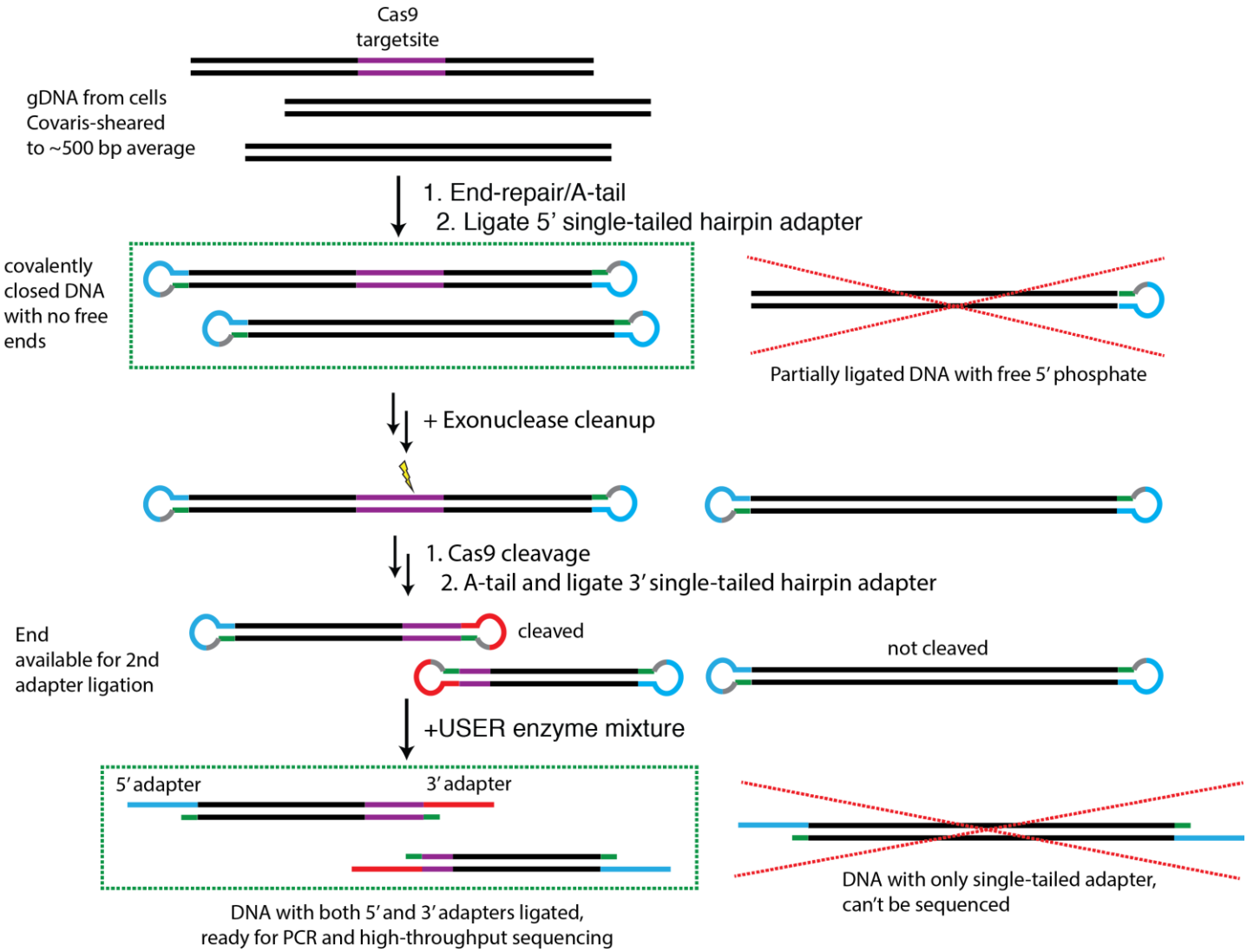
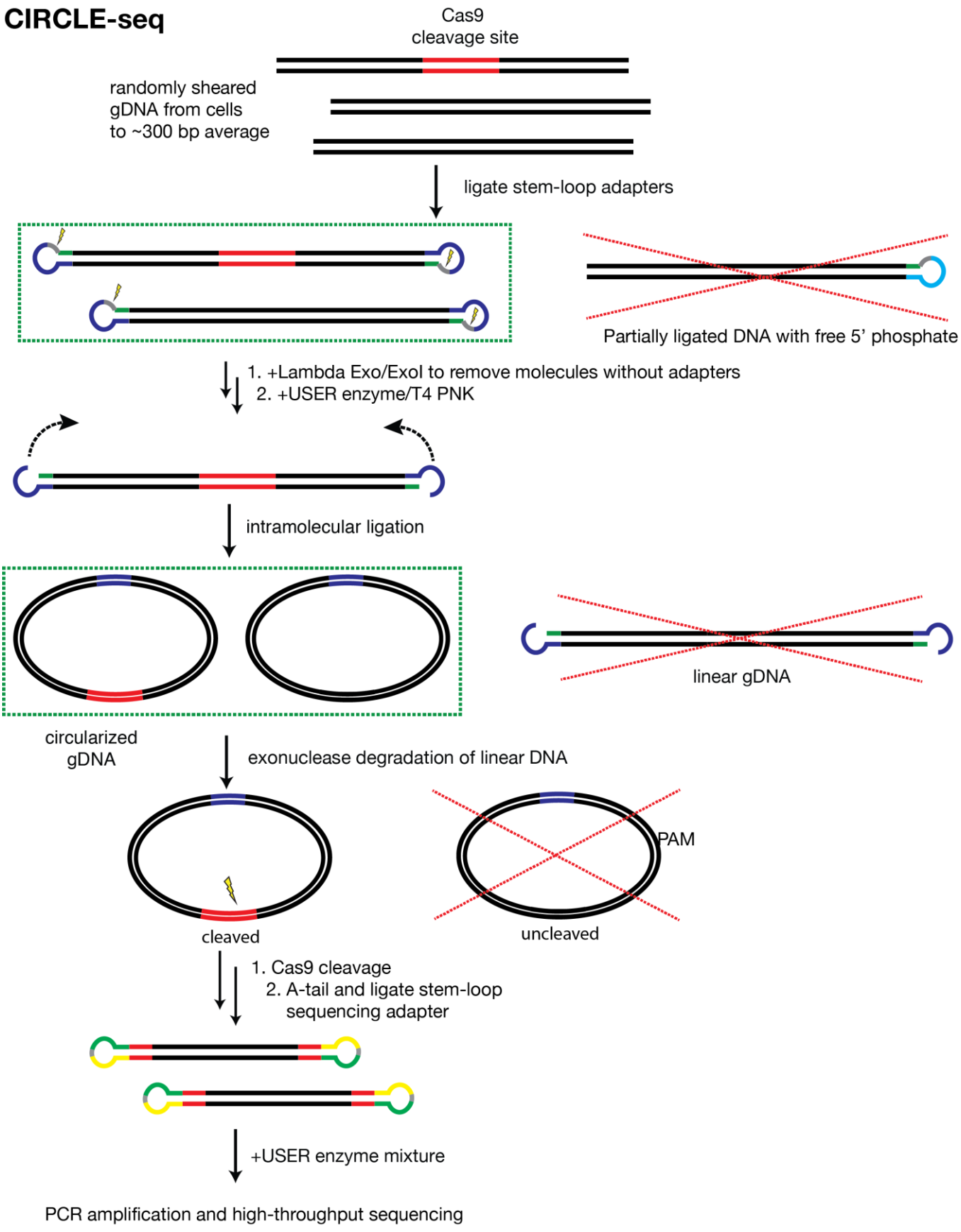
linear stem loop method



Supplementary Figure 1

Detailed schematic overview of linear stem-loop method

Genomic DNA is randomly sheared to an average of ~300 bp, end-repaired, A-tailed, and ligated to uracil-containing stem loop adapter 1. Covalently closed DNA molecules with stem-loop adapters ligated to both ends are selected for by treatment with a mixture of Lambda exonuclease and *E. coli* Exonuclease I, and then treated with Cas9-sgRNA complex. Cleaved molecules will have a newly available end for subsequent ligation of stem-loop adapter 2. Ligation of both stem loop adapters provides required 5' and 3' sequences for PCR and high-throughput sequencing.

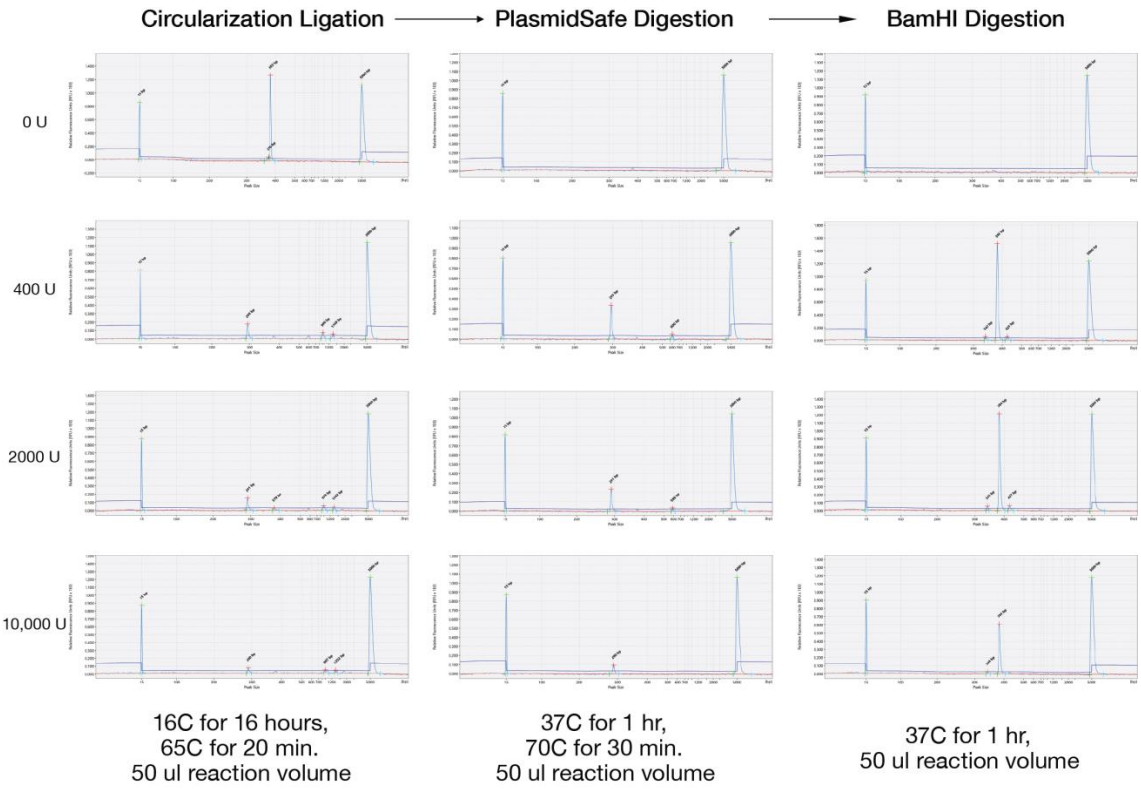


Nature Methods: doi:10.1038/nmeth.4278

Supplementary Figure 2

Detailed schematic overview of CIRCLE-seq method.

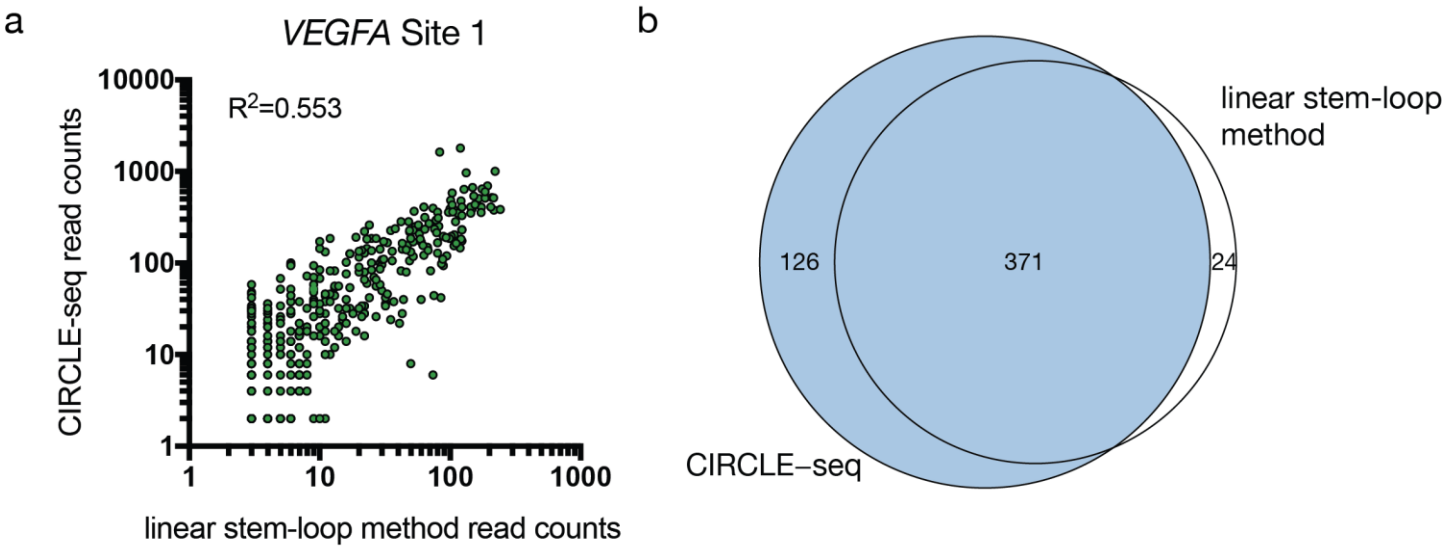
Genomic DNA is randomly sheared to an average of ~300 bp, end-repaired, A-tailed, and ligated to uracil-containing stem-loop adapters. Covalently closed DNA molecules with stem-loop adapters ligated to both ends are selected for by treatment with a mixture of Lambda exonuclease and *E. coli* Exonuclease I. 4 bp overhangs are released with a mixture of USER enzyme and T4 PNK, and DNA molecules are circularized at low concentrations favoring intramolecular ligation. Unwanted linear DNA is degraded with Plasmid-Safe ATP-dependent DNase. Circular DNA is treated with Cas9–sgRNA complex and cleaved, linearized DNA is ligated to sequencing adapters and amplified for high-throughput sequencing.



Supplementary Figure 3

Optimization of *in vitro* circularization conditions with uracil-containing stem loop adapters and a PCR amplicon.

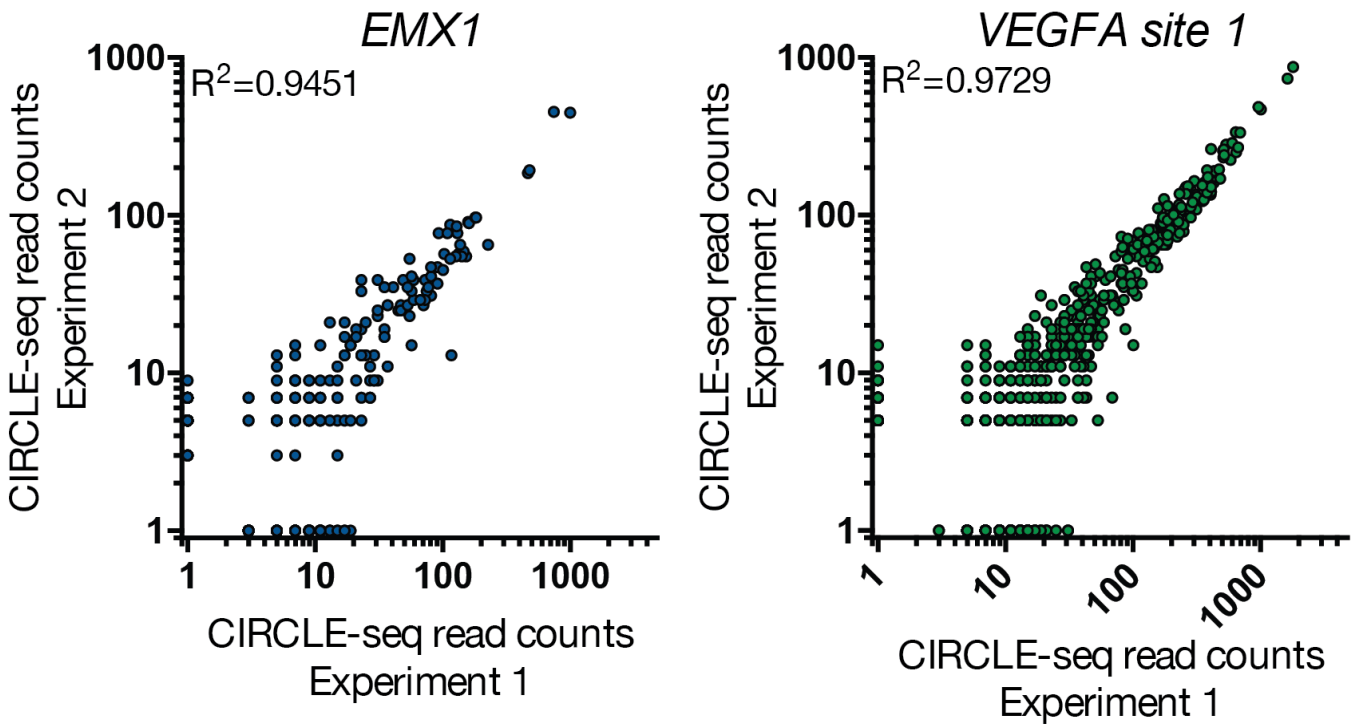
Qiaxcel capillary electrophoretic traces of intramolecular ligation, exonuclease treatment, and restriction enzyme digestion. The observed electrophoretic mobility shift is consistent with circularization. An exonuclease-resistant population of circular DNA molecules is observed after Plasmid-Safe treatment. Digestion with BamHI restriction enzyme linearizes the circularized DNA and results in the expected shift in mobility.



