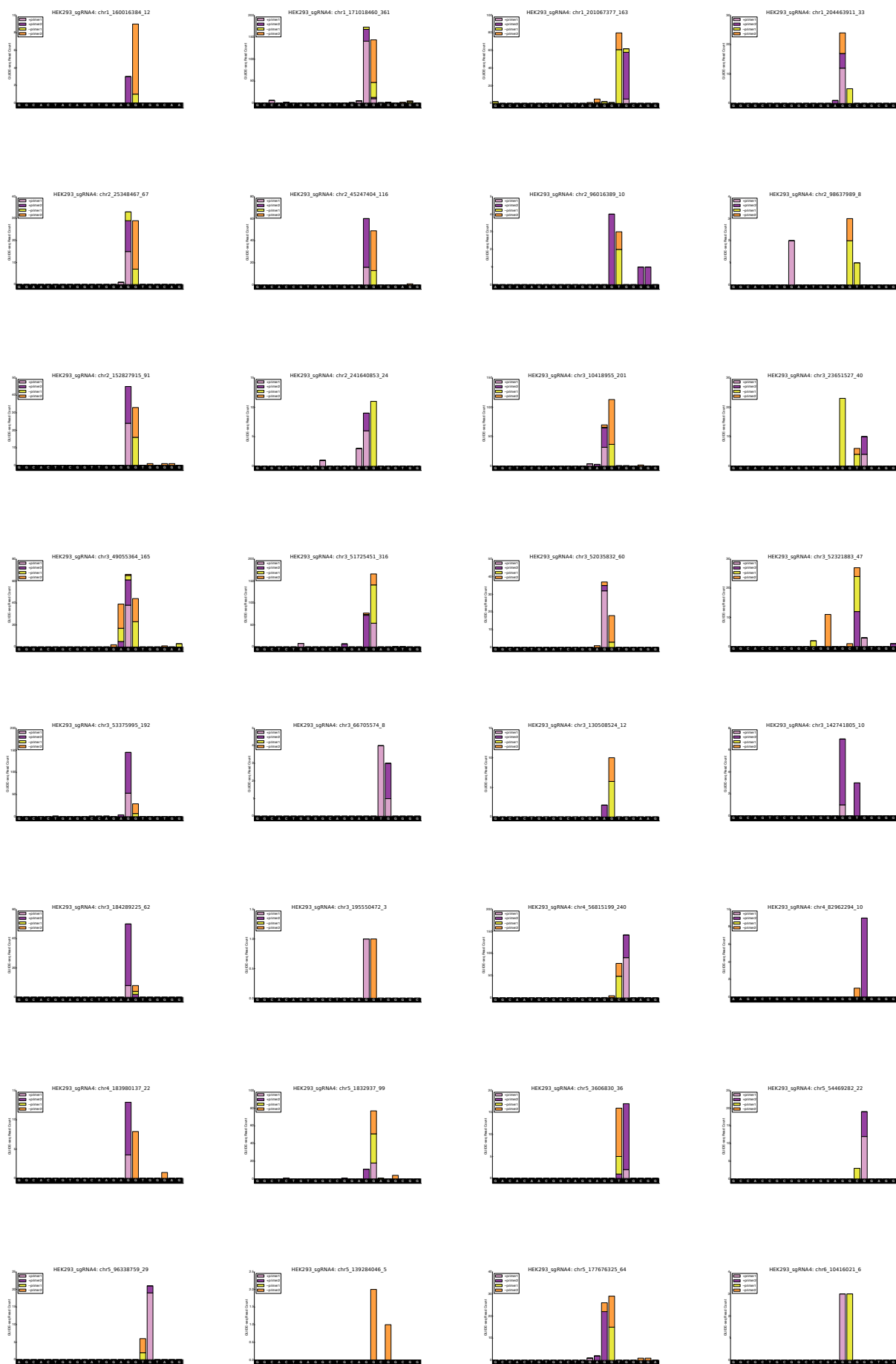


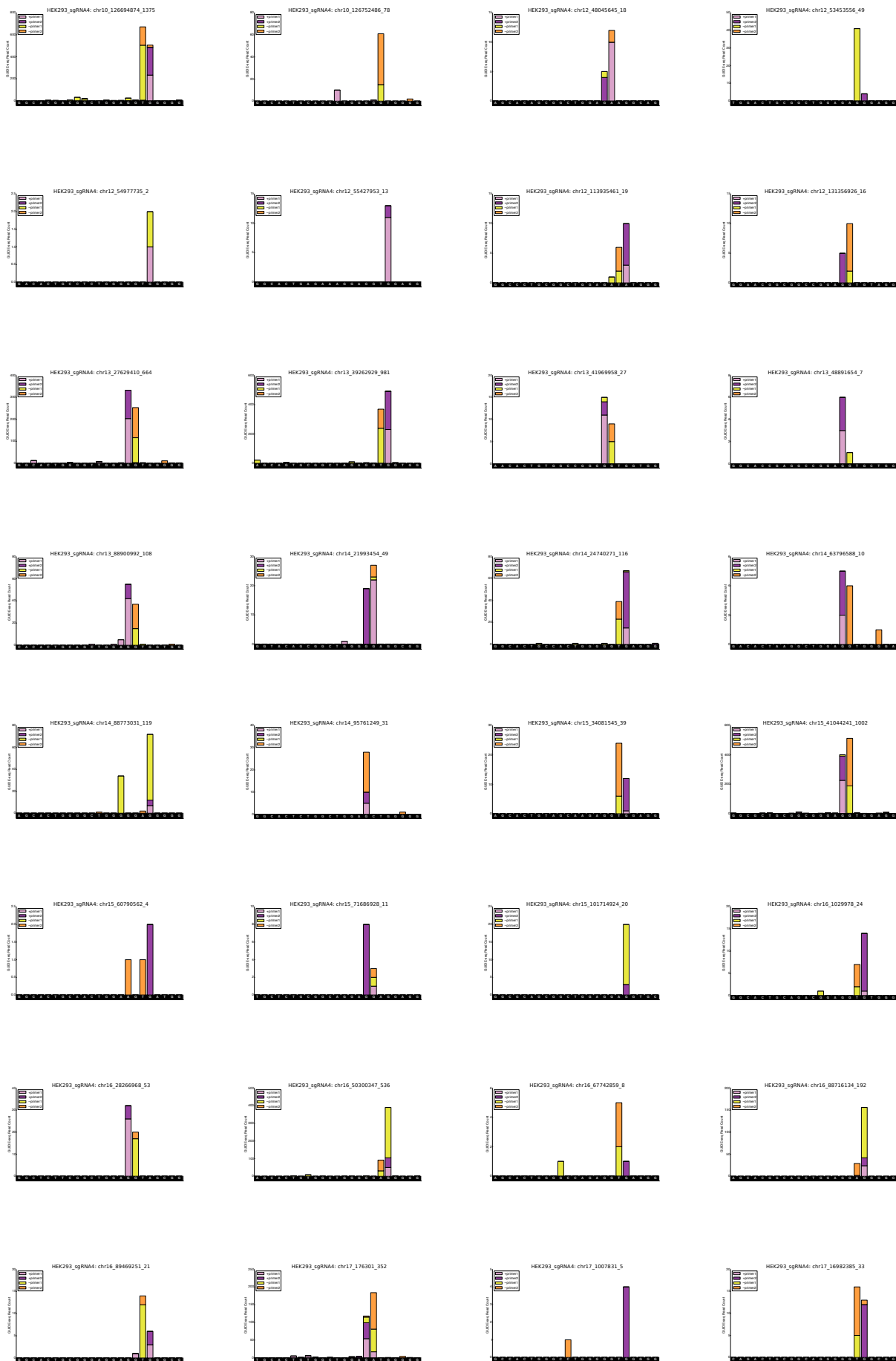
Details for both protocols can be found in Supplementary Methods.

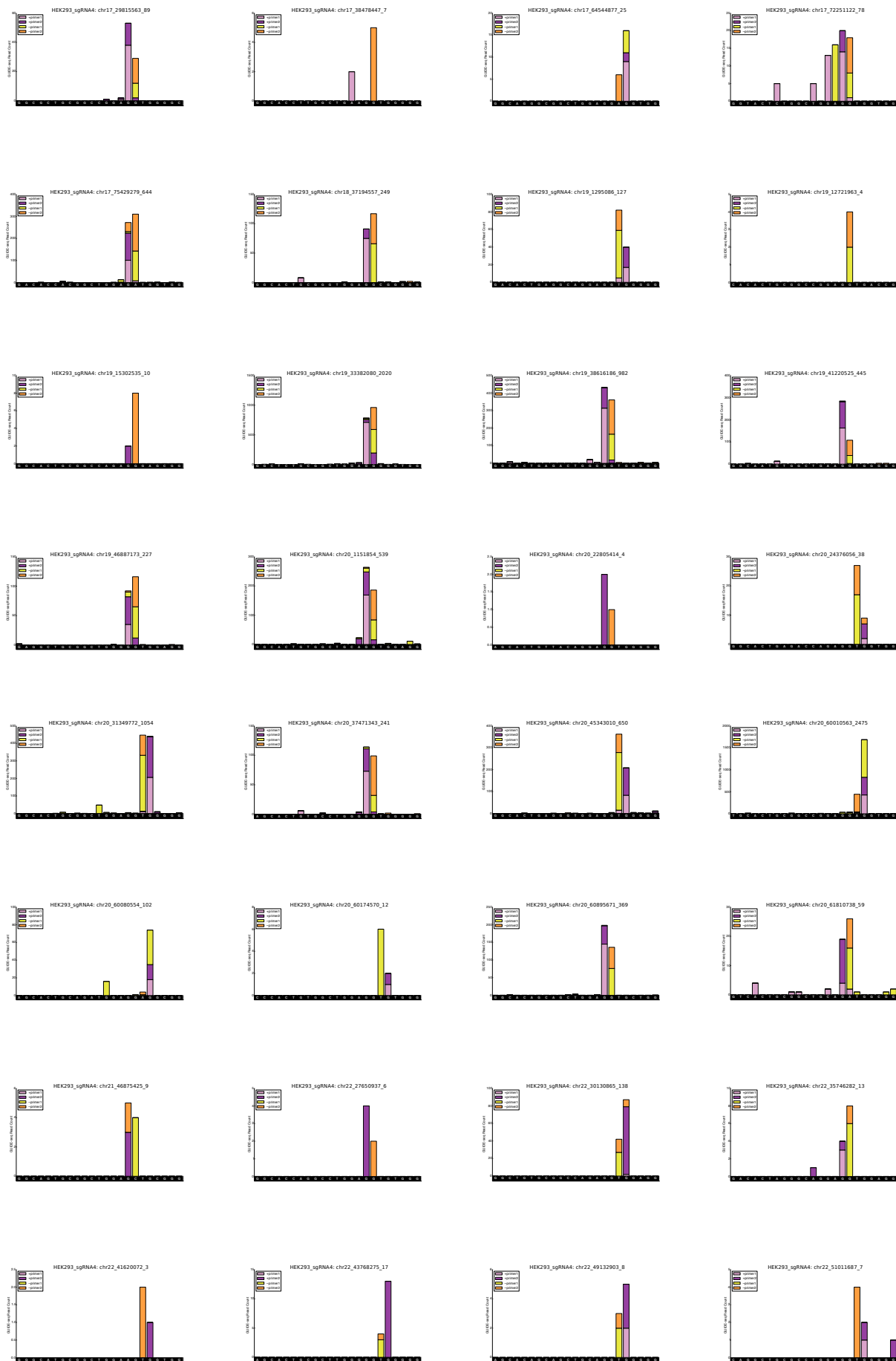
Supplementary Fig. 2 GUIDE-Seq-based identification of DSBs induced by RGNs directed by full-length gRNAs

