Journal: Nature Biotechnology

Article Title:	**High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity**
Author:	Vikram Pattanayak,[1] Steven Lin,[2] John P. Guilinger,[1]Enbo Ma[2], Jennifer A. Doudna[2,3,4]* & David R. Liu[1]*

Supplementary Figure S23	Effects on Cas9:sgRNA concentration on specificity
Supplementary Figure S24	Effects of sgRNA architecture on specificity
Supplementary Figure S25	Cas9:guide RNA cleavage of off-target DNA sequences *in vitro*
Supplementary Figure S26	Effect of guide RNA architecture and Cas9:sgRNA concentration on *in vitro* cleavage of an off-target site.
Supplementary Table S2	CLTA1 genomic off-target sequences
Supplementary Table S3	CLTA4 genomic off-target sequences
Supplementary Table S4	Genomic coordinates of CLTA1 and CLTA4 off-target sites
Supplementary Table S5	Cellular modification induced by Cas9:CLTA1 sgRNA
Supplementary Table S6	CLTA1 genomic off-target indel sequences
Supplementary Table S7	CLTA4 genomic off-target indel sequences
Supplementary Table S8	Oligonucleotides used in this study
Algorithms for Data Analysis	Sequence binning
Algorithms for Data Analysis	NHEJ sequence calling

SUPPLEMENTARY TEXT

Pre- and Post-Selection Library Composition

The pre-selection libraries for CLTA1, CLTA2, CLTA3, and CLTA4 had observed mean mutation rates of 4.82 (n = 1,129,593), 5.06 (n = 847,618), 4.66 (n = 692,997), and 5.00 (n = 951,503) mutations per 22-base pair target site, including the two-base pair PAM, respectively. The post-selection libraries treated under enzyme-limiting conditions with Cas9 plus CLTA1, CLTA2, CLTA3, or CLTA4 v.2.1 sgRNAs contained means of 1.14 (n= 1,206,268), 1.21 (n = 668,312), 0.91 (n = 1,138,568), and 1.82 (n= 560,758) mutations per 22-base pair target site. Under enzyme-excess conditions, the mean number of mutations among sequences surviving selection increased to 1.61 (n= 640,391), 1.86 (n = 399,560), 1.46 (n = 936,414), and 2.24 (n= 506,179) mutations per 22-base pair target site, respectively, for CLTA1, CLTA2, CLTA3, or CLTA4 v2.1 sgRNAs. These results reveal that the selection significantly enriched library members with fewer mutations for all Cas9:sgRNA complexes tested, and that enzyme-excess conditions resulted in the putative cleavage of more highly mutated library members compared with enzyme-limiting conditions (**Supplementary Figure S3**).

Specificity at the Non-PAM End of the Target Site

To assess the ability of Cas9:v2.1 sgRNA under enzyme-excess conditions to tolerate multiple mutations distal to the PAM, we calculated maximum specificity scores at each position for sequences that contained mutations only in the region of one to 12 base pairs at the end of the target site most distal from the PAM (**Supplementary Figures S8-S15**).

The results of this analysis show no selection (maximum specificity score ~ 0) against sequences with up to three mutations, depending on the target site, at the end of the molecule farthest from the PAM when the rest of the sequence contains no mutations. For example, when only the three base pairs farthest from the PAM are allowed to vary (indicated by red bars in **Supplementary Figure S9c**) in the CLTA2 target site, the maximum specificity scores at each of the three variable positions are close to zero, indicating that there was no selection for any of the four possible base pairs at each of the three variable positions. However, when the eight base pairs farthest from the PAM are allowed to vary (**Supplementary Figure S9h**), the maximum specificity scores at positions 4-8 are all greater than +0.4, indicating that the Cas9:sgRNA has a sequence preference at these positions even when the rest of the substrate contains preferred, on-target base pairs.

We also calculated the distribution of mutations (**Supplementary Figures S12-15**), in both pre-selection and v2.1 sgRNA-treated post-selection libraries under enzyme-excess conditions, when only the first 1-12 base pairs of the target site are allowed to vary. There is significant overlap between the pre-selection and post-selection libraries for only a subset of the data (**Supplementary Figures S12-15, a-c**, demonstrating minimal to no selection in the post-selection library for sequences with mutations only in the first three base pairs of the target site. These results collectively show that Cas9:sgRNA can tolerate a small number of mutations (~one to three) at the end of the sequence farthest from the PAM when provided with maximal sgRNA:DNA interactions in the rest of the target site.

Specificity at the PAM End of the Target Site

We plotted positional specificity as the sum of the magnitudes of the specificity scores for all four base pairs at each position of each target site, normalized to the same sum for the most highly specified position (**Supplementary Figure S16-S18**). Under both enzyme-limiting and enzyme-excess conditions, the PAM end of the target site is highly specified. Under enzyme-limiting conditions, the PAM end of the molecule is almost absolutely specified (specificity score $\geq$ +0.9 for guide RNA-specified base pairs) by CLTA1, CTLA2, and CLTA3 guide RNAs (**Figure 2** and **Supplementary Figures S4-S7**), and highly specified by CLTA4 guide RNA (specificity score of +0.7 to +0.9). Within this region of high specificity, specific single mutations, consistent with wobble pairing between the

guide RNA and target DNA, that are tolerated. For example, under enzyme-limiting conditions for single-mutant sequences, a dA:dT off-target base pair and a guide RNA-specified dG:dC base pair are equally tolerated at position 17 out of 20 (relative to the non-PAM end of the target site) of the CLTA3 target site. At this position, an rG:dT wobble RNA:DNA base pair may be formed, with minimal apparent loss of cleavage activity.

Effects of Cas9:sgRNA Concentration on DNA Cleavage Specificity

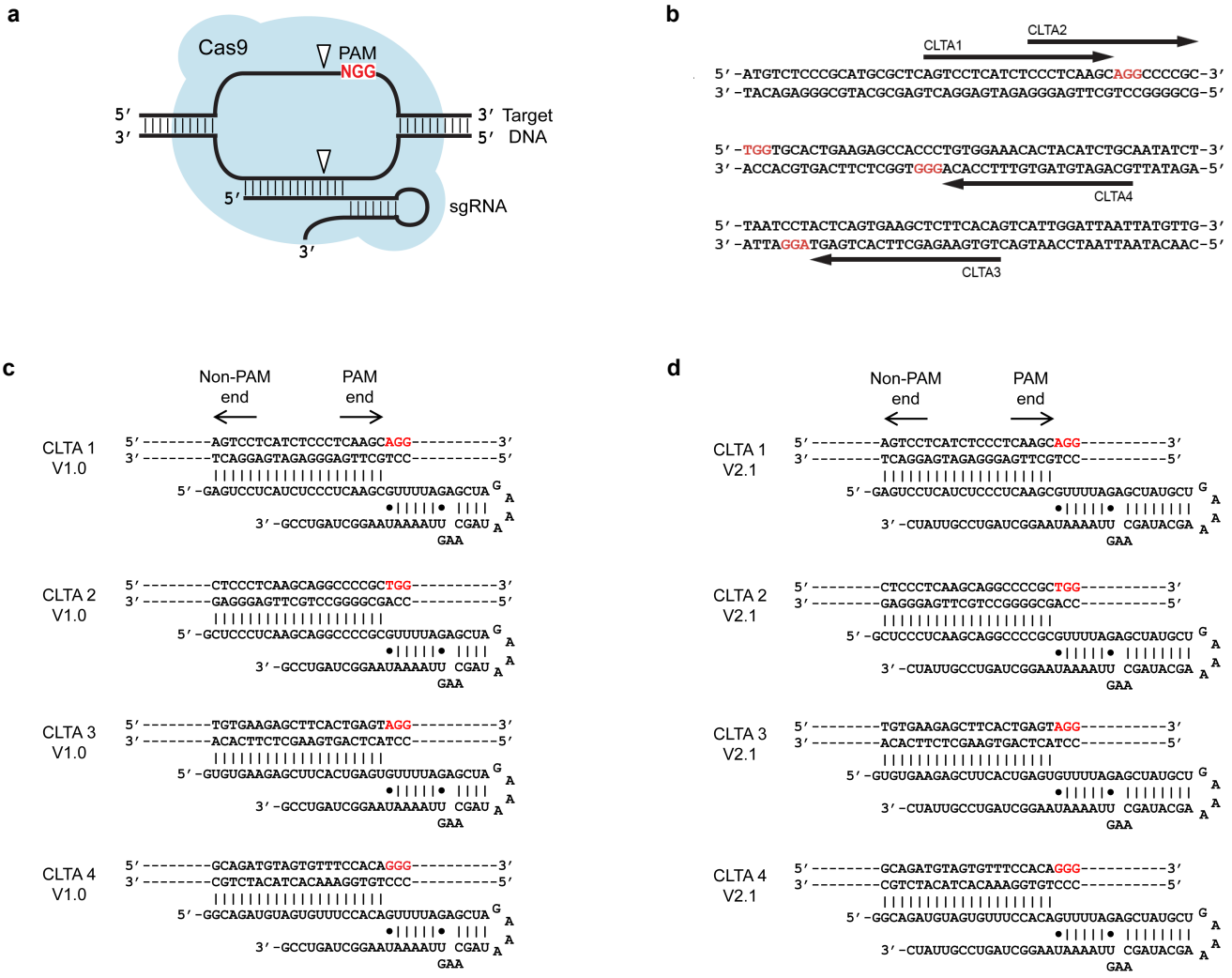
To assess the effect of enzyme concentration on patterns of specificity for the four target sites tested, we calculated the concentration-dependent difference in positional specificity and compared it to the maximal possible change in positional specificity (**Supplementary Figure S23**). In general, specificity was higher under enzyme-limiting conditions than enzyme-excess conditions. A change from enzyme-excess to enzyme-limiting conditions generally increased the specificity at the PAM end of the target by ≥ 80% of the maximum possible change in specificity. Although a decrease in enzyme concentration generally induces small (~30%) increases in specificity at the end of the target sites farthest from the PAM, concentration decreases induce much larger increases in specificity at the end of the target site nearest the PAM. For CLTA4, a decrease in enzyme concentration is accompanied by a small (~30%) decrease in specificity at some base pairs near the end of the target site farthest from the PAM.

Specificity of PAM Nucleotides

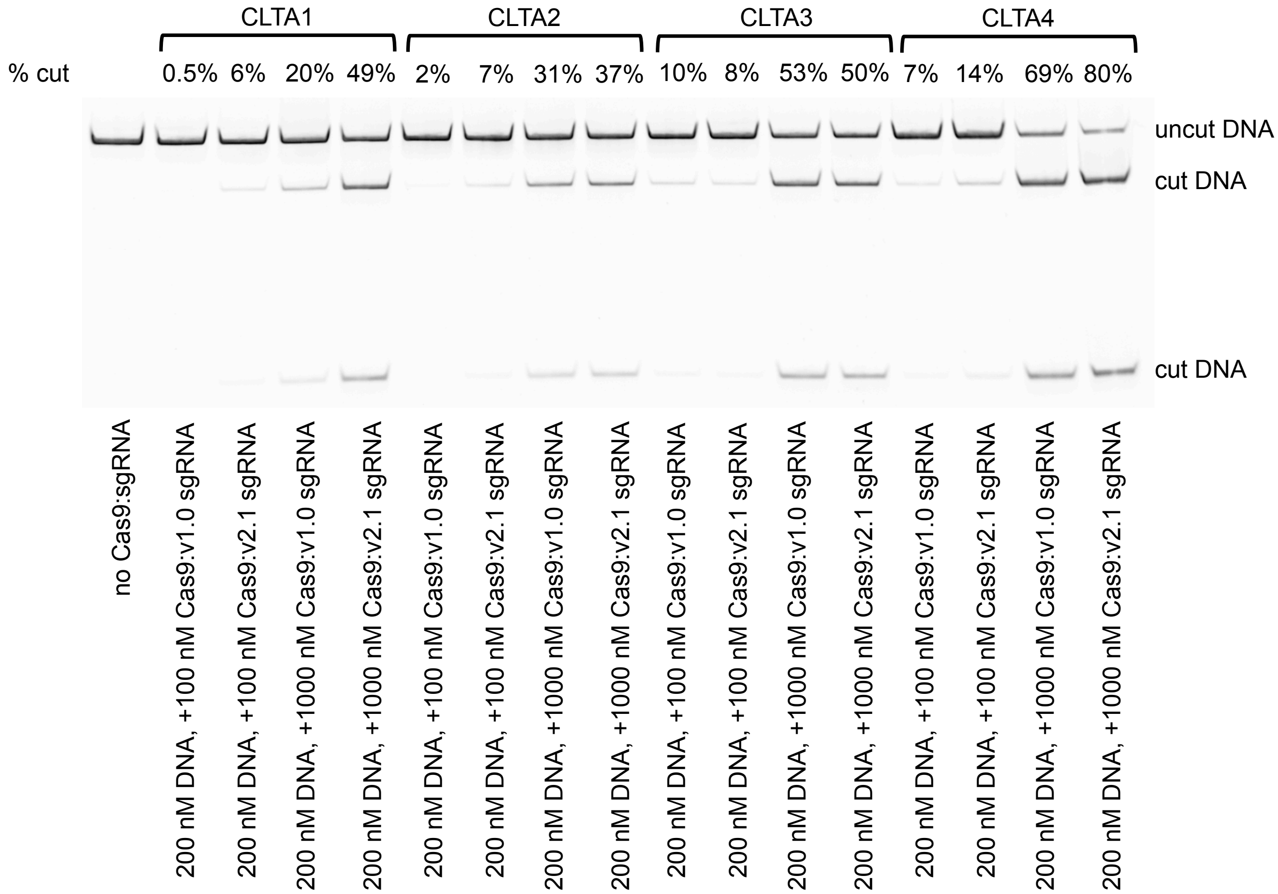
To assess the contribution of the PAM to specificity, we calculated the abundance of all 16 possible PAM dinucleotides in the pre-selection and post-selection libraries, considering all observed post-selection target site sequences (**Supplementary Figure S19**) or considering only post-selection target site sequences that contained no mutations in the 20 base pairs specified by the guide RNA (**Supplementary Figure S20**). Considering all observed post-selection target site sequences, under enzyme-limiting conditions, GG dinucleotides represented 99.8%, 99.9%, 99.8%, and 98.5% of the post-selection PAM dinucleotides for selections with CLTA1, CLTA2, CLTA3, and CLTA4 v2.1 sgRNAs, respectively. In contrast, under enzyme-excess conditions, GG dinucleotides represented 97.7%, 98.3%, 95.7%, and 87.0% of the post-selection PAM dinucleotides for selections with CLTA1, CLTA2, CLTA3, and CLTA4 v2.1 sgRNAs, respectively. These data demonstrate that an increase in enzyme concentration leads to increased cleavage of substrates containing non-canonical PAM dinucleotides.

To account for the pre-selection library distribution of PAM dinucleotides, we calculated specificity scores for the PAM dinucleotides (**Supplementary Figures S21** and **S22**). When only on-target post-selection sequences are considered under enzyme-excess conditions (**Supplementary Figure S22**), non-canonical PAM dinucleotides with a single G rather than two Gs are relatively tolerated. Under enzyme-excess conditions, Cas9:CLTA4 sgRNA 2.1 exhibited the highest tolerance of non-canonical PAM dinucleotides of all the Cas9:sgRNA combinations tested. AG and GA dinucleotides were the most tolerated, followed by GT, TG, and CG PAM dinucleotides. In selections with Cas9:CLTA1, 2, or 3 sgRNA 2.1 under enzyme-excess conditions, AG was the predominate non-canonical PAM (**Supplementary Figures S21** and **S22**). Our results are consistent with another recent study of PAM specificity, which shows that Cas9:sgRNA can recognize AG PAM dinucleotides[1]. In addition, our results show that under enzyme-limiting conditions, GG PAM dinucleotides are highly specified, and under enzyme-excess conditions, non-canonical PAM dinucleotides containing a single G can be tolerated, depending on the guide RNA context.

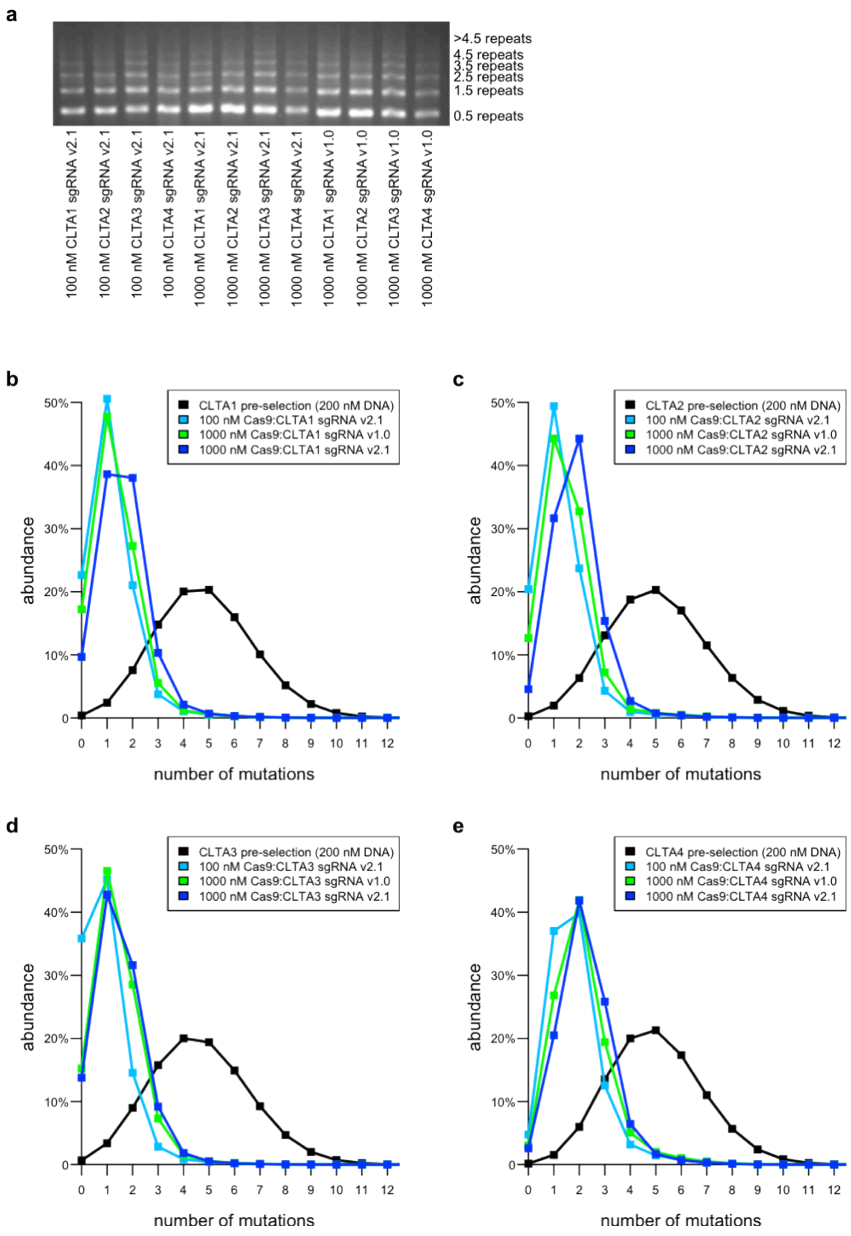
SUPPLEMENTARY FIGURES AND TABLES



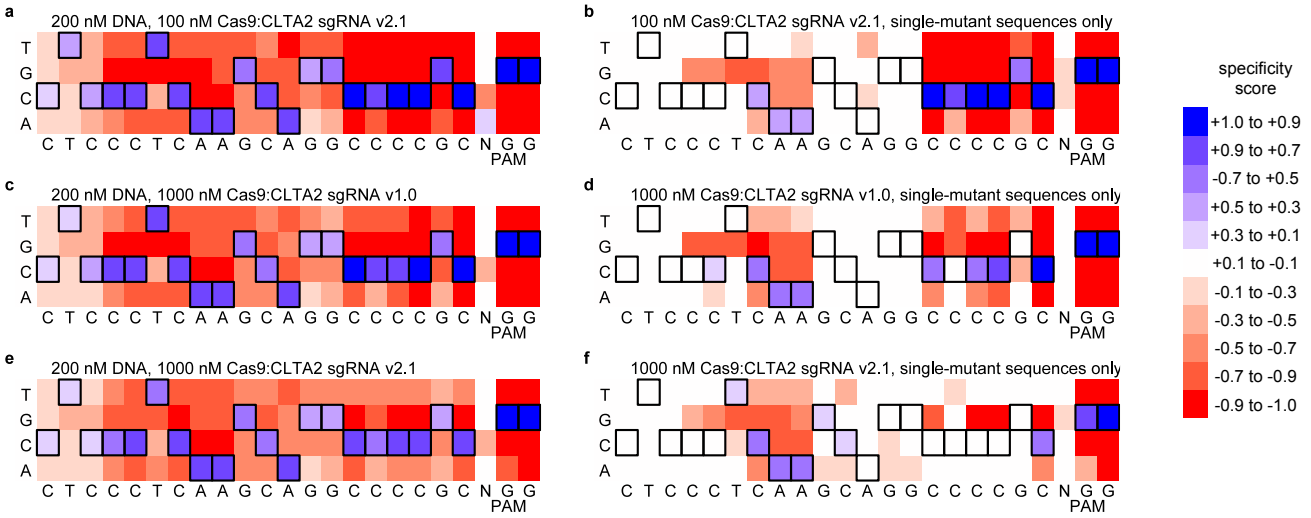
Supplementary Figure S1. Target sites profiled in this study. (**a**) The 5' end of the sgRNA has 20 nucleotides that are complementary to the target site. The target site contains an NGG motif (PAM) adjacent to the region of RNA:DNA complementarity. (**b**) Four human clathrin gene (CLTA) target sites are shown. (**c, d**) Four human clathrin gene (CLTA) target sites are shown with sgRNAs. sgRNA v1.0 is shorter than sgRNA v2.1. The PAM is shown in red for each site. The non-PAM end of the target site corresponds to the 5' end of the sgRNA.