Supplementary Figure 4

Comparison of CIRCLE-seq and covalently closed linear stem-loop strategies for identifying nuclease-induced off-target effects.

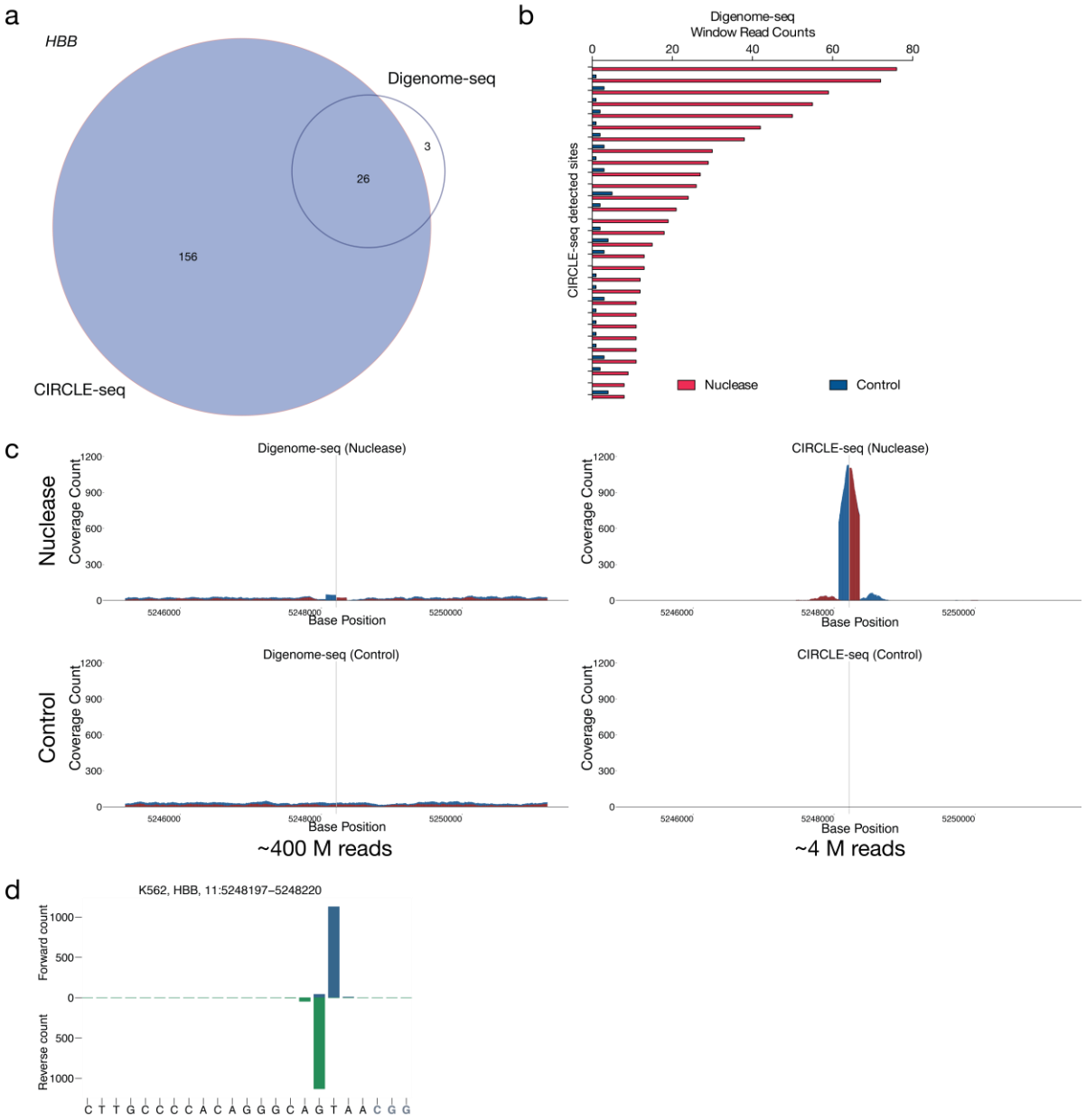
(a) Scatterplot of read counts for linear stem-loop and circular (CIRCLE-seq) methods for detecting Cas9 nuclease-induced off-target sites for a sgRNA targeted against *VEGFA* site 1. (b) Venn diagram showing overlap of sites detected by CIRCLE-seq and alternative linear stem-loop method.



Supplementary Figure 5

CIRCLE-seq read counts are highly reproducible.

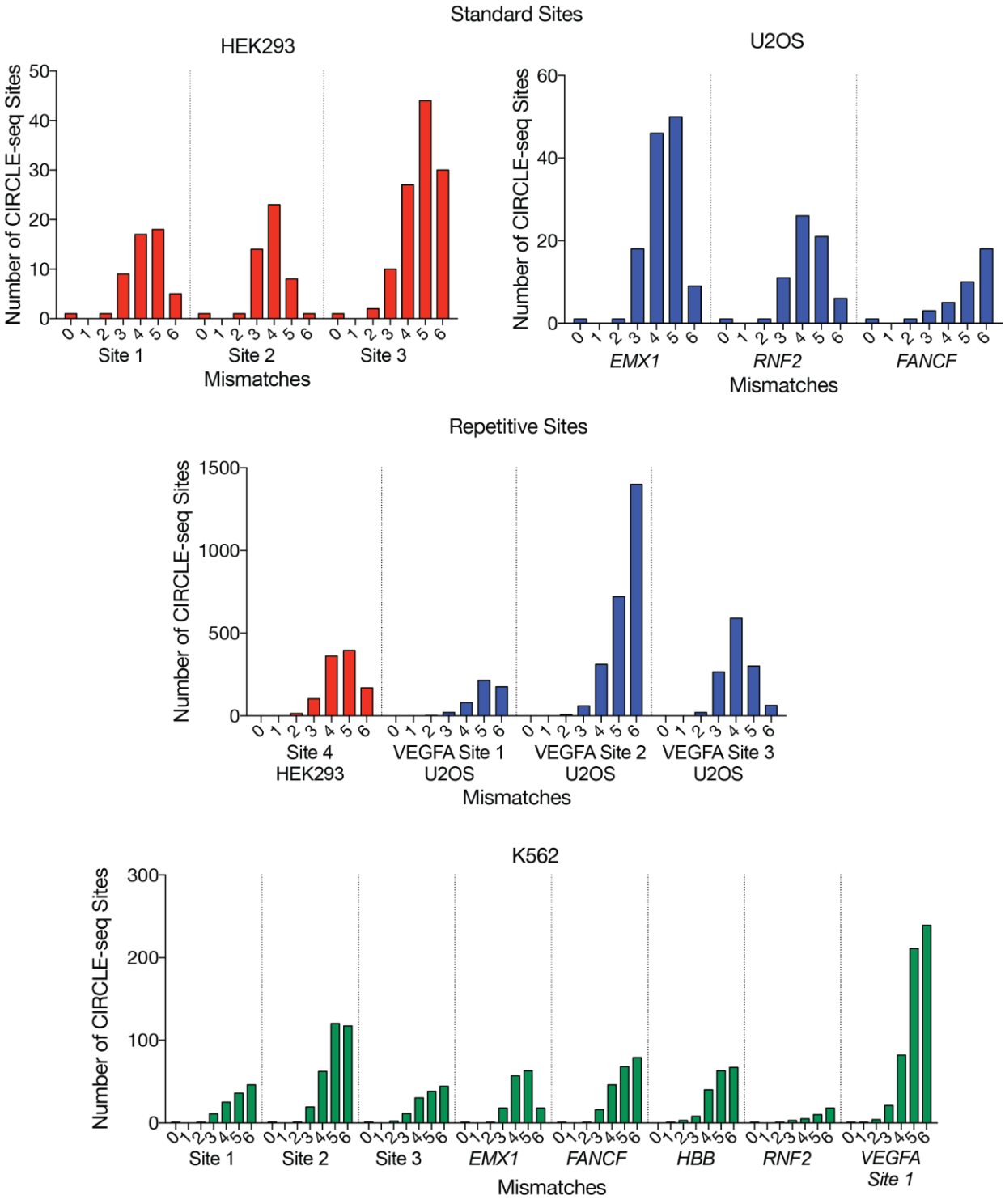
Scatterplots of CIRCLE-seq read counts between two independent CIRCLE-seq libraries prepared from the same source of genomic DNA (human U2OS cells) for sgRNAs targeted against *EMX1* and *VEGFA* site 1.



Supplementary Figure 6

Comparison of CIRCLE-seq with Digenome-seq.

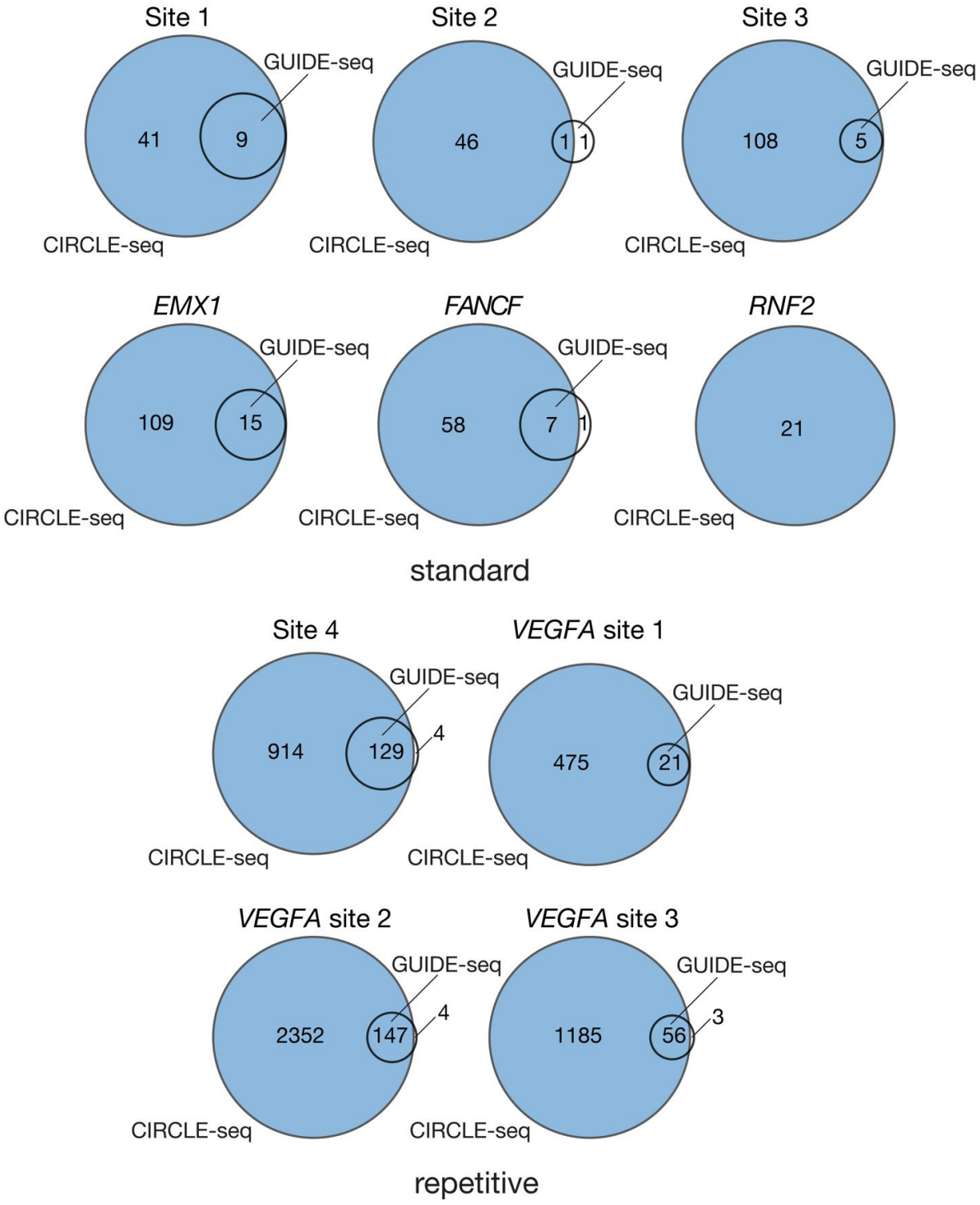
(a) Venn diagram showing intersections of off-target sites of Cas9 and a sgRNA targeted against the *HBB* gene detected by CIRCLE-seq (blue) and Digenome-seq (clear). (b) CIRCLE-seq reads observed at 3 sites that are called by Digenome-seq but not CIRCLE-seq. Integrated Genome Viewer (IGV) plots showing supporting CIRCLE-seq read alignments mapped to human reference genome (GrCh37). Reads mapping to the reverse strand are colored in blue, reads mapping to the forward strand are colored in red. (c) Barplot of Digenome-seq start mapping read counts at off-target cleavage positions identified by CIRCLE-seq but not called by Digenome-seq for nuclease-treated (red) and control (blue) HAP1 genomic DNA. (d) Plots comparing mapping of sequencing reads for CIRCLE-seq and Digenome-seq at the on-target site of a sgRNA targeted to the *HBB* locus. Both nuclease-treated and control samples are shown. A thin grey line indicates expected cleavage site position; read coverage for forward reads is colored in red, and reverse reads in blue. (e) CIRCLE-seq start mapping position plot at the on-target site for the *HBB* sgRNA used in (c). (+) strand mapping reads are colored in blue, (-) strand mapping reads are colored in green.



Supplementary Figure 7

Histogram of number of mismatches for CIRCLE-seq off-target sites.

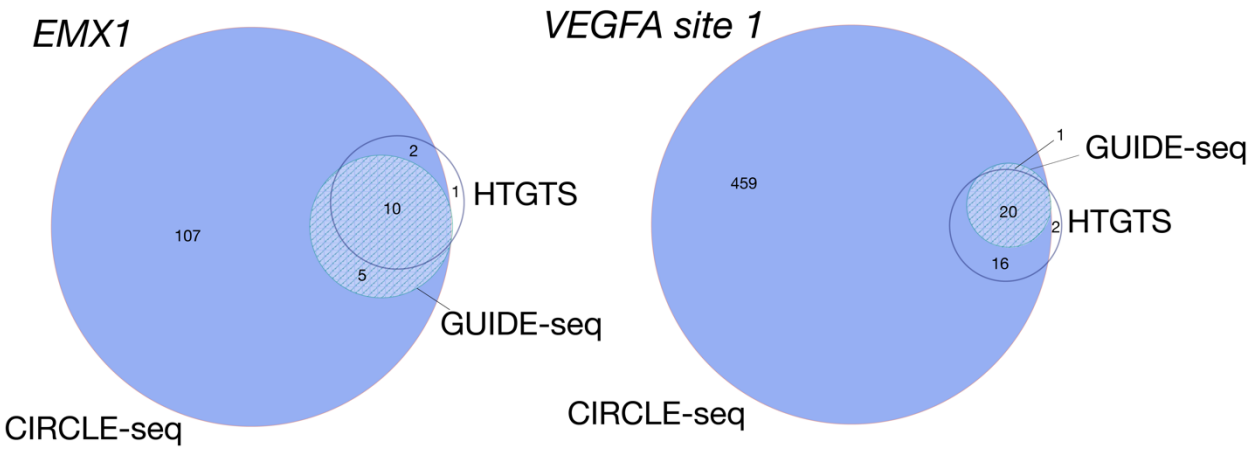
Number of mismatches in CIRCLE-seq detected off-target sites relative to the intended target site of sgRNAs targeted against standard sites in HEK293 & U2OS cells, repetitive sites in HEK293 & U2OS cells, and sites in K562 cells.