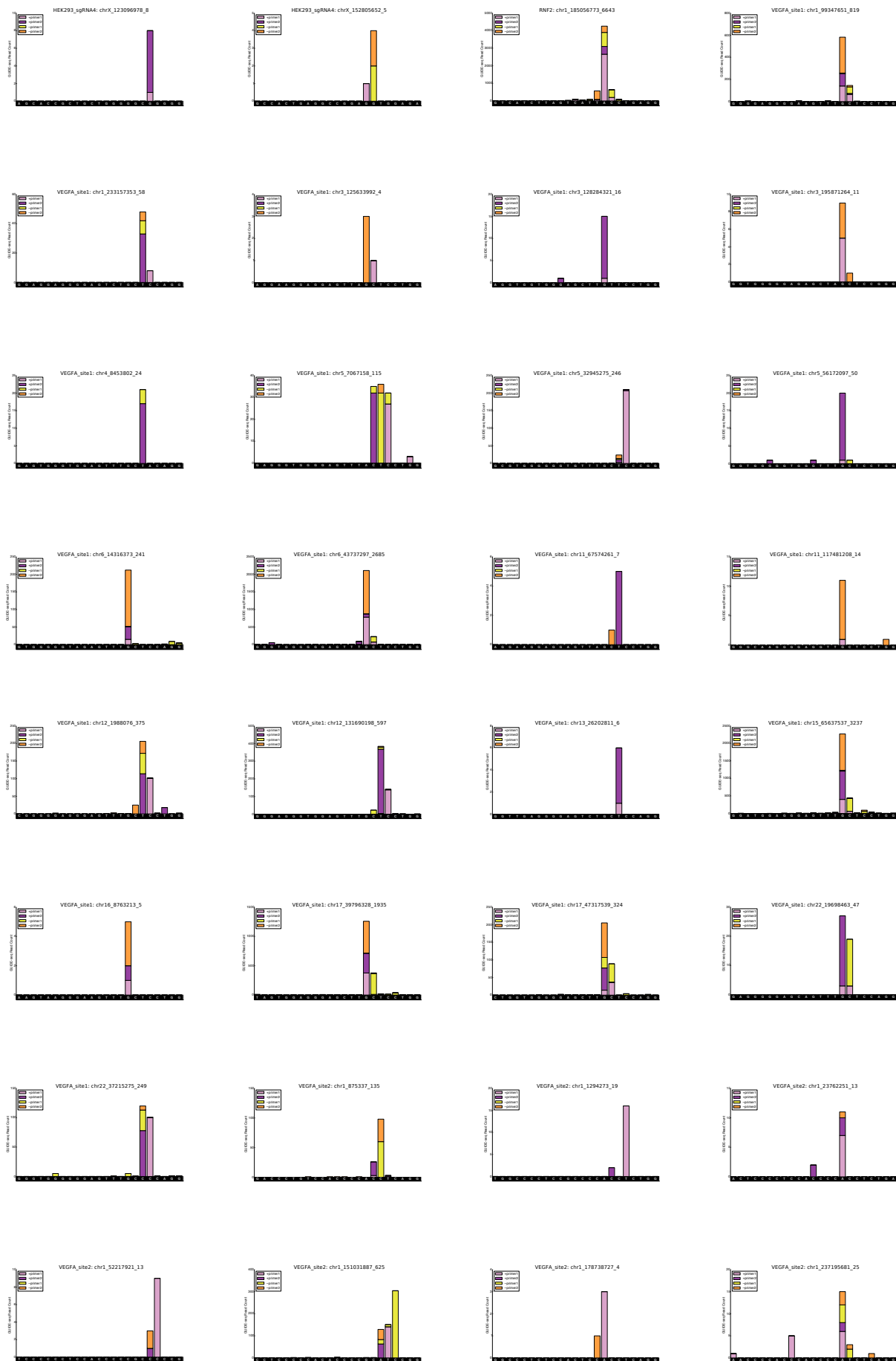


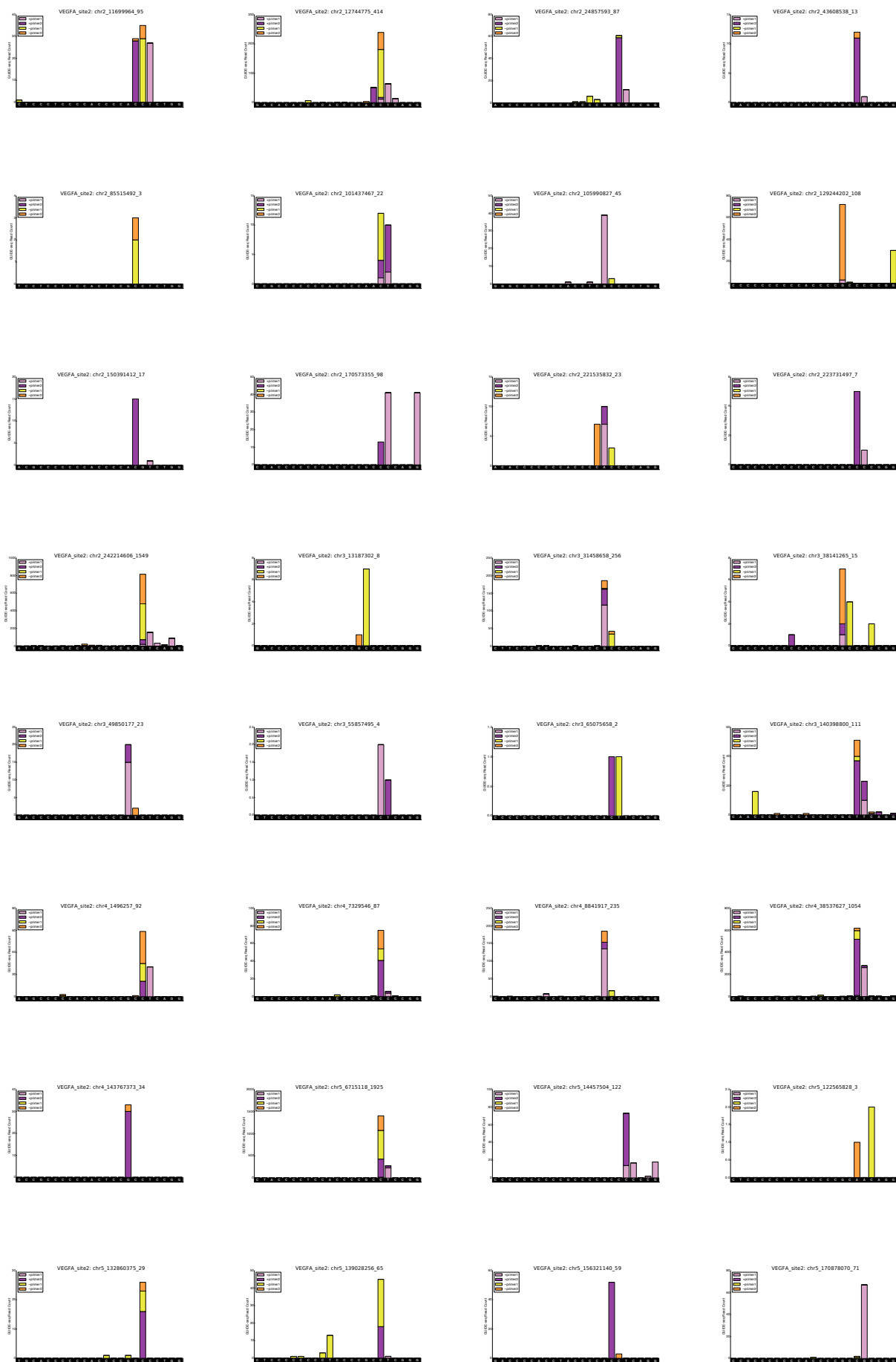


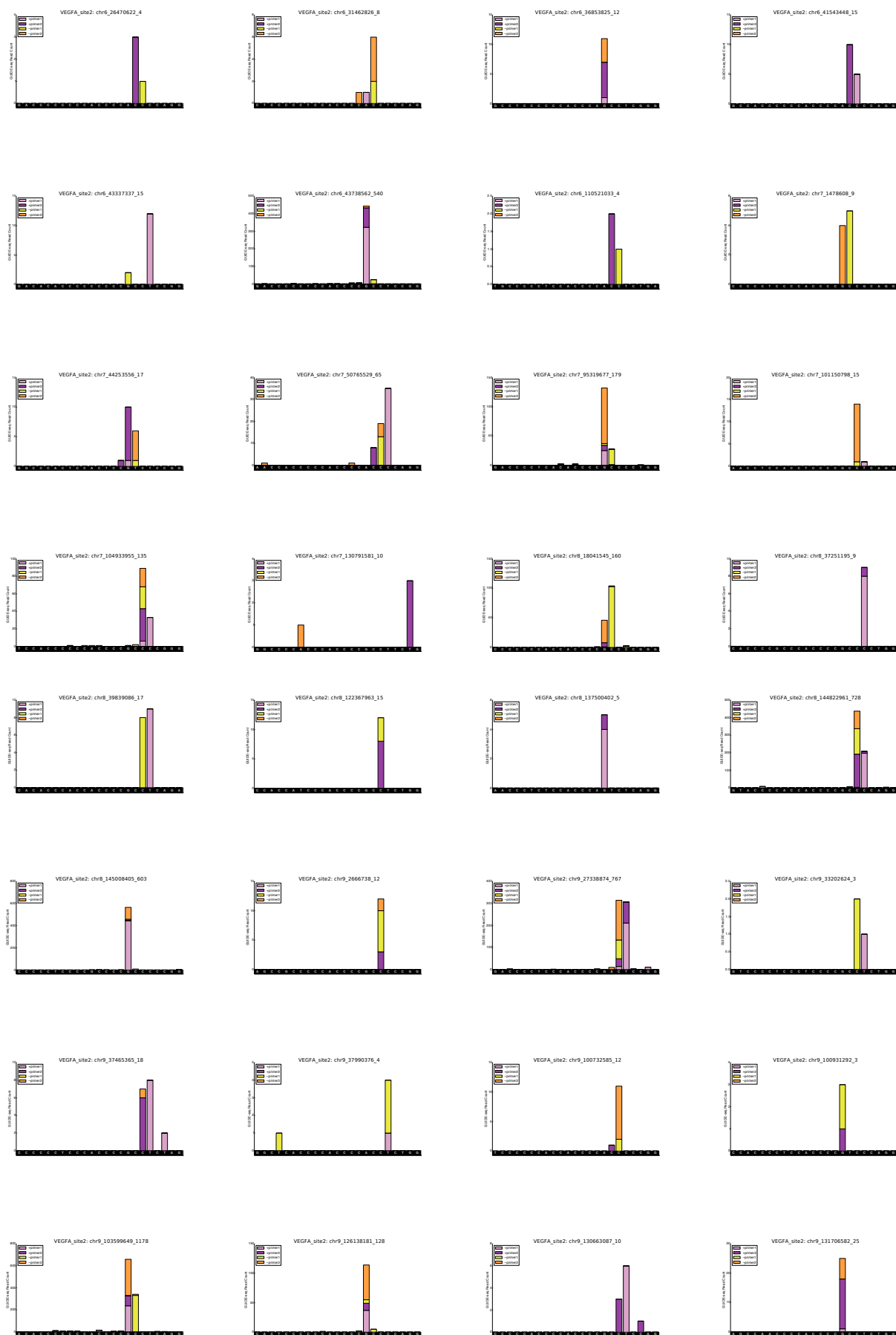


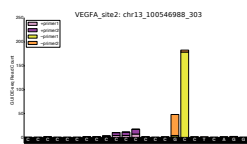
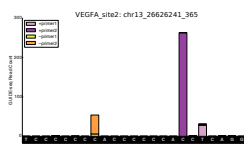
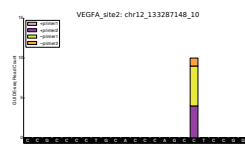
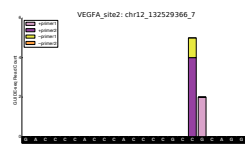
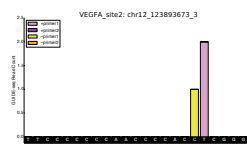
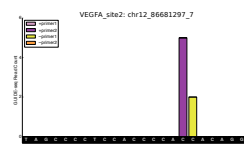
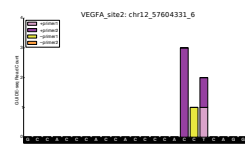
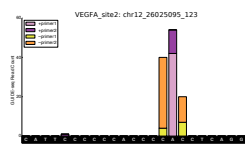
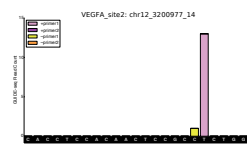
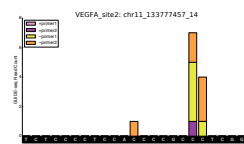
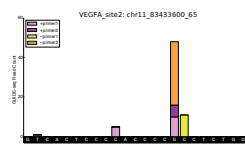
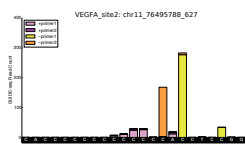
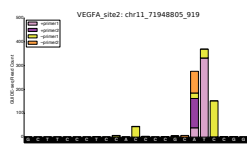
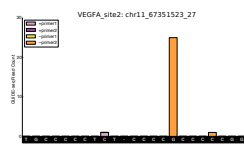
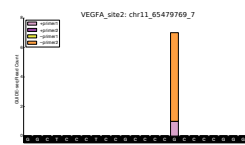
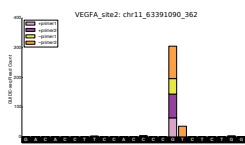
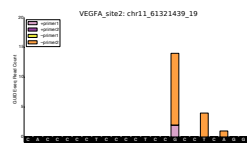
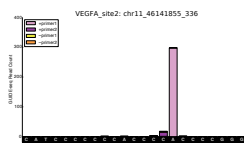
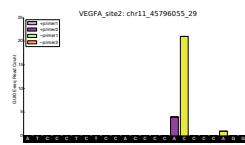
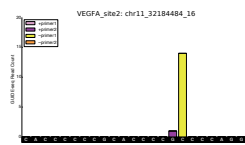
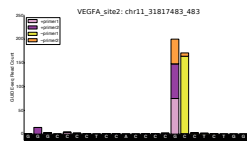
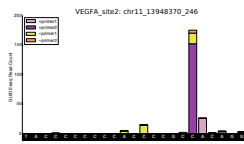
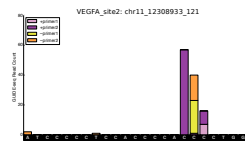
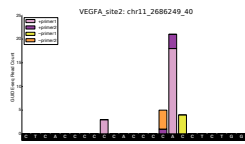
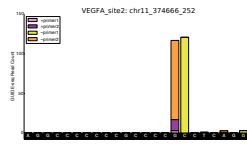
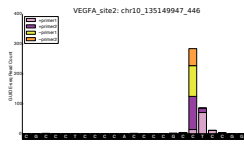
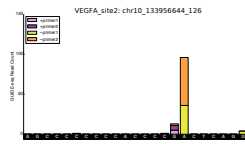
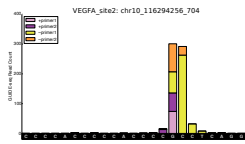
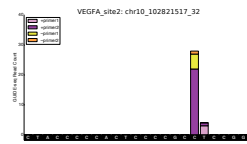
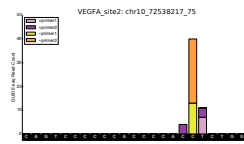
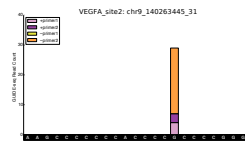
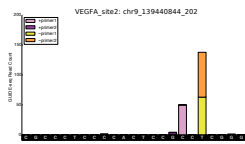