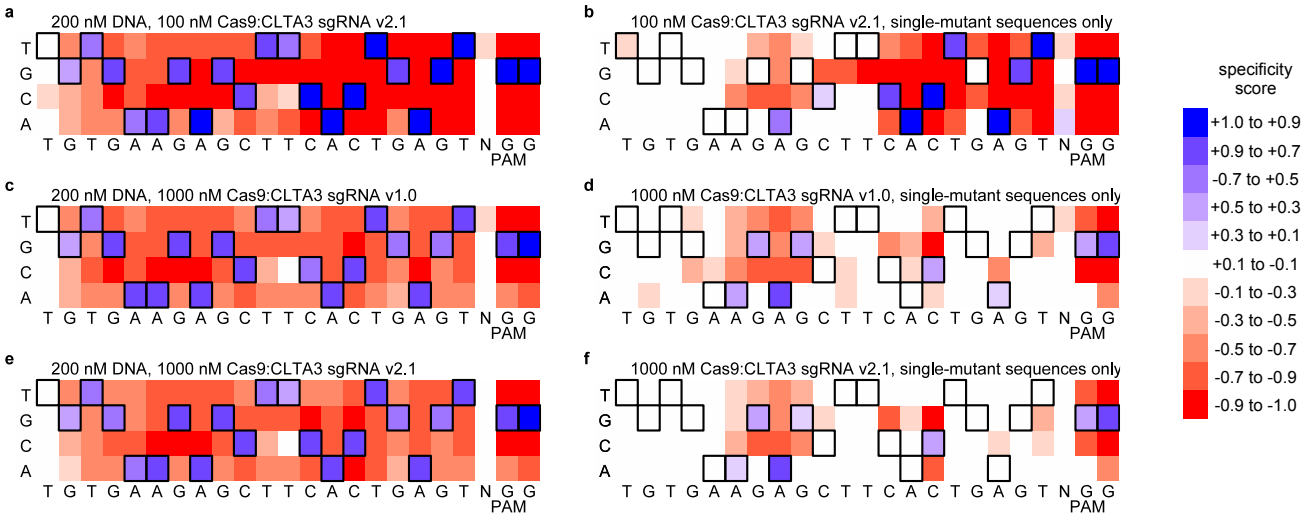
Supplementary Figure S2. **Cas9:guide RNA cleavage of on-target DNA sequences *in vitro***. Discrete DNA cleavage assays on an approximately 1-kb linear substrate were performed with 200 nM on-target site and 100 nM Cas9:v1.0 sgRNA, 100 nM Cas9:v2.1 sgRNA, 1000 nM Cas9:v1.0 sgRNA, and 1000 nM Cas9:v2.1 sgRNA for each of four CLTA target sites. For CLTA1, CLTA2, and CLTA4, Cas9:v2.1 sgRNA shows higher activity than Cas9:v1.0 sgRNA. For CLTA3, the activities of the Cas9:v1.0 sgRNA and Cas9:v2.1 sgRNA were comparable.



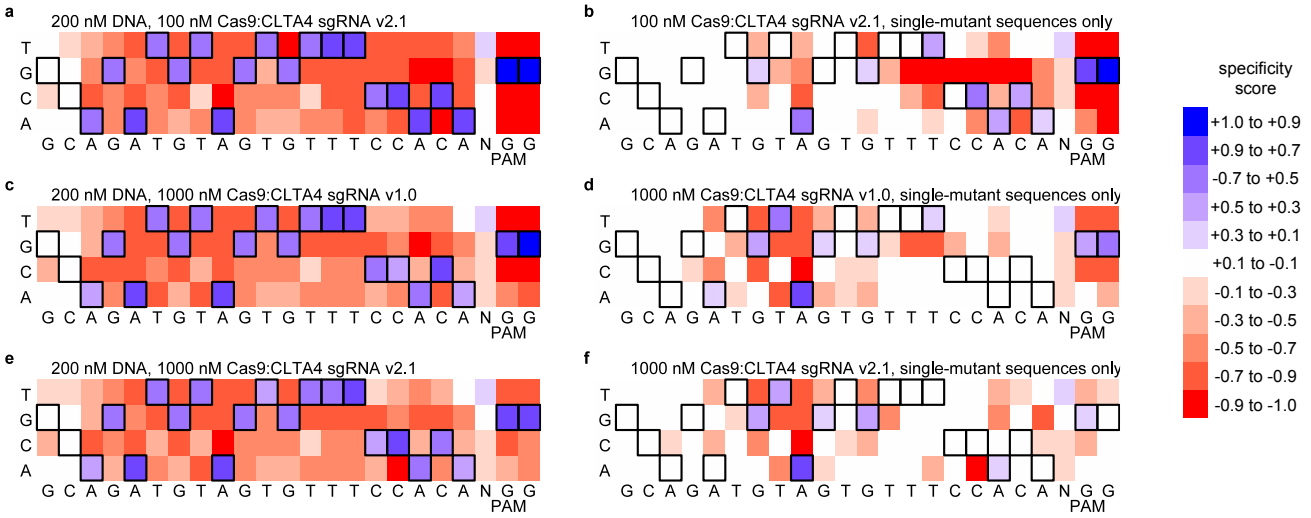
Supplementary Figure S3. ***In vitro*** **selection results for four target sites.** *In vitro* selections were performed on 200 nM pre-selection library with 100 nM Cas9:sgRNA v2.1 (light blue), 1000 nM Cas9:sgRNA v1.0 (green), or 1000 nM Cas9:sgRNA v2.1 (dark blue). (**a**) Post-selection PCR products are shown for the 12 selections performed. DNA containing 1.5 repeats were quantified for each selection and pooled in equimolar amounts before gel purification and sequencing. (**b**-**e**) Distributions of mutations are shown for pre-selection (black) and post-selection libraries (colored). The post-selection libraries are enriched for sequences with fewer mutations than the pre-selection libraries. Mutations are counted from among the 20 base pairs specified by the sgRNA and the two-base pair PAM. *P*-values are < 0.01 for all pairwise comparisons between distributions in each panel. *P*-values were calculated using t-tests, assuming unequal size and unequal variance.



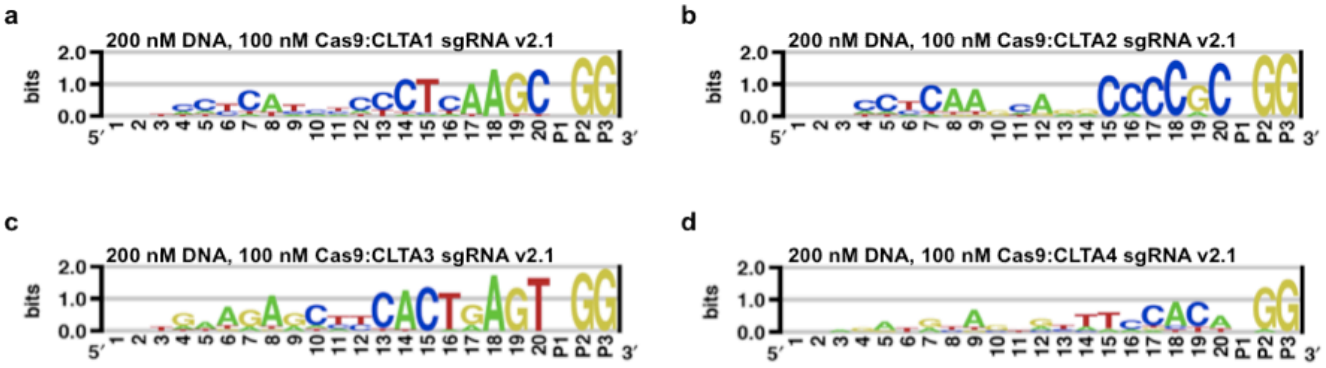
Supplementary Figure S4. ***In vitro*** **selection results for Cas9:CLTA2 sgRNA**. Heat maps[2] show the specificity profiles of Cas9:CLTA2 sgRNA v2.1 under enzyme-limiting conditions (**a, b**), Cas9:CLTA2 sgRNA v1.0 under enzyme-excess conditions (**c, d**), and Cas9:CLTA2 sgRNA v2.1 under enzyme-excess conditions (**e, f**). Heat maps show all post-selection sequences (**a, c, e**) or only those sequences containing a single mutation in the 20-base pair sgRNA-specified target site and two-base pair PAM (**b, d, f**). Specificity scores of 1.0 (dark blue) and -1.0 (dark red) corresponds to 100% enrichment for and against, respectively, a particular base pair at a particular position. Black boxes denote the intended target nucleotides.



Supplementary Figure S5. ***In vitro*** **selection results for Cas9:CLTA3 sgRNA**. Heat maps[2] show the specificity profiles of Cas9:CLTA3 sgRNA v2.1 under enzyme-limiting conditions (**a, b**), Cas9:CLTA3 sgRNA v1.0 under enzyme-excess conditions (**c, d**), and Cas9:CLTA3 sgRNA v2.1 under enzyme-saturating conditions (**e, f**). Heat maps show all post-selection sequences (**a, c, e**) or only those sequences containing a single mutation in the 20-base pair sgRNA-specified target site and two-base pair PAM (**b, d, f**). Specificity scores of 1.0 (dark blue) and -1.0 (dark red) corresponds to 100% enrichment for and against, respectively, a particular base pair at a particular position. Black boxes denote the intended target nucleotides.



Supplementary Figure S6. ***In vitro*** **selection results for Cas9:CLTA4 sgRNA**. Heat maps[2] show the specificity profiles of Cas9:CLTA4 sgRNA v2.1 under enzyme-limiting conditions (**a, b**), Cas9:CLTA4 sgRNA v1.0 under enzyme-excess conditions (**c, d**), and Cas9:CLTA4 sgRNA v2.1 under enzyme-saturating conditions (**e, f**). Heat maps show all post-selection sequences (**a, c, e**) or only those sequences containing a single mutation in the 20-base pair sgRNA-specified target site and two-base pair PAM (**b, d, f**). Specificity scores of 1.0 (dark blue) and -1.0 (dark red) corresponds to 100% enrichment for and against, respectively, a particular base pair at a particular position. Black boxes denote the intended target nucleotides.



Supplementary Figure S7. *In vitro* selection results as sequence logos. Information content is plotted[3] for each target site position (1-20) specified by CLTA1 (**a**), CLTA2 (**b**), CLTA3 (**c**), and CLTA4 (**d**) sgRNA v2.1 under enzyme-limiting conditions. Positions in the PAM are labelled "P1," "P2," and "P3." Information content is plotted in bits. 2.0 bits indicates absolute specificity and 0 bits indicates no specificity.

100 nM Cas9:CLTA1 v2.1 sgRNA

position	-4	-3	-2	-1	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	PAM1	PAM2	PAM3	+4	+5	+6	+7
A	212906	240335	195549	240068	1.04E+06	72751	40206	62972	41734	17376	18710	1.17E+06	24455	83195	46063	33528	8551	9668	4582	32237	1.19E+06	1.20E+06	2032	4237	261056	1386	574	235167	223887	222343	301956
C	285295	248395	263973	260202	37925	32496	24822	1.10E+06	1.12E+06	42444	1.16E+06	5339	22096	1.06E+06	48105	1.14E+06	1.14E+06	1.18E+06	4090	1.13E+06	4363	828	969	1.19E+06	210095	167	152	211027	273777	264354	309690
G	214213	219078	220275	189578	61062	1.04E+06	25785	11117	9125	5423	5745	5121	8080	14905	8906	3732	3294	3867	3597	7260	3400	2474	1.19E+06	1301	238989	1.20E+06	1.21E+06	205765	222282	240526	217260
T	493854	498460	526471	516420	64694	59173	1.12E+06	35336	34236	1.14E+06	20532	24018	1.15E+06	50488	1.10E+06	32417	57980	13065	1.19E+06	36826	8959	3966	9354	8065	496128	475	211	554309	486322	479045	377362

1000 nM Cas9:CLTA1 v1.0 sgRNA

position	-4	-3	-2	-1	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	PAM1	PAM2	PAM3	+4	+5	+6	+7
A	154613	184336	154288	177436	805105	66777	43354	56461	32941	15531	19465	904223	19696	56566	35200	26674	7925	9269	4859	32891	910633	925527	3595	5976	183479	1390	413	182704	171051	174062	221899
C	227144	201856	215667	220894	30269	30133	24249	825333	865486	35164	889622	5488	17340	828521	36975	876790	880022	908816	4419	859691	5694	776	2120	920211	180463	120	88	180657	220438	211411	245967
G	163868	174062	177891	148150	47940	784264	26342	17972	10299	6332	5785	5938	9185	11560	10641	3020	2819	3185	2994	6763	3631	2894	916417	1415	193418	932808	934044	172551	172071	176484	161703
T	389059	374430	386838	388204	51370	53510	840739	34918	25958	877657	19812	19035	888463	38037	851868	28200	43918	13414	922412	35339	14726	5487	12552	7082	377324	366	139	398772	371124	372727	305115

1000 nM Cas9:CLTA1 v2.1 sgRNA

position	-4	-3	-2	-1	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	PAM1	PAM2	PAM3	+4	+5	+6	+7
A	104782	127116	103361	124521	554601	40232	29541	38710	23659	10435	11462	618404	14608	41826	27762	19590	8961	9434	9549	35083	604115	607264	4685	16515	125225	10391	2519	125288	114575	120476	149847
C	154144	136337	145670	146754	20057	19440	17922	569754	590426	25233	612203	3834	15297	561351	26392	592757	594469	616112	11645	553993	13212	4438	5146	590160	116022	329	138	123802	154249	146572	166531
G	113998	119668	120741	103026	32861	547445	18468	9314	6346	3908	4295	3719	5851	10887	15360	5605	3378	3517	5896	22551	8658	12770	613580	3712	121392	628464	637588	118800	113560	118464	111278
T	267467	257270	270619	266090	32872	33274	574460	22613	19960	600815	12431	14434	604635	26327	570877	22439	33583	1328	613301	28764	14406	15919	17000	30004	277752	1207	146	272501	258007	254879	212735

CLTA1 pre-selection library

position	-4	-3	-2	-1	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	PAM1	PAM2	PAM3	+4	+5	+6	+7
A	241543	217144	209045	198284	943175	103452	76259	106919	124476	59762	108373	937511	65477	110282	67774	96299	88029	109977	62686	119399	931093	908362	64248	111479	190574	97896	104002	183367	178912	198049	219754
C	254366	269805	276090	322860	52984	65855	58943	834238	812029	52168	839963	54708	43285	831610	50109	861358	841819	817157	51676	797914	60106	52998	42317	813253	239201	56843	59450	289074	295400	289007	284268
G	230024	196574	210445	180859	60496	857631	66783	89366	85315	67098	77499	59257	71824	89579	68090	66121	86080	96496	81367	104949	52143	77389	918970	96000	192652	879150	870948	196672	202194	196499	202544
T	403590	446000	433943	427520	72868	102585	927538	99000	107703	950495	103688	78047	948937	98052	943550	105745	113595	105893	933794	107261	86181	90774	103988	108791	507096	95634	95123	460410	453017	445968	422957

100 nM Cas9:CLTA2 v2.1 sgRNA

position	-4	-3	-2	-1	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	PAM1	PAM2	PAM3	+4	+5	+6	+7
A	109129	135587	94032	141748	5.74E+04	44802	48284	24464	11611	16668	6282	6.58E+05	655917	28909	24210	656617	59160	36601	2974	12980	3.27E+03	1.09E+03	17686	689	193742	284	129	143150	165553	136708	146056
C	155710	138970	207735	220443	529643	24503	566049	6.27E+05	6.46E+05	19040	6.52E+05	2351	2577	1.30E+04	617274	2.64E+03	1.48E+04	9.12E+03	660929	6.50E+05	660305	666122	1314	6.65E+05	42664	48	43	162563	111729	143442	177253
G	136555	142038	118241	105620	39991	2.11E+04	26481	3756	3627	2889	2488	3025	3202	609865	8312	5889	581322	606454	1564	2134	1819	89	6.44E+05	505	137388	6.68E+05	6.68E+05	103305	146355	139972	124772
T	266918	251717	248304	200501	41277	577893	2.75E+04	13008	7318	6.30E+05	7487	4920	6.62E+03	16554	1.85E+04	3165	13024	16134	2.85E+03	3253	2918	1016	4886	2608	294518	146	42	259294	244675	248190	220231

1000 nM Cas9:CLTA2 v1.0 sgRNA

position	-4	-3	-2	-1	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	PAM1	PAM2	PAM3	+4	+5	+6	+7
A	94138	115628	85485	120876	52411	41438	46093	22399	9066	14310	5351	567337	565061	24132	23848	558483	49577	39401	5425	30774	6408	5055	36081	2573	148145	782	243	132801	126862	118528	122897
C	140695	125708	179224	191394	452192	21517	481298	538392	557549	16233	562576	1973	2127	11807	525901	4992	13617	13316	563557	535780	560658	567693	4938	569653	46472	70	45	133402	123970	130555	148756
G	113243	118054	101836	91048	35101	18969	22797	3440	2802	2960	2526	2895	2793	526655	9738	8100	495166	496382	1789	3325	1846	166	519782	520	125177	575395	576103	118877	108849	104210	103370
T	228367	217053	209898	173125	36739	494519	26255	12212	7026	542940	5990	4238	6462	13849	16956	4868	18093	27344	5672	6564	7531	3529	15642	3697	256649	196	52	191363	216762	223150	201420

1000 nM Cas9:CLTA2 v2.1 sgRNA

position	-4	-3	-2	-1	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	PAM1	PAM2	PAM3	+4	+5	+6	+7
A	64249	81812	58977	85387	35172	29833	33434	19419	9272	13136	4907	391675	389930	19852	16657	383605	32780	22855	9722	25181	12518	17950	28198	5471	100745	4933	834	89339	87351	82615	85108
C	96983	87918	124642	127760	316077	14548	327166	364874	380987	11360	387025	1694	1815	8124	363374	5168	9569	9710	374342	355544	373485	370343	11652	378841	40532	238	34	93621	87920	91380	105625
G	77913	80500	68612	64299	23522	15748	19664	3856	3035	2752	2062	2398	2439	360755	7431	6019	344511	350245	1559	5882	1339	391	331376	1034	74803	393760	398660	79776	75927	74068	70435
T	160415	149330	147329	122114	24789	339431	19296	11411	6266	372312	5566	3793	5376	10829	12098	4768	12700	16750	13937	12953	12218	10876	28334	14214	183480	629	32	136824	148362	151497	138392

CLTA2 pre-selection library

position	-4	-3	-2	-1	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	PAM1	PAM2	PAM3	+4	+5	+6	+7
A	203147	173899	167999	170203	89970	73770	88239	88611	76114	78589	75016	726091	712150	96111	90307	728931	91515	84764	79586	86205	87337	85547	92983	100316	177716	84144	88017	177831	180209	176904	174190
C	181430	214835	246369	272618	632831	41977	641062	644565	670872	40877	649838	38931	44691	46591	628706	32296	49519	46571	641958	624548	637703	635473	51727	594349	136372	41282	41689	216880	206368	210039	235263
G	177090	153006	151178	140868	58664	49976	60827	56077	52341	46259	55484	39801	38939	630670	55013	38368	627263	642878	59549	55292	53056	57979	616575	66553	158929	656315	654970	162242	160704	157741	138890
T	285951	305878	282072	263929	66153	681895	57490	58365	48291	678893	67280	42795	51838	74246	73592	48023	79321	73405	66525	81573	69522	68619	86333	86400	374601	65877	62942	290665	300337	302934	299275

100 nM Cas9:CLTA3 v2.1 sgRNA

position	-4	-3	-2	-1	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	PAM1	PAM2	PAM3	+4	+5	+6	+7
A	212836	248582	202151	249368	9.13E+04	77392	19048	39738	1078520	1106930	46196	1.12E+06	64461	11912	30992	21158	6465	1130430	4097	5750	4.71E+04	1.14E+06	6151	2047	305062	1993	394	213566	240851	230230	252637
C	233270	241259	274819	305120	37894	35918	13930	5.61E+03	1.22E+04	3774	6.35E+03	4063	6018	1.11E+06	27501	4.68E+04	1.12E+06	1.96E+03	1129400	1.82E+03	3421	167	1451	6.66E+02	261609	103	82	319990	253055	261338	293644
G	211701	187534	185281	196614	66632	9.88E+05	26572	1074020	12936	9205	1066570	7418	1050360	3828	3949	2231	2504	2471	1726	2881	1081680	876	1.13E+06	600	228865	1.14E+06	1.14E+06	142425	192720	220683	227840
T	480761	461193	476317	387466	942707	37284	1.08E+06	19204	34885	1.87E+04	19450	11145	1.77E+04	13689	1.08E+06	1068370	9829	3709	3.34E+03	1128120	6398	1320	4480	1135280	343032	211	69	462587	451942	426317	364447