Supplementary Figure 8

Venn diagrams showing intersection of CIRCLE-seq and GUIDE-seq detected genomic off-target cleavage sites.

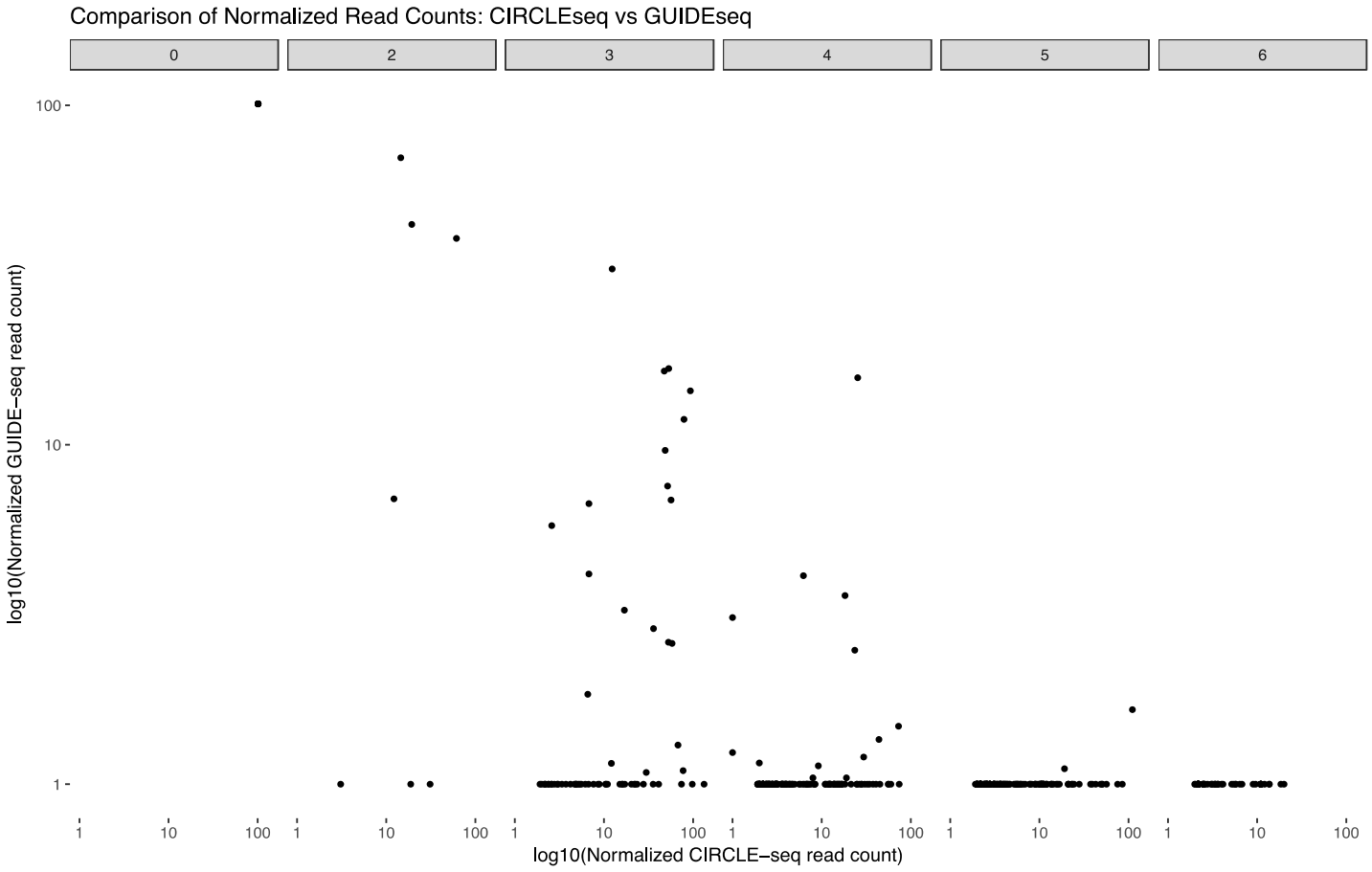
CIRCLE-seq sites are indicated in blue and GUIDE-seq sites with clear circles. The top six comparisons are for sgRNAs targeted against standard genomic sites, and the bottom four comparison are targeted against more repetitive sites.



Supplementary Figure 9

Venn diagrams showing overlap between sets of off-target cleavage sites detected between CIRCLE-seq, GUIDE-seq, and HTGTS.

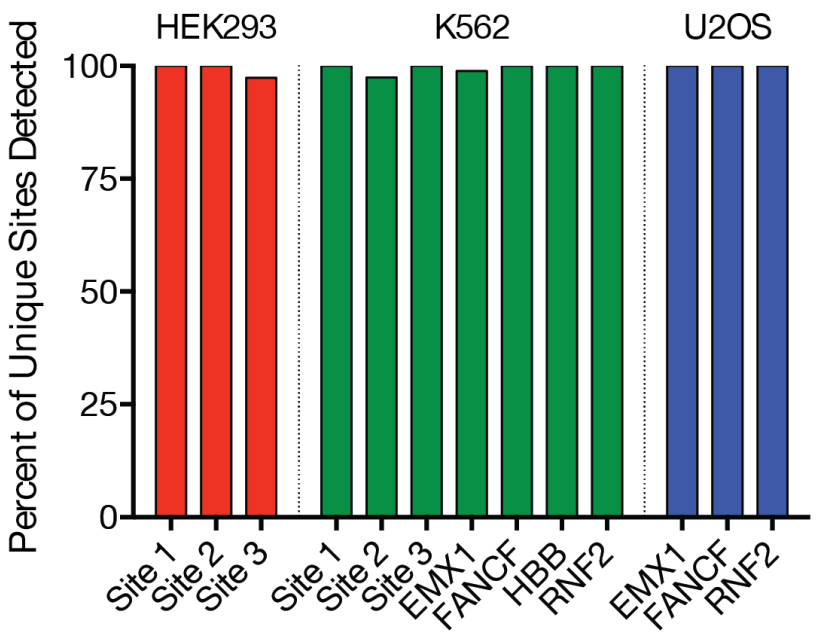
CIRCLE-seq (solid blue) detects virtually all off-target cleavage sites detected by both GUIDE-seq (hatched blue) and HTGTS (clear).



Supplementary Figure 10

CIRCLE-seq read count percentile vs. GUIDE-seq read count.

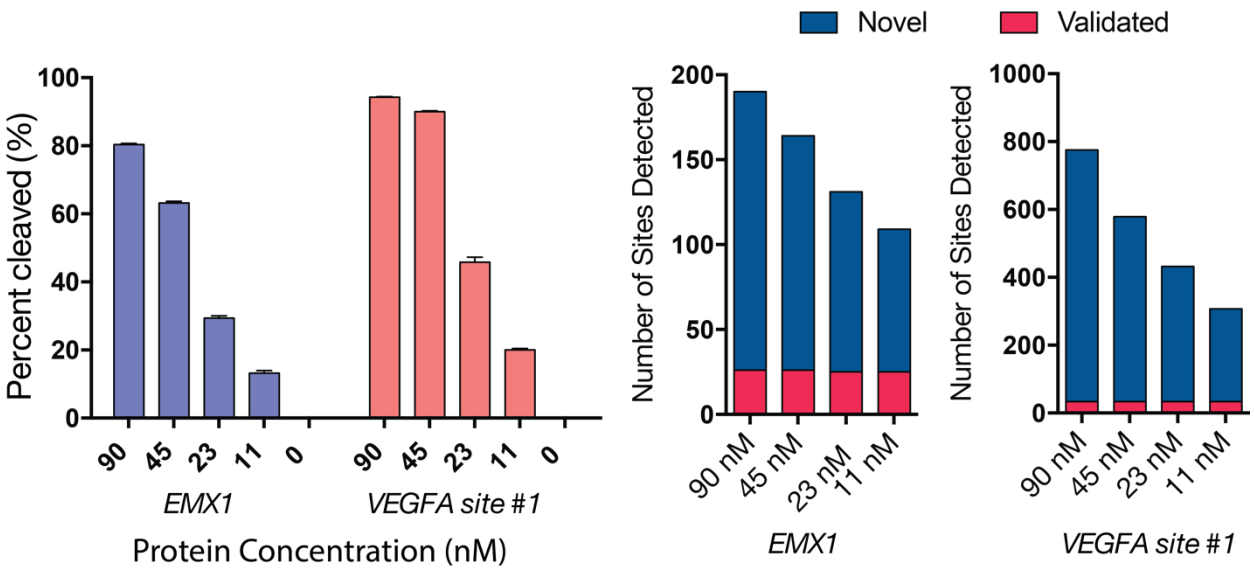
Normalized GUIDE-seq read counts plotted against normalized CIRCLE-seq reads grouped by mismatch numbers between 0 and 6.



Supplementary Figure 11

CIRCLE-seq sites detected by reference-free site discovery algorithm.

Percentage of unique cleavage sites that can be found using a reference-independent site discovery algorithm, for CIRCLE-seq experiments performed with sgRNAs targeting non-repetitive sites in HEK293 (red), K562 (green), and U2OS genomic DNA (blue).



Supplementary Figure 12

Effects of titrating Cas9 protein concentration on *in vitro* cleavage efficiency and number of CIRCLE-seq sites detected.

(a) Barplot of percent *in vitro* cleavage of a targetsite containing PCR amplicon by Cas9 at different concentrations. (b) Number of sites detected by CIRCLE-seq at different concentrations of Cas9:sgRNA complex. Validated sites include both those detected by GUIDE-seq and by confirmatory targeted tag sequencing.

Supplementary Note 1

Optimization of CIRCLE-seq

To achieve restriction-enzyme independent circularization of genomic DNA, we tested a strategy based on ligation of an uracil-containing stem loop adapter to an end-repaired, A-tailed PCR amplicon. We enzymatically selected for covalently-closed DNA molecules that had stem-loop adapters ligated to both sides with a mixture of Lambda exonuclease and *E. coli* exonuclease I. 4 bp overhangs were released using a mixture of USER enzyme and T4 PNK, ligation was performed with T4 DNA ligase under conditions favoring intramolecular ligation, and successful circularization was measured by capillary electrophoresis (**Supplementary Fig. 3**). The conditions resulting in highest circularization efficiency (400 U T4 DNA ligase, 2.5 ng/ul DNA concentration) were used for circularization in all subsequent CIRCLE-seq experiments.

To determine which concentration of Cas9 ribonucleoprotein complex could fully cleave a PCR amplicon containing the corresponding gRNA target site *in vitro*, we performed *in vitro* cleavage assays at varying RNP concentrations. We found that near-complete cleavage of the target amplicon was achieved only with the highest concentration (90 nM Cas9, 9 nM DNA) (**Supplementary Fig. 12**).

We subsequently conducted CIRCLE-seq on two target sites at 4 protein concentrations and found that CIRCLE-seq remains sensitive even with these lower concentrations of nuclease, though the total number of off-target sites is reduced. However, one off-target site previously detected by GUIDE-seq was not identified in these lower concentration experiments, suggesting that CIRCLE-seq at the higher protein concentration is likely to yield the most comprehensive

Nature Methods: doi:10.1038/nmeth.4278

results (**Supplementary Fig. 12**). This 10:1 RNP:DNA ratio was used for all other CIRCLE-seq experiments described.

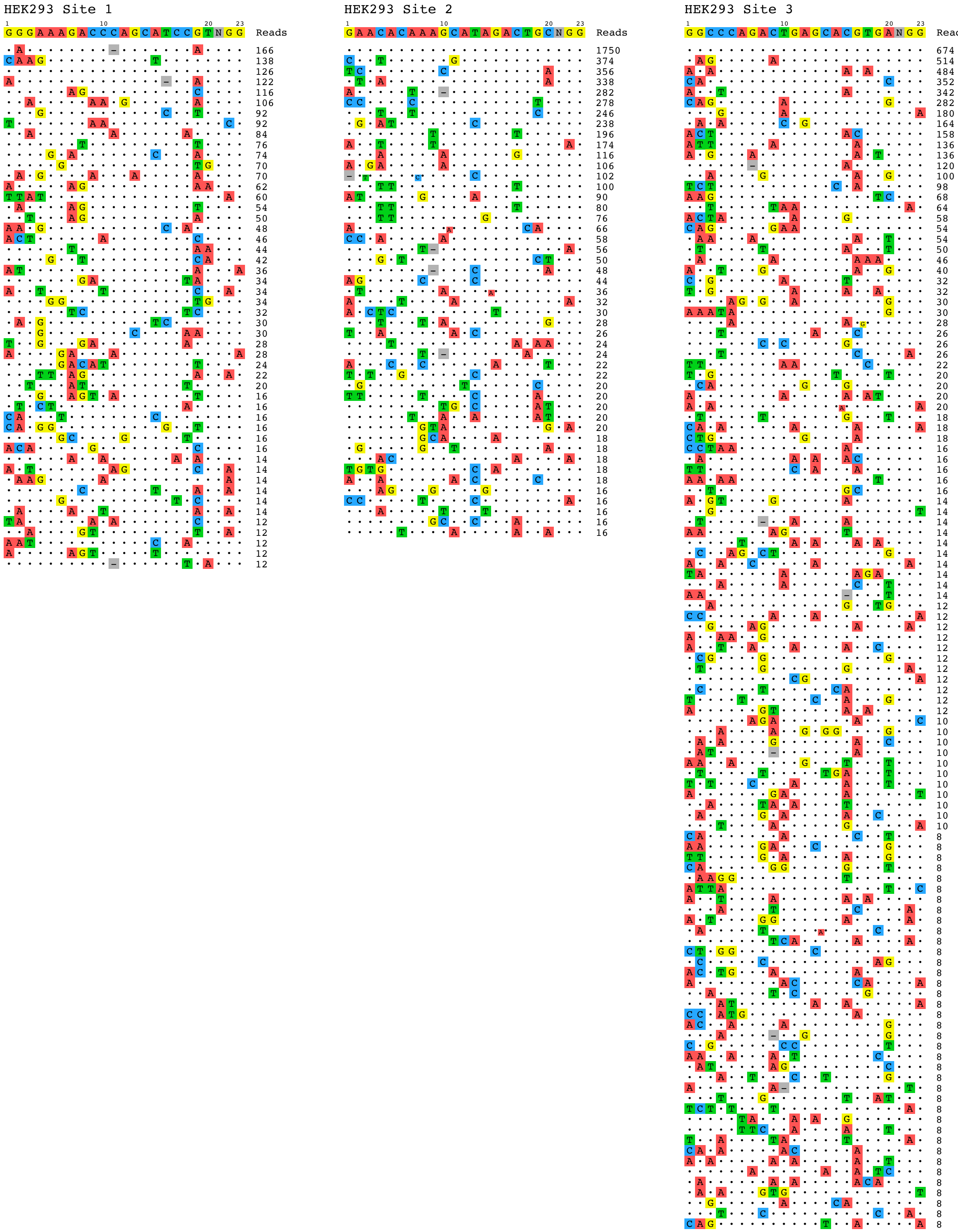
To characterize the technical reproducibility of CIRCLE-seq, we performed independent library preparations from the same source of U2OS genomic DNA. We observed strong CIRCLE-seq read count correlations in independent technical replicates (**Supplementary Fig. 4**).