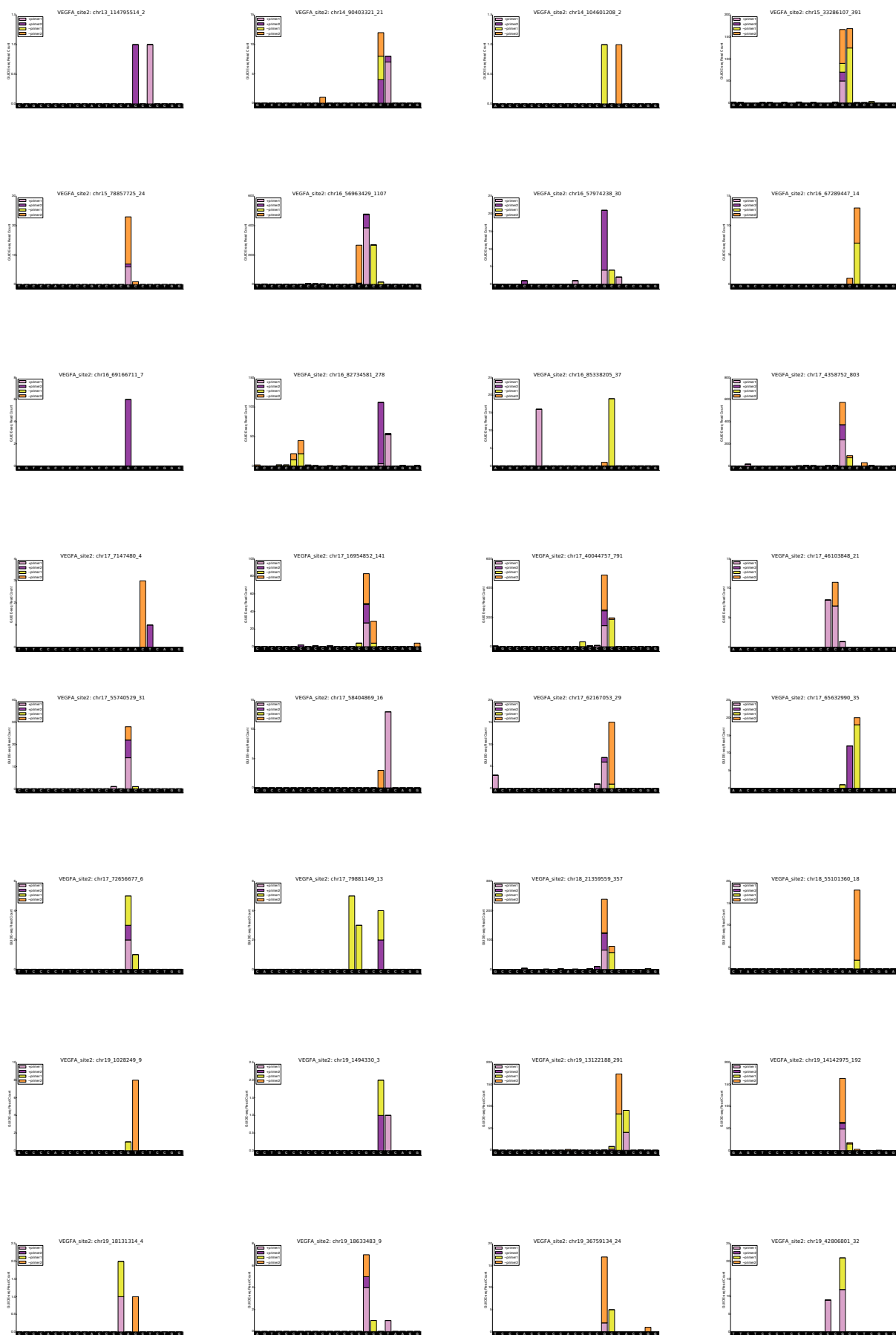


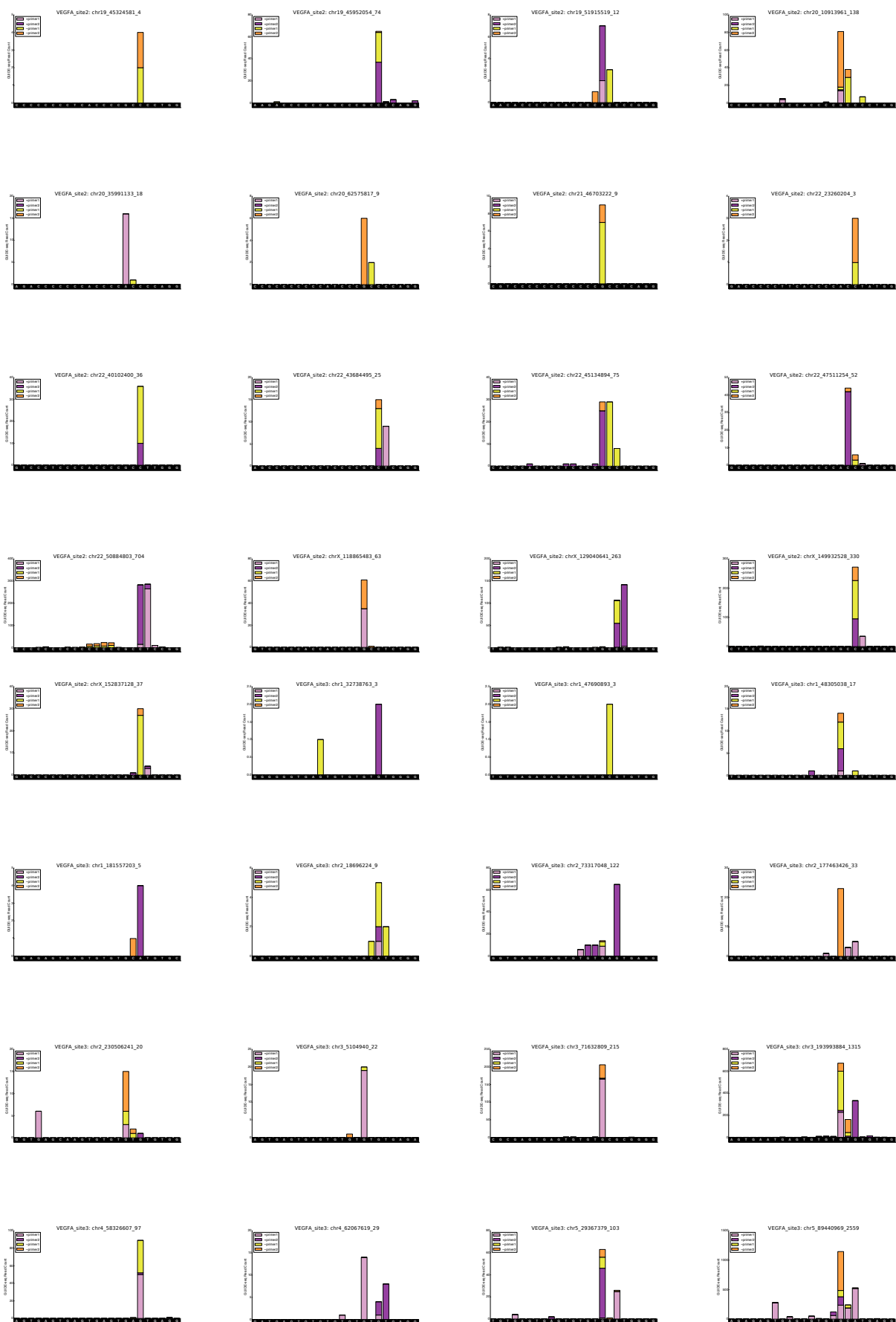


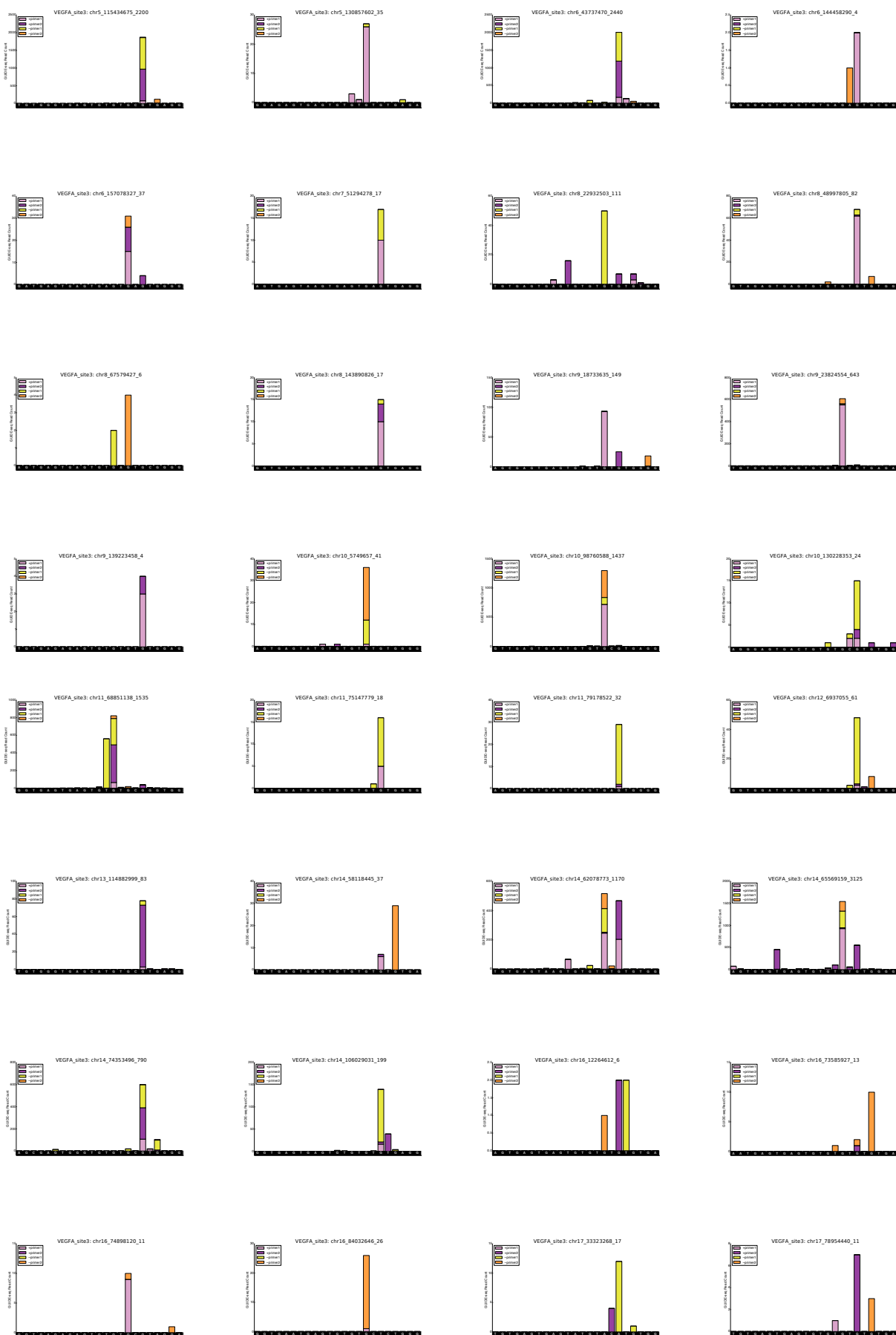


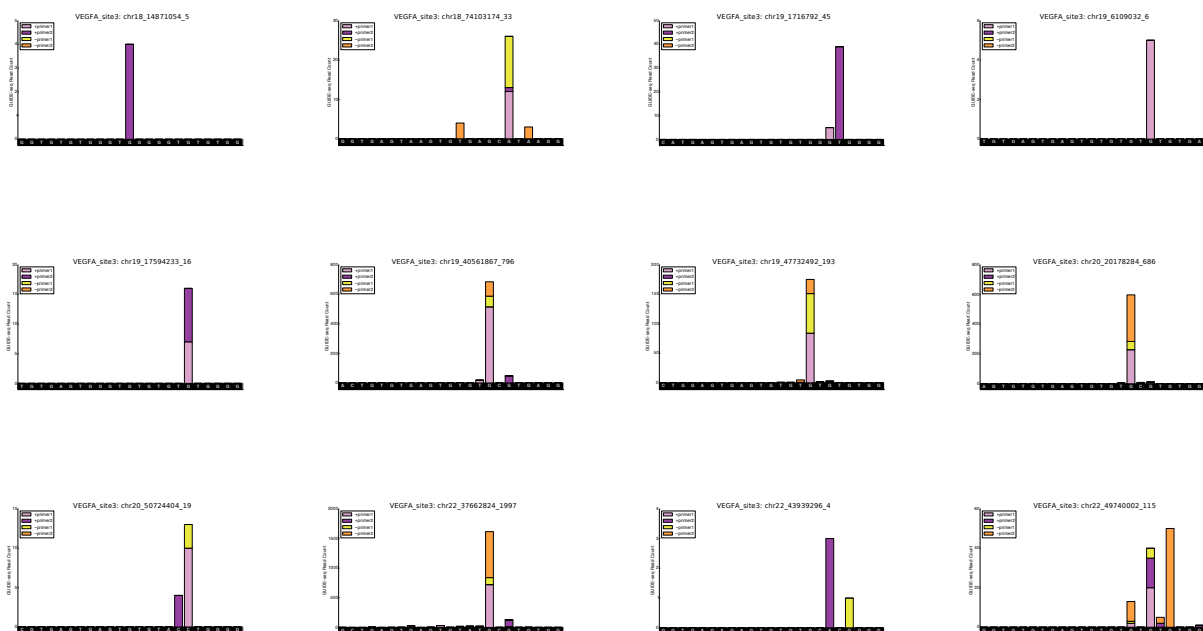






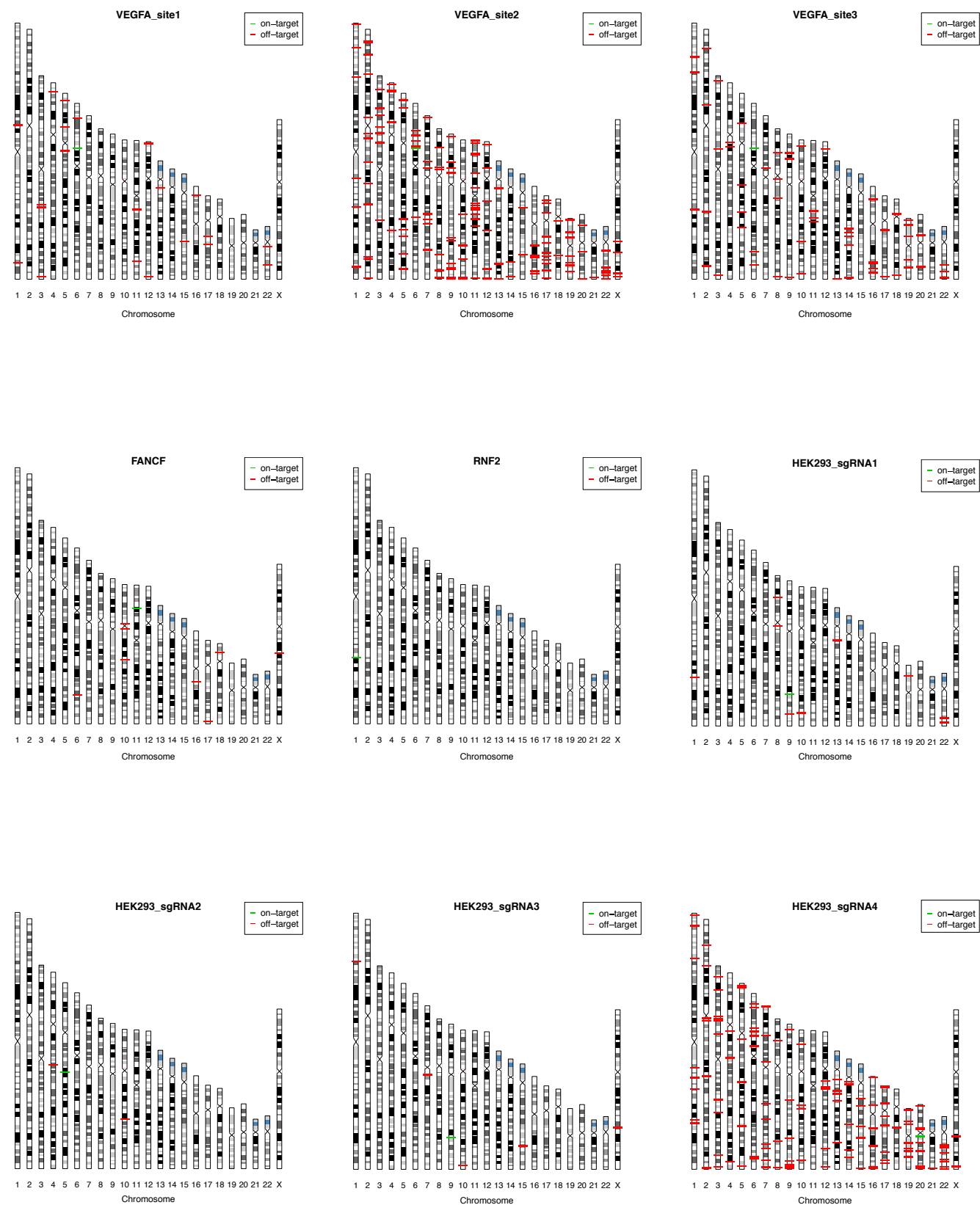




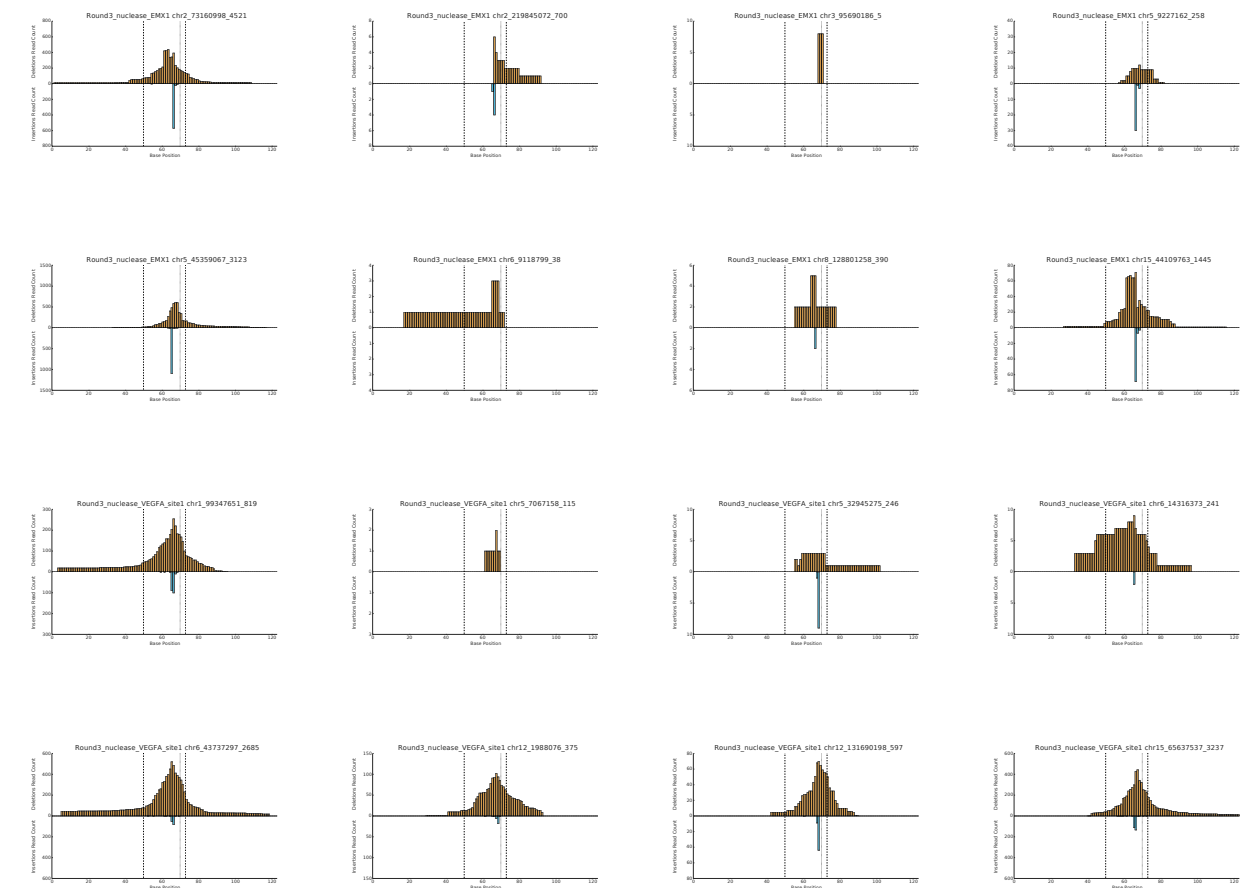


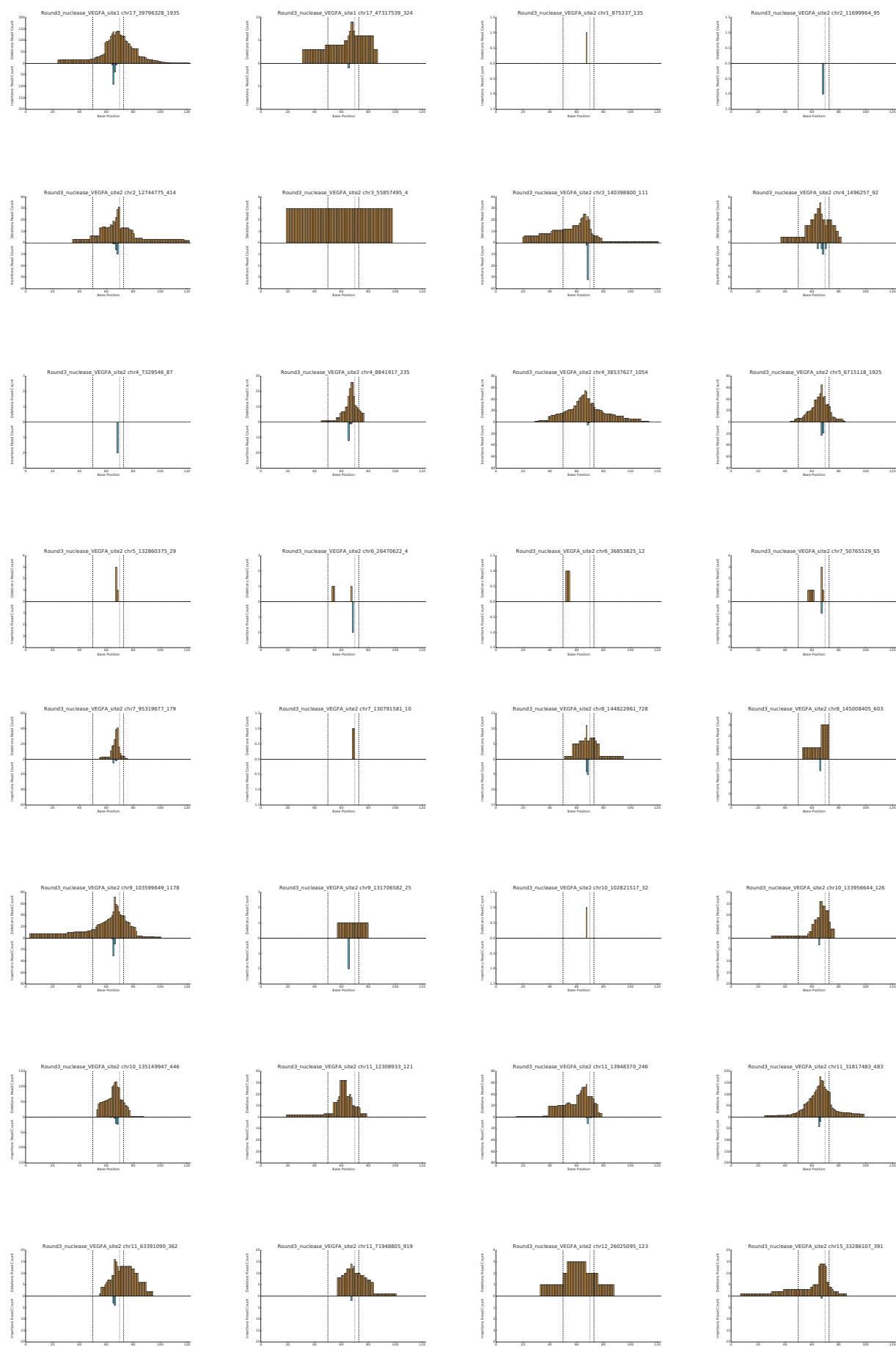
Start sites of GUIDE-Seq reads mapped back to the genome enable localization of the DSB to within a few base pairs. Mapped reads for the on- and off-target sites of the ten RGNs directed by full-length gRNAs we assessed by GUIDE-Seq are shown. In all cases, the target site sequence is shown with the 20 bp protospacer sequence to the left and the PAM sequence to the right on the x-axis.