1000 nM Cas9:CLTA3 v1.0 sgRNA

position	-4	-3	-2	-1	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	PAM1	PAM2	PAM3	+4	+5	+6	+7
A	219833	263464	207913	264018	97886	78562	20663	39724	1136320	1151200	42966	1156400	49443	18669	44652	44644	44771	1152540	16264	30980	71714	1156700	47106	27658	276285	36304	12701	219034	239515	244440	255360
C	240570	261247	311444	333414	39996	40484	13961	5323	11099	5475	10323	6501	8456	1126310	36792	56203	1096280	8437	1156840	8448	25120	4351	24685	9473	297135	1331	939	354289	298216	277740	292917
G	221683	206195	199246	215583	76580	1032080	24785	1126840	12654	12465	1114450	12075	1113930	12078	19275	9014	7707	9466	2708	17195	1053760	10278	1085310	10308	238545	1148550	1174510	171862	193096	217301	239319
T	506611	457791	470094	375682	974235	37571	1129290	16811	28626	19560	20956	13723	16864	31636	1087980	1078840	39940	18250	12883	1132070	38103	17372	31596	1141260	376732	2514	550	443512	457870	449216	401101

1000 nM Cas9:CLTA3 v2.1 sgRNA

position	-4	-3	-2	-1	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	PAM1	PAM2	PAM3	+4	+5	+6	+7
A	169775	206549	166197	201768	75243	67150	20449	36549	876154	898360	39901	911344	44415	13218	37301	33080	26409	893670	6315	20807	52541	903619	33690	20904	205940	26623	9880	172210	182986	187305	196429
C	197800	209445	243688	264177	32775	34540	14250	7885	14793	4878	7791	4636	7510	890591	28425	46269	870864	7991	910584	5931	19923	4977	18171	6508	229797	1163	693	283240	240802	224453	236469
G	174766	158928	158824	168325	58121	801768	26558	866689	13343	12052	868394	8837	867980	7923	14022	6553	3393	7912	1499	12906	836022	9011	859000	8302	190011	906628	925513	132620	153591	172169	187623
T	394073	361492	367705	302144	770275	32956	875157	25291	32124	21124	20328	11597	16509	24682	856666	850512	35748	28841	18016	896770	27928	18807	24953	900700	310666	2000	328	348344	359035	352487	315893

CLTA3 pre-selection library

position	-4	-3	-2	-1	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	PAM1	PAM2	PAM3	+4	+5	+6	+7
A	173122	135327	138244	142599	50365	69486	37040	66315	575295	566722	70249	528947	72610	41265	61770	56547	75555	586476	61203	51740	70943	569277	70484	50807	130402	57527	61702	110207	118993	126967	127707
C	143788	158534	162646	177240	25902	40142	28129	34668	38933	36129	61591	52201	46032	559715	32233	34830	519328	30904	540977	24982	45344	35359	44014	35778	174938	42259	46083	201434	190347	184768	207347
G	137601	132826	130592	128304	42860	534378	42217	531723	29873	34068	479149	49753	501888	41949	43243	30118	38922	34282	34082	37275	515778	36956	516177	45203	137307	539445	527404	113323	119846	118423	127230
T	238486	266310	261515	244854	573870	48991	585611	60291	48896	56078	82008	62096	72467	50068	555751	571502	59192	41335	56735	579000	60932	51405	62322	561209	250350	53766	57808	268033	263811	262839	230713

100 nM Cas9:CLTA4 v2.1 sgRNA

position	-4	-3	-2	-1	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	PAM1	PAM2	PAM3	+4	+5	+6	+7
A	55030	78101	78867	81833	8.09E+04	58148	525585	29962	544918	19446	54151	2.59E+04	550200	29521	34194	38891	26542	23991	15243	25122	5.36E+03	5.51E+05	1994	540029	47731	4642	1401	77633	56902	63224	54815
C	168401	162082	139480	130495	22088	428628	4498	1.21E+04	5.14E+03	15601	7.10E+03	35217	2481	2.35E+04	16846	2.03E+04	3.69E+04	8.47E+03	5182	5.22E+05	547711	5715	546119	3.02E+03	152056	655	473	141123	164035	146401	190955
G	89302	75785	82959	133275	415632	4.70E+04	14868	504358	6156	9951	493432	14899	4528	498832	27411	497382	6729	3344	3716	3926	3162	554	1.45E+03	4637	72296	5.55E+05	5.58E+05	84257	77627	75123	91454
T	248025	244790	259452	215155	42090	26956	1.58E+04	14300	4541	5.16E+05	6071	484788	3.55E+03	8877	4.82E+05	4222	490573	524958	5.37E+05	9437	4528	3692	11194	13069	288675	911	495	257745	262194	276010	223534

1000 nM Cas9:CLTA4 v1.0 sgRNA

position	-4	-3	-2	-1	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	PAM1	PAM2	PAM3	+4	+5	+6	+7
A	95188	141261	145156	141850	151224	116745	928773	50295	975924	29201	95476	30383	980248	50181	65094	77253	42674	41050	32933	55244	39984	942989	19900	887311	80159	28536	12390	142460	96664	110844	99920
C	305024	297215	260676	243819	34420	745345	8606	17266	7541	29948	10779	47831	5069	32501	30389	29610	61641	25910	21400	887446	900777	34590	940504	23749	257985	2556	4791	252462	297152	258929	338099
G	159888	139073	153474	225343	742232	85777	29776	907007	9285	13455	883325	19640	8303	902733	44730	879985	16677	7879	8429	12432	17373	4103	7346	20095	139488	964013	976818	154302	139784	136512	165750
T	438973	421524	439767	388061	71197	51206	31918	24505	6323	926469	9493	901219	5453	13658	858860	12225	878081	924234	936311	43951	40939	17391	31323	67918	521441	3968	5074	449849	465473	492788	395304

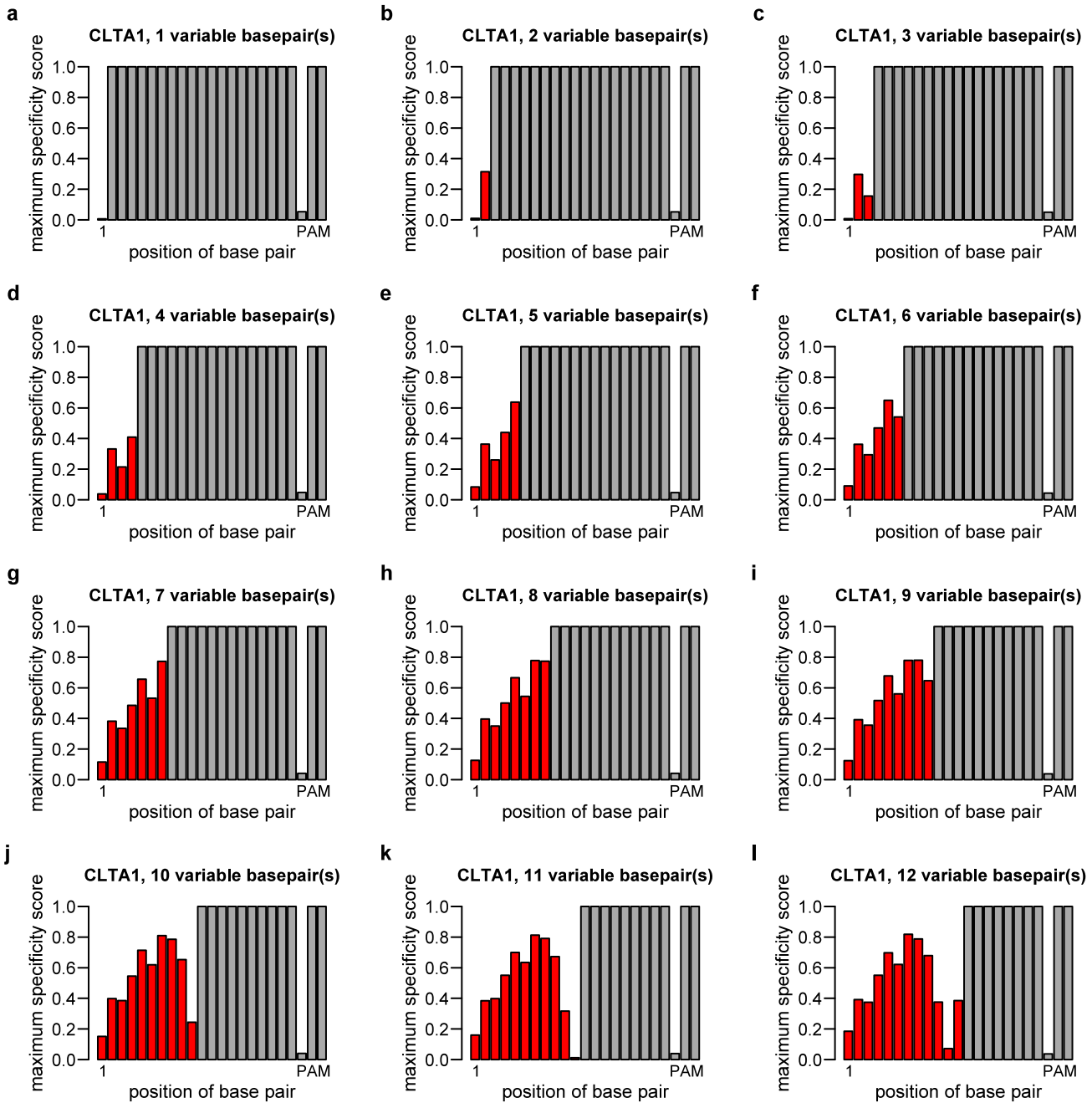
1000 nM Cas9:CLTA4 v2.1 sgRNA

position	-4	-3	-2	-1	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	PAM1	PAM2	PAM3	+4	+5	+6	+7
A	47674	70467	71535	72698	72554	54587	471218	27627	493315	16818	47470	17728	498471	29769	40021	41618	24741	23050	16409	27974	2697	478335	12667	451298	36128	22041	16967	68943	49017	56451	51102
C	154985	151636	133622	122579	18730	384037	4452	10916	4303	16232	5436	28594	1961	19017	19152	18001	35213	12845	13497	445302	480543	15631	469503	11832	122541	3529	8965	126313	153105	134293	171499
G	80869	69972	76726	118084	379024	42360	14989	453870	5084	6863	448794	10260	3281	450120	23076	439828	7741	5091	5456	7558	7112	3083	5302	10184	87517	474540	471647	85849	72063	71600	85239
T	222651	214104	224296	192818	35871	25195	15520	13766	3477	466266	4489	449597	2466	7273	423930	6732	438484	465193	470817	25345	15827	9130	18707	32865	259993	6069	8600	225074	231994	243835	198339

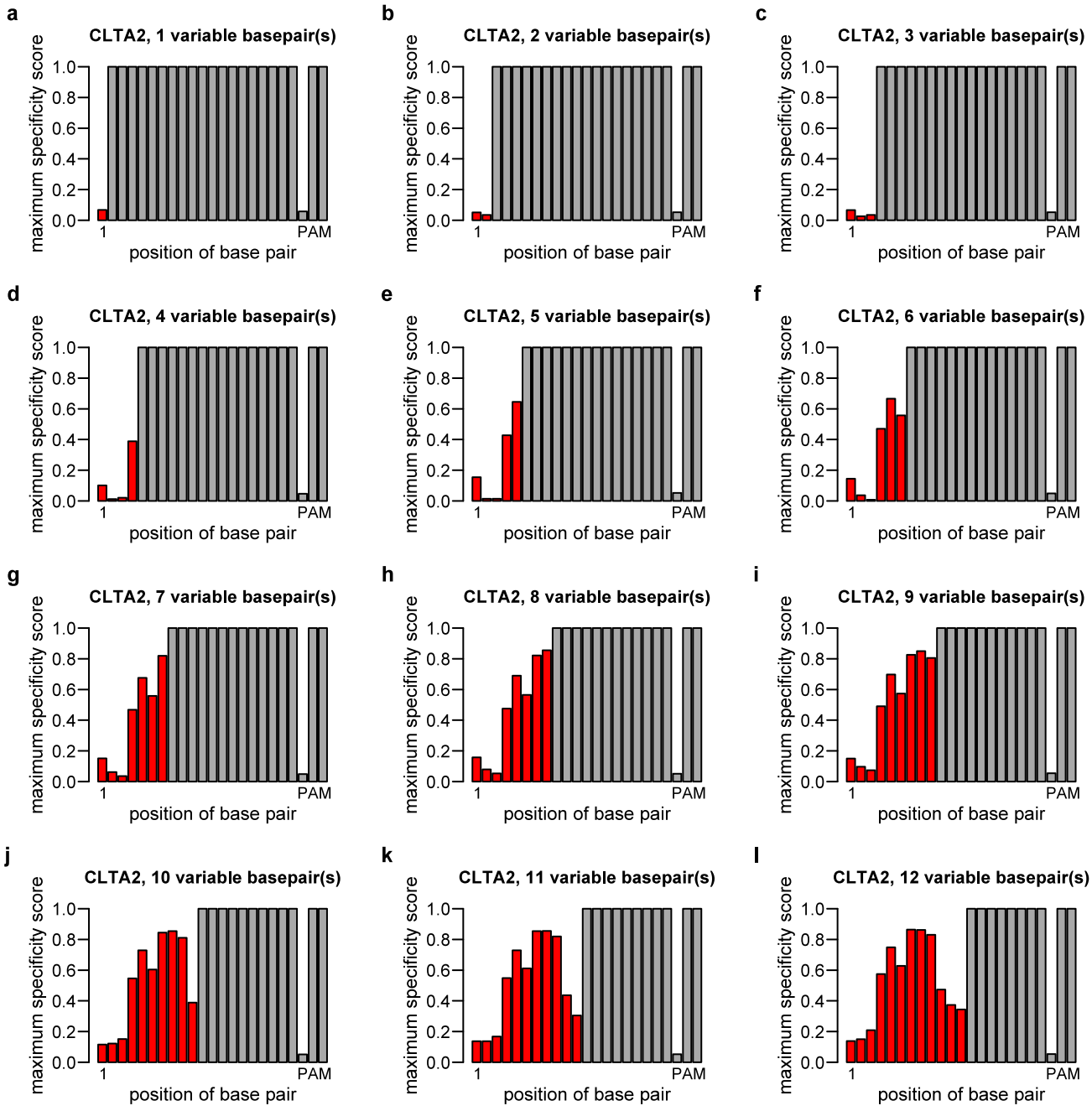
CLTA4 pre-selection library

position	-4	-3	-2	-1	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	PAM1	PAM2	PAM3	+4	+5	+6	+7
A	106798	131577	131941	132368	127160	103294	820923	103844	840417	99163	133349	123366	824537	126564	115133	122618	108492	107761	96384	99908	76163	806675	75877	793806	87755	82110	83605	111015	103082	109315	101198
C	304597	297419	277233	283453	50833	722264	29748	65558	44890	59551	73916	77470	45318	84973	73106	90384	78280	76978	66776	738550	776738	55522	754283	42188	278802	57603	55530	266156	281433	258029	295144
G	146240	137027	134399	183111	695802	68240	51484	708098	30709	62837	673752	89897	49093	672860	88125	663922	67768	53472	58440	47550	41427	42574	54424	59162	151536	740525	732891	163824	158224	146268	151560
T	393868	385480	407930	352571	77708	57705	49348	74003	35487	729952	70486	660770	32555	67106	675139	74579	696963	713292	729903	65495	57175	46732	66919	56347	433410	71265	79477	410508	408764	437891	403601

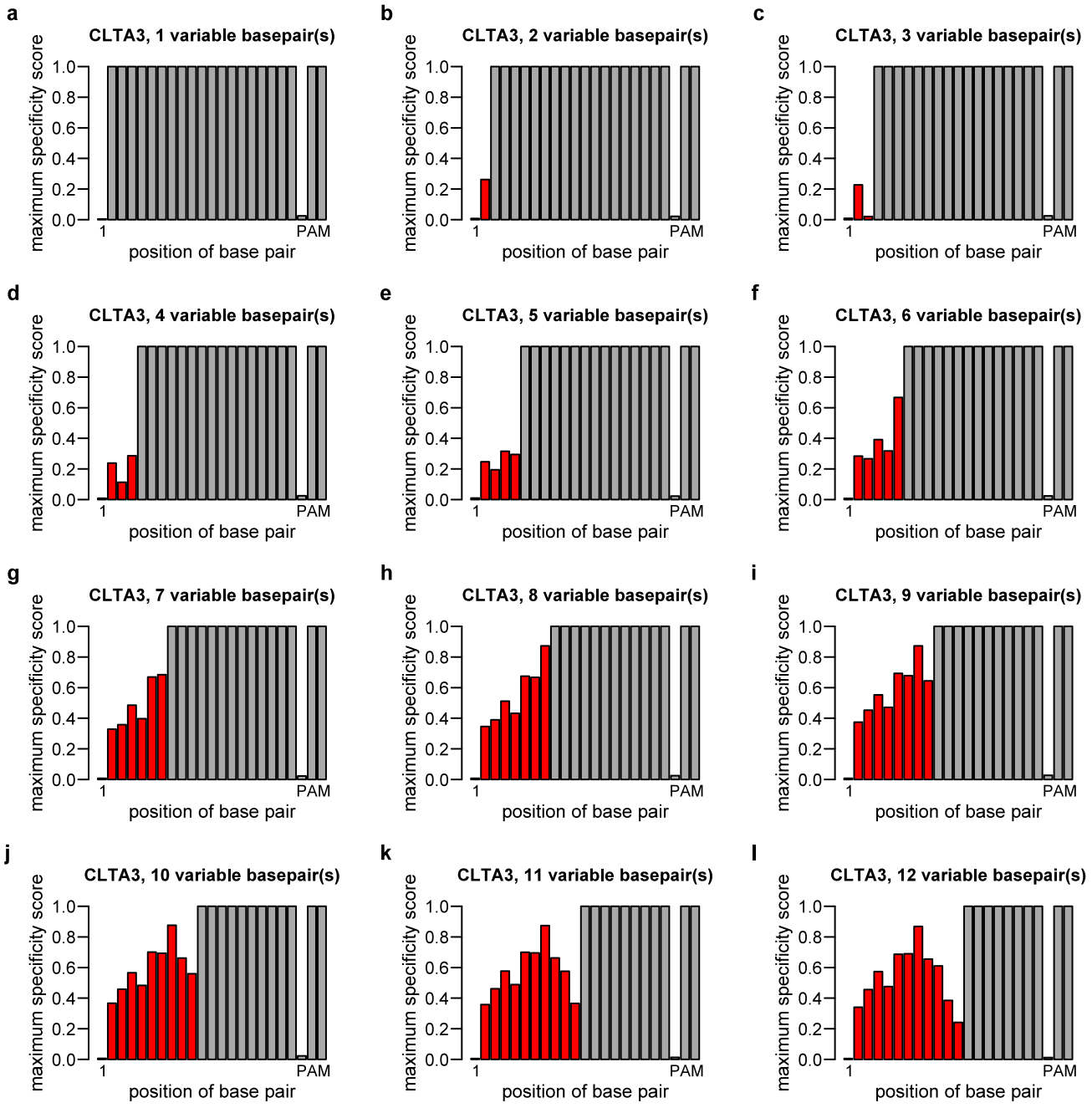
Supplementary Table S1. Raw selection sequence counts. Positions -4 to -1 are the four nucleotides preceeding the 20-base pair target site. PAM1, PAM2, and PAM3 are the PAM positions immediately following the target site. Positions +4 to +7 are the four nucleotides immediately following the PAM.



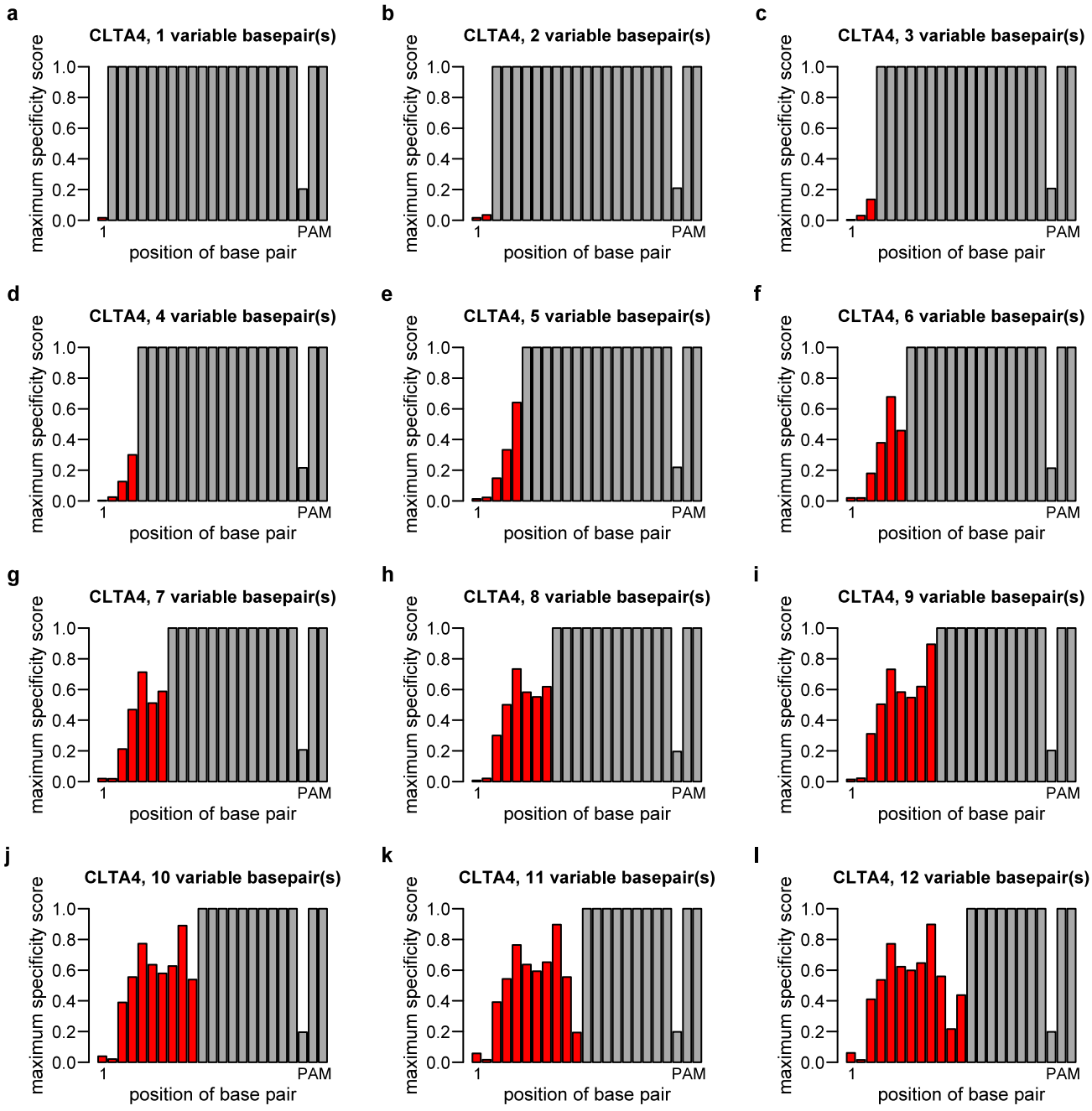
Supplementary Figure S8. Tolerance of mutations distal to the PAM for CLTA1. The maximum specificity scores at each position are shown for the Cas9:CLTA1 v2.1 sgRNA selections when considering only those sequences with on-target base pairs in gray, while allowing mutations in the first 1-12 base pairs (**a-l**). The positions that are not constrained to on-target base pairs are indicated by red bars. Higher specificity score values indicate higher specificity at a given position. The positions that were not allowed to contain any mutations (gray) were plotted with a specificity score of +1. For all panels, specificity scores were calculated from pre-selection library sequences and post-selection library sequences with an $n \geq 5{,}130$ and $n \geq 74{,}538$, respectively.