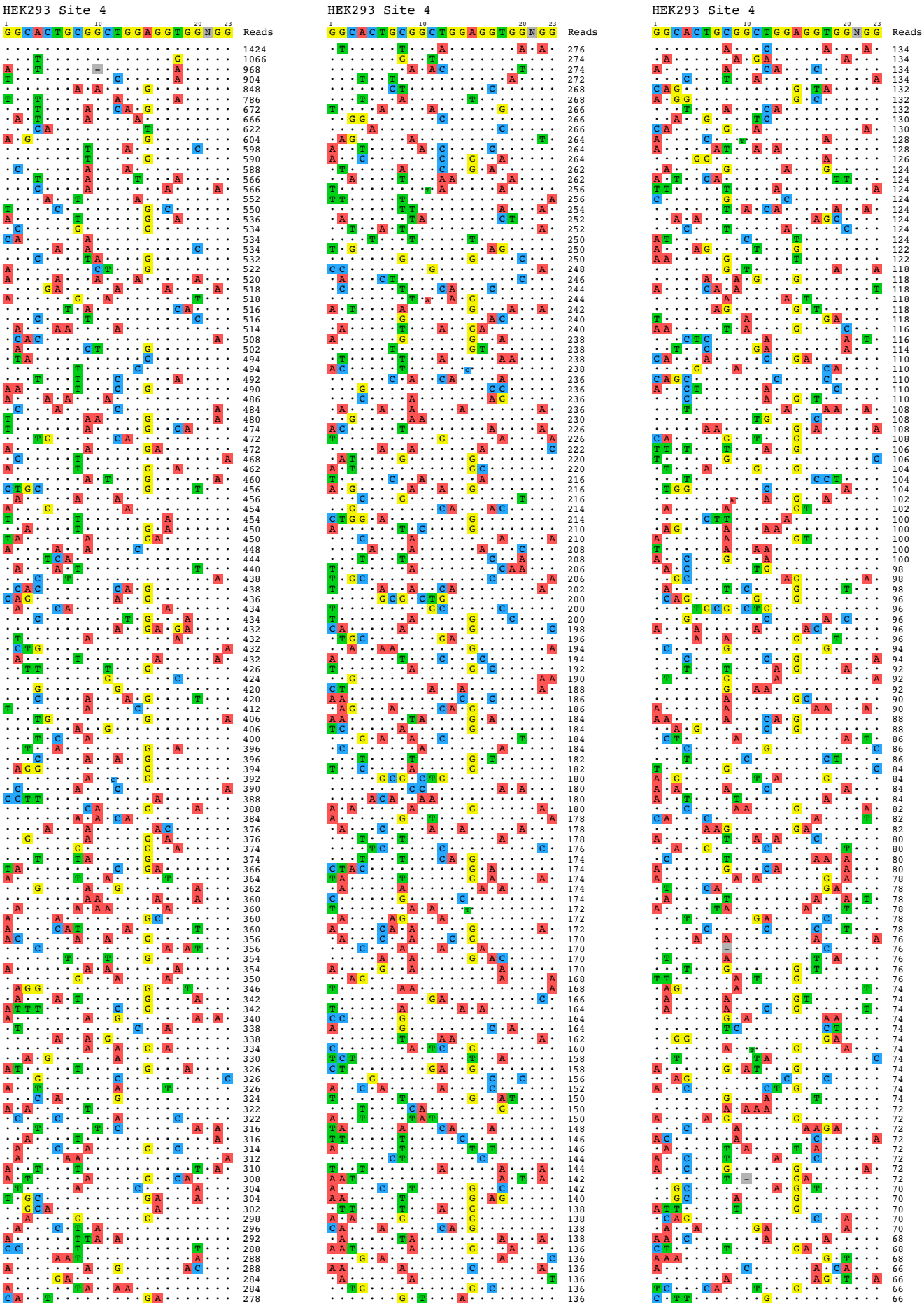
CIRCLE-seq on Repetitive Target Sites

To provide a more challenging test of CIRCLE-seq, we also profiled SpCas9 with four additional gRNAs targeted to repetitive sequences that had also been previously characterized by GUIDE-seq. Due to the repetitive nature of their targets, these four gRNAs have a relatively larger number of closely matched sites in the human genome (**Supplementary Table 1**) and, not unsurprisingly, have had been shown by GUIDE-seq to induce a large number of off-target effects in human cells[30]. As expected, CIRCLE-seq also identified a much larger number of off-target sites, ranging in number from 496 to 2503 for each of the four gRNAs (**Supplementary Table 2**) and distributed throughout the human genome. Included among these were 353 of the 364 off-target sites previously identified by GUIDE-seq experiments (**Supplementary Fig. 8**). For 9 of the 11 sites found by GUIDE-seq but not identified by CIRCLE-seq, evidence of supporting reads could be found in the CIRCLE-seq data but not of a sufficiently high number to meet our statistical threshold, once again suggesting that greater sequencing read depth should would enable detection of these sites.

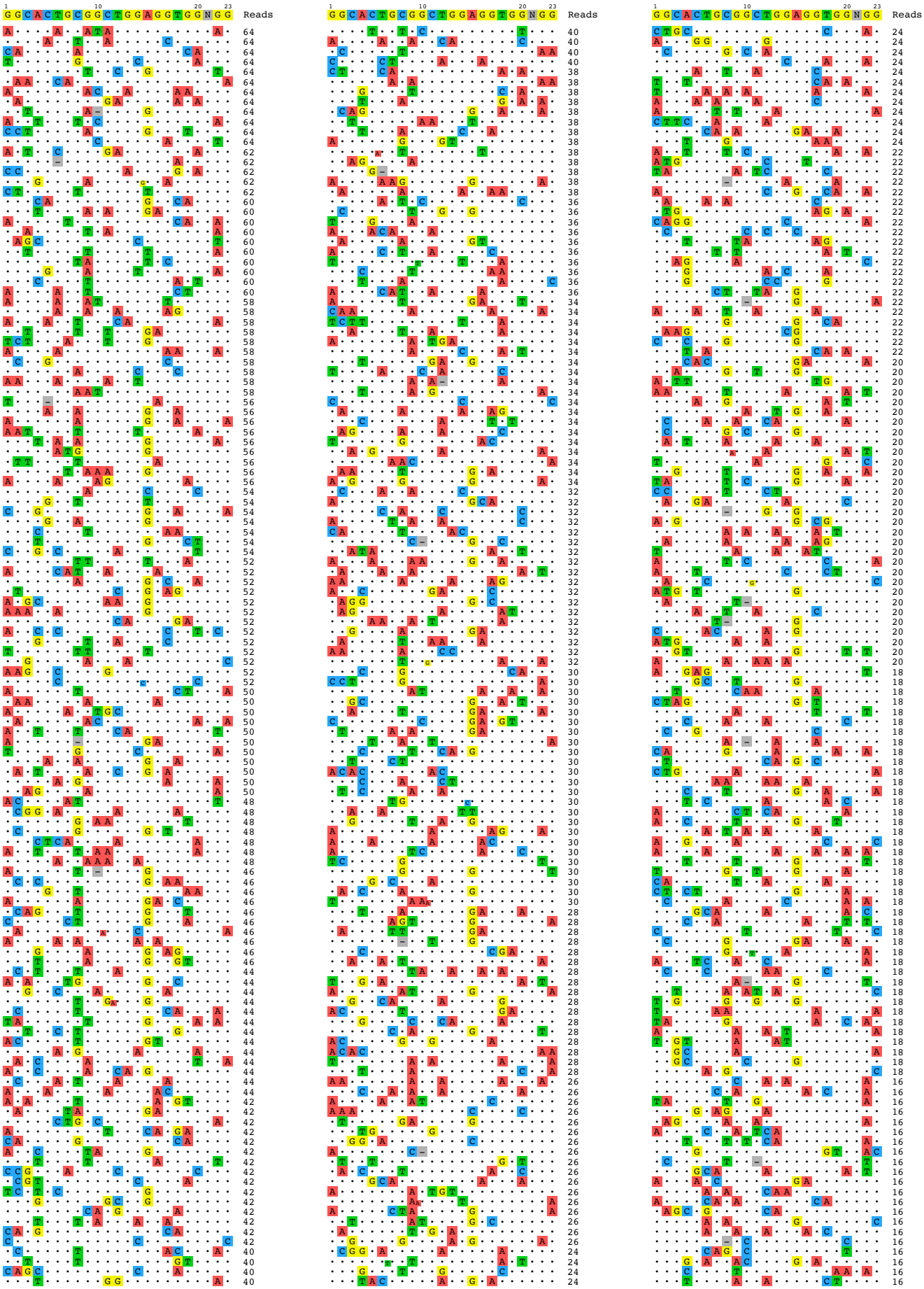
Supplementary Note 2. Visualization of off-target sites detected by CIRCLE-seq.

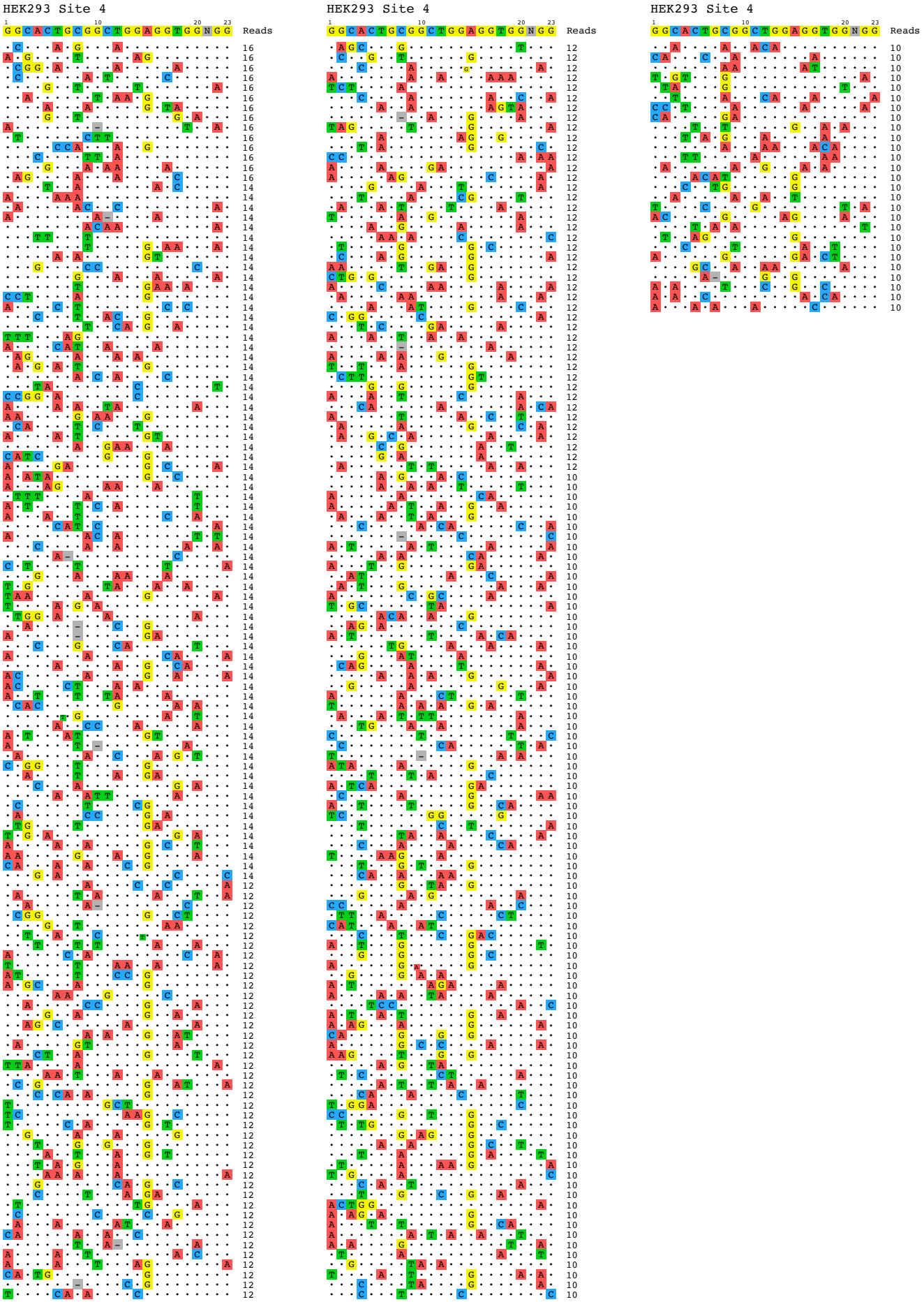
The intended target site sequence is on top. Off-target sites are ordered by CIRCLE-seq read count, matches to the intended target sequence are indicated with a dot and mismatches with colored nucleotides.



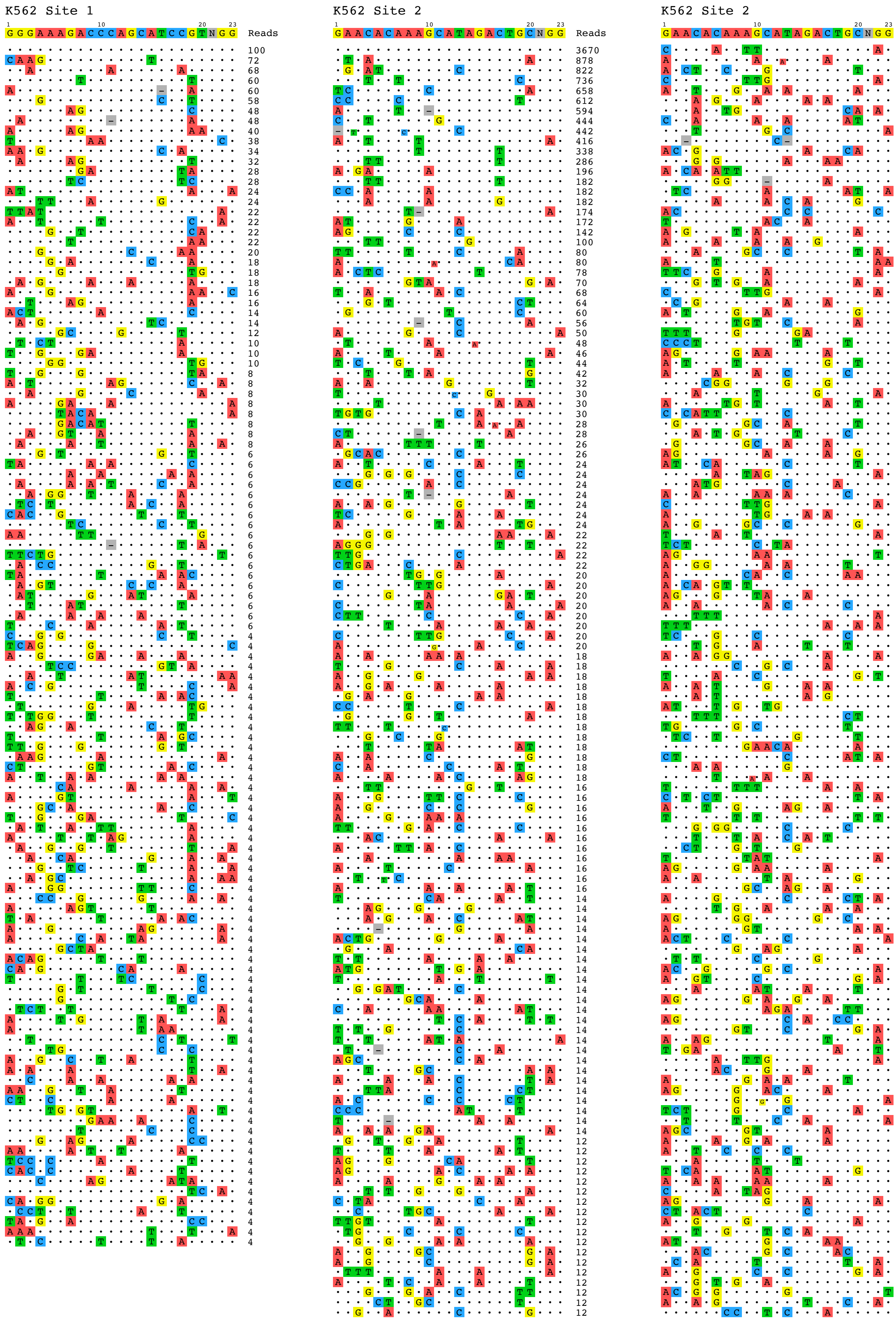


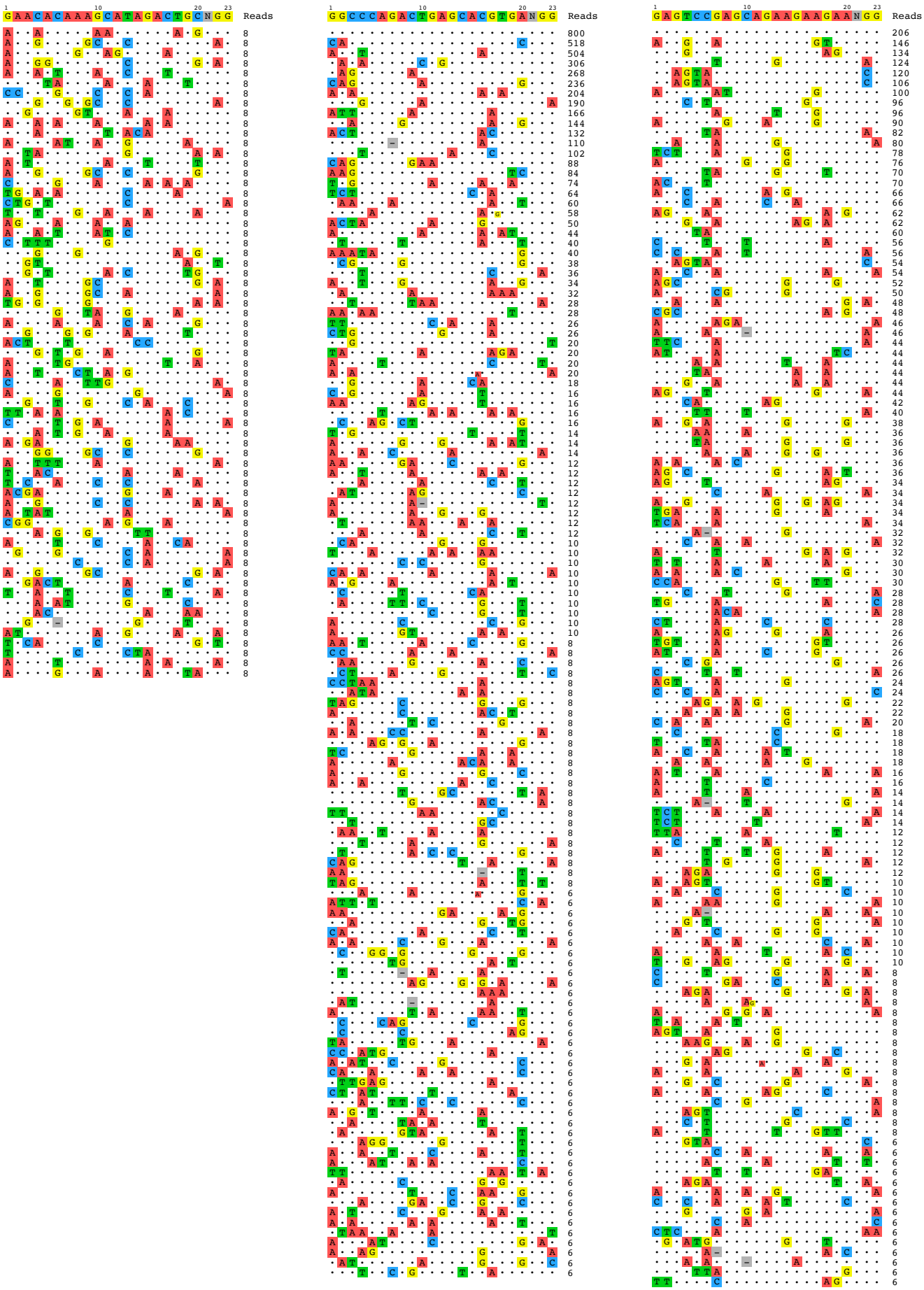
Nature Methods: doi:10.1038/nmeth.4278


HEK293 Site 4
1 10 20 23
G G C A C T G C G G C T G G A G G T G G N G G
Reads
HEK293 Site 4
1 10 20 23
G G C A C T G C G G C T G G A G G T G G N G G
Reads
HEK293 Site 4
1 10 20 23
G G C A C T G C G G C T G G A G G T G G N G G
Reads