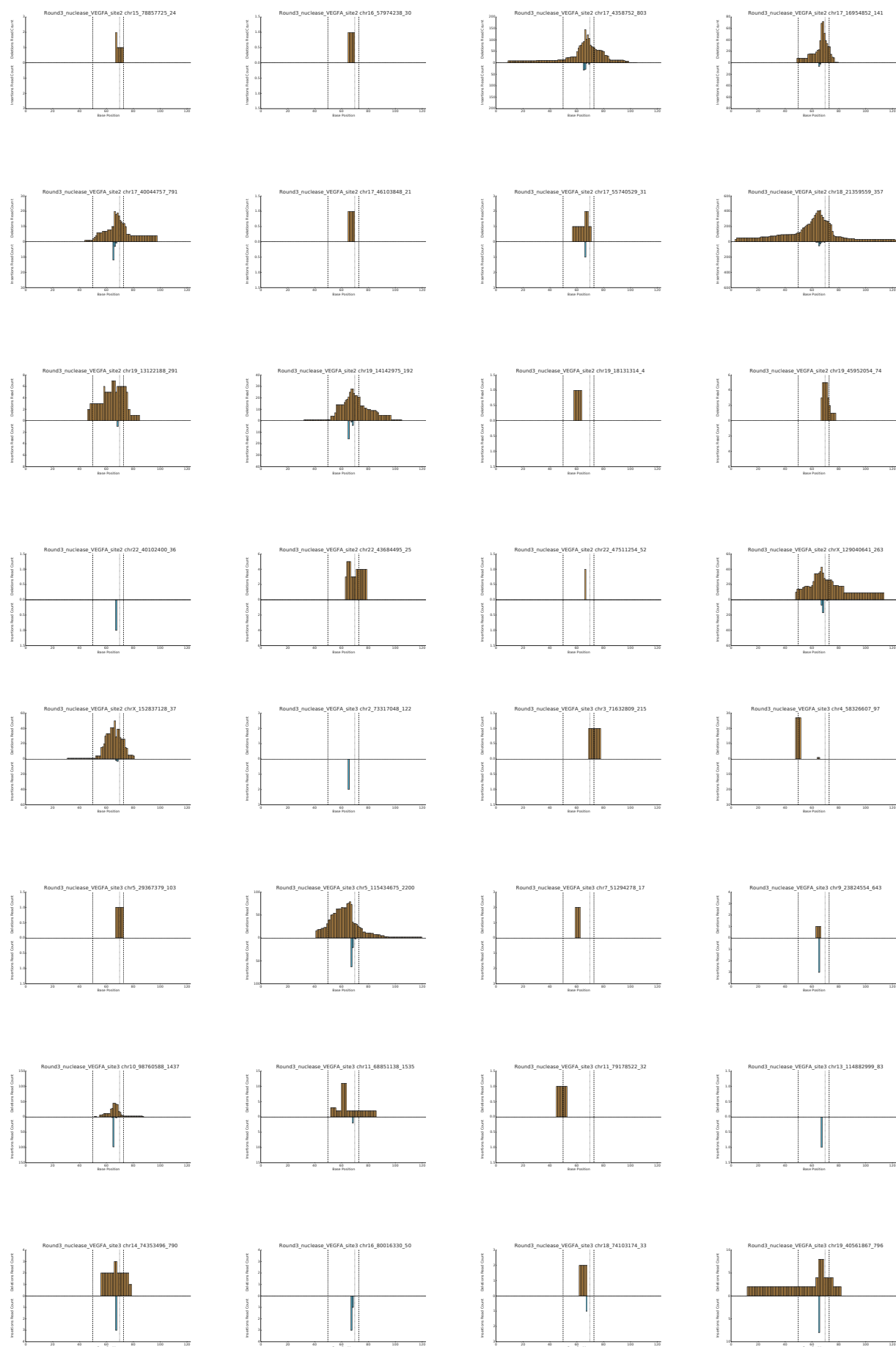
Supplementary Fig. 3 Chromosome ideograms of CRISPR/Cas9 on- and off-target sites for all ten RGNs evaluated by GUIDE-Seq

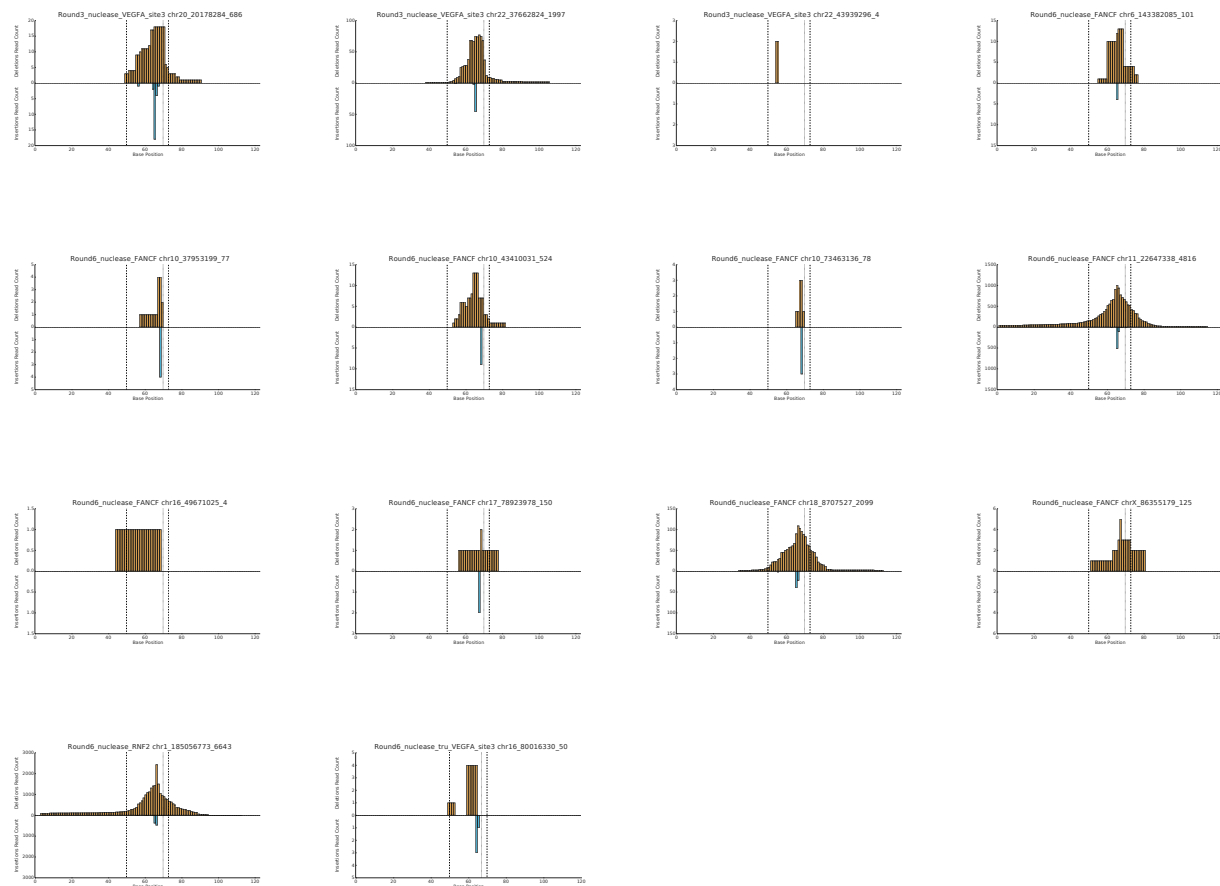


Supplementary Fig. 4 Histogram plots of mapped indel mutations for six RGN on-target and off-target sites