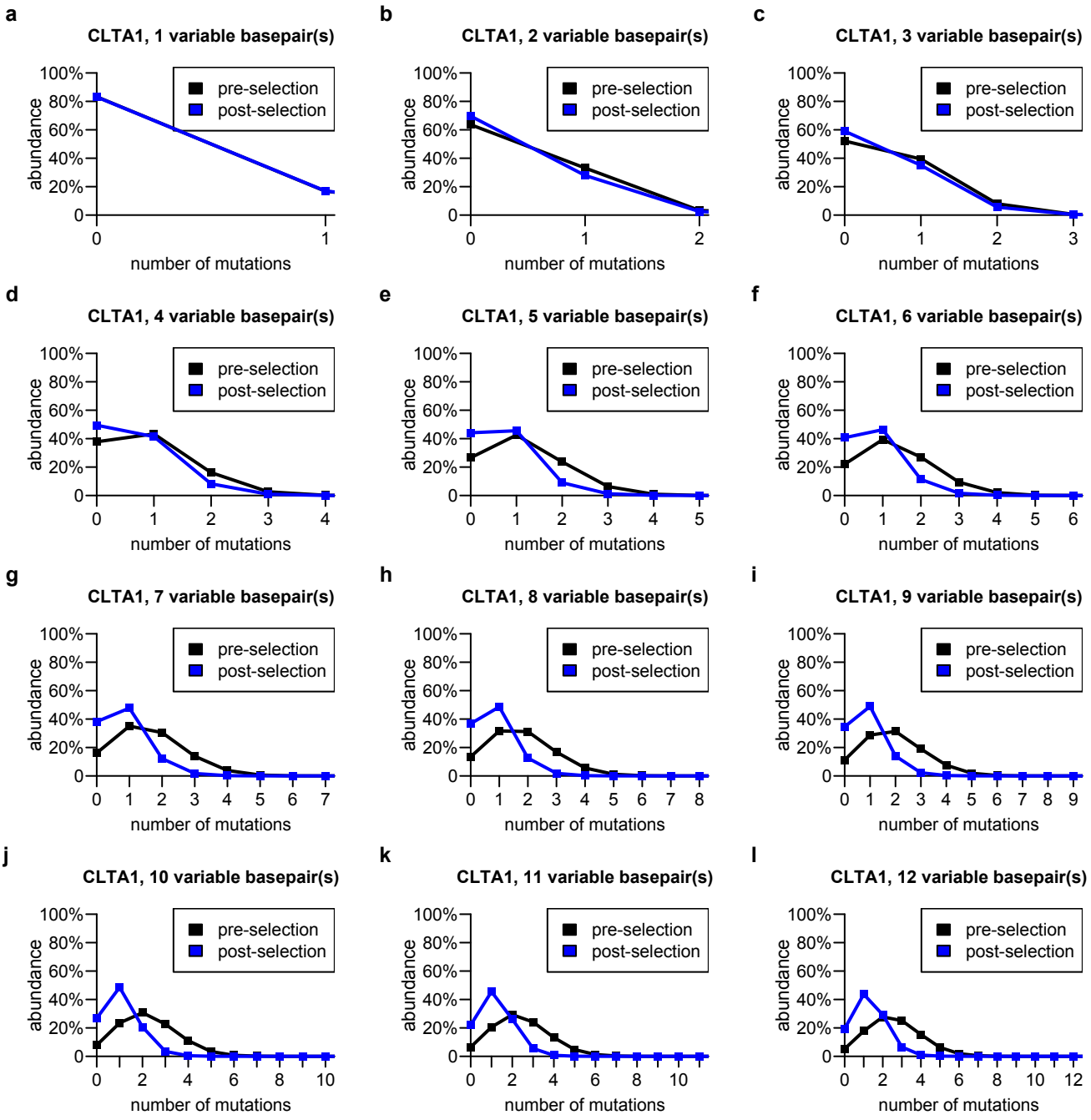
Supplementary Figure S9. Tolerance of mutations distal to the PAM for CLTA2. The maximum specificity scores at each position are shown for the Cas9:CLTA2 v2.1 sgRNA selections when considering only those sequences with on-target base pairs in gray, while allowing mutations in the first 1-12 base pairs (**a-l**). The positions that are not constrained to on-target base pairs are indicated by red bars. Higher specificity score values indicate higher specificity at a given position. The positions that were not allowed to contain any mutations (gray) were plotted with a specificity score of +1. For all panels, specificity scores were calculated from pre-selection library sequences and post-selection library sequences with an $n \geq 3{,}190$ and $n \geq 25{,}365$, respectively.



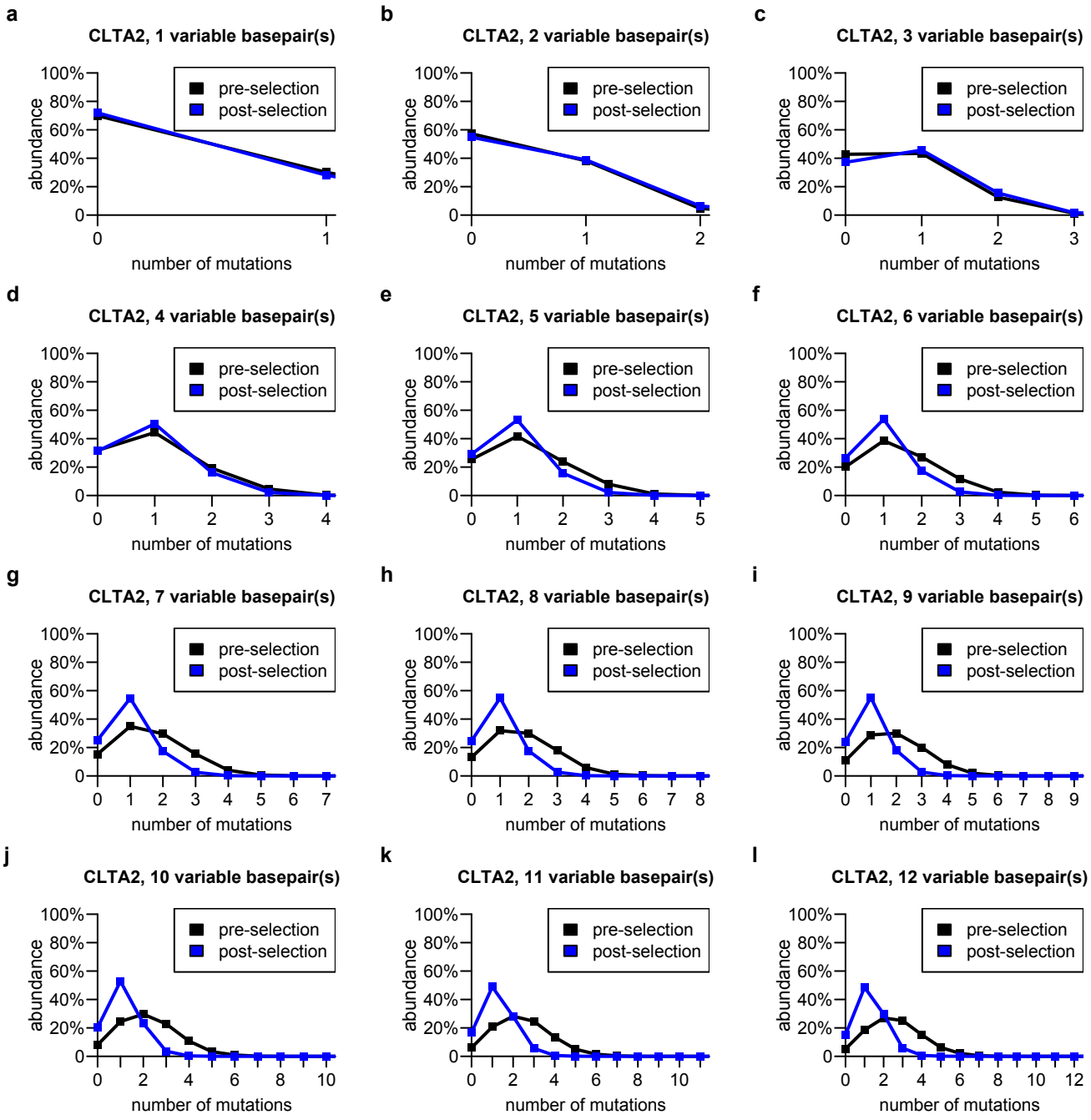
Supplementary Figure S10. Tolerance of mutations distal to the PAM for CLTA3. The maximum specificity scores at each position are shown for the Cas9:CLTA3 v2.1 sgRNA selections when considering only those sequences with on-target base pairs in gray, while allowing mutations in the first 1-12 base pairs (**a-l**). The positions that are not constrained to on-target base pairs are indicated by red bars. Higher specificity score values indicate higher specificity at a given position. The positions that were not allowed to contain any mutations (gray) were plotted with a specificity score of +1. For all panels, specificity scores were calculated from pre-selection library sequences and post-selection library sequences with an $n \geq 5{,}604$ and $n \geq 158{,}424$, respectively.



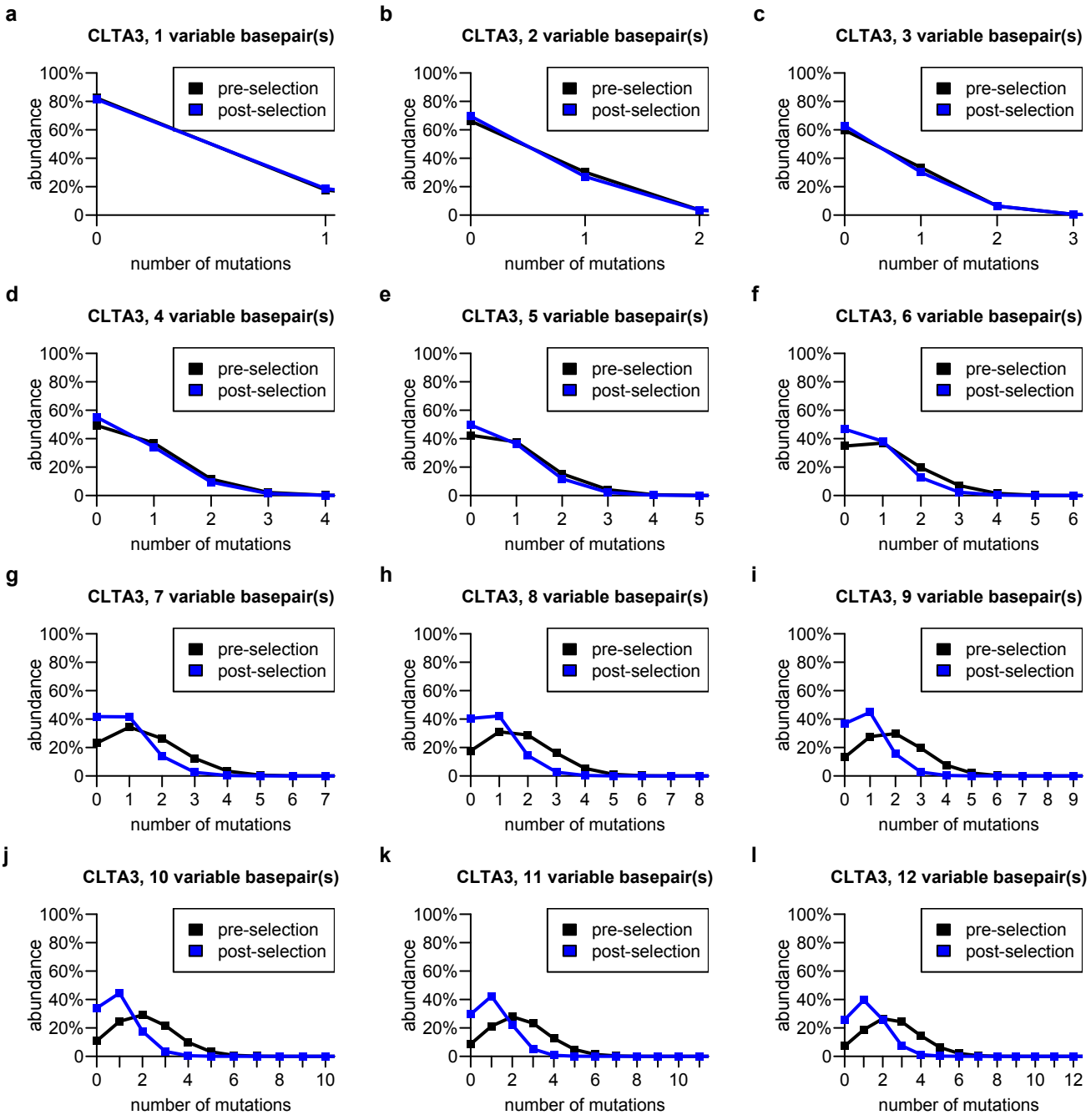
Supplementary Figure S11. Tolerance of mutations distal to the PAM for CLTA4. The maximum specificity scores at each position are shown for the Cas9:CLTA4 v2.1 sgRNA selections when considering only those sequences with on-target base pairs in gray, while allowing mutations in the first 1-12 base pairs (**a-i**). The positions that are not constrained to on-target base pairs are indicated by red bars. Higher specificity score values indicate higher specificity at a given position. The positions that were not allowed to contain any mutations (gray) were plotted with a specificity score of +1. For all panels, specificity scores were calculated from pre-selection library sequences and post-selection library sequences with an $n \geq 2{,}323$ and $n \geq 21{,}819$, respectively.



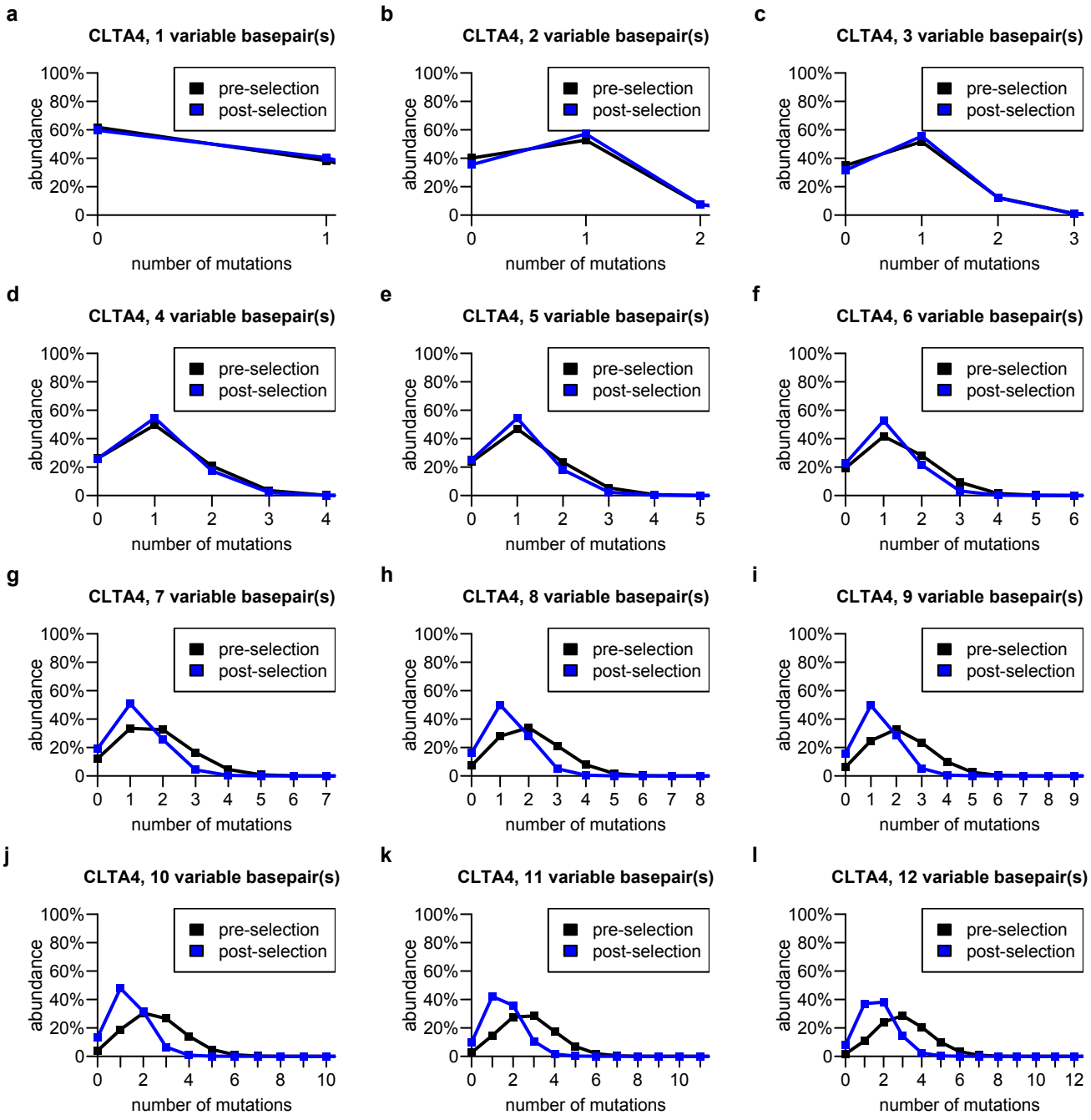
Supplementary Figure S12. Tolerance of mutations distal to the PAM in CLTA1 target sites. Distributions of mutations are shown for *in vitro* selection on 200 nM pre-selection library (black) with 1000 nM Cas9:CLTA1 sgRNA v2.1 (blue). The number of mutations shown are in a 1-12 base pair target site subsequence farthest from the PAM (**a-l**) when the rest of the target site, including the PAM, contains only on-target base pairs. The pre-selection and post-selection distributions are similar for up to three base pairs, demonstrating tolerance for target sites with mutations in the three base pairs farthest from the PAM when the rest of the target sites have optimal interactions with the Cas9:sgRNA. For all panels, graphs were generated from pre-selection library sequences and post-selection library sequences with an $n \geq 5{,}130$ and $n \geq 74{,}538$, respectively.