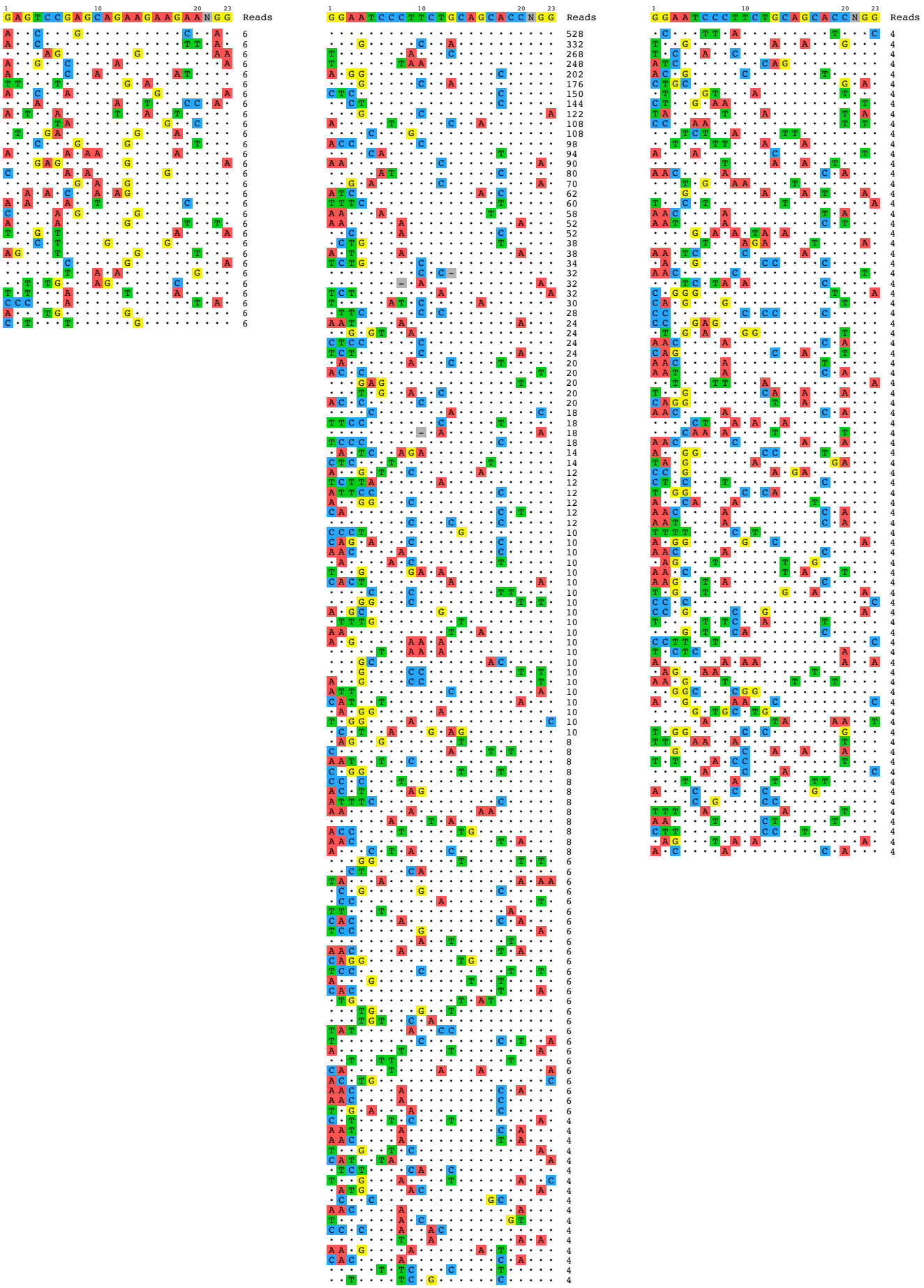


Nature Methods: doi:10.1038/nmeth.4278

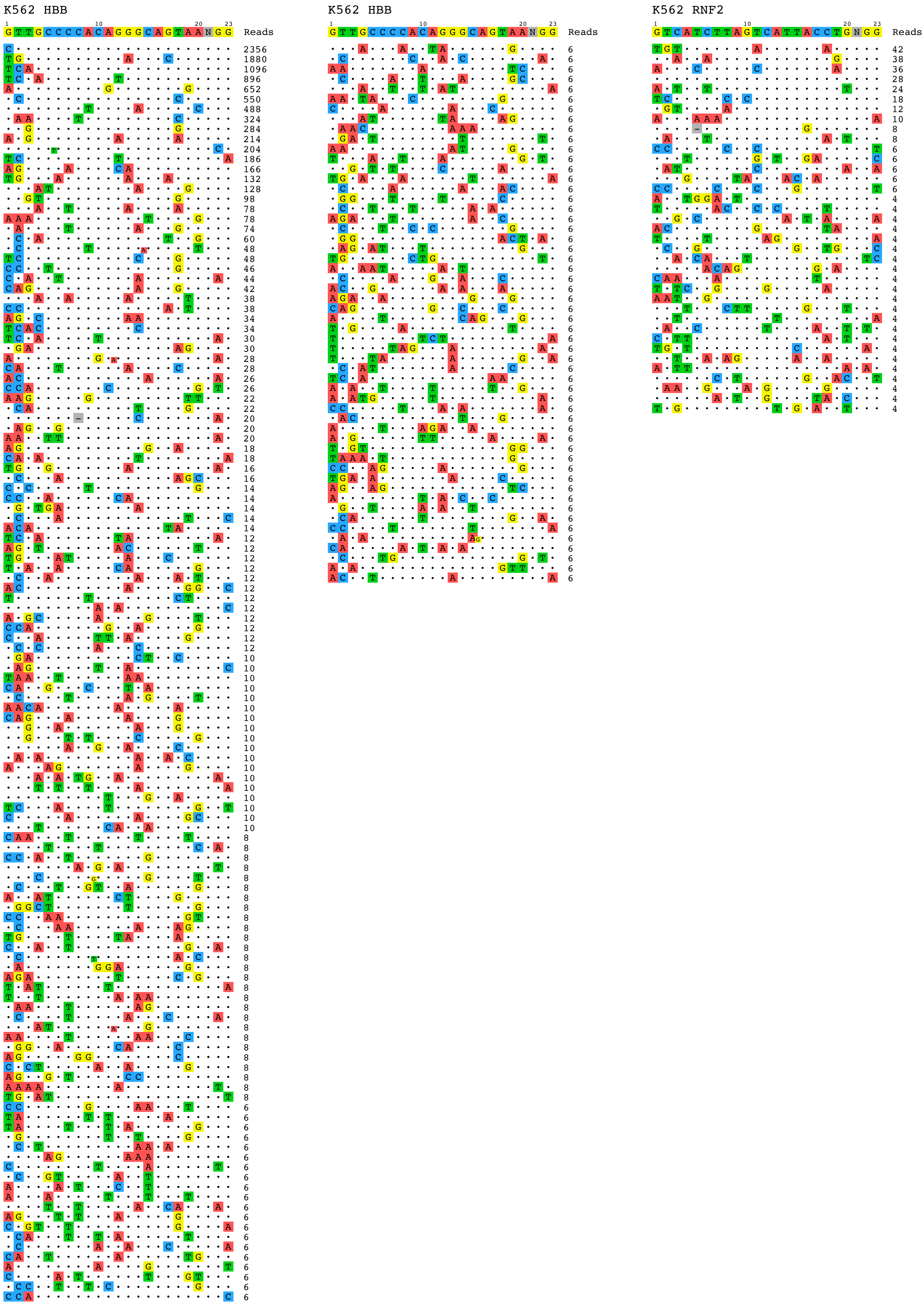

K562 Site 1
GGGAAAGACCCAGCATCCGTNGG
Reads
K562 Site 2
GAACACAAAGCATAGACTGCNGG
Reads
K562 Site 2
GAACACAAAGCATAGACTGCNGG
Reads