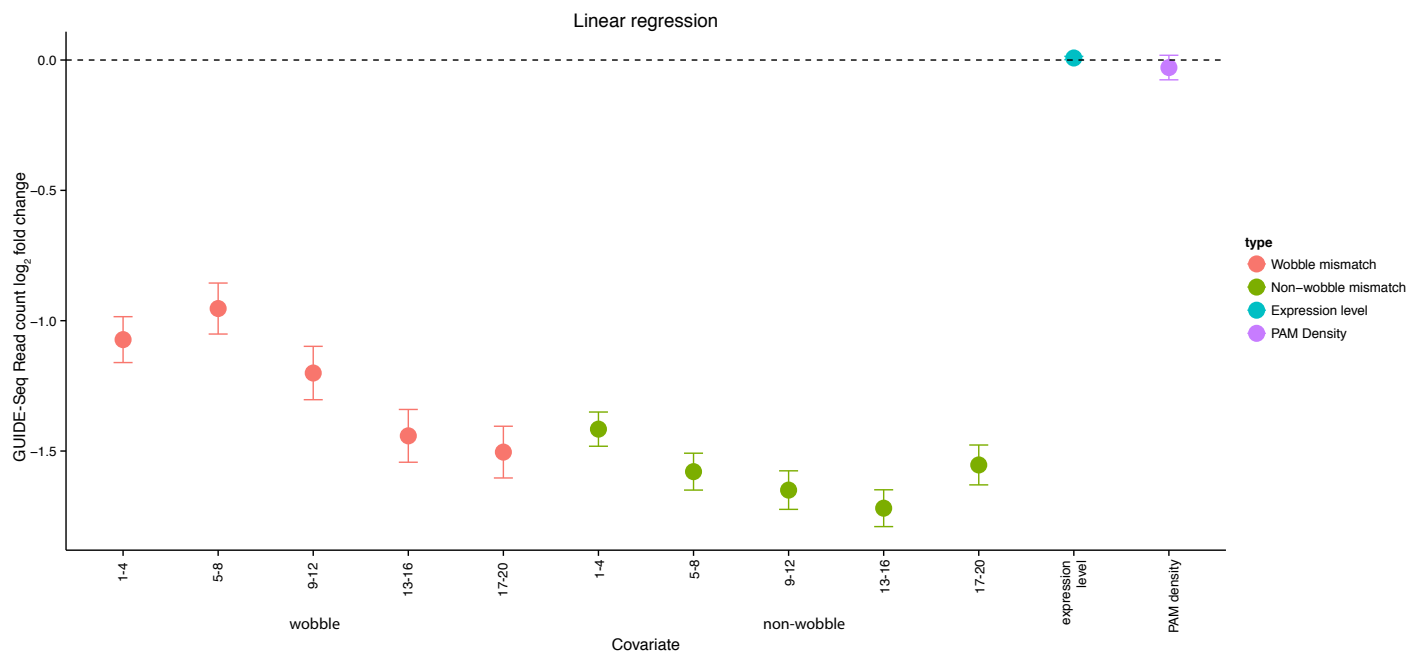




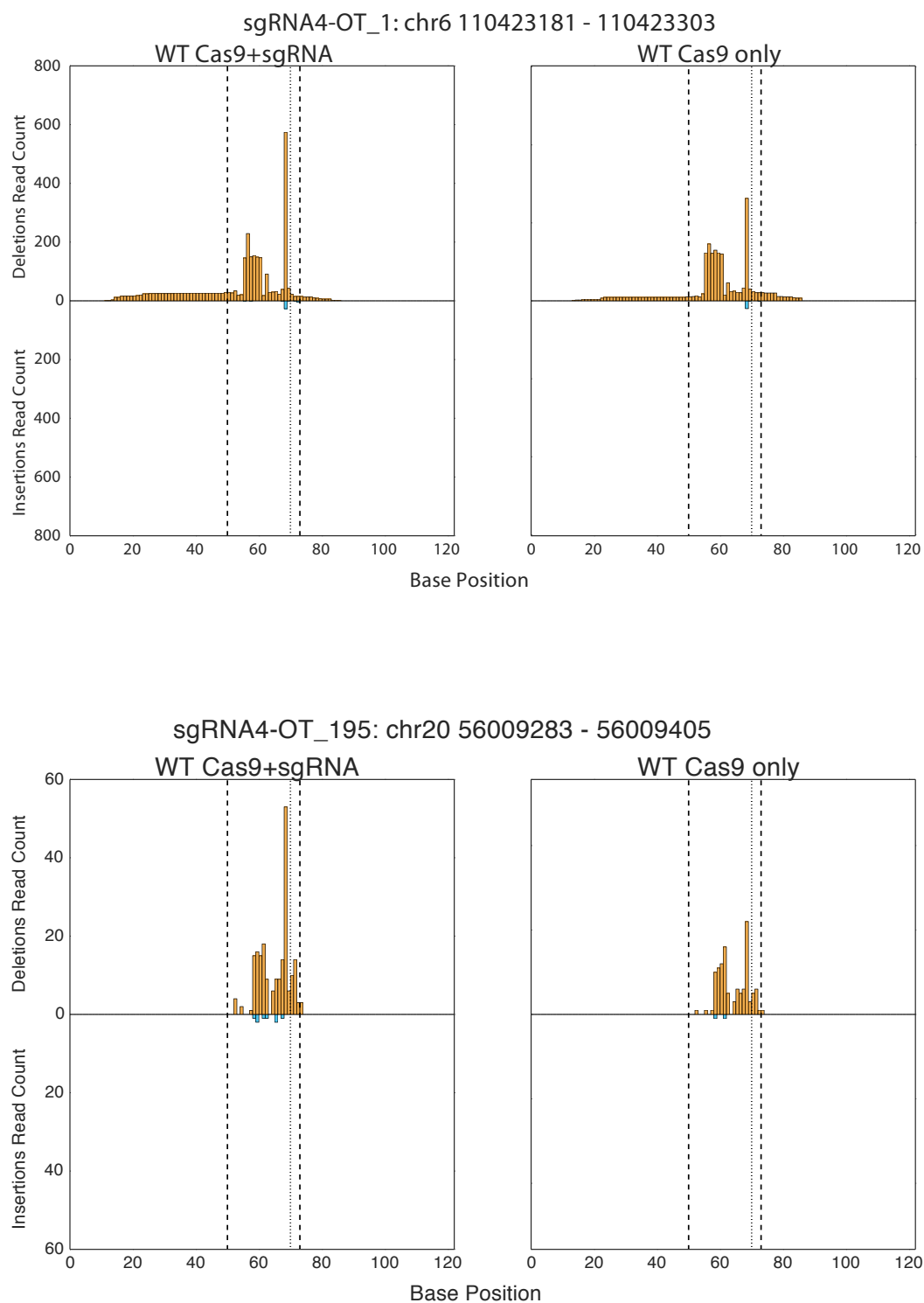


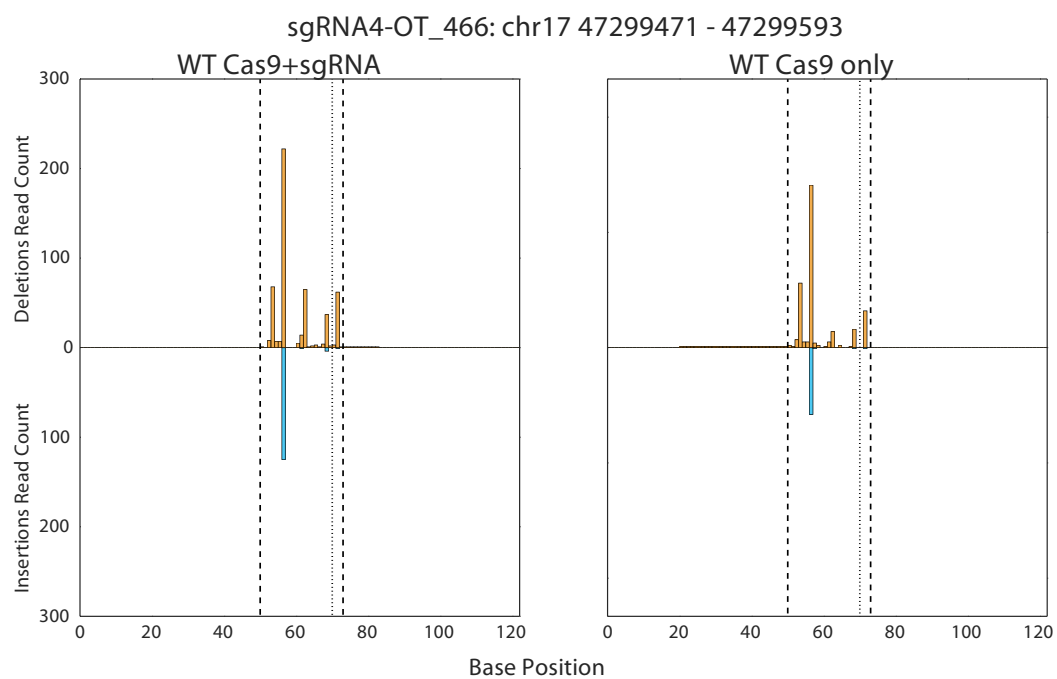
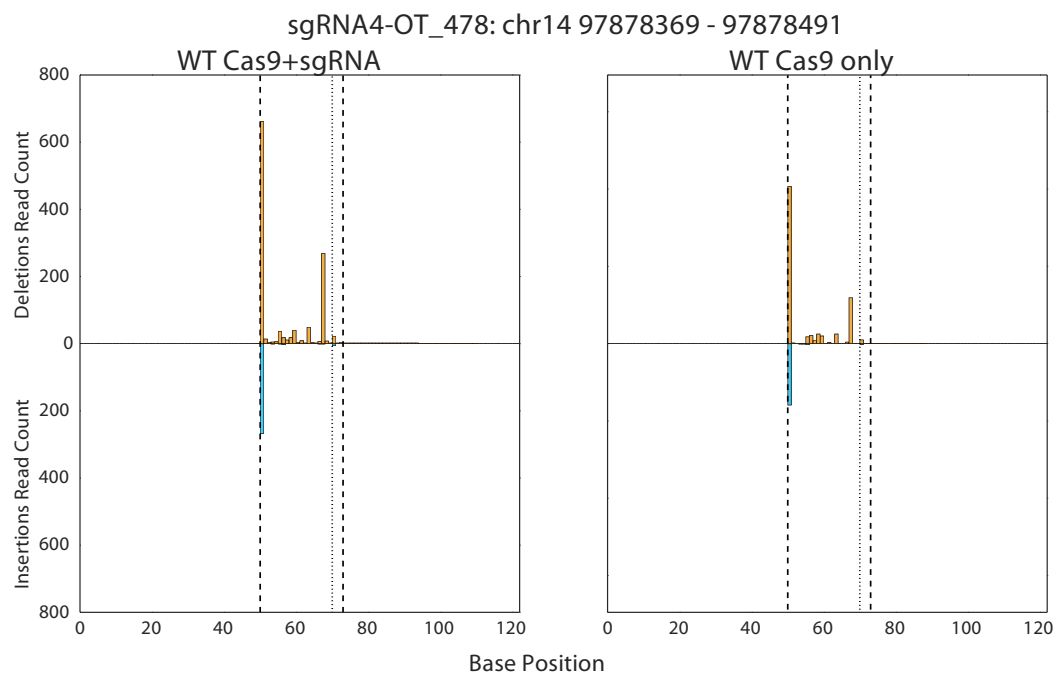
Deletions are shown above the X-axis whereas insertions are shown below. Boundaries of the overall target site (i.e., protospacer and PAM sequence) are shown with dotted lines and the boundary between the protospacer and PAM sequence is shown as a dotted line between the other two. RGN cleavage is predicted to occur 3 to 4 bps from the 5' edge of the PAM sequence.

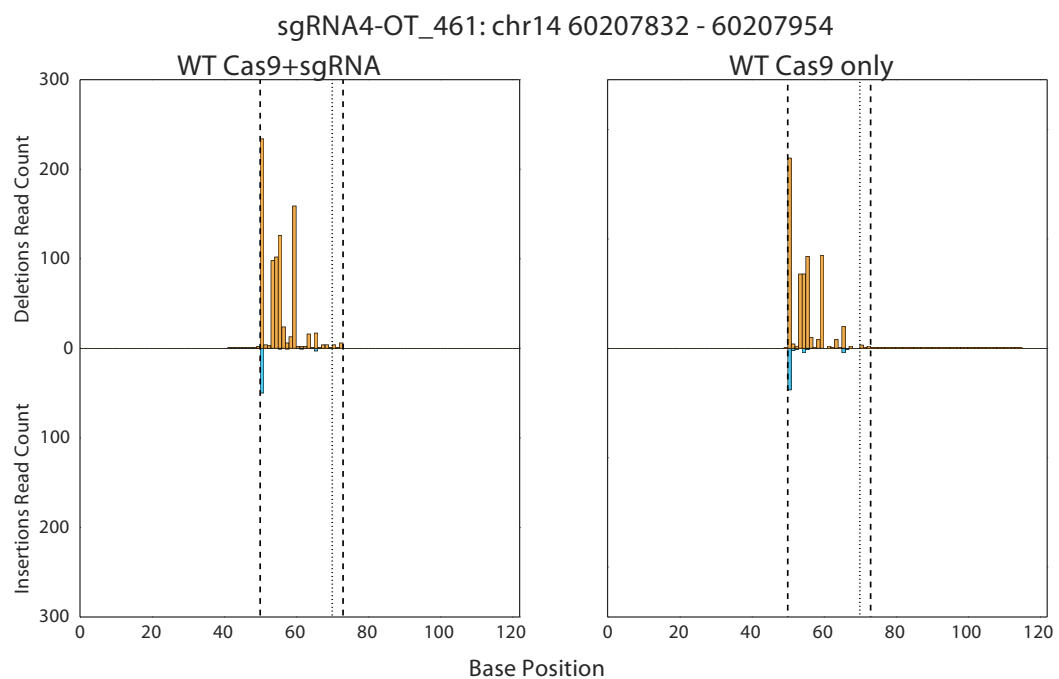
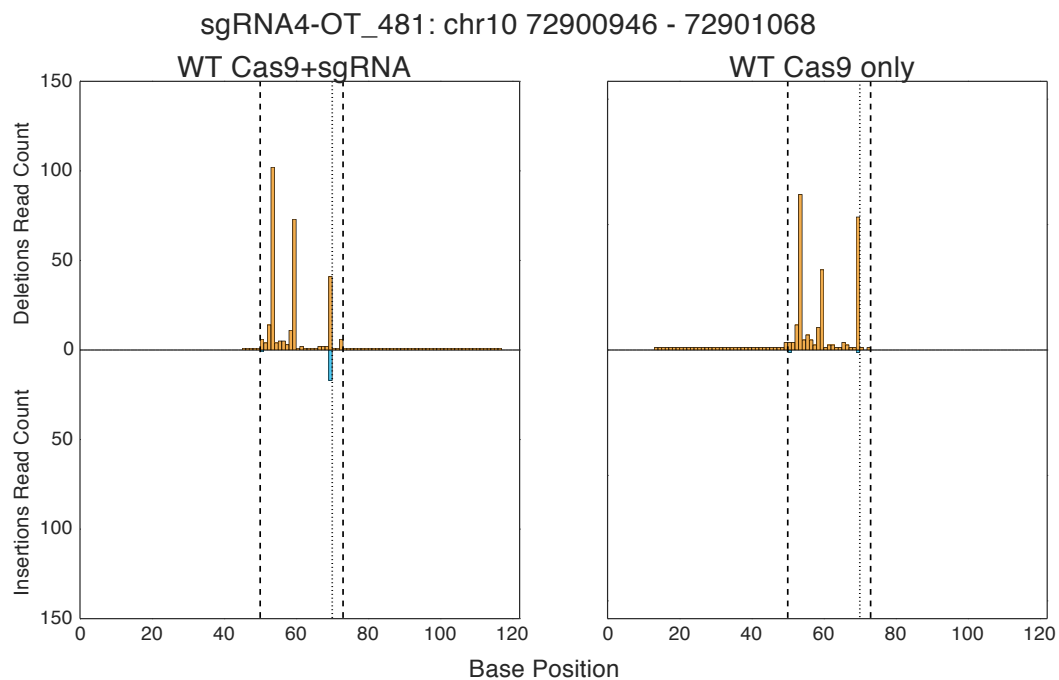
Supplementary Fig. 5 Multi-factor linear regression model to show independent effects of factors on GUIDE-Seq read count

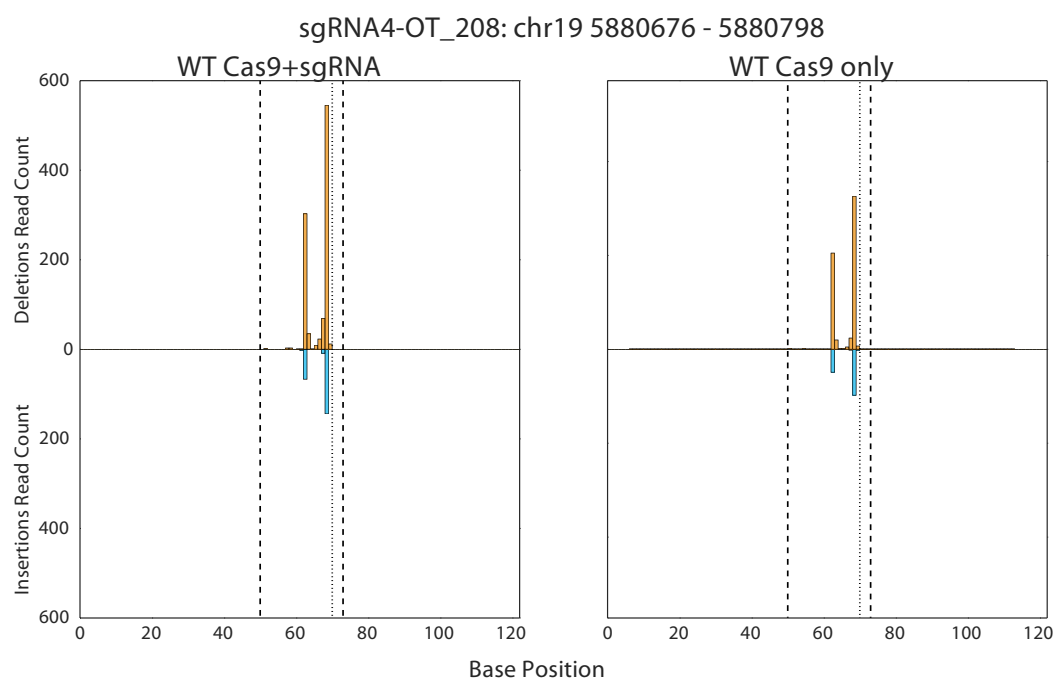


Supplementary Fig. 6 Histogram plots of mapped indel mutations for seven ChIP-Seq binding sites previously characterized as off-target cleavage sites