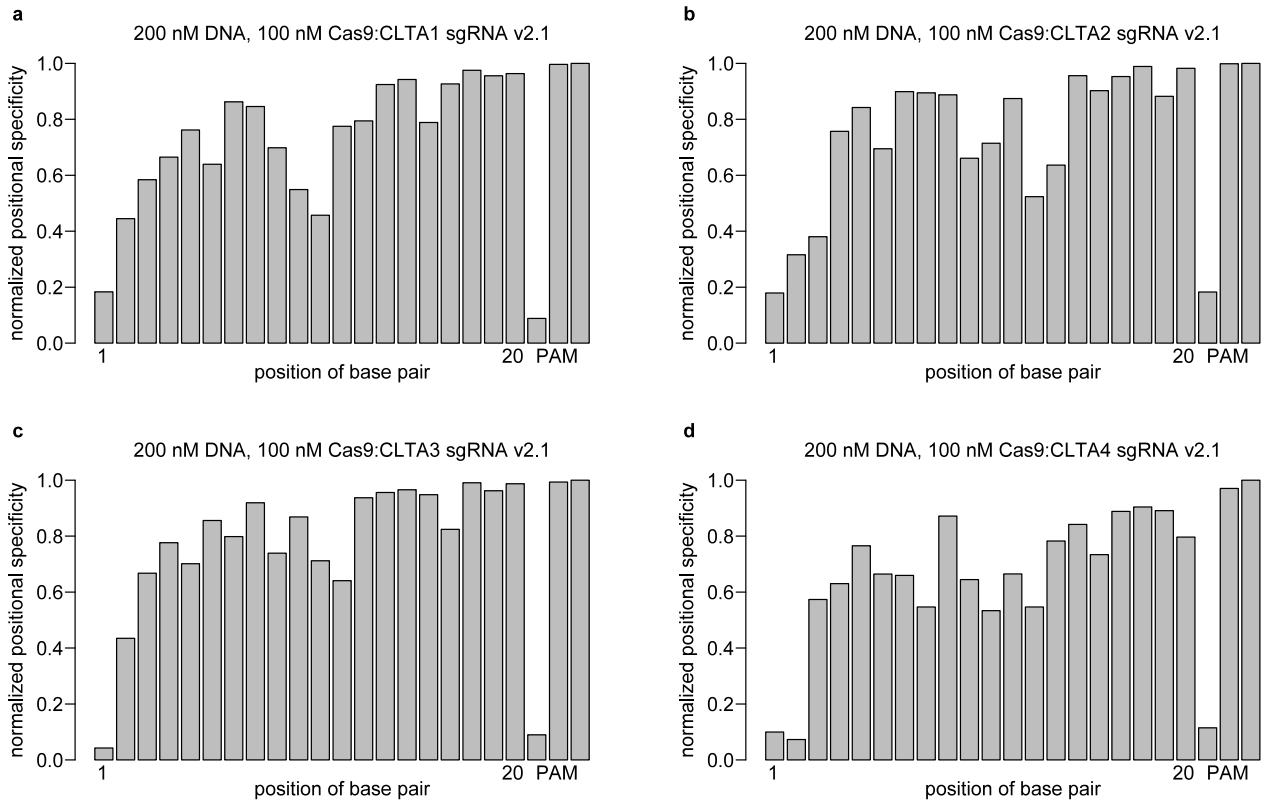
Supplementary Figure S13. Tolerance of mutations distal to the PAM in CLTA2 target sites. Distributions of mutations are shown for *in vitro* selection on 200 nM pre-selection library (black) with 1000 nM Cas9:CLTA2 sgRNA v2.1 (blue). The number of mutations shown are in a 1-12 base pair target site subsequence farthest from the PAM (**a-l**) when the rest of the target site, including the PAM, contains only on-target base pairs. The pre-selection and post-selection distributions are similar for up to three base pairs, demonstrating tolerance for target sites with mutations in the three base pairs farthest from the PAM when the rest of the target sites have optimal interactions with the Cas9:sgRNA. For all panels, graphs were generated from pre-selection library sequences and post-selection library sequences with an $n \geq 3{,}190$ and $n \geq 21{,}265$, respectively.



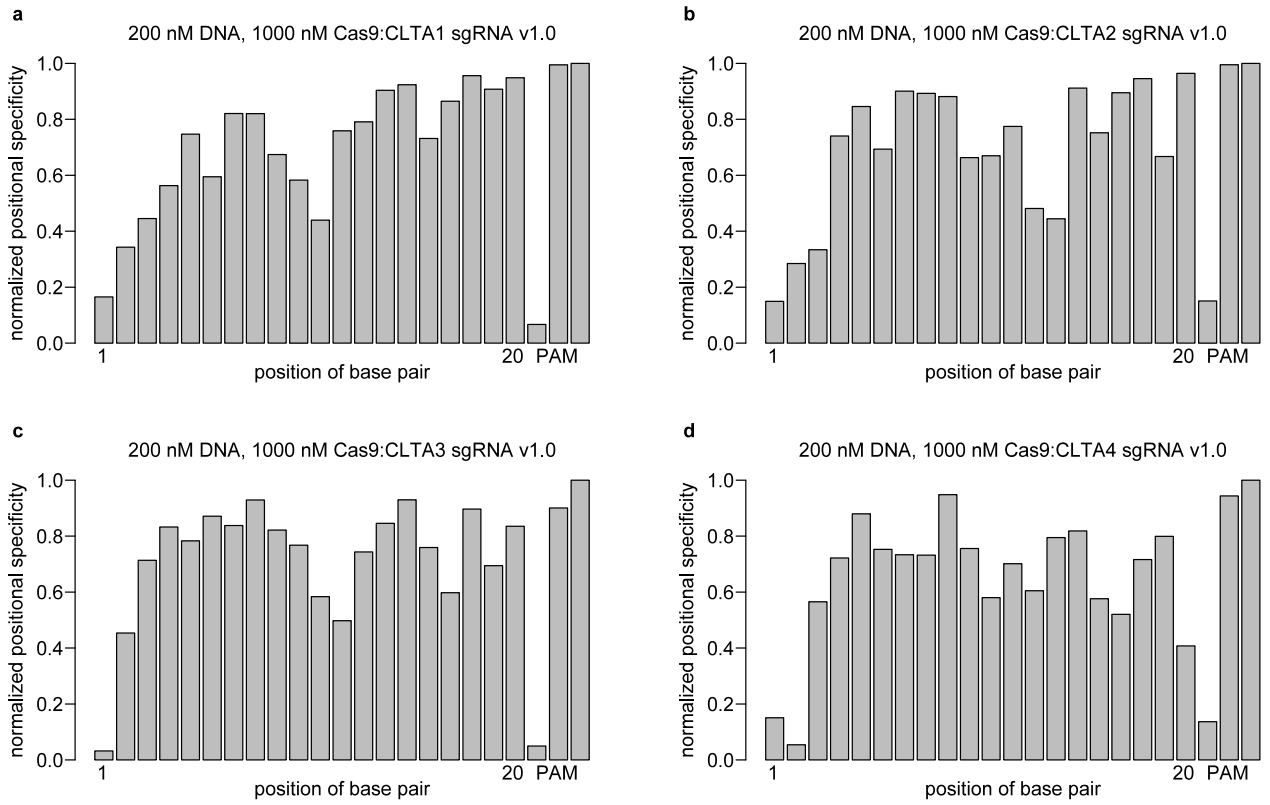
Supplementary Figure S14. Tolerance of mutations distal to PAM in CLTA3 target sites. Distributions of mutations are shown for *in vitro* selection on 200 nM pre-selection library (black) with 1000 nM Cas9:CLTA3 sgRNA v2.1 (blue). The number of mutations shown are in a 1-12 base pair target site subsequence farthest from the PAM (**a-l**) when the rest of the target site, including the PAM, contains only on-target base pairs. The pre-selection and post-selection distributions are similar for up to three base pairs, demonstrating tolerance for target sites with mutations in the three base pairs farthest from the PAM when the rest of the target sites have optimal interactions with the Cas9:sgRNA. For all panels, graphs were generated from pre-selection library sequences and post-selection library sequences with an $n \geq 5{,}604$ and $n \geq 158{,}424$, respectively.



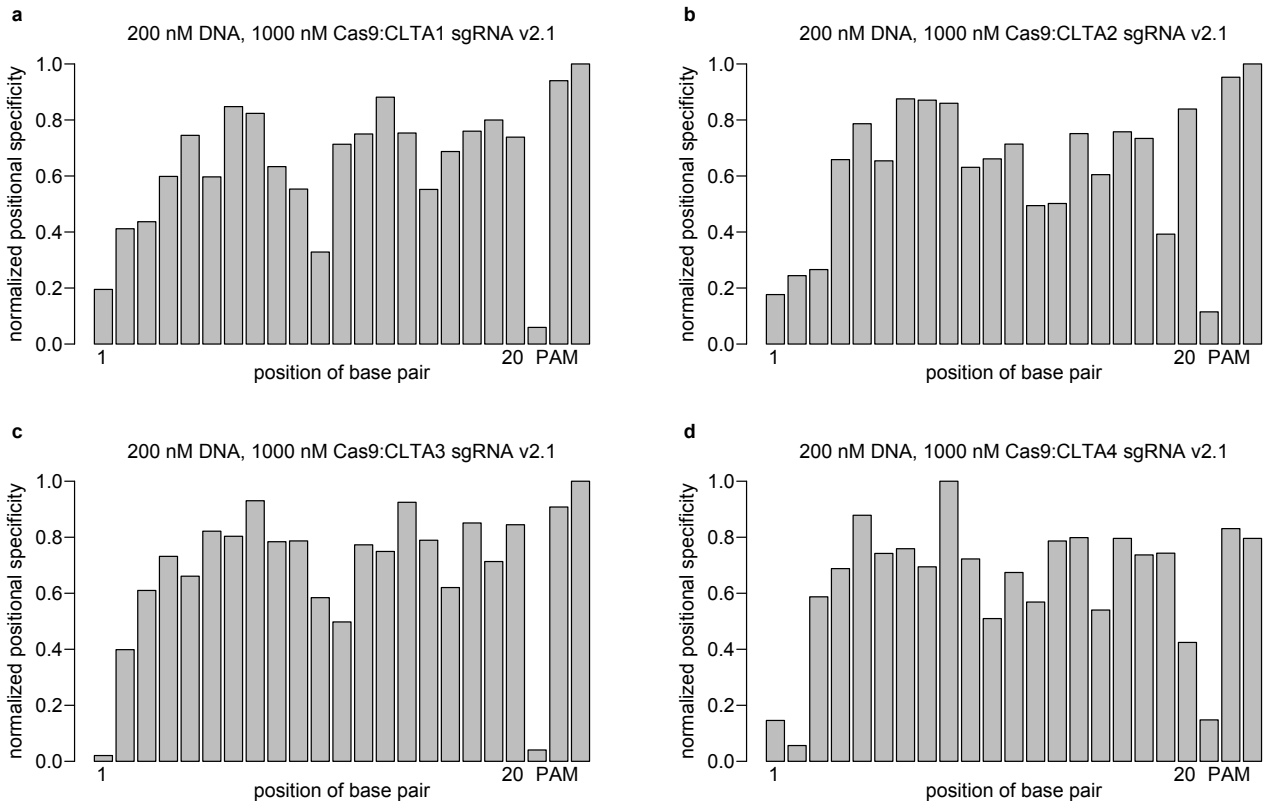
Supplementary Figure S15. Tolerance of mutations distal to PAM in CLTA4 target sites. Distributions of mutations are shown for *in vitro* selection on 200 nM pre-selection library (black) with 1000 nM Cas9:CLTA4 sgRNA v2.1 (blue). The number of mutations shown are in a 1-12 base pair target site subsequence farthest from the PAM (**a-l**) when the rest of the target site, including the PAM, contains only on-target base pairs. The pre-selection and post-selection distributions are similar for up to three base pairs, demonstrating tolerance for target sites with mutations in the three base pairs farthest from the PAM when the rest of the target sites have optimal interactions with the Cas9:sgRNA. For all panels, graphs were generated from pre-selection library sequences and post-selection library sequences with an $n \geq 2{,}323$ and $n \geq 21{,}819$, respectively.



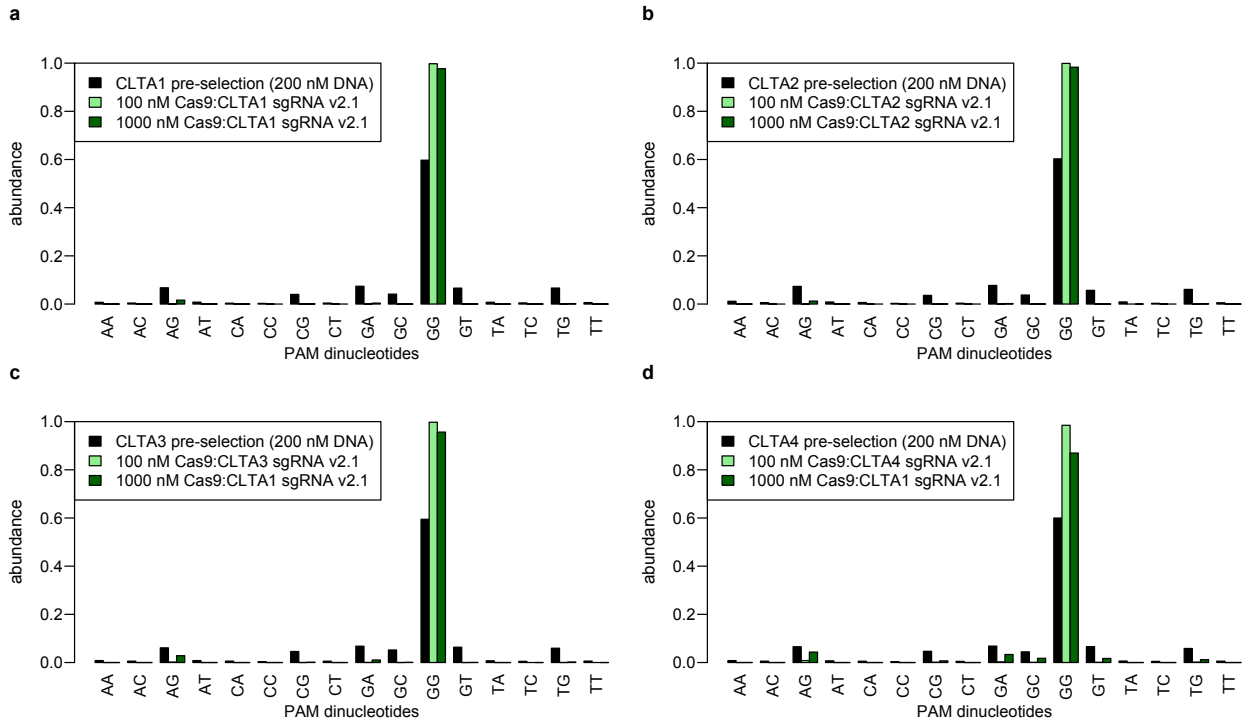
Supplementary Figure S16. **Positional specificity patterns for 100 nM Cas9:sgRNA v2.1**. Positional specificity, defined as the sum of the magnitude of the specificity score for each of the four possible base pairs recognized at a certain position in the target site, is plotted for each target site under enzyme-limiting conditions for sgRNA v2.1. The positional specificity is shown as a value normalized to the maximum positional specificity value of the target site. Positional specificity is highest at the end of the target site proximal to the PAM and is lowest in the middle of the target site and in the several nucleotides most distal to the PAM.



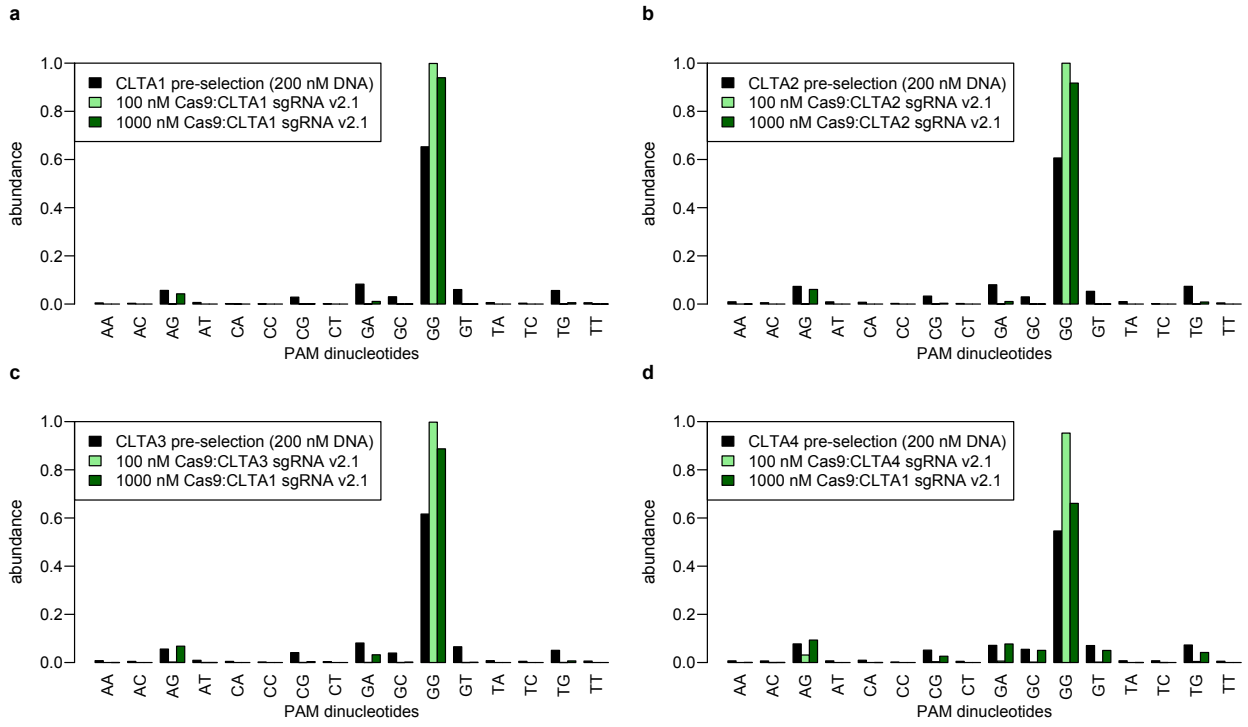
Supplementary Figure S17. Positional specificity patterns for 1000 nM Cas9:sgRNA v1.0. Positional specificity, defined as the sum of the magnitude of the specificity score for each of the four possible base pairs recognized at a certain position in the target site, is plotted for each target site under enzyme-excess conditions with sgRNA v1.0. The positional specificity is shown as a value normalized to the maximum positional specificity value of the target site. Positional specificity is relatively constant across the target site but is lowest in the middle of the target site and in the several nucleotides most distal to the PAM.



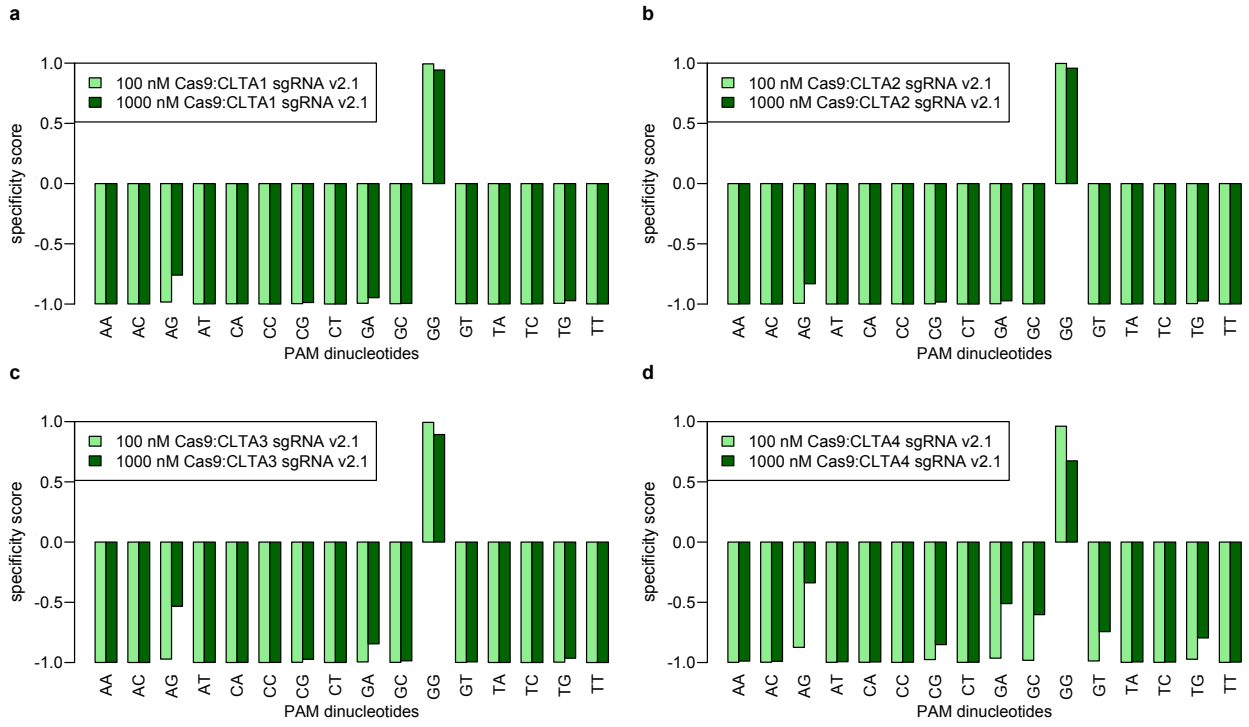
Supplementary Figure S18. Positional specificity patterns for 1000 nM Cas9:sgRNA v2.1. Positional specificity, defined as the sum of the magnitude of the specificity score for each of the four possible base pairs recognized at a certain position in the target site, is plotted for each target site under enzyme-excess conditions with sgRNA v2.1. The positional specificity is shown as a value normalized to the maximum positional specificity value of the target site. Positional specificity is relatively constant across the target site but is lowest in the middle of the target site and in the several nucleotides most distal to the PAM.