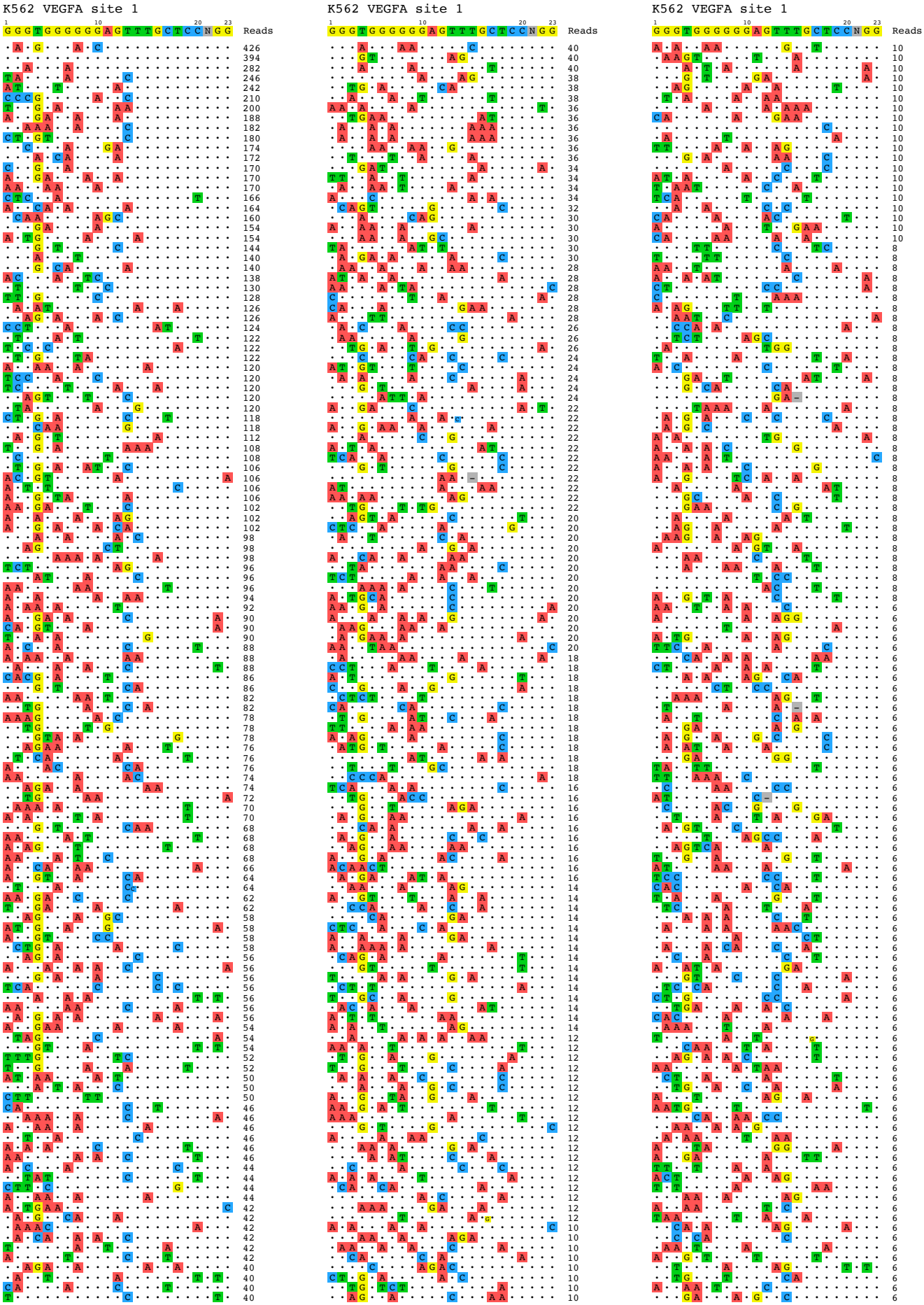
Nature Methods: doi:10.1038/nmeth.4278



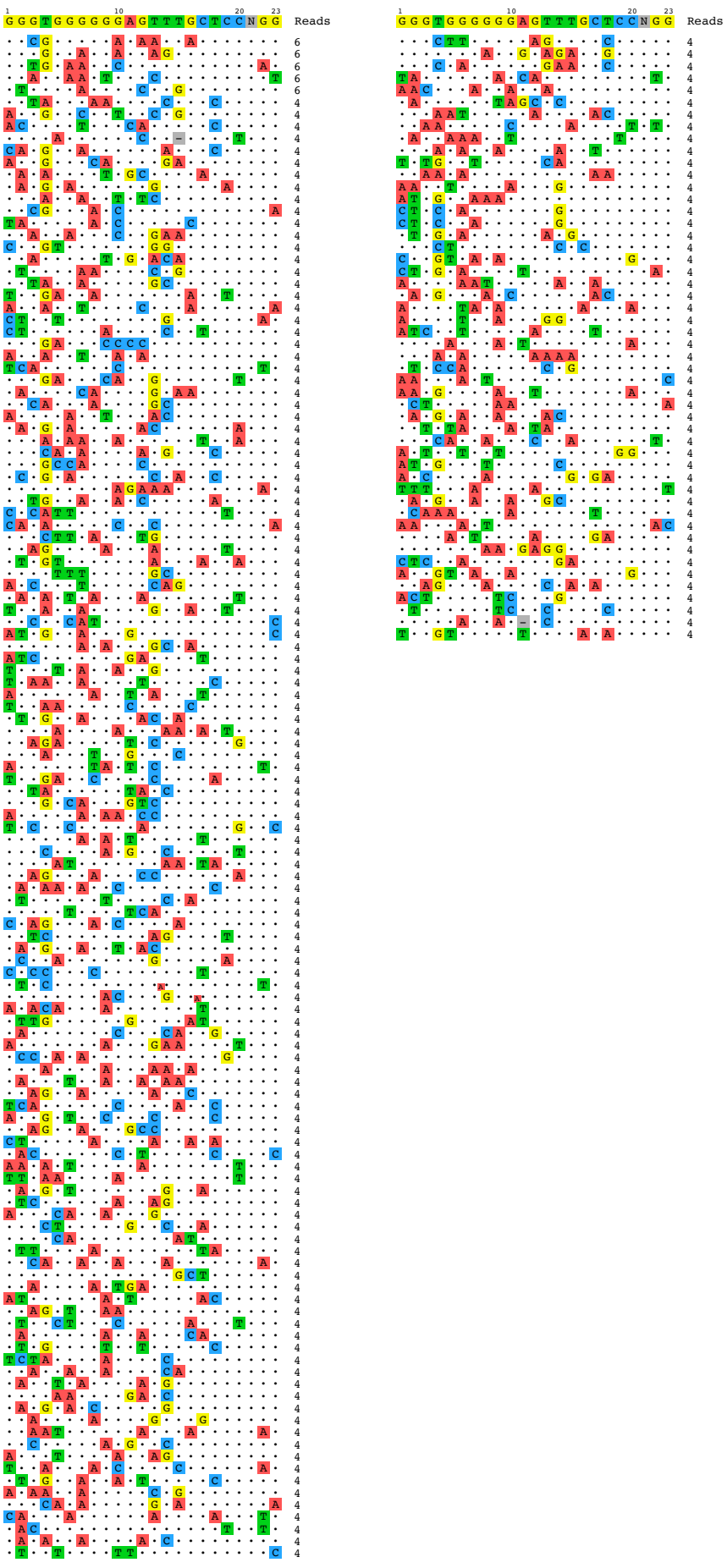
Nature Methods: doi:10.1038/nmeth.4278

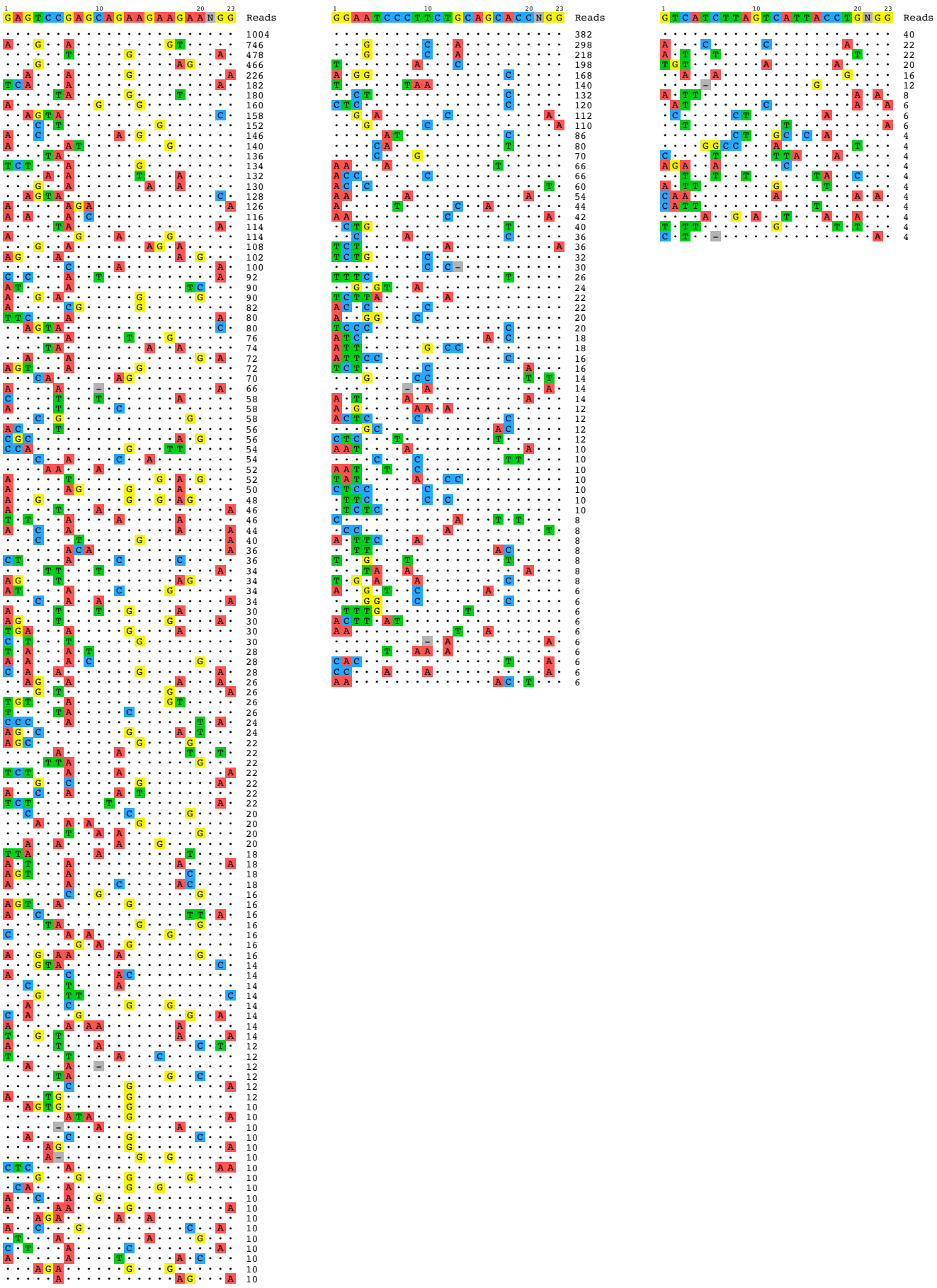


K562 HBB
GTTGCCCCACAGGGCAGTAANGG
Reads
K562 HBB
GTTGCCCCACAGGGCAGTAANGG
Reads
K562 RNF2
GTCATCTTAGTCATTACCTGNGG
Reads

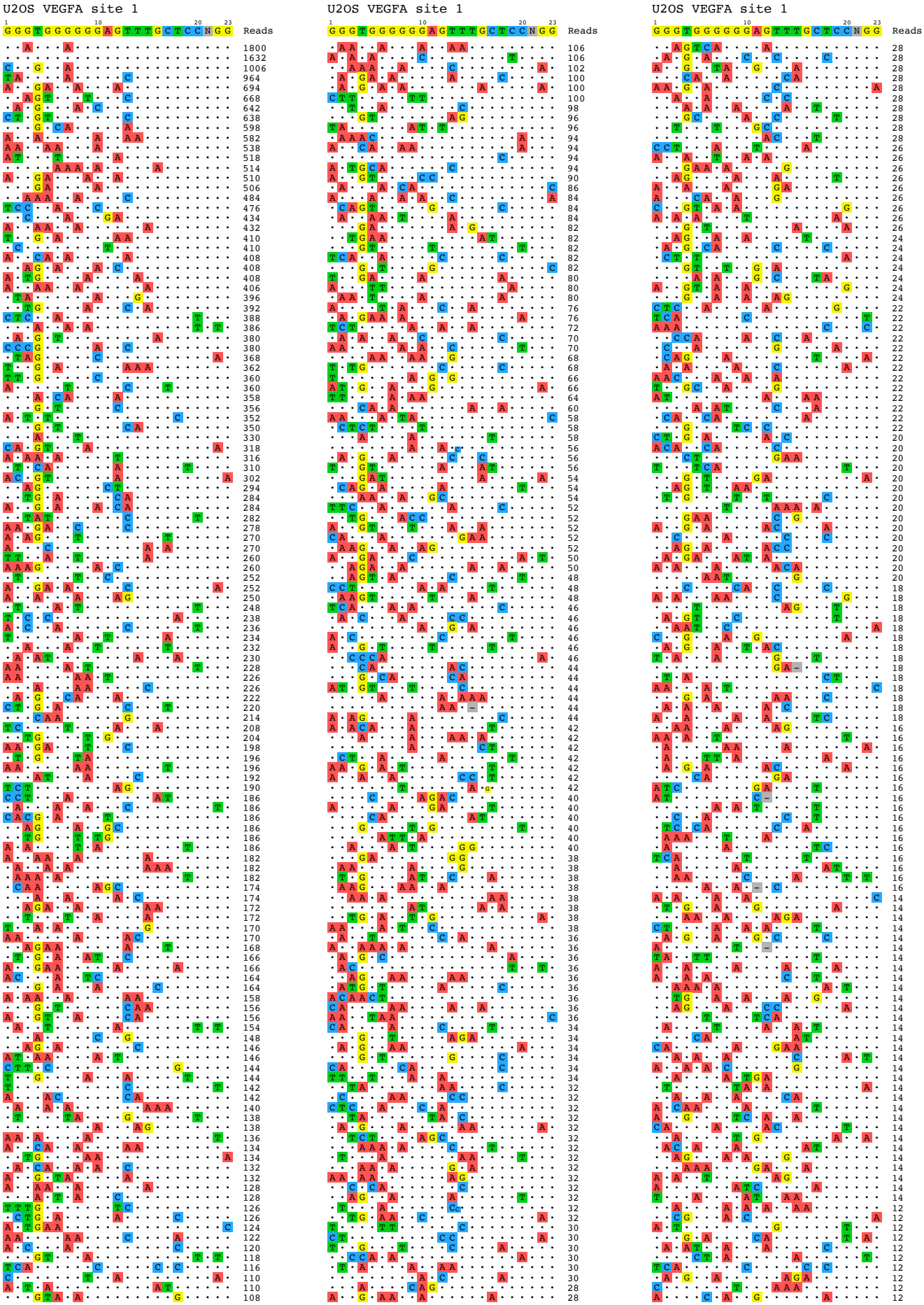


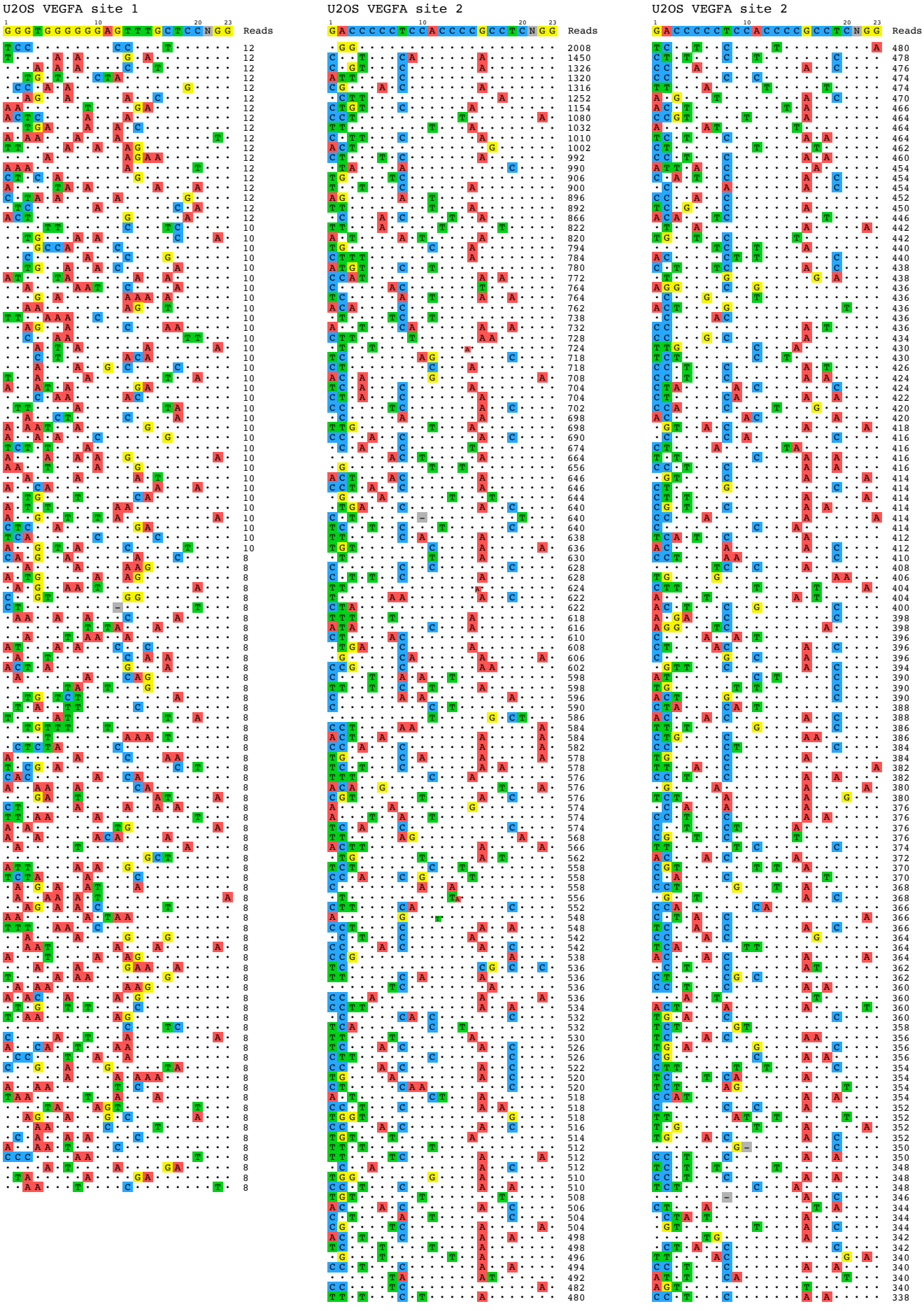
Nature Methods: doi:10.1038/nmeth.4278


K562 VEGFA site 1
1 10 20 23
G G G T G G G G G G A G T T T G C T C C N G G
Reads
K562 VEGFA site 1
1 10 20 23
G G G T G G G G G G A G T T T G C T C C N G G
Reads