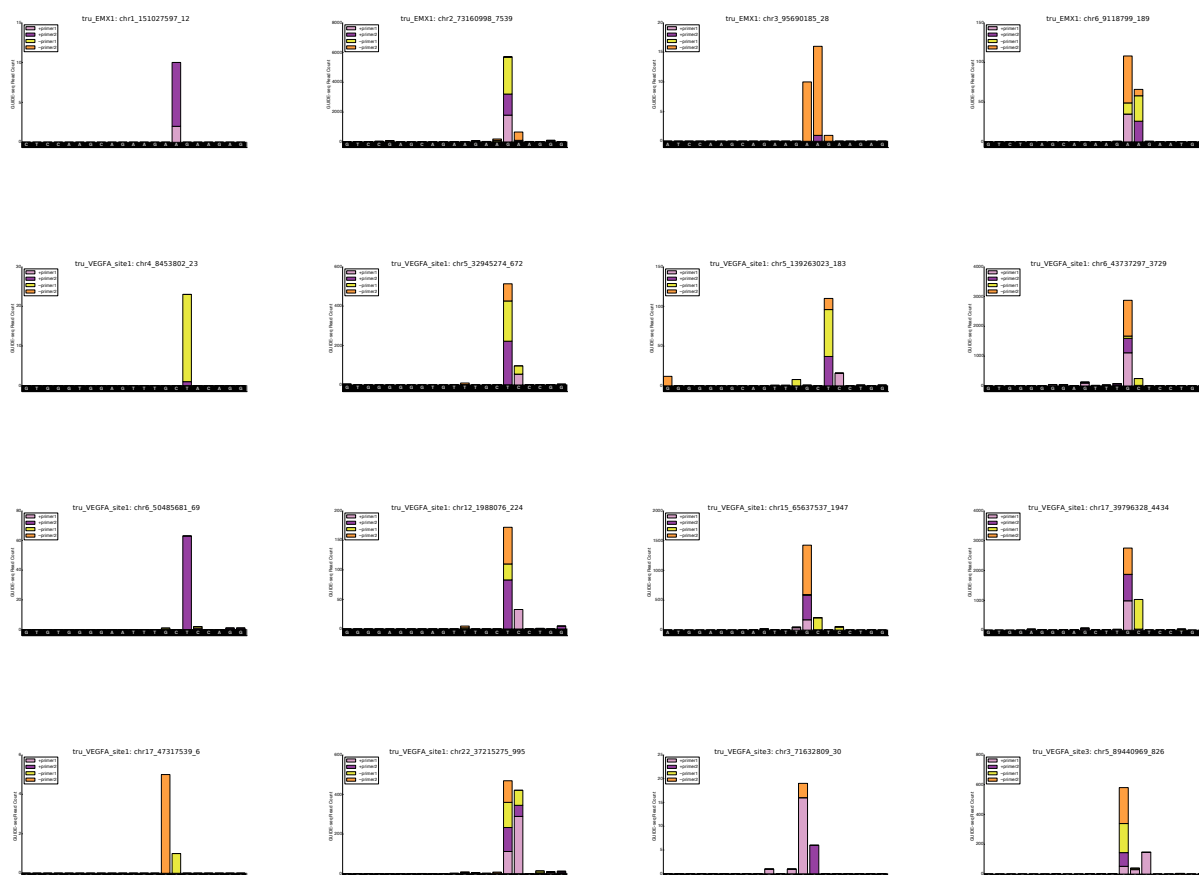


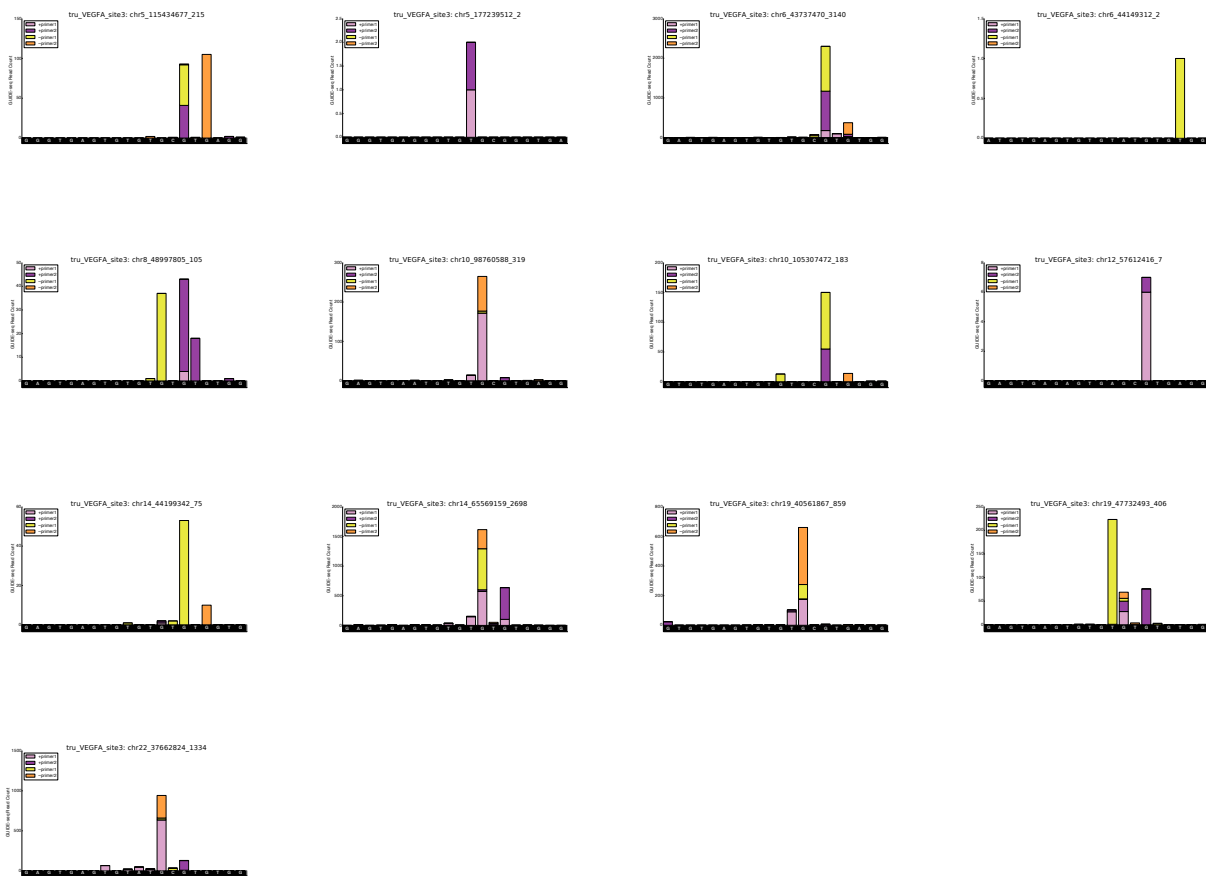




Experimental and control samples are shown side-by-side for each site. Indels and target sites are plotted as in Supplementary Fig. 4.

Supplementary Fig. 7 GUIDE-Seq-based identification of DSBs induced by RGNs directed by tru-gRNAs





Start sites of GUIDE-Seq reads mapped back to the genome enable localization of the DSB to within a few base pairs. Mapped reads for the on- and off-target sites of the three RGNs directed by truncated gRNAs we assessed by GUIDE-Seq are shown. In all cases, the target site sequence is shown with the protospacer sequence to the left and the PAM sequence to the right on the x-axis.

Supplementary Table 1 Target Sites and Cell Lines Used

Target site name	Cells	Target Site (5' → 3')
EMX1	U2OS	GAGTCCGAGCAGAAGAAGAANGG
VEGFA site1	U2OS	GGGTGGGGGGAGTTTGCTCCNGG
VEGFA site2	U2OS	GACCCCTCCACCCGCCTCNGG
VEGFA site3	U2OS	GGTGAGTGAGTGTGTGCGTGNGG
RNF2	U2OS	GTCATCTTAGTCATTACCTGNGG
FANCF	U2OS	GGAATCCCTTCTGCAGCACCNGG
HEK293 site 1	293	GGGAAAGACCCAGCATCCGTNGG
HEK293 site 2	293	GAACACAAAGCATAGACTGCNGG
HEK293 site 3	293	GGCCCAGACTGAGCACGTGANGG
HEK293 site 4	293	GGCACTGCGGCTGGAGGTGGNGG
truncated VEGFA site 1	U2OS	GTGGGGGGAGTTTGCTCCNGG
truncated VEGFA site 3	U2OS	GAGTGAGTGTGTGCGTGNGG
truncated EMX1	U2OS	GTCCGAGCAGAAGAAGAANGG

Supplementary Table 2 Genomic locations of all GUIDE-Seq detected RGN-induced cleavage sites

See attached Excel file.

Supplementary Table 3 Summary of RGN-Independent Breakpoint Hotspots in Human U2OS and HEK293 Cells

Cells	Chromosome	Start	End	Interval (bp)
U2OS	chr1	121484547	121485429	882
U2OS	chr1	236260170	236260754	584
U2OS	chr3	197900267	197900348	81
U2OS	chr4	191044096	191044100	4
U2OS	chr5	10020	10477	457
U2OS	chr7	16437577	16439376	1799
U2OS	chr7	158129486	158129491	5
U2OS	chr9	140249964	140249977	13
U2OS	chr9	140610510	140610516	6
U2OS	chr10	42599569	42599575	6
U2OS	chr11	129573467	129573469	2
U2OS	chr11	134946499	134946506	7
U2OS	chr12	95427	95683	256
U2OS	chr12	29944278	29946544	2266
U2OS	chr16	83984266	83984271	5
U2OS	chr17	63965908	63967122	1214
U2OS	chr18	63765	63769	4
U2OS	chr18	37381409	37381971	562
U2OS	chr2	9877829	9877857	28
U2OS	chr2	182140586	182140587	1
U2OS	chr2	209041635	209041637	2
U2OS	chr2	242838677	242838859	182
U2OS	chr22	49779897	49782342	2445
U2OS	chr22	49780337	49780338	1
U2OS	chrX	155260204	155260352	148
HEK293	chr1	121484526	121485404	878
HEK293	chr6	58778207	58779300	1093
HEK293	chr7	61968971	61969378	407
HEK293	chr10	42385171	42385189	18
HEK293	chr10	42400389	42400394	5
HEK293	chr10	42597212	42599582	2370
HEK293	chr19	27731978	27731991	13

Supplementary Table 4 Common Primers Needed for GUIDE-Seq

P7 Adapters		Sequence (5' → 3')
P701		CAAGCAGAAGACGGCATACGAGATTCGCCTTAGTGACTGGAGTCCTCTCTATGGGCAGTCGGTGA
P702		CAAGCAGAAGACGGCATACGAGATCTAGTACGGTGACTGGAGTCCTCTCTATGGGCAGTCGGTGA
P703		CAAGCAGAAGACGGCATACGAGATTTCTGCCTGTGACTGGAGTCCTCTCTATGGGCAGTCGGTGA
P704		CAAGCAGAAGACGGCATACGAGATGCTCAGGAGTGACTGGAGTCCTCTCTATGGGCAGTCGGTGA
P705		CAAGCAGAAGACGGCATACGAGATAGGAGTCCGTGACTGGAGTCCTCTCTATGGGCAGTCGGTGA
P706		CAAGCAGAAGACGGCATACGAGATCATGCCTAGTGACTGGAGTCCTCTCTATGGGCAGTCGGTGA
P707		CAAGCAGAAGACGGCATACGAGATGTAGAGAGGTGACTGGAGTCCTCTCTATGGGCAGTCGGTGA
P708		CAAGCAGAAGACGGCATACGAGATCCTCTCTGGTGACTGGAGTCCTCTCTATGGGCAGTCGGTGA
P5 Adapter Primers		Sequence (5' → 3')
P5_1		AATGATACGGCGACCACCGAGATCTA
P5_2		AATGATACGGCGACCACCGAGATCTACAC
Custom Sequencing Primers		Sequence (5' → 3')
Index1		ATCACCGACTGCCCATAGAGAGGACTCCAGTCAC
Read2		GTGACTGGAGTCCTCTCTATGGGCAGTCGGTGAT
Illumina Y-adapters 1-16 (with Molecular Index tag NNWNNWNN)		Sequence (5' → 3')
Miseq Common Adapter		[Phos]GATCGGAAGAGC*C*A
A01		AATGATACGGCGACCACCGAGATCTACACTAGATCGCANNWNNWNNACACTCTTTCCCTACACGACGCTCTTCCGATC*T
A02		AATGATACGGCGACCACCGAGATCTACACCTCTCTATNNWNNWNNACACTCTTTCCCTACACGACGCTCTTCCGATC*T
A03		AATGATACGGCGACCACCGAGATCTACACTATCCTCTNNWNNWNNACACTCTTTCCCTACACGACGCTCTTCCGATC*T
A04		AATGATACGGCGACCACCGAGATCTACACAGAGTAGANNWNNWNNACACTCTTTCCCTACACGACGCTCTTCCGATC*T
A05		AATGATACGGCGACCACCGAGATCTACACGTAAGGAGNNWNNWNNACACTCTTTCCCTACACGACGCTCTTCCGATC*T
A06		AATGATACGGCGACCACCGAGATCTACACACTGCATANNWNNWNNACACTCTTTCCCTACACGACGCTCTTCCGATC*T
A07		AATGATACGGCGACCACCGAGATCTACACAAGGAGTANNWNNWNNACACTCTTTCCCTACACGACGCTCTTCCGATC*T
A08		AATGATACGGCGACCACCGAGATCTACACCTAAGCCTNNWNNWNNACACTCTTTCCCTACACGACGCTCTTCCGATC*T
A09		AATGATACGGCGACCACCGAGATCTACACGACATTGTNNWNNWNNACACTCTTTCCCTACACGACGCTCTTCCGATC*T
A10		AATGATACGGCGACCACCGAGATCTACACACTGATGGNNWNNWNNACACTCTTTCCCTACACGACGCTCTTCCGATC*T
A11		AATGATACGGCGACCACCGAGATCTACACGTACCTAGNNWNNWNNACACTCTTTCCCTACACGACGCTCTTCCGATC*T
A12		AATGATACGGCGACCACCGAGATCTACACCAGAGCTANNWNNWNNACACTCTTTCCCTACACGACGCTCTTCCGATC*T