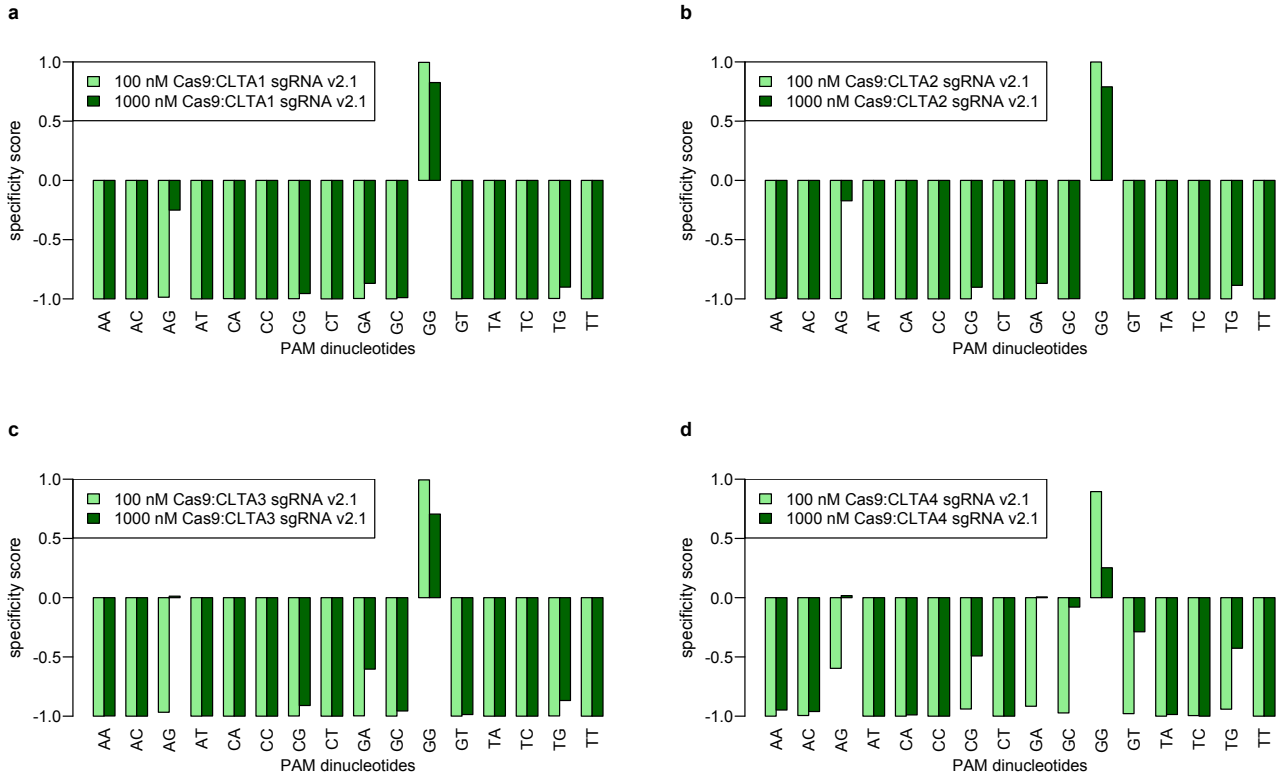
Supplementary Figure S19. PAM nucleotide preferences. The abundance in the pre-selection library (black) and post-selection libraries under enzyme-limiting (light green) or enzyme-excess (dark green) conditions are shown for all 16 possible PAM dinucleotides for selections with CLTA1 (**a**), CLTA2 (**b**), CLTA3 (**c**), and CLTA4 (**d**) sgRNA v2.1. GG dinucleotides increased in abundance in the post-selection libraries, while the other possible PAM dinucleotides decreased in abundance after the selection.



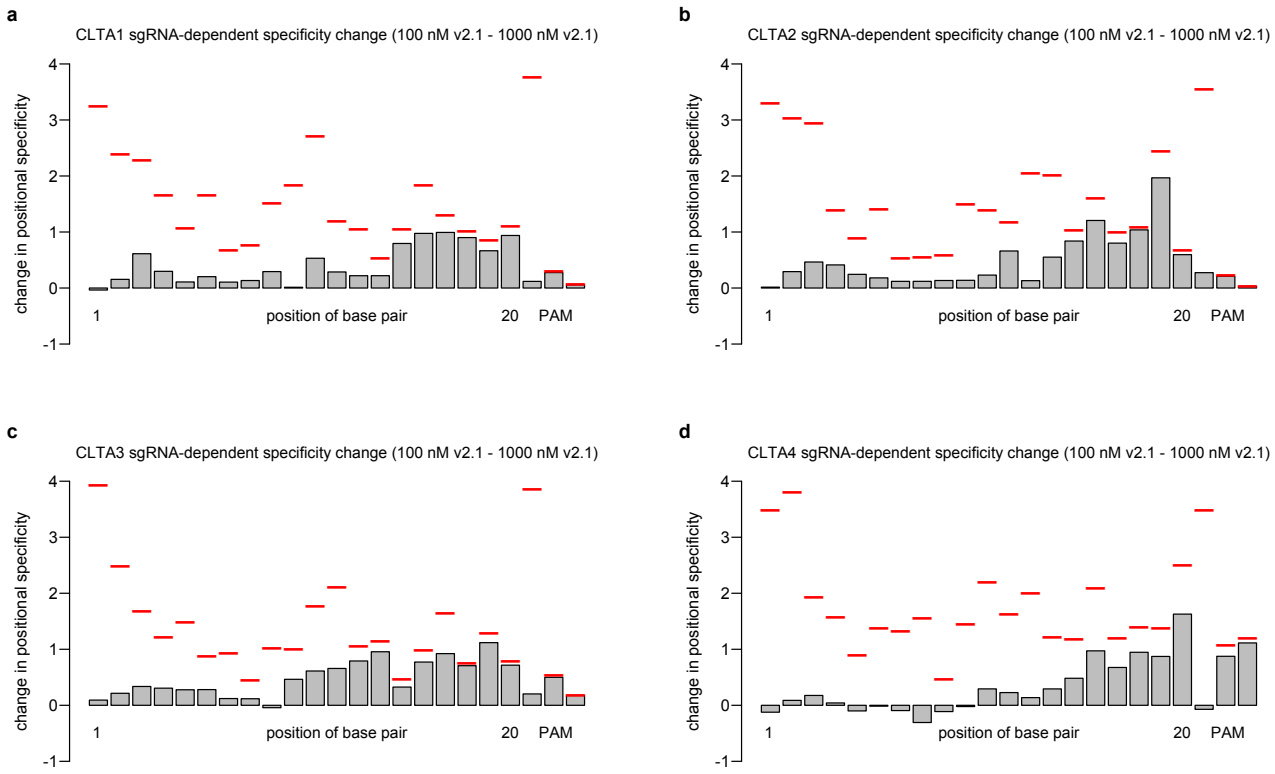
Supplementary Figure S20. PAM nucleotide preferences for on-target sites. Only post-selection library members containing no mutations in the 20 base pairs specified by the guide RNAs were included in this analysis. The abundance in the pre-selection library (black) and post-selection libraries under enzyme-limiting (light green) and enzyme-excess conditions (dark green) are shown for all 16 possible PAM dinucleotides for selections with CLTA1 (a), CLTA2 (b), CLTA3 (c), and CLTA4 (d) sgRNA v2.1. GG dinucleotides increased in abundance in the post-selection libraries, while the other possible PAM dinucleotides generally decreased in abundance after the selection, although this effect for the enzyme-excess concentrations of Cas9:sgRNA was modest or non-existent for many dinucleotides.



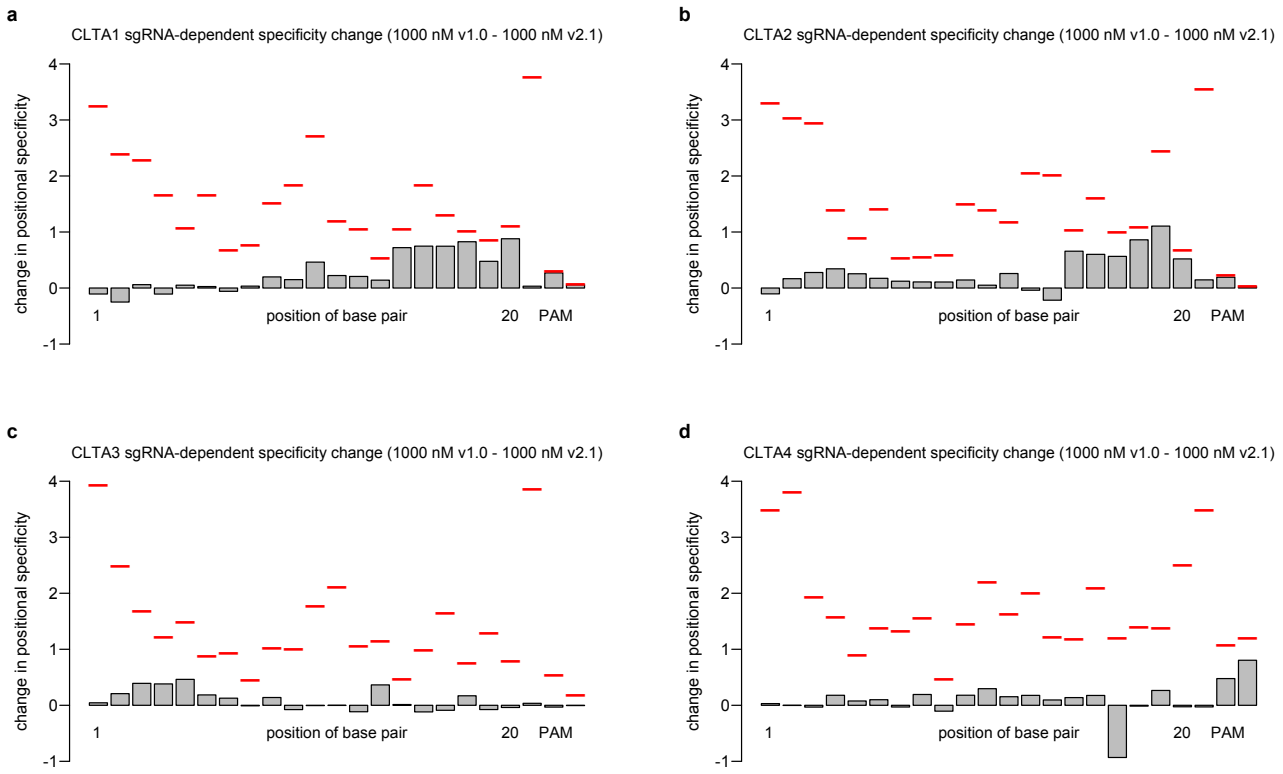
Supplementary Figure S21. PAM dinucleotide specificity scores. The specificity scores under enzyme-limiting (light green) and enzyme-excess conditions (dark green) are shown for all 16 possible PAM dinucleotides (positions 2 and 3 of the three-nucleotide NGG PAM) for selections with CLTA1 (**a**), CLTA2 (**b**), CLTA3 (**c**), and CLTA4 (**d**) sgRNA v2.1. The specificity score indicates the enrichment of the PAM dinucleotide in the post-selection library relative to the pre-selection library, normalized to the maximum possible enrichment of that dinucleotide. A specificity score of +1.0 indicates that a dinucleotide is 100% enriched in the post-selection library, and a specificity score of -1.0 indicates that a dinucleotide is 100% de-enriched. GG dinucleotides were the most enriched in the post-selection libraries, and AG, GA, GC, GT, and TG show less relative de-enrichment compared to the other possible PAM dinucleotides.



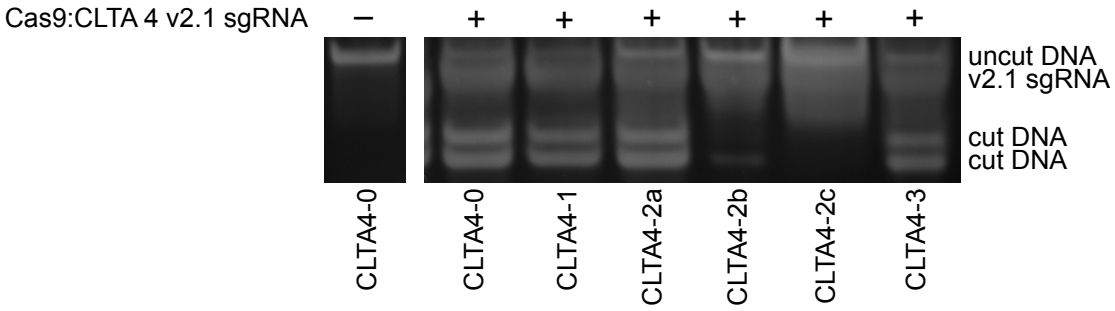
Supplementary Figure S22. PAM dinucleotide specificity scores for on-target sites. Only post-selection library members containing no mutations in the 20 base pairs specified by the guide RNAs were included in this analysis. The specificity scores under enzyme-limiting (light green) and enzyme-excess conditions (dark green) are shown for all 16 possible PAM dinucleotides (positions 2 and 3 of the three-nucleotide NGG PAM) for selections with CLTA1 (a), CLTA2 (b), CLTA3 (c), and CLTA4 (d) sgRNA v2.1. The specificity score indicates the enrichment of the PAM dinucleotide in the post-selection library relative to the pre-selection library, normalized to the maximum possible enrichment of that dinucleotide. A specificity score of +1.0 indicates that a dinucleotide is 100% enriched in the post-selection library, and a specificity score of -1.0 indicates that a dinucleotide is 100% de-enriched. GG dinucleotides were the most enriched in the post-selection libraries, AG and GA nucleotides were neither enriched or de-enriched in at least one selection condition, and GC, GT, and TG show less relative de-enrichment compared to the other possible PAM dinucleotides.



Supplementary Figure S23. **Effects of Cas9:sgRNA concentration on specificity.** Positional specificity changes between enzyme-limiting (200 nM DNA, 100 nM Cas9:sgRNA v2.1) and enzyme-excess (200 nM DNA, 1000 nM Cas9:sgRNA v2.1) conditions are shown for selections with sgRNAs targeting CLTA1 (**a**), CLTA2 (**b**), CLTA3 (**c**), and CLTA4 (**d**) target sites. Red lines indicate the maximum possible change in positional specificity for a given position. The highest changes in specificity occur proximal to the PAM as enzyme concentration is increased.

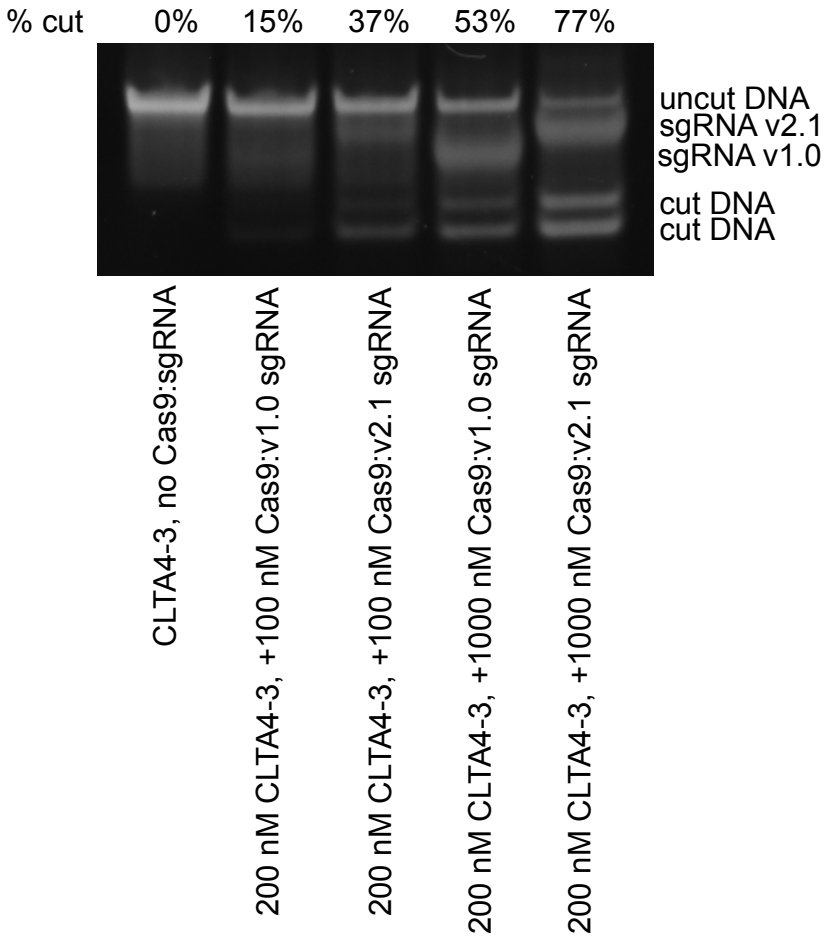


Supplementary Figure S24. Effects of sgRNA architecture on specificity. Positional specificity changes between Cas9:sgRNA v1.0 and Cas9:sgRNA v2.1 under enzyme-excess (200 nM DNA, 1000 nM Cas9:sgRNA v2.1) conditions are shown for selections with sgRNAs targeting CLTA1 (a), CLTA2 (b), CLTA3 (c), and CLTA4 (d) target sites. Red lines indicate the maximum possible change in positional specificity for a given position.



	sequence	*In vitro* selection enrichment value	% cut
CLTA4-0	GCAGATGTAGTGTTTCCACAGGG	7.9	85%
CLTA4-1	GaAGATGTAGTGTTTCCACAGGG	27.5	84%
CLTA4-2a	GaAGATGTAGTGTTTCCACtGGG	43.9	79%
CLTA4-2b	GCAGATGgAGgGTTTCCACAGGG	1.0	35%
CLTA4-2c	GCAGATGTAGTGTTaCCAgAGGG	0.064	none detected
CLTA4-3	GggGATGTAGTGTTTCCACtGGG	95.9	72%

Supplementary Figure S25. **Cas9:guide RNA cleavage of off-target DNA sequences *in vitro*.** Discrete DNA cleavage assays on a 96-bp linear substrate were performed with 200 nM DNA and 1000 nM Cas9:CLTA4 v2.1 sgRNA for the on-target CLTA4 site (CLTA4-0) and five CLTA4 off-target sites identified by *in vitro* selection. Enrichment values shown are from the *in vitro* selection with 1000 nM Cas9:CLTA4 v2.1 sgRNA. CLTA4-1 and CLTA4-3 were the most highly enriched sequences under these conditions. CLTA4-2a, CLTA4-2b, and CLTA4-2c are two-mutation sequences that represent a range of enrichment values from high enrichment to no enrichment to high de-enrichment. Red lowercase letters indicate mutations relative to the on-target CLTA4 site. The enrichment values are qualitatively consistent with the observed amount of cleavage *in vitro*.



Supplementary Figure S26. **Effect of guide RNA architecture and Cas9:sgRNA concentration on *in vitro* cleavage of an off-target site**. Discrete DNA cleavage assays on a 96-bp linear substrate were performed with 200 nM DNA and 100 nM Cas9:v1.0 sgRNA, 100 nM Cas9:v2.1 sgRNA, 1000 nM Cas9:v1.0 sgRNA, or 1000 nM Cas9:v2.1 sgRNA for the CLTA4-3 off-target site (5' GggGATGTAGTGTTTCCACtGGG - mutations are shown in lowercase letters). DNA cleavage is observed under all four conditions tested, and cleavage rates are higher under enzyme-excess conditions, or with v2.1 sgRNA compared with v1.0 sgRNA.

			no sgRNA		v1.0 sgRNA		v2.1 sgRNA	
	m	sequence	modified sequences	total sequences	modified sequences	total sequences	modified sequences	total sequences
CLTA1-0-1	0	AGTCCTCATCTCCCTCAAGCAGG	2	58889	18	42683	178	52845
CLTA1-1-1	1	AGTCCTCAaCTCCCTCAAGCAGG	1	39804	9	29000	37	40588
CLTA1-2-1	2	AGcCCTCATtTCCCTCAAGCAGG	0	16276	0	15032	0	18277
CLTA1-2-2	2	AcTCCTCATCcCCCTCAAGCCGG	3	21267	1	20042	33	22579
CLTA1-2-3	2	AGTCaTCATCTCCCTCAAGCAGa	0	0	0	0	0	0
CLTA1-3-1	3	cGTCCTCcTCTCCCcCAAGCAGG	2	53901	0	42194	0	52205
CLTA1-3-2	3	tGTCCTCtTCTCCCTCAAGCAGa	0	14890	0	14231	0	15937
CLTA1-4-1	4	AagCtTCATCTCtCTCAAGCTGG	0	49579	2	31413	0	41234
CLTA1-4-2	4	AGTaCTCtTtTCCCTCAgGCTGG	2	30013	1	23470	4	26999
CLTA1-4-3	4	AGTCtTaAatTCCCTCAAGCAGG	2	63792	0	52321	1	73007
CLTA1-4-4	4	AGTgCTCATCTaCCagAAGCTGG	1	12585	0	11339	0	12066
CLTA1-4-5	4	ccTCCTCATCTCCCTgcAGCAGG	4	30568	1	23810	0	27870
CLTA1-4-6	4	ctaCaTCATCTCCCTCAAGCTGG	0	13200	1	12886	2	12843
CLTA1-4-7	4	gGTCCTCATCTCCCTaAAaCAGa	1	8697	3	8188	0	8783
CLTA1-4-8	4	tGTCCTCATCggCCTCAgGCAGG	0	13169	0	8805	2	12830
CLTA1-5-1	5	AGaCacCATCTCCCTtgAGCTGG	0	46109	1	32515	2	35567
CLTA1-5-2	5	AGgCaTCATCTaCaTCAAGtTGG	0	41280	0	28896	0	35152
CLTA1-5-3	5	AGTaaTCActTCCaTCAAGCCGG	0	0	0	0	0	0
CLTA1-5-4	5	tccCCTCAcCTCCCTaAAGCAGG	2	24169	5	17512	1	23483
CLTA1-5-5	5	tGTCtTtATtTCCCTCtAGCTGG	0	11527	0	10481	1	11027
CLTA1-6-1	6	AGTCCTCATCTCCCTCAAGCAGG	0	6537	0	5654	0	6741