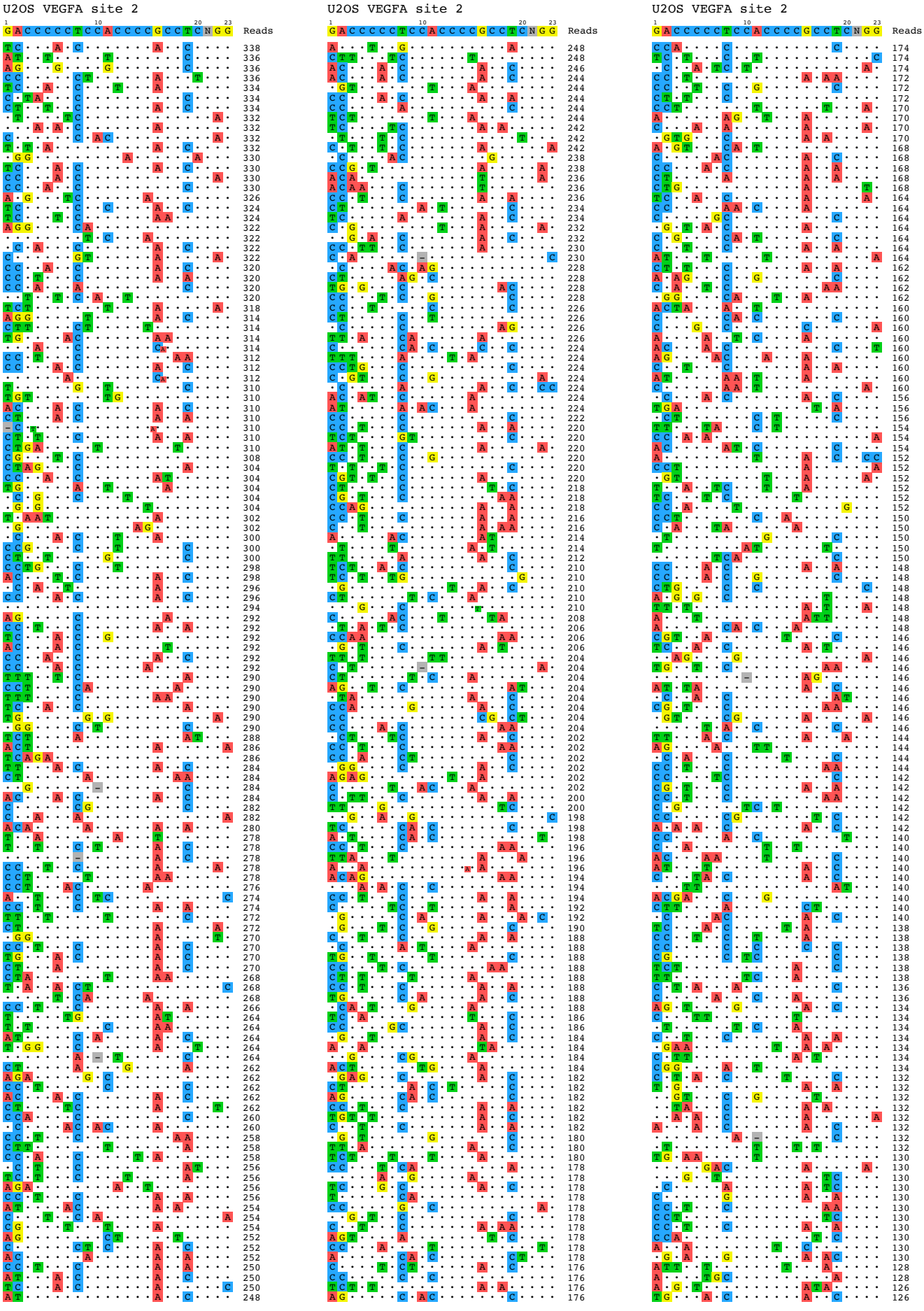


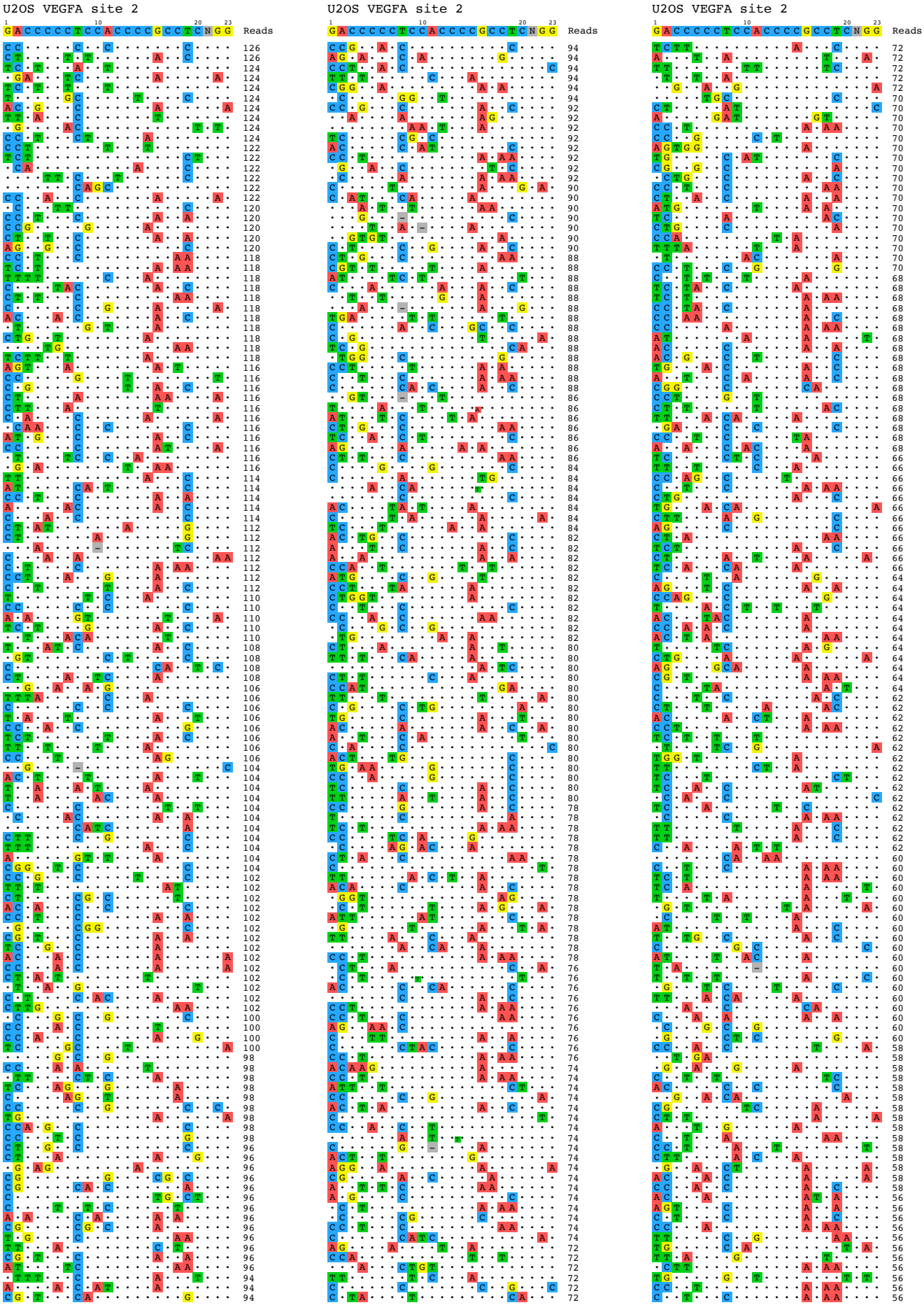
Nature Methods: doi:10.1038/nmeth.4278


U2OS VEGFA site 1
GGGTGGGGGGAGTTTGCTCCNGG
Reads
1800
1632
1006
964
694
668
642
638
598
582
538
518
514
510
506
484
476
434
432
410
410
408
408
408
406
396
392
388
386
380
380
368
362
360
360
358
356
352
350
330
318
316
310
302
294
284
284
282
278
270
270
260
260
252
252
250
248
238
236
234
232
230
228
226
226
222
220
214
208
204
198
196
196
192
190
186
186
186
186
186
182
182
174
174
172
172
170
170
168
166
166
164
164
158
156
156
154
148
146
146
144
144
142
142
140
138
138
136
134
134
132
132
128
128
126
126
124
122
120
118
116
110
110
108
U2OS VEGFA site 1
GGGTGGGGGGAGTTTGCTCCNGG
Reads
106
106
102
100
100
100
98
96
96
94
94
94
90
86
84
84
82
82
82
82
80
80
76
72
70
70
68
68
66
66
64
60
58
58
58
56
56
56
54
54
52
52
52
52
50
50
48
48
46
46
46
46
44
44
44
44
42
42
42
42
42
40
40
40
40
38
38
38
38
38
36
36
36
36
36
34
34
34
34
34
32
32
32
32
32
32
32
32
32
32
32
30
30
30
30
28
28
U2OS VEGFA site 1
GGGTGGGGGGAGTTTGCTCCNGG
Reads
28
28
28
28
28
28
28
28
28
28
26
26
26
26
26
26
26
24
24
24
24
24
24
22
22
22
22
22
22
22
22
22
20
20
20
20
20
20
20
20
20
18
18
18
18
18
18
18
18
18
16
16
16
16
16
16
16
16
16
16
14
14
14
14
14
14
14
14
14
14
14
14
14
14
14
14
14
14
14
14
14
12
12
12
12
12
12

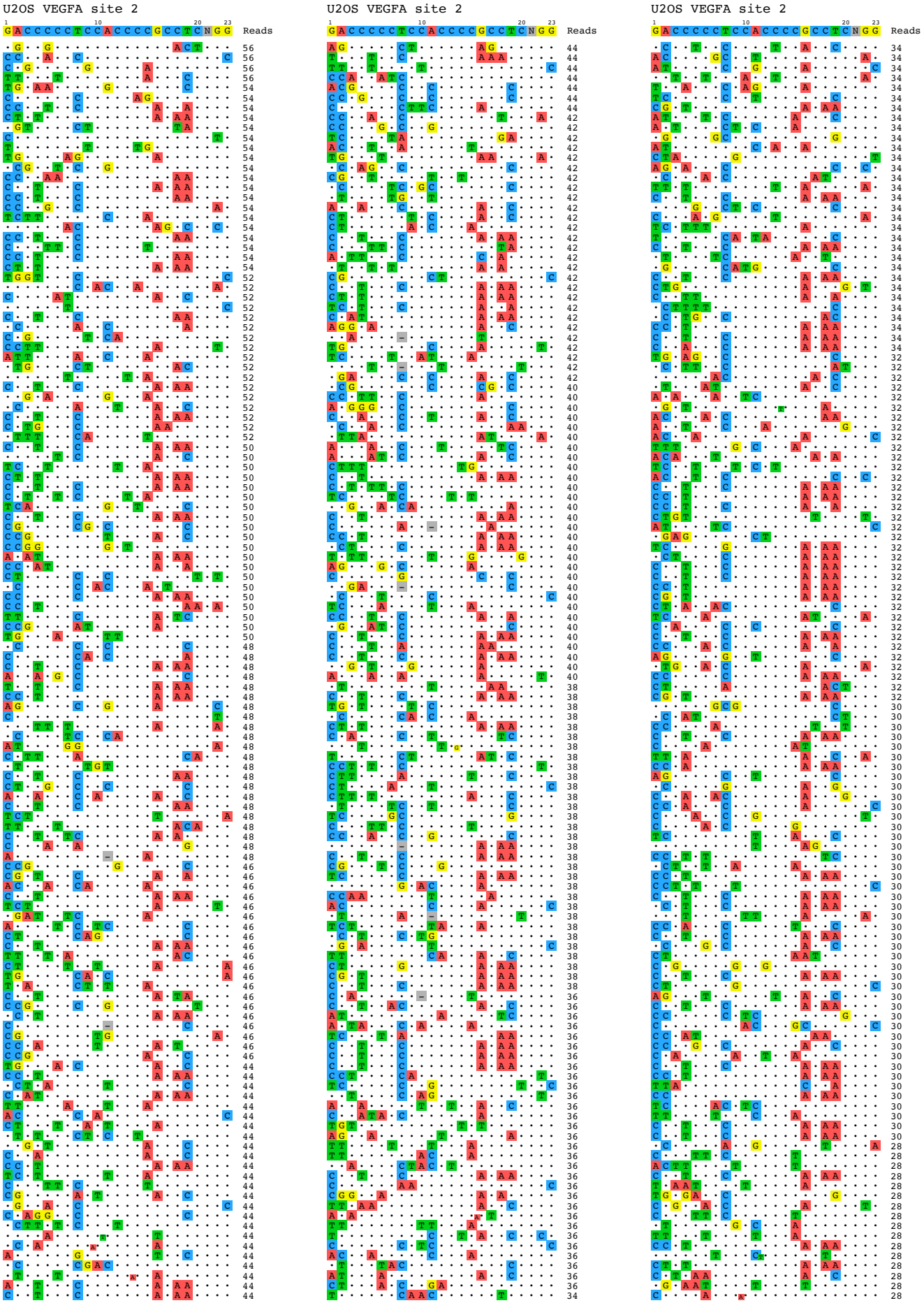


Nature Methods: doi:10.1038/nmeth.4278

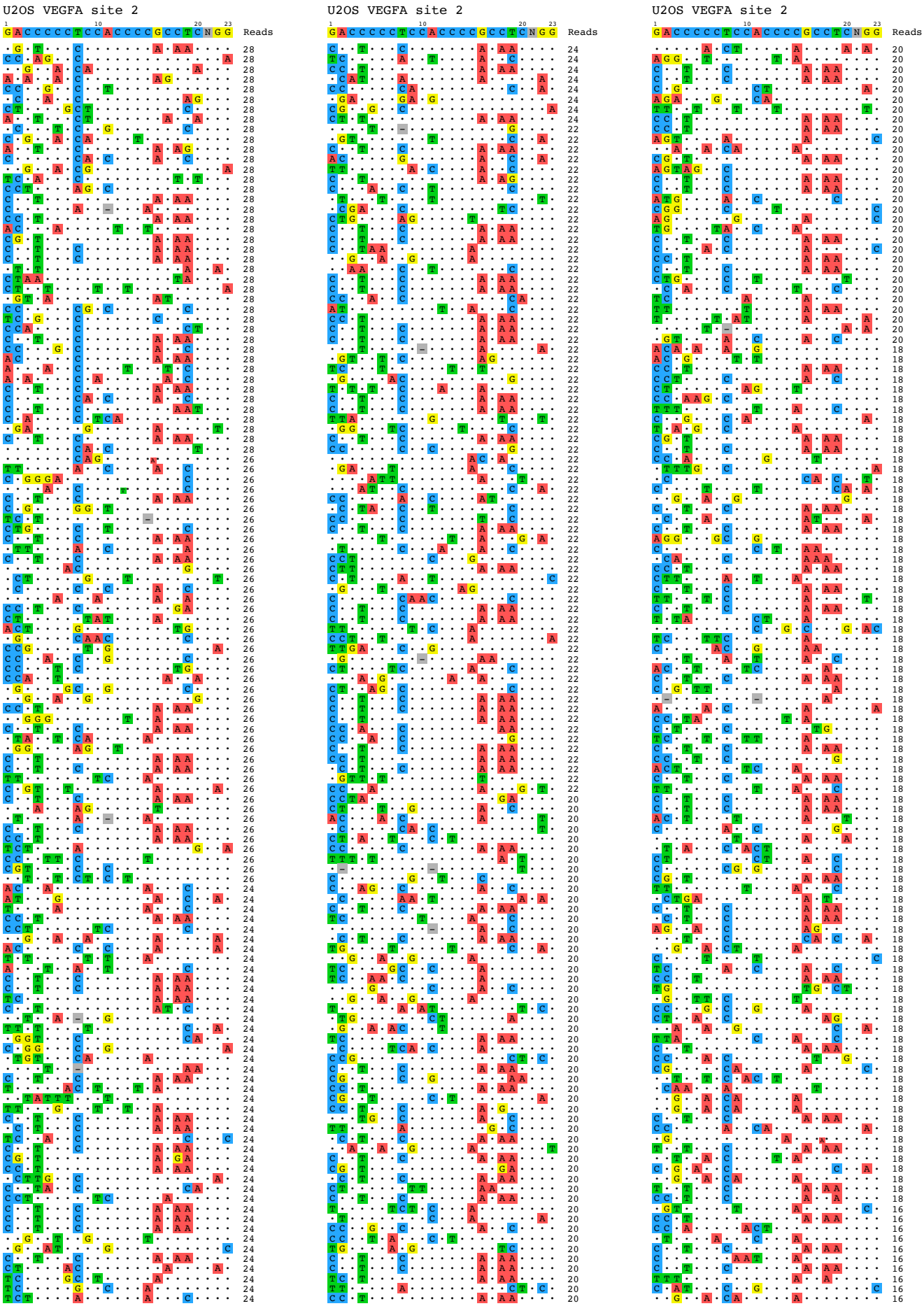

U2OS VEGFA site 2
GACCCCCTCCACCCCGCCTCNGG
Reads
U2OS VEGFA site 2
GACCCCCTCCACCCCGCCTCNGG
Reads
U2OS VEGFA site 2
GACCCCCTCCACCCCGCCTCNGG
Reads