Illumina Y-adapters 1-16 (with Molecular Index tag NNWNNWNN)		Sequence (5' → 3')
A13		AATGATACGGCGACCACCGAGATCTACACCATAGTGANNWNNWNNACACTCTTTCCCTACACGACGCTCTTCCGATC*T
A14		AATGATACGGCGACCACCGAGATCTACACTACCTAGTNNWNNWNNACACTCTTTCCCTACACGACGCTCTTCCGATC*T
A15		AATGATACGGCGACCACCGAGATCTACACCGCATATNNWNNWNNACACTCTTTCCCTACACGACGCTCTTCCGATC*T
A16		AATGATACGGCGACCACCGAGATCTACACTGGATTGTNNWNNWNNACACTCTTTCCCTACACGACGCTCTTCCGATC*T

*** Indicates a Phosphorothioate Bond Modification**

Primer Name	Sequence (5' → 3')	Strand/Direction
Nuclease_off+_GSP1	GGATCTCGACGCTCTCCCTGTTTAATTGAGTTGTCATATGTTAATAAC	+
Nuclease_off_-_GSP1	GGATCTCGACGCTCTCCCTATACCGTTATTAACATATGACA	-
Nuclease_off+_GSP2	CCTCTCTATGGGCAGTCGGTGATACATATGACAACTCAATTAAAC	+
Nuclease_off_-_GSP2	CCTCTCTATGGGCAGTCGGTGATTGAGTTGTCATATGTTAATAACGGTA	-

Supplementary Results

Our re-analysis of sequencing data from the previous study of Kuscu et al.¹ suggested that the vast majority of indel mutations found at those sites were more likely caused by PCR or sequencing errors rather than by RGN cleavage activity. Kuscu et al. performed genome-wide ChIP-Seq on cells expressing dCas9 and a gRNA they refer to as sgRNA4. sgRNA 4 is the same as the gRNA we use in our current study that targets what we call HEK293 site 4. Kuscu et al. identified seven ChIP-Seq binding sites at which they reported that active Cas9 directed by sgRNA4 could induce indel mutations. Based on their observations, they concluded that these seven sites represent *bona fide* off-target Cas9 cleavage sites for sgRNA4.

However, our independent re-analysis of the Kuscu et al. data strongly suggests that the indel mutations they observed are unlikely to have been induced by Cas9 and sgRNA4. For many of the sites with statistically significant differences (FDR < 0.05) in indel percentages between their control and Cas9/sgRNA4 samples, the distribution of indels is not consistent with a Cas9-induced cleavage event (**Supplementary Fig. 6**). Cas9-mediated cleavage would be expected to introduce indels that primarily localize within 3 to 4 bps of the 5' edge of the PAM sequence. However, for nearly all of the seven sites, many of the indels localize to other parts of the protospacer. Examination of the protospacer sequences suggests that the observed indels may arise due the presence of homopolymer stretches, a known artifact with the Illumina next-generation sequencing platform. In addition, the patterns of indels at all seven of these sites are nearly identical between the control and the nuclease samples, again consistent with the idea that these indels are largely, if not wholly, sequencing artifacts. We conclude that it is unlikely that these seven sites represent *bona fide* Cas9/sgRNA4 cleavage sites.

1. Kuscu, C., Arslan, S., Singh, R., Thorpe, J. & Adli, M. Genome-wide analysis reveals characteristics of off-target sites bound by the Cas9 endonuclease. *Nat Biotechnol* **32**, 677-683 (2014).

Supplementary Discussion

Potential reasons for failure to detect two previously described off-target sites of tru-gRNA-directed RGNs

One of the two tru-gRNA off-target sites not detected by GUIDE-Seq showed evidence of capture in our raw data but was filtered out by our read calling algorithm because the sequencing reads were only unidirectional and originated from just one primer (**Online Methods**). (The lack of bidirectional mapping reads for this site might be due to a repetitive region on one side of the off-target site that makes it challenging to map the reads accurately.) The other undetected tru-gRNA off-target site has been previously reported to be mutagenized at a low frequency and therefore it is possible that it was not detected in our experiment due to insufficient sampling of sequences.

Supplementary Methods – Detailed GUIDE-Seq Protocol

Store at Room Temperature

Item	Vendor
Covaris S220 microTube,	Covaris
Ethanol, 200-proof (100%)	Sigma Aldrich
MicroAmp Optical 96-well Plates	Applied Biosystems
Nuclease-free H ₂ O	Promega
Qubit Assay Tubes, 500 tubes/pack	Invitrogen
Qubit dsDNA BR Kit – 500 Assays	Invitrogen
- dsDNA BR Reagent	*store component at RT, protect from light
- dsDNA BR Buffer	*store at RT
- dsDNA Standard 1: 0 ng/μL	*store at 4°C
- dsDNA Standard 2: 100 ng/μL	*store at 4°C
TMAC Buffer, 5M	Sigma Aldrich
-Tetramethylammonium Chloride	
1X TE Buffer/10mM Tris-HCl, pH 8.0	Invitrogen
UltraPure 0.5M EDTA, pH 8.0 (Gibco) (4x100 mL)	Life Technologies

Store at 4°C

Item	Vendor
Agencourt AMPure XP Beads- 60 mL	Beckman Coulter

Store at -20°C

Item	Catalog #
25mM dNTP Solution Mix	Enzymatics, Inc.
Slow ligation buffer	Enzymatics, Inc.
End-repair mix (low concentration)	Enzymatics, Inc.
T4 DNA Ligase	Enzymatics, Inc.
- 10X T4 DNA Ligase Buffer (Slow Ligation Buffer)	
Platinum® Taq DNA Polymerase	Life Technologies
-10X PCR Buffer (no MgCl ₂ +) - 50mM MgCl₂+	
qPCR Illumina Library Quantification Kits	KAPA Biosystems, Inc.

Equipment

96-well Plate Magnetic Stand	Invitrogen
Qubit Fluorometer 2.0	Life Technologies
Covaris S-2 Focused Ultra-sonicator™ Instrument	Covaris
Tabletop centrifuge	Thermo Scientific
Tabletop vortexer	Thermo Scientific
Thermocycler	Eppendorf
Miseq	Illumina

Protocol for Guide-SEQ Library Preparation

Y-adapter Preparation

The Y-adapter is made by annealing the Miseq common oligo with each of the sample barcode adapters (A01 to A16, see **Supplementary Table 4**). The adapters also contain 8-mer NNWNNWNN (N = A, C, T, or G; W = A or T) molecular indexes.

1X TE Buffer	80.0 μ L
A## (100 μ M)	10.0 μ L
MiSeq Common Adapter_MI (100 μ M)	10.0 μ L
Total	100.0 μL

Annealing program: 95°C for 1 s; 60°C for 1s; slow ramp down (approximately -2°C/min) to 4°C; hold at 4°C. Store in -20°C.

Input Quantification and Shearing

1. dsDNA is quantified by Qubit and brought to a final volume of 120 μ L using 1X TE Buffer.
2. Each sample is sheared to an average length of 500 bp according to the standard operating protocol for the Covaris S2.
3. A cleanup with 120 μ L of AMPure XP SPRI beads (1X ratio) is performed according to manufacturer protocol, and eluted in 15 μ L of 1X TE Buffer.

End repair, A-tailing and Ligation

End Repair

4. To a 200 μ L PCR tube or well in a 96-well plate, add the following (per reaction):

Nuclease-free H ₂ O	0.5 μ L
dNTP mix, 5mM	1.0 μ L
SLOW Ligation Buffer, 10X	2.5 μ L
End-repair mix (low concentration)	2.0 μ L
Buffer for Taq Polymerase, 10X (Mg ²⁺ + free)	2.0 μ L
Taq Polymerase (non-hot start)	0.5 μ L
Total	8.5 μL
+ DNA sample (from Step 3)	14.0 μ L
Total	22.5 μL

End Repair Thermocycler Program: 12°C for 15min, 37°C for 15min; 72°C for 15min; hold at 4°C

Adapter Ligation

5. To the sample reaction tube or well, add the following reagents *in order* (mix by pipetting):

Annealed Y adapter_MI (10 μ M)	1.0 μ L
T4 DNA Ligase	2.0 μ L
+ DNA sample (from step #3)	22.5 μ L
Total	25.5 μL

Adapter Ligation Thermocycler Program: 16°C for 30min, 22°C for 30min, hold at 4°C

6. 0.9X SPRI clean (22.95 μ L Ampure XP beads), elute in 12 μ L of 1X TE buffer.

PCRs

PCR 1 (oligo tag primer [Discovery] or large primer pool [Deep-sequencing Validation])

7. Prepare the following master mix:

Nuclease-free H ₂ O	11.9 µL
Buffer for Taq Polymerase, 10X (Mg ²⁺ + free)	3.0 µL
dNTP mix, 10mM	0.6 µL
Mg ²⁺ , 50 mM	1.2 µL
Platinum Taq polymerase, 5 U/µl	0.3 µL
GSP1 Primer (10uM)/Primer Pool (*)	1.0 µL**
TMAC (0.5M)	1.5 µL
P5_1, 10 µM	0.5 µL
Total	20.0 µL
+ DNA sample (from Step 6)	10.0 µL
Total	30.0 µL

* For Discovery, make separate master mixes for +/(sense) and –/(antisense) reactions, and proceed with separate PCR reactions.

* For deep-sequencing validation, one master mix can be made. Primer Pool should be normalized to a total amount of 30 pmol in the 30 ul reaction.

Thermocycler Program: 95°C for 5 min, 14 cycles of [95°C for 30 s, 20% ramping down to 65°C, 65°C for 5 min], 72°C for 5 min, 4°C hold

8. 1.2X SPRI clean (36.0 uL), elute in 15 ul of 1X TE Buffer.

PCR 2 (oligo tag primer [Discovery] or large primer pool [Deep-sequencing Validation])

9. Prepare the following master mix:

Nuclease-free H ₂ O	5.4 µL
Buffer for Taq Polymerase, 10X (Mg ²⁺ + free)	3.0 µL
dNTP mix, 10mM	0.6 µL
Mg ²⁺ , 50 mM	1.2 µL
Platinum Taq polymerase, 5 U/µl	0.3 µL
GSP2 Primer (10uM)/Primer Pool (*)	1.0 µL
TMAC (0.5M)	1.5 µL
P5_2, 10 µM	0.5 µL
Total	13.5 µL
+ P7_# (10uM)*	1.5 µL
+ DNA sample with beads (from Step 8)	15.0 µL
Total	30.0 µL

* Primer concentrations should follow the specifications described in PCR1

* For the P7_#, at least 4 should be used in one sequencing run for good image registration on Illumina sequencer (e.g. P701 – P704 or P705 – P708)

Thermocycler Program: 95°C for 5 min, 14 cycles of [95°C for 30 s, 20% ramping down to 65°C, 65°C for 5 min], 72°C for 5 min, 4°C hold

10. 0.7X SPRI clean (21.0 uL), elute in 30 ul of 1X TE Buffer.

Library quantification by qPCR and Sequencing

qPCR Quantification

11. Quantitate library using Kapa Biosystems kit for Illumina Library Quantification kit, according to manufacturer instruction.

Normalization and Sequencing

12. Using the mean quantity estimate of number of molecules per uL given by the qPCR run for each sample, proceed to normalize the total set of libraries to 1.2×10^{10} molecules, divided by the number of libraries to be pooled together for sequencing. This will give a by molecule input for each sample, and also a by volume input for each sample.

After pooling, Speedvac (dry down) the library with a Vacufuge to a final volume of 10 uL for sequencing.

Denature the library and load onto the Miseq according to Illumina's standard protocol for sequencing with an Illumina Miseq Reagent Kit V2 - 300 cycle (2 x 150 bp paired end), except:

- 1) Add 3 ul of 100 μ M custom sequencing primer Index 1 to Miseq Reagent cartridge position 13 (Index Primer Mix). Add 3 ul of 100 μ M custom sequencing primer Read 2 to Miseq Reagent cartridge position 14 (Read 2 Primer Mix).
- 2) Sequence with the following number of cycles "151 | 8 | 16 | 151" with the paired-end Nextera sequencing protocol.

Submit sequencing data in either bcl or fastq format to relevant pipelines for downstream bioinformatics analysis.