Supplementary Table S2. CLTA1 genomic off-target sequences. 20 human genomic DNA sequences were identified that were enriched in the Cas9:CLTA1 v2.1 sgRNA *in vitro* selections under enzyme-limiting or enzyme-excess conditions. “m” refers to number of mutations from on-target sequence with mutations shown in lower case. Sites shown in red contain insertions or deletions (indels) that are consistent with significant Cas9:sgRNA-mediated modification in HEK293T cells. Human genome coordinates are shown for each site (assembly GRCh37). CLTA1-0-1 is present at two loci, and sequence counts were pooled from both loci. Sequence counts are shown for amplified and sequenced DNA for each site from HEK293T cells treated with Cas9 without sgRNA (“no sgRNA”), Cas9 with CLTA1 v1.0 sgRNA, or Cas9 with CLTA1 v2.1 sgRNA.

			no sgRNA		v1.0 sgRNA		v2.1 sgRNA	
	m	sequence	modified sequences	total sequences	modified sequences	total sequences	modified sequences	total sequences
CLTA4-0-1	0	GCAGATGTAGTGTTTCCACAGGG	6	29191	2005	18640	14970	19661
CLTA4-3-1	3	aCAtATGTAGTaTTTCCACAGGG	2	34165	11	20018	3874	16082
CLTA4-3-2	3	GCAtATGTAGTGTTTCCAaATGt	3	17923	0	11688	2	13880
CLTA4-3-3	3	cCAGATGTAGTaTTcCCACAGGG	0	16559	0	12007	52	11082
CLTA4-3-4	3	GCAGtTtTAGTGTTTtCACAGGG	0	21722	0	12831	0	15726
CLTA4-3-5	3	GCAGAgtTAGTGTTTCCACACaG	1	21222	2	13555	3	16425
CLTA4-3-6	3	GCAGATGgAGgGTTTtCACAGGG	3	20342	3	12804	3	14068
CLTA4-3-7	3	GgAaATtTAGTGTTTCCACAGGG	2	38894	3	24017	1	29347
CLTA4-4-1	4	aaAGAaGTAGTaTTTCCACATGG	0	0	0	0	0	0
CLTA4-4-2	4	aaAGATGTAGTcaTTCCACAAGG	1	27326	0	17365	1	18941
CLTA4-4-3	4	aaAtATGTAGTcTTTCCACAGGG	2	46232	3	32264	0	32638
CLTA4-4-4	4	atAGATGTAGTGTTTCCAaAGGa	9	27821	1	16223	8	15388
CLTA4-4-5	4	cCAGAgGTAGTGcTcCCACAGGG	1	20979	1	15674	1	15086
CLTA4-4-6	4	cCAGATGTgagGTTTCCACAAGG	4	22021	0	15691	1	14253
CLTA4-4-7	4	ctAcATGTAGTGTTTCCAtATGG	2	35942	0	23076	1	11867
CLTA4-4-8	4	ctAGATGaAGTGcTTCCACATGG	1	10692	1	7609	59	8077
CLTA4-4-9	4	GaAaATGgAGTGTTTaCACATGG	0	34616	0	22302	1	24671
CLTA4-4-10	4	GCAaATGaAGTGTcaCCACAAGG	1	25210	0	16187	0	16974
CLTA4-4-11	4	GCAaATGTAtTaTTTCCACtAGG	0	34144	1	24770	0	22547
CLTA4-4-12	4	GCAGATGTAGctTTTgtACATGG	0	14254	0	9616	0	9994
CLTA4-4-13	4	GCAGcTtaAGTGTTTtCACATGG	8	39466	1	7609	5	16525
CLTA4-4-14	4	ttAcATGTAGTGTTTaCACACGG	0	0	0	22302	0	0
CLTA4-5-1	5	GaAGAgGaAGTGTTTgCcCAGGG	1	27616	1	16319	1	16140
CLTA4-5-2	5	GaAGATGTgGaGTTgaCACATGG	1	22533	0	14292	0	15013
CLTA4-5-3	5	GCAGAaGTAcTGTTgttACAAGG	1	44243	1	29391	1	29734
CLTA4-5-4	5	GCAGATGTgGaaTTaCaACAGGG	0	27321	0	13640	0	14680
CLTA4-5-5	5	GCAGtcaTAGTGTaTaCACATGG	1	26538	0	18449	1	20559
CLTA4-5-6	5	taAGATGTAGTaTTTCCAaAAGt	1	15145	1	8905	0	7911
CLTA4-6-1	6	GCAGcTGgcaTtTcTCCACACGG	0	2	0	0	0	0
CLTA4-6-2	6	GgAGATcTgaTGgTTCtACAAGG	2	27797	0	19450	2	21709
CLTA4-6-3	6	taAaATGcAGTGTaTCCAtATGG	4	27551	0	18424	0	18783
CLTA4-7-1	7	GCcagaaTAGTtTTTCaACAAGG	0	20942	0	13137	1	13792
CLTA4-7-2	8	ttgtATtTAGaGaTTgCACAAGG	0	28470	0	18104	0	20416

Supplementary Table S3. CLTA4 genomic off-target sequences. 33 human genomic DNA sequences were identified that were enriched in the Cas9:CLTA4 v2.1 sgRNA *in vitro* selections under enzyme-limiting or enzyme-excess conditions. "m" refers to number of mutations from on-target sequence with mutations shown in lower case. Sites shown in red contain insertions or deletions (indels) that are consistent with significant Cas9:sgRNA-mediated modification in HEK293T cells. Human genome coordinates are shown for each site (assembly GRCh37). Sequence counts are shown for amplified and sequenced DNA for each site from HEK293T cells treated with Cas9 without sgRNA ("no sgRNA"), Cas9 with CLTA4 v1.0 sgRNA, or Cas9 with CLTA4 v2.1 sgRNA.

Off-target site	Human genome coordinates
CLTA1-0-1	9(+): 36,211,732-36,211,754 12(+): 7,759,893-7,759,915
CLTA1-1-1	8(-): 15,546,437-15,546,459
CLTA1-2-1	3(-): 54,223,111-54,223,133
CLTA1-2-2	15(+): 89,388,670-89,388,692
CLTA1-2-3	5(+): 88716920-88,716,942
CLTA1-3-1	21(+): 27,972,462-27,972,484
CLTA1-3-2	4(-): 17,179,924-17,179,946
CLTA1-4-1	1(+): 147,288,742-147,288,764
CLTA1-4-2	10(+): 97,544,444-97,544,466
CLTA1-4-3	2(-): 161,873,870-161,873,892
CLTA1-4-4	1(+): 196,172,702-196,172,724
CLTA1-4-5	13(+): 56,574,636-56,574,658
CLTA1-4-6	2(+): 241,357,827-241,357,849
CLTA1-4-7	3(+): 121,248,627-121,248,649
CLTA1-4-8	12(+): 132,937,319-132,937,341
CLTA1-5-1	9(-): 80,930,919 -80,930,941
CLTA1-5-2	2(+): 140,901,875-14,0901,897
CLTA1-5-3	3(+): 45,016,841-45,016,863
CLTA1-5-4	X(+): 40,775,684-40,775,706
CLTA1-5-5	2(-): 185,151,622-185,151,644
CLTA1-6-1	X(+): 150,655,097-150,655,119
CLTA4-0-1	9(-): 36,211,779-36,211,801
CLTA4-3-1	12(-): 50,679,419-50,679,441
CLTA4-3-2	X(-): 143,939,483-143,939,505
CLTA4-3-3	11(-): 47,492,611-47,492,633
CLTA4-3-4	3(-): 162,523,715-162,523,737
CLTA4-3-5	11(+): 30,592,975-30,592,997
CLTA4-3-6	4(-): 155,252,699-155,252,721
CLTA4-3-7	18(+): 39,209,441-39,209,463
CLTA4-4-1	17(-): 36,785,650-36,785,672
CLTA4-4-2	1(-): 241,537,119-241,537,141
CLTA4-4-3	8(-): 120,432,103-120,432,125
CLTA4-4-4	6(-): 106,204,600-106,204,622
CLTA4-4-5	8(+): 102,527,804-102,527,826
CLTA4-4-6	8(-): 94,685,538-94,685,560
CLTA4-4-7	2(+): 35,820,054-35,820,076
CLTA4-4-8	3(-): 36,590,352-36,590,374
CLTA4-4-9	12(+): 100,915,498-100,915,520
CLTA4-4-10	21(+): 33,557,705-33,557,727
CLTA4-4-11	8(+): 10,764,183-10,764,205
CLTA4-4-12	19(+): 37,811,645-37,811,667
CLTA4-4-13	13(-): 26,832,673-26,832,695
CLTA4-4-14	6(+): 19,349,572-19,349,594
CLTA4-5-1	11(-): 502,300-502,322
CLTA4-5-2	8(-): 28,389,683-28,389,705
CLTA4-5-3	2(-): 118,557,405-118,557,427
CLTA4-5-4	2(-): 103,248,360-103,248,382
CLTA4-5-5	21(-): 42,929,085-42,929,107
CLTA4-5-6	13(-): 83,097,278-83,097,300
CLTA4-6-1	2(+): 43,078,423-43,078,445
CLTA4-6-2	7(-): 11,909,384-11,909,406
CLTA4-6-3	5(-): 69,775,482-69,775,504
CLTA4-7-1	16(+): 30,454,945-30,454,967
CLTA4-7-2	9(-): 77,211,328-77,211,350

Supplementary Table S4. Genomic coordinates of CLTA1 and CLTA4 off-target sites. 54 human genomic DNA sequences were identified that were enriched in the Cas9:CLTA1 v2.1 sgRNA and Cas9:CLTA4 v2.1 sgRNA *in vitro* selections under enzyme-limiting or enzyme-excess conditions. Human genome coordinates are shown for each site (assembly GRCh37).

				in vitro enrichment		modification frequency in HEK293T cells		
	number of mutations	sequence	gene	v1.0	v2.1	no sgRNA	v1.0	v2.1
CLTA1-0-1	0	AGTCCTCATCTCCCTCAAGCAGG	CLTA	41.4	23.3	0.003%	0.042%	0.337%
CLTA1-1-1	1	AGTCCTCAaCTCCCTCAAGCAGG	TUSC3	25.9	14	0.003%	0.031%	0.091%
CLTA1-2-1	2	AGcCCTCATtTCCCTCAAGCAGG	CACNA2D3	15.4	26.2	0%	0%	0%
CLTA1-2-2	2	AcTCCTCATCcCCCTCAAGCCGG	ACAN	29.2	18.8	0.014%	0.005%	0.146%
CLTA1-2-3	2	AGTCaTCATCTCCCTCAAGCAGa		0.06	1.27	n.t.	n.t.	n.t.
CLTA1-3-1	3	cGTCCTCcTCTCCCcCAAGCAGG		0	2.07	0.004%	0%	0%
CLTA1-3-2	3	tGTCCTCtTCTCCCTCAAGCAGa	BC029598	0	1.47	0%	0%	0%
CLTA1-4-1	4	AagCtTCATCTCtCTCAAGCTGG				0%	0.006%	0%
CLTA1-4-2	4	AGTaCTCtTtTCCCTCAgGCTGG	ENTPD1			0.007%	0.004%	0.015%
CLTA1-4-3	4	AGTCtTaAatTCCCTCAAGCAGG				0.003%	0%	0.001%
CLTA1-4-4	4	AGTgCTCATCTaCCagAAGCTGG				0.008%	0%	0%
CLTA1-4-5	4	ccTCCTCATCTCCCTgcAGCAGG				0.013%	0.004%	0%
CLTA1-4-6	4	ctaCaTCATCTCCCTCAAGCTGG				0%	0.008%	0.016%
CLTA1-4-7	4	gGTCCTCATCTCCCTaAAaCAGa	POLQ (coding)			0.011%	0.037%	0%
CLTA1-4-8	4	tGTCCTCATCggCCTCAgGCAGG				0%	0%	0.016%
CLTA1-5-1	5	AGaCacCATCTCCCTtgAGCTGG	PSAT1			0%	0.003%	0.006%
CLTA1-5-2	5	AGgCaTCATCTaCaTCAAGtTGG				0%	0%	0%
CLTA1-5-3	5	AGTaaTCActTCCaTCAAGCCGG	ZDHHC3, EXOSC7			n.t.	n.t.	n.t.
CLTA1-5-4	5	tccCCTCAcCTCCCTaAAGCAGG				0.008%	0.029%	0.004%
CLTA1-5-5	5	tGTCtTtATtTCCCTCtAGCTGG				0%	0%	0.009%
CLTA1-6-1	6	AGTCCTCATCTCCCTCAAGCAGG				0%	0%	0%

Supplementary Table S5. Cellular modification induced by Cas9:CLTA1 sgRNA. 20 human genomic DNA sequences were identified that were enriched in the Cas9:CLTA1 v2.1 sgRNA *in vitro* selections under enzyme-limiting or enzyme-excess conditions. Sites shown in red contain insertions or deletions (indels) that are consistent with significant Cas9:sgRNA-mediated modification in HEK293T cells. *In vitro* enrichment values for selections with Cas9:CLTA1 v1.0 sgRNA or Cas9:CLTA1 v2.1 sgRNA are shown for sequences with three or fewer mutations. Enrichment values were not calculated for sequences with four or more mutations due to low numbers of *in vitro* selection sequence counts. Modification frequencies (number of sequences with indels divided by total number of sequences) in HEK293T cells treated with Cas9 without sgRNA ("no sgRNA"), Cas9 with CLTA1 v1.0 sgRNA, or Cas9 with CLTA1 v2.1 sgRNA. P-values of sites that show significant modification in v1.0 sgRNA- or v2.1 sgRNA-treated cells compared to cells treated with Cas9 without sgRNA were 1.1E-05 (v1.0) and 6.9E-55 (v2.1) for CLTA1-0-1, 2.6E-03 (v1.0) and 2.0E-10 (v2.1) for CLTA1-1-1, and 4.6E-08 (v2.1) for CLTA1-2-2. P-values were calculated using a one-sided Fisher exact test. "Not tested (n.t.)" indicates that the site was not tested or PCR of the genomic sequence failed to provide specific amplification products.

CLTA1-0-1

		# of sequences		
	sequence	no sgRNA	v1.0 sgRNA	v2.1 sgRNA
ref	AGTCCTCATCTCCCTCAAGCAGG	58,887	42,665	52,667
	AGTCCTCATCTCCCTCAAAGCAGG	0	0	66
	AGTCCTCATCTCCCTC-AGCAGG	0	2	28
	AGTCCTCAT--------------	0	0	13
	AGTCCTCATCTCCCTCATAGCAGG	0	0	11
	AGTCCTCAT--------AGCAGG	0	0	9
	AGTCCTCATCT------AGCAGG	0	0	8
	AGTCCTCA---------AGCAGG	0	0	6
	AGTCCTCATCTCCCTCAAAGGCAGTGTTTGTTACTTGAGTTTGTCAGCAGG	0	0	4
	AGTCCTCATCTCCCTCATTAGCAGG	0	0	4
	AGTCCTCATCTCCCTCAGGGCTTGTTTACAGCTCACCTTTGAATTTGCACAAGCGTGCAAGCAGG	0	0	3
	AGTCCTCATCTCCCT--AGCAGG	0	11	0
	AGTCCTCATCCCTC--AAGCAGG	0	3	0
	AGTCCTCATCTCCCT-AAGCAGG	1	2	0
	other	1	0	26
	modified total	2	18 (0.042%)	178 (0.34%)

CLTA1-1-1

		# of sequences		
	sequence	no sgRNA	v1.0 sgRNA	v2.1 sgRNA
ref	AGTCCTCAaCTCCCTCAAGCAGG	39,803	28,991	40,551
	AGTCCTCAaCTCCCTCAAAGCAGG	0	4	13
	AGTCCTCAaCTCCCTCA------	0	0	12
	AGTCCTCAaCTCCCTC-AGCAGG	0	2	4
	AGTCCTCAaCTCCCTCAGAAAGGTGTTGAAAATCAGAAAGAGAGAAACAAGCAGG	0	0	3
	AGTCCTCAaCTCCCTCAATCTACGGTCCATTCCCGTTTCCACTCACCTTGCGCCGCAGCAGG	0	0	2
	AGTCCTCAaCTCCCT-AAGCAGG	0	3	1
	AGTCCTCAaCTCCCTCAACCAACTTTAACATCCTGCTGGTTCTGTCATTAATAAGTTGAAAGCAGG	0	0	1
	AGTCCTCAaCTCCCTCACAGCAAATAAAAAAGTTGTTTATGCATATTCAGATAAGCAAAGCAGG	0	0	1
	AGTCCTCAaCTCCC--AAGCAGG	1	0	0
	modified total	1	9 (0.031%)	37 (0.091%)

CLTA1-2-2

		# of sequences		
	sequence	no sgRNA	v1.0 sgRNA	v2.1 sgRNA
ref	AcTCCTCATCcCCCTCAAGCCGG	21,264	20,041	22,546
	AcTCCTCATCcCCCTCAAAGCCGG	0	0	8
	AcTCCTCATCcCCCTCAGAGCCGG	0	0	7
	AcTCCTC----------AGCCGG	0	0	5
	AcTCCTCATCcCCCTCAAAAGCCGG	0	0	2
	AcTCCTCATCcCCCTCAGCAGCCGG	0	0	2

	AcTCCTCATCcCCCTCATAGCCGG	0	0	2
	AcTCCTCATCcCCCTCATCCCCGG	0	0	2
	AcTCCTCATCcC-----AGCCGG	0	0	2
	AcTCCTCATCcCCCTA-AGCCGG	3	1	1
	AcTCCTCATCcCCCTCAATAGCCGG	0	0	1
	AcTCCTCACCcCCCTCAGCAGCCGG	0	0	1
	modified total	3	1	33 (0.15%)

Supplementary Table S6. CLTA1 genomic off-target indel sequences. Insertion and deletion-containing sequences from cells treated with amplified and sequenced DNA for the on-target genomic sequence (CLTA1-0-1) and each modified off-target site from HEK293T cells treated with Cas9 without sgRNA ("no sgRNA"), Cas9 with CLTA1 v1.0 sgRNA, or Cas9 with CLTA1 v2.1 sgRNA. "ref" refers to the human genome reference sequence for each site, and the modified sites are listed below. Mutations relative to the on-target genomic sequence are shown in lowercase letters. Insertions and deletions are shown in red letters or red dashes, respectively. Modification percentages are shown for those conditions (v1.0 sgRNA or v2.1 sgRNA) that show statistically significant enrichment of modified sequences compared to the control (no sgRNA).

CLTA4-0-1

		# of sequences		
	sequence	control	v1.0 sgRNA	v2.1 sgRNA
ref	GCAGATGTAGTGTTTCCACAGGG	29,185	16,635	17,555
	GCAGATGTAGTGTTTC-ACAGGG	1	891	5,937
	GCAGATGTAGTGTTTCCCACAGGG	0	809	5,044
	GCAGATGTAGTG----CACAGGG	0	14	400
	GCAGATGTAGTGTTTCC-CAGGG	0	19	269
	GCAGATGTAC-------ACAGGG	0	17	262
	GCAGATGTAGTGTCA---CAGGG	2	6	254
	GCAGATGTAGTGTTCA--CAGGG	0	21	229
	GCAGATGTAGTGTTTC--CAGGG	1	14	188
	GCAGATGTAGT-----CACAGGG	0	0	152
	GCAGATGT----------AGGG	0	6	129
	other	2	208	2,106
	modified total	6	2,005 (11%)	14,970 (76%)

CLTA4-3-1

		# of sequences		
	sequence	control	v1.0 sgRNA	v2.1 sgRNA
ref	aCAtATGTAGTaTTTCCACAGGG	34,163	20,007	12,208
	aCAtATGTAGTaTTTCCCACAGGG	0	8	1779
	aCAtATGTAGTaTTTCA-CAGGG	1	0	293
	aCAtATGTAGTaTTTC--CAGGG	1	0	227
	aCAtAT----------CACAGGG	0	0	117
	a----------CAGGG	0	0	96
	aCAt----------CACAGGG	0	0	78
	aCAtATGTAGT-----CACAGGG	0	0	77
	aCAtATGTAGTaTTTCC------	0	0	76
	aCAtATGT----------AGGG	0	0	68
	aCAtATGTAG------CACAGGG	0	0	64
	other	0	3	999
	modified total	2	11 (0.055%)	3874 (24%)

CLTA4-3-3

		# of sequences		
	sequence	control	v1.0 sgRNA	v2.1 sgRNA
ref	cCAGATGTAGTaTTcCCACAGGG	16,559	12,007	11,030
	cCAGATGTAGTaTTcCCCACAGGG	0	0	35
	cCAGATGTAGTaT----ACAGGG	0	0	5
	cCAGATGTAGTaT---CACAGGG	0	0	3
	cCAGATGTAGTaTTcCCAACACAGGG	0	0	2
	cCAGATGTAGTaTT--CACAGGG	0	0	2
	cCAGATGTAGTaTTcC--CAGGG	0	0	2
	cCAGATGTA----------	0	0	2
	cCAGATGTAGTaTTcC-ACAGGG	0	0	1
	modified total	0	0	52 (0.47%)

CLTA4-4-8

		# of sequences		
	sequence	control	v1.0 sgRNA	v2.1 sgRNA
ref	ctAGATGaAGTGcTTCCACATGG	10,691	7,608	8,018
	ctAGATGaAGTGcTTCCCACATGG	0	0	49
	ctAGATGaAGTGcTTC-ACATGG	0	0	6
	ctAGATGaAGTG----------	0	0	2
	ctAGATGaAGTGcTTCCACACATGG	0	0	1
	ctAGATGaAGTGcTTC--CATGG	1	0	0
	ctAGATGaAGTGcTTCC-CATGG	0	1	0
	modified total	1	1	59 (0.73%)