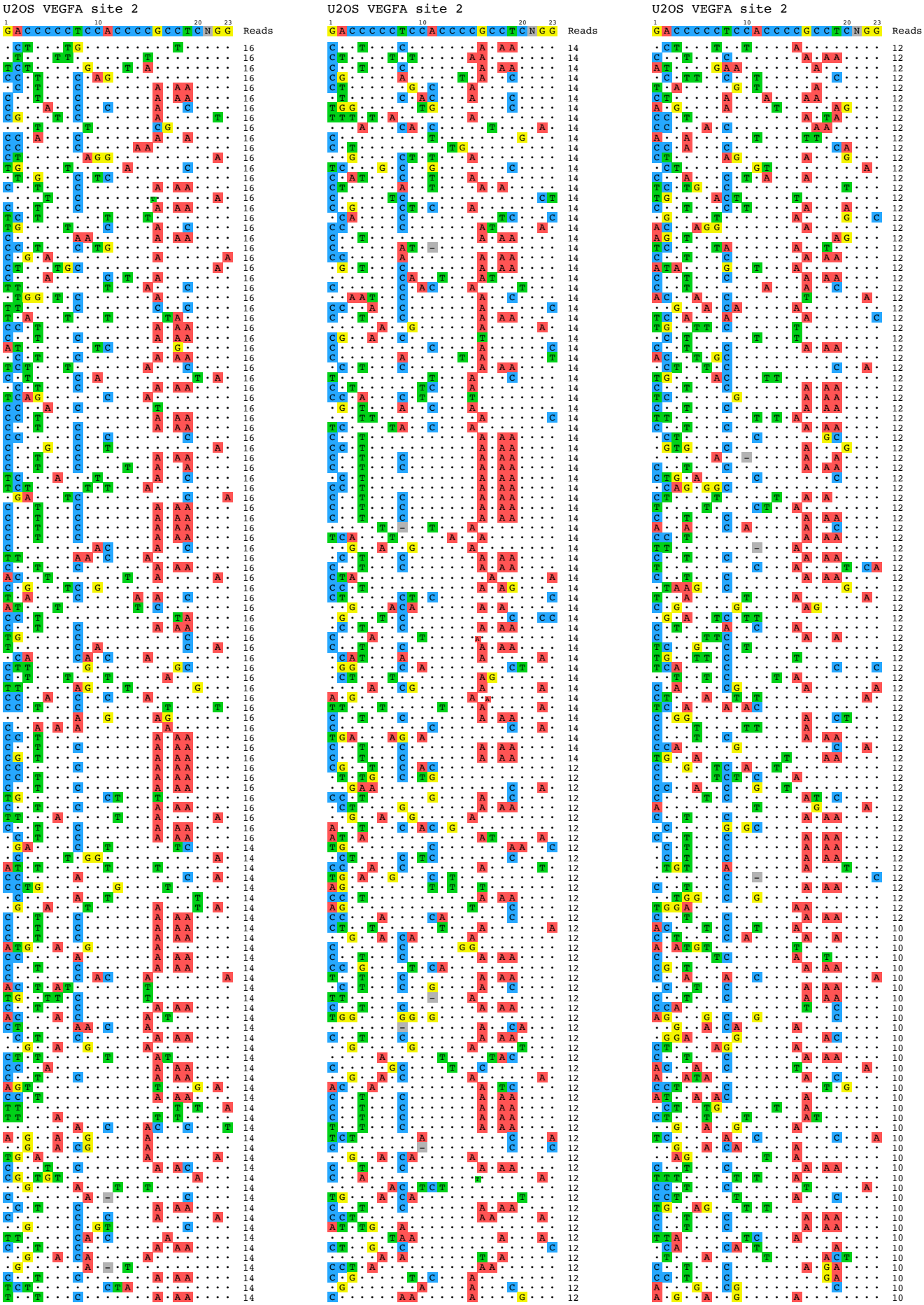
Nature Methods: doi:10.1038/nmeth.4278



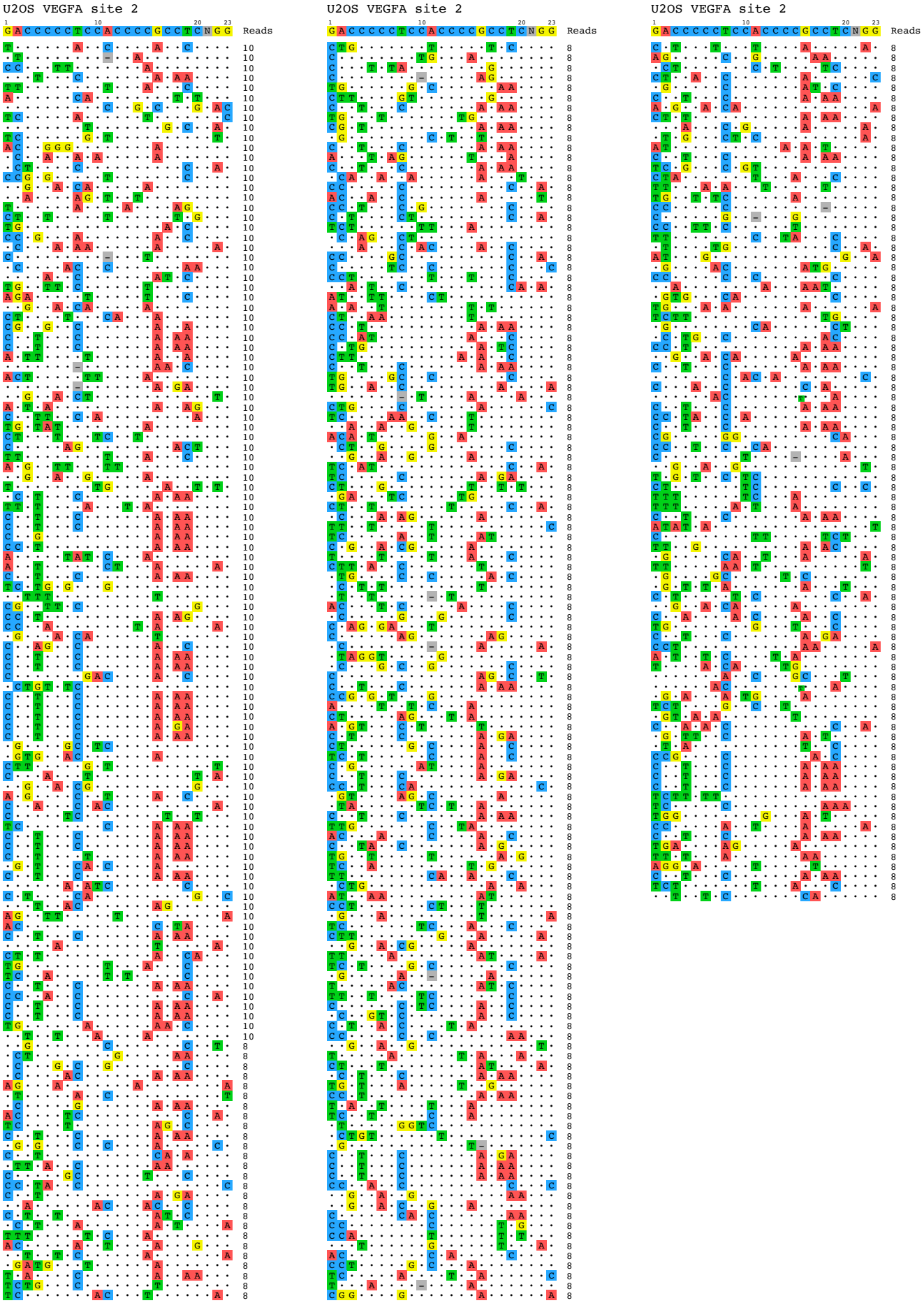
Nature Methods: doi:10.1038/nmeth.4278

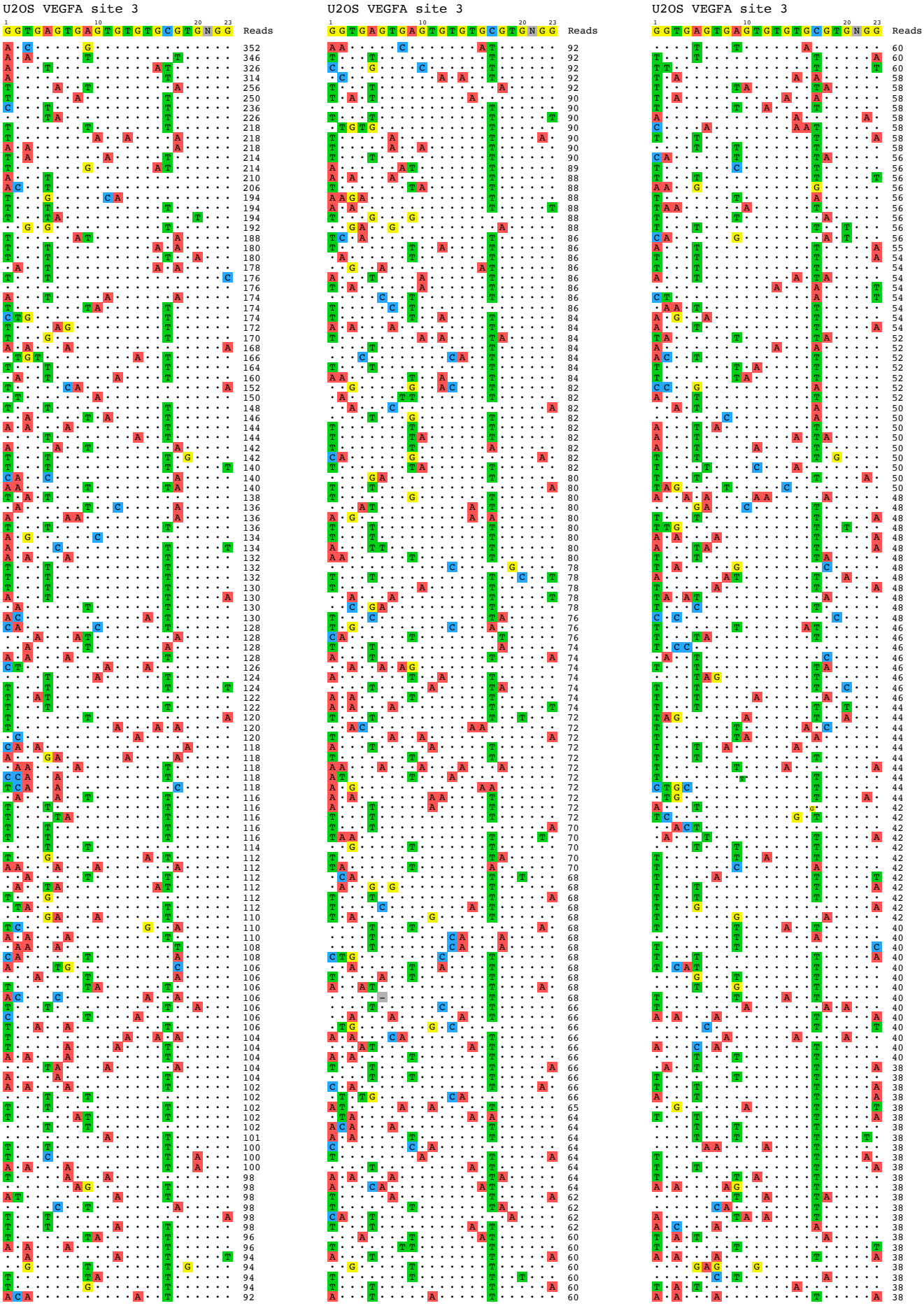


Nature Methods: doi:10.1038/nmeth.4278



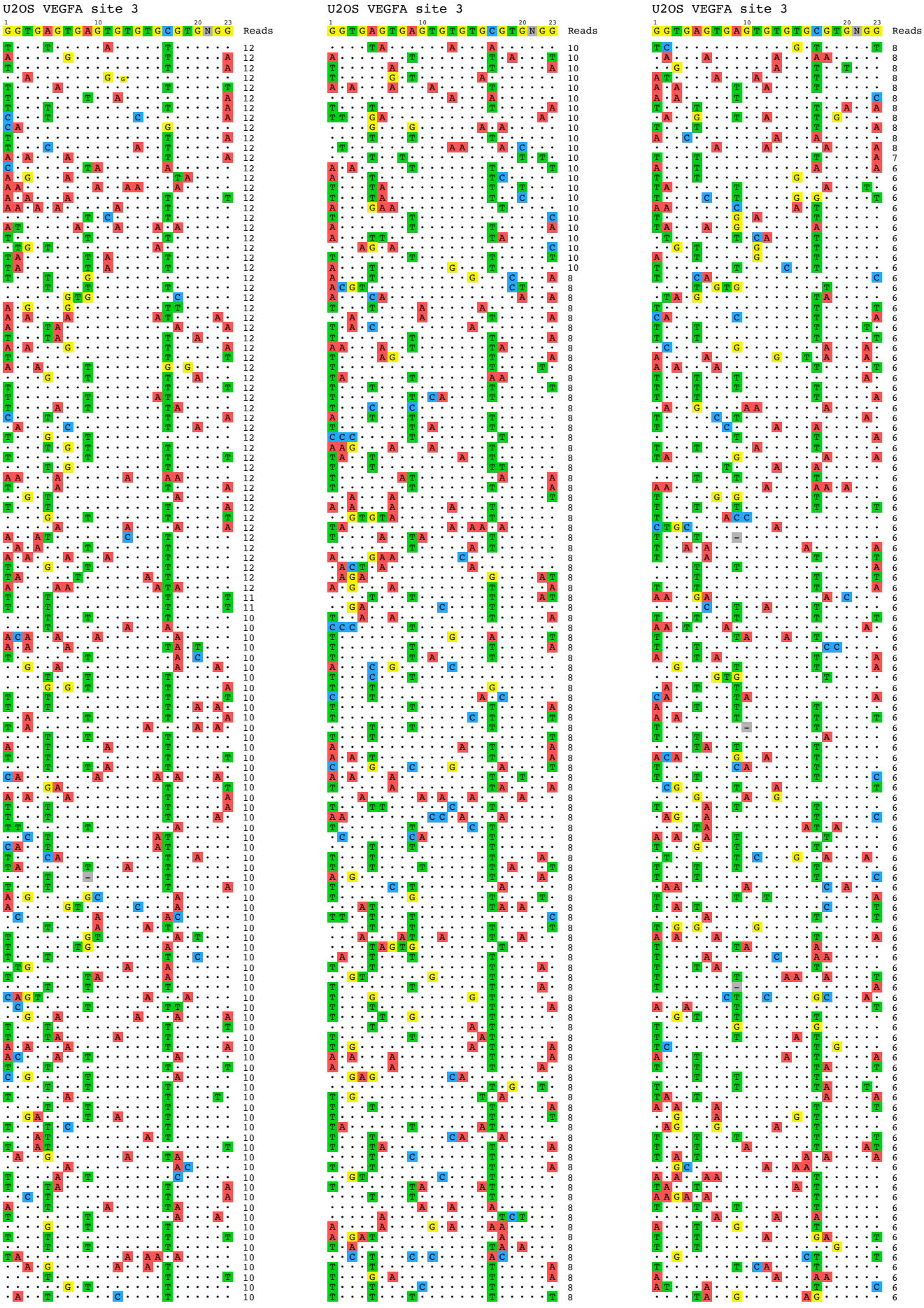
Nature Methods: doi:10.1038/nmeth.4278


U2OS VEGFA site 2
GACCCCCTCCACCCCGCCTCNGG
Reads
U2OS VEGFA site 2
GACCCCCTCCACCCCGCCTCNGG
Reads
U2OS VEGFA site 2
GACCCCCTCCACCCCGCCTCNGG
Reads

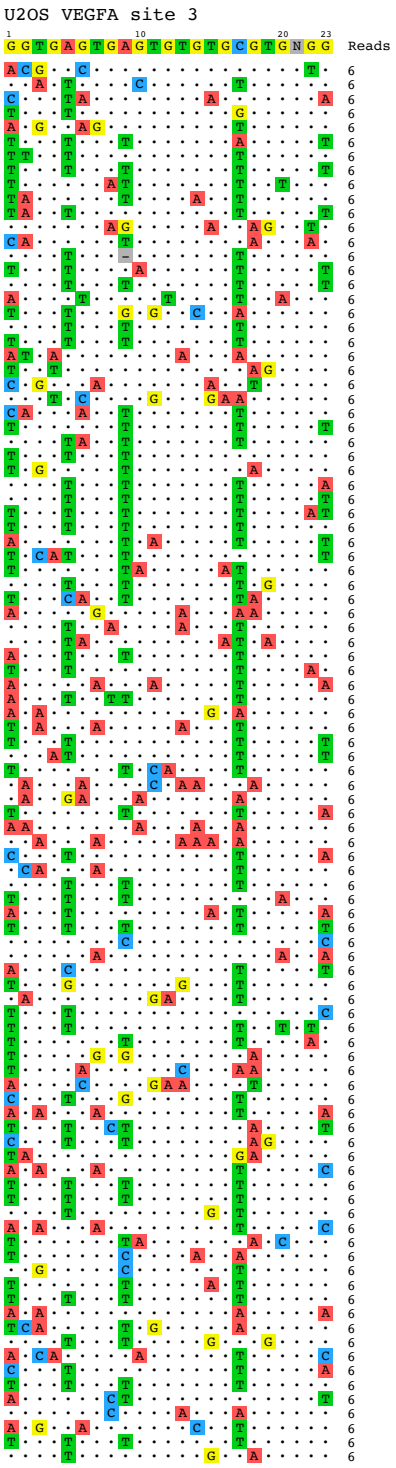


Nature Methods: doi:10.1038/nmeth.4278


U2OS VEGFA site 3
1 10 20 23
G G T G A G T G A G T G T G T G C G T G N G G
Reads
U2OS VEGFA site 3
1 10 20 23
G G T G A G T G A G T G T G T G C G T G N G G
Reads
U2OS VEGFA site 3
1 10 20 23
G G T G A G T G A G T G T G T G C G T G N G G
Reads



Nature Methods: doi:10.1038/nmeth.4278


U2OS VEGFA site 3
1
10
20
23
G G T G A G T G A G T G T G T G C G T G N G G
Reads
6

Supplementary Table 1. Table of numbers of *in silico* off-target sites predicted in the human genome.

Target Site Sequence	Targetsite	0	1	2	3	4	5	6	7	8
GAGTCCGAGCAGAAGAAGAANGG	EMX1	1	1	2	27	421	4313	34761	218047	1156729
GGAATCCCTTCTGCAGCACCNGG	FANCF	1	1	3	33	449	3155	21793	135144	724696
GTCATCTTAGTCATTACCTGNGG	RNF2	1	1	1	11	204	2029	18023	138077	830825
GGGAAAGACCCAGCATCCGTNGG	Site_1	1	1	2	14	132	1499	13410	99120	627262
GAACACAAAGCATAGACTGCNGG	Site_2	1	1	2	16	239	3075	27129	180822	1026201
GGCCCAGACTGAGCACGTGANGG	Site_3	1	1	2	16	156	1831	15689	112679	645364
GGCACTGCGGCTGGAGGTGGNGG	Site_4	1	1	10	125	1231	9452	56139	297118	1471381
GGGTGGGGGGAGTTTGCTCCNGG	VEGFA_site_1	1	2	6	51	442	3870	28723	178630	929570
GACCCCCTCCACCCCGCCTCNGG	VEGFA_site_2	1	1	10	58	726	7636	51673	305299	1469770
GGTGAGTGAGTGTGTGCGTGNGG	VEGFA_site_3	1	2	37	1077	24857	530932	921004	1538579	2944099

Supplementary Table 2. List of all CIRCLE-seq detected off-target sites.

See attached file.

Supplementary Table 3. List of CIRCLE-seq read counts and HTGTS scores for off-target sites detected for Cas9 and gRNAs targeted against *EMX1* and *VEGFA* site 1.

See attached file.

Supplementary Table 4. Deep sequencing read counts for targeted tag integration sequencing of off-target cleavage sites of Cas9 and gRNAs targeted against *EMX1* and *VEGFA* site 1.

See attached file.

Supplementary Table 5. Listing of cell-type specific SNPs in protospacer or PAM of off-target cleavage sites detected by CIRCLE-seq.

See attached file.

Supplementary Table 6. Primers used in target tag integration sequencing.

See attached file.