Supplementary Table S7. CLTA4 genomic off-target indel sequences. Insertion and deletion-containing sequences from cells treated with amplified and sequenced DNA for the on-target genomic sequence (CLTA4-0-1) and each modified off-target site from HEK293T cells treated with Cas9 without sgRNA ("no sgRNA"), Cas9 with CLTA4 v1.0 sgRNA, or Cas9 with CLTA4 v2.1 sgRNA. "ref" refers to the human genome reference sequence for each site, and the modified sites are listed below. Mutations relative to the on-target genomic sequence are shown in lowercase letters. Insertions and deletions are shown in red letters or red dashes, respectively. Modification percentages are shown for those conditions (v1.0 sgRNA or v2.1 sgRNA) that show statistically significant enrichment of modified sequences compared to the control (no sgRNA).

oligonucleotide name	oligonucleotide sequence (5'->3')
CLTA1 v2.1 template fwd	TAA TAC GAC TCA CTA TAG GAG TCC TCA TCT CCC TCA AGC GTT TTA GAG CTA TGC TG
CLTA2 v2.1 template fwd	TAA TAC GAC TCA CTA TAG GCT CCC TCA AGC AGG CCC CGC GTT TTA GAG CTA TGC TG
CLTA3 v2.1 template fwd	TAA TAC GAC TCA CTA TAG GTG TGA AGA GCT TCA CTG AGT GTT TTA GAG CTA TGC TG
CLTA4 v2.1 template fwd	TAA TAC GAC TCA CTA TAG GGC AGA TGT AGT GTT TCC ACA GTT TTA GAG CTA TGC TG
v2.1 template rev	GAT AAC GGA CTA GCC TTA TTT TAA CTT GCT ATG CTT TTC AGC ATA GCT CTA AAA C
CLTA1 v1.0 template	CGG ACT AGC CTT ATT TTA ACT TGC TAT TTC TAG CTC TAA AAC GCT TGA GGG AGA TGA GGA CTC CTA TAG TGA GTC GTA TTA
CLTA2 v1.0 template	CGG ACT AGC CTT ATT TTA ACT TGC TAT TTC TAG CTC TAA AAC GCG GGG CCT GCT TGA GGG AGC CTA TAG TGA GTC GTA TTA
CLTA3 v1.0 template	CGG ACT AGC CTT ATT TTA ACT TGC TAT TTC TAG CTC TAA AAC ACT CAG TGA AGC TCT TCA CAC CTA TAG TGA GTC GTA TTA
CLTA4 v1.0 template	CGG ACT AGC CTT ATT TTA ACT TGC TAT TTC TAG CTC TAA AAC TGT GGA AAC ACT ACA TCT GCC CTA TAG TGA GTC GTA TTA
T7 promoter oligo	TAA TAC GAC TCA CTA TAG G
CLTA1 lib	/5Phos/AAC ACA NNN NC*C* NG*C* T*T*G* A*G*G* G*A*G* A*T*G* A*G*G* A*C*T* NNN NAC CTG CCG AGA ACA CA
CLTA2 lib	/5Phos/TCT TCT NNN NC*C* NG*C* G*G*G* G*C*C* T*G*C* T*T*G* A*G*G* G*A*G* NNN NAC CTG CCG AGT CTT CT
CLTA3 lib	/5Phos/AGA GAA NNN NC*C* NA*C* T*C*A* G*T*G* A*A*G* C*T*C* T*T*C* A*C*A* NNN NAC CTG CCG AGA GAG AA
CLTA4 lib	/5Phos/TTG TGT NNN NC*C* NT*G* T*G*G* A*A*A* C*A*C* T*A*C* A*T*C* T*G*C* NNN NAC CTG CCG AGT TGT GT
CLTA1 site fwd	CTA GCA GTC CTC ATC TCC CTC AAG CAG GC
CLTA1 site rev	AGC TGC CTG CTT GAG GGA GAT GAG GAC TG
CLTA2 site fwd	CTA GTC TCC CTC AAG CAG GCC CCG CTG GT
CLTA2 site rev	AGC TAC CAG CGG GGC CTG CTT GAG GGA GA
CLTA3 site fwd	CTA GCT GTG AAG AGC TTC ACT GAG TAG GA
CLTA3 site rev	AGC TTC CTA CTC AGT GAA GCT CTT CAC AG
CLTA4 site fwd	CTA GTG CAG ATG TAG TGT TTC CAC AGG GT
CLTA4 site rev	AGC TAC CCT GTG GAA ACA CTA CAT CTG CA
test fwd	GCG ACA CGG AAA TGT TGA ATA CTC AT
test rev	GGA GTC AGG CAA CTA TGG ATG AAC G
off-target CLTA4-0 fwd	ACT GTG AAG AGC TTC ACT GAG TAG GAT TAA GAT ATT GCA GAT GTA GTG TTT CCA CAG GGT
off-target CLTA4-1 fwd	ACT GTG AAG AGC TTC ACT GAG TAG GAT TAA GAT ATT GAA GAT GTA GTG TTT CCA CAG GGT
off-target CLTA4-2a fwd	ACT GTG AAG AGC TTC ACT GAG TAG GAT TAA GAT ATT GAA GAT GTA GTG TTT CCA CTG GGT
off-target CLTA4-2b fwd	ACT GTG AAG AGC TTC ACT GAG TAG GAT TAA GAT ATT GCA GAT GGA GGG TTT CCA CAG GGT
off-target CLTA4-2c fwd	ACT GTG AAG AGC TTC ACT GAG TAG GAT TAA GAT ATT GCA GAT GTA GTG TTA CCA GAG GGT
off-target CLTA4-3 fwd	ACT GTG AAG AGC TTC ACT GAG TAG GAT TAA GAT ATT GGG GAT GTA GTG TTT CCA CTG GGT
off-target CLTA4-0 rev	TCC CTC AAG CAG GCC CCG CTG GTG CAC TGA AGA GCC ACC CTG TGG AAA CAC TAC ATC TGC
off-target CLTA4-1 rev	TCC CTC AAG CAG GCC CCG CTG GTG CAC TGA AGA GCC ACC CTG TGG AAA CAC TAC ATC TTC
off-target CLTA4-2a rev	TCC CTC AAG CAG GCC CCG CTG GTG CAC TGA AGA GCC ACC CAG TGG AAA CAC TAC ATC TTC
off-target CLTA4-2b rev	TCC CTC AAG CAG GCC CCG CTG GTG CAC TGA AGA GCC ACC CTG TGG AAA CCC TCC ATC TGC
off-target CLTA4-2c rev	TCC CTC AAG CAG GCC CCG CTG GTG CAC TGA AGA GCC ACC CTC TGG TAA CAC TAC ATC TGC
off-target CLTA4-3 rev	TCC CTC AAG CAG GCC CCG CTG GTG CAC TGA AGA GCC ACC CAG TGG AAA CAC TAC ATC CCC

adapter1(AACA)	AAT GAT ACG GCG ACC ACC GAG ATC TAC ACT CTT TCC CTA CAC GAC GCT CTT CCG ATC TAA CA
adapter2(AACA)	TGT TAG ATC GGA AGA GCG TCG TGT AGG GAA AGA GTG TAG ATC TCG GTG G
adapter1(TTCA)	AAT GAT ACG GCG ACC ACC GAG ATC TAC ACT CTT TCC CTA CAC GAC GCT CTT CCG ATC TTT CA
adapter2(TTCA)	TGA AAG ATC GGA AGA GCG TCG TGT AGG GAA AGA GTG TAG ATC TCG GTG G
adapter1	AAT GAT ACG GCG ACC ACC GAG ATC TAC ACT CTT TCC CTA CAC GAC GCT CTT CCG ATC T
adapter2	AGA TCG GAA GAG CGT CGT GTA GGG AAA GAG TGT AGA TCT CGG TGG
lib adapter1	GAC GGC ATA CGA GAT
CLTA1 lib adapter2	AAC AAT CTC GTA TGC CGT CTT CTG CTT G
CLTA2 lib adapter2	TCT TAT CTC GTA TGC CGT CTT CTG CTT G
CLTA3 lib adapter2	AGA GAT CTC GTA TGC CGT CTT CTG CTT G
CLTA4 lib adapter2	TTG TAT CTC GTA TGC CGT CTT CTG CTT G
CLTA1 sel PCR	CAA GCA GAA GAC GGC ATA CGA GAT TGT GTT CTC GGC AGG T
CLTA2 sel PCR	CAA GCA GAA GAC GGC ATA CGA GAT AGA AGA CTC GGC AGG T
CLTA3 sel PCR	CAA GCA GAA GAC GGC ATA CGA GAT TTC TCT CTC GGC AGG T
CLTA4 sel PCR	CAA GCA GAA GAC GGC ATA CGA GAT ACA CAA CTC GGC AGG T
PE2 short	AAT GAT ACG GCG ACC ACC GA
CLTA1 lib seq PCR	AAT GAT ACG GCG ACC ACC GAG ATC TAC ACT CTT TCC CTA CAC GAC GCT CTT CCG ATC TNN NNA CCT ACC TGC CGA GAA CAC A
CLTA2 lib seq PCR	AAT GAT ACG GCG ACC ACC GAG ATC TAC ACT CTT TCC CTA CAC GAC GCT CTT CCG ATC TNN NNA CCT ACC TGC CGA GTC TTC T
CLTA3 lib seq PCR	AAT GAT ACG GCG ACC ACC GAG ATC TAC ACT CTT TCC CTA CAC GAC GCT CTT CCG ATC TNN NNA CCT ACC TGC CGA GAG AGA A
CLTA4 lib seq PCR	AAT GAT ACG GCG ACC ACC GAG ATC TAC ACT CTT TCC CTA CAC GAC GCT CTT CCG ATC TNN NNA CCT ACC TGC CGA GTT GTG T
lib fwd PCR	CAA GCA GAA GAC GGC ATA CGA GAT
CLTA1-0-1 (Chr. 9) fwd	ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT CAA GTC TAG CAA GCA GGC CA
CLTA1-0-1 (Chr. 12) fwd	ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT CAG GCA CTG AGT GGG AAA GT
CLTA1-1-1 fwd	ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT TAA CCC CAA GTC AGC AAG CA
CLTA1-2-1 fwd	ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT TTG CTG GTC AAT ACC CTG GC
CLTA1-2-2 fwd	ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT TGA GTA CCC CTG AAA TGG GC
CLTA1-3-1 fwd	ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT TCG CTA CCA ATC AGG GCT TT
CLTA1-3-2 fwd	ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT CCA TTG CCA CTT GTT TGC AT
CLTA1-4-1 fwd	ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT CCT ACC CCC ACA ACT TTG CT
CLTA1-4-2 fwd	ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT GTG TAC ATC CAG TGC ACC CA
CLTA1-4-3 fwd	ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT TCG GAA AGG ACT TTG AAT ACT TGT
CLTA1-4-4 fwd	ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT CGG CCC AAG ACC TCA TTC AC
CLTA1-4-5 fwd	ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT GTC CTC TCT GGG GCA GAA GT
CLTA1-4-6 fwd	ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT AGC TGA GTC ATG AGT TGT CTC C
CLTA1-4-7 fwd	ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT CTG CCA GCT TCT CAC ACC AT
CLTA1-4-8 fwd	ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT CTG AAG GAC AAA GGC GGG AA
CLTA1-5-1 fwd	ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT AAG GTG CTA AAG GCT CCA CG
CLTA1-5-2 fwd	ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT GAC CAT TGG TGA GCC CAG AG
CLTA1-5-3 fwd	ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT TTT TTC GGG CAA CTG CTC AC
CLTA1-5-4 fwd	ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT GCA AGC CTT CTC TCC TCA GA
CLTA1-5-5 fwd	ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT ACA CAA ACT TCC CTG AGA CCC
CLTA1-6-1 fwd	ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT TGA GTT AGC CCT GCT GTT CA
CLTA4-0-1 fwd	ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT TGA AGA GCT TCA CTG AGT AGG A
CLTA4-3-1 fwd	ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT TCC CCT TAC AGC CAA TTT CGT
CLTA4-3-2 fwd	ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT TGC TGA TGA AAT GCA ATT AAG AGG T
CLTA4-3-3 fwd	ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT GGT CCC TGC AAG CCA GTA TG
CLTA4-3-4 fwd	ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT ATC AAA GCC TTG TAT CAC AGT T
CLTA4-3-5 fwd	ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT CCC AAA TAA TGC AGG AGC CAA
CLTA4-3-6 fwd	ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT CTG CCT TTA GTG GGA CAG ACT T

CLTA4-3-7 fwd	ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT AGT AAC CCT AGT AGC CCT CCA
CLTA4-4-1 fwd	ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT CAT TGC AGT GAG CCG AGA TTG
CLTA4-4-2 fwd	ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT TGG CAA AGT TCA CTT CCA TGT
CLTA4-4-3 fwd	ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT TGC TCT GTG ATG TCT GCC AC
CLTA4-4-4 fwd	ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT TGT GTA GGA TTG TGA ACC AGC A
CLTA4-4-5 fwd	ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT TCC CAG CCC AGC ATT TTT CT
CLTA4-4-6 fwd	ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT AGG TTG CTT TGT GCA CAG TC
CLTA4-4-7 fwd	ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT CCT GGC TTG GGA TGT TGG AA
CLTA4-4-8 fwd	ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT TTG CCC AAG GTC ATA CTG CT
CLTA4-4-9 fwd	ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT ACC CAC TAG GTA GCC ATA ATC CA
CLTA4-4-10 fwd	ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT CGG TCA TGT CGC TTG GAA GA
CLTA4-4-11 fwd	ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT TTG GCC CAT ATT GCT TTA TGC TG
CLTA4-4-12 fwd	ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT ATT AGG GGT TGG CTG CAT GA
CLTA4-4-13 fwd	ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT CCA AGA CGT GTT GCA TGC TG
CLTA4-4-14 fwd	ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT TGG GAG GTG ATA AAT TCC CTA AAT
CLTA4-5-1 fwd	ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT CCA GAG ACA AAG GTG GGG AG
CLTA4-5-2 fwd	ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT TCA TAC AGA AGA GCA AAG TAC CA
CLTA4-5-3 fwd	ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT CAA AGA GGG GTA TCG GGA GC
CLTA4-5-4 fwd	ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT AAA TGG AAG AAC CAA GTA GAT GAA
CLTA4-5-5 fwd	ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT TTT TGG TTG ACA GAT GGC CAC A
CLTA4-5-6 fwd	ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT TCT TAC TTG TGT GAT TTT AGA ACA A
CLTA4-6-1 fwd	ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT GAT GGT TCA TGC AGA GGG CT
CLTA4-6-2 fwd	ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT GCT GGT CTT TCC TGA GCT GT
CLTA4-6-3 fwd	ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT CTC CAT CAG ATA CCT GTA CCC A
CLTA4-7-1 fwd	ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT GGG AAA ACA CTC TCT CTC TGC T
CLTA4-7-2 fwd	ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT GGA GGC CAC GAC ACA CAA TA
CLTA1-0-1 (Chr. 9) rev	GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATCT CAC AGG GTG GCT CTT CAG TG
CLTA1-0-1 (Chr. 12) rev	GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATCT TGC ACA TGT TTC CAC AGG GT
CLTA1-1-1 rev	GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATCT AGT GTT TCC AGG AGC GGT TT
CLTA1-2-1 rev	GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATCT AAG CCT CAG GCA CAA CTC TG
CLTA1-2-2 rev	GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATCT TAG GGG AGG GGC AAA GAC A
CLTA1-3-1 rev	GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATCT GGG AAC AGT GGT ATG CTG GT
CLTA1-3-2 rev	GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATCT AGT GTG GAC ACT GAC AAG GAA
CLTA1-4-1 rev	GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATCT TCA CTG CCT GGG TGC TTT AG
CLTA1-4-2 rev	GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATCT TAC CCC AGC CTC CAG CTT TA
CLTA1-4-3 rev	GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATCT TGA CTA CTG GGG AGC GAT GA
CLTA1-4-4 rev	GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATCT AGG CTG TTA TGC AGG AAA GGA A
CLTA1-4-5 rev	GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATCT GCG GTT GAG GTG GAT GGA AG
CLTA1-4-6 rev	GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATCT GGC AGC ATC CCT TAC ATC CT
CLTA1-4-7 rev	GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATCT AGA AAA AGC TTC CCC AGA AAG GA
CLTA1-4-8 rev	GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATCT CTG CAC CAA CCT CTA CGT CC
CLTA1-5-1 rev	GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATCT CTG GAG AGG GCA TAG TTG GC
CLTA1-5-2 rev	GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATCT TGG AAG GCT CTT TGT GGG TT
CLTA1-5-3 rev	GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATCT TTC CTA GCG GGA ACT GGA AA
CLTA1-5-4 rev	GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATCT AGG CTA ATG GGG TAG GGG AT
CLTA1-5-5 rev	GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATCT TGT CCA TGT TGG CTG AGG TG
CLTA1-6-1 rev	GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATCT CAG GCC AAC CTT GAC AAC TT
CLTA4-0-1 rev	GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATCT AGC AGG CCA AAG ATG TCT CC
CLTA4-3-1 rev	GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATCT TCT GCT CTT GAG GTT ATT TGT CC
CLTA4-3-2 rev	GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATCT GGG ACC AAT TTG CTA CTC ATG G
CLTA4-3-3 rev	GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATCT TGG AGG CTG TAA ACG TCC TG
CLTA4-3-4 rev	GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATCT TGC TAT GAT TTG CTG AAT TAC TCC T
CLTA4-3-5 rev	GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATCT GCA ATT TTG CAG ACC ACC ATC
CLTA4-3-6 rev	GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATCT GGC AGC TTG CAA CCT TCT TG
CLTA4-3-7 rev	GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATCT TCA TGA GAG TTT CCC CAA CA
CLTA4-4-1 rev	GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATCT ACT TGA GGG GGA AAA AGT TTC TTA
CLTA4-4-2 rev	GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATCT TGG TCC CTG TCT GTC ATT GG
CLTA4-4-3 rev	GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATCT AAG CGA GTG ACT GTC TGG GA

CLTA4-4-4 rev	GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATCT CAT GGG TGG GAC ACG TAG TT
CLTA4-4-5 rev	GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATCT GGC TTT CCT GGA CAC CCT ATC
CLTA4-4-6 rev	GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATCT AGA GCG AGG GAG CGA TGT A
CLTA4-4-7 rev	GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATCT TTG TGG ACC ACT GCT TAG TGC
CLTA4-4-8 rev	GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATCT CAA CTA CCC TGA GGC CAC C
CLTA4-4-9 rev	GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATCT GGT CAG CAC TCC TCA GCT TT
CLTA4-4-10 rev	GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATCT TGG AGG ATG CAT GCC ACA TT
CLTA4-4-11 rev	GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATCT CCC AGC CTC TTT GAC CCT TC
CLTA4-4-12 rev	GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATCT CCC ACA CCA GGC TGT AAG G
CLTA4-4-13 rev	GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATCT TAG ATA TAT GGG TGT GTC TGT ACG
CLTA4-4-14 rev	GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATCT TTC CAA AGT GGC TGA ACC AT
CLTA4-5-1 rev	GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATCT CCC ACA GGG CTG ATG TTT CA
CLTA4-5-2 rev	GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATCT TTG TAA TGC AAC CTC TGT CAT GC
CLTA4-5-3 rev	GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATCT CCA GCT CCA GCA ATC CAT GA
CLTA4-5-4 rev	GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATCT TTT GGG AAA GAT AGC CCT GGA
CLTA4-5-5 rev	GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATCT CAA TGA AAC AGC GGG GAG GT
CLTA4-5-6 rev	GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATCT ACA ATC ACG TGT CCT TCA CT
CLTA4-6-1 rev	GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATCT CAG ATC CCT CCT GGG CAA TG
CLTA4-6-2 rev	GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATCT GTC AGG AGG CAA GGA GGA AC
CLTA4-6-3 rev	GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATCT ACT TCC TTC CTT TTG AGA CCA AGT
CLTA4-7-1 rev	GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATCT GCG GCA GAT TCC TGG TGA TT
CLTA4-7-2 rev	GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATCT GGT CAC CAT CAG CAC AGT CA
PE1-barcode1	CAA GCA GAA GAC GGC ATA CGA GAT ATA TCA GTG TGA CTG GAG TTC AGA CGT GTG CT
PE1-barcode2	CAA GCA GAA GAC GGC ATA CGA GAT TTT CAC CGG TGA CTG GAG TTC AGA CGT GTG CT
PE1-barcode3	CAA GCA GAA GAC GGC ATA CGA GAT CCA CTC ATG TGA CTG GAG TTC AGA CGT GTG CT
PE1-barcode4	CAA GCA GAA GAC GGC ATA CGA GAT TAC GTA CGG TGA CTG GAG TTC AGA CGT GTG CT
PE1-barcode5	CAA GCA GAA GAC GGC ATA CGA GAT CGA AAC TCG TGA CTG GAG TTC AGA CGT GTG CT
PE1-barcode6	CAA GCA GAA GAC GGC ATA CGA GAT ATC AGT ATG TGA CTG GAG TTC AGA CGT GTG CT
PE2-barcode1	AAT GAT ACG GCG ACC ACC GAG ATC TAC ACA TTA CTC GAC ACT CTT TCC CTA CAC GAC
PE2-barcode2	AAT GAT ACG GCG ACC ACC GAG ATC TAC ACT CCG GAG AAC ACT CTT TCC CTA CAC GAC
PE2-barcode3	AAT GAT ACG GCG ACC ACC GAG ATC TAC ACC GCT CAT TAC ACT CTT TCC CTA CAC GAC

Supplementary Table S8. **Oligonucleotides used in this study**. All oligonucleotides were purchased from Integrated DNA Technologies. An asterisk (*) indicates that the preceding nucleotide was incorporated as a hand mix of phosphoramidites consisting of 79 mol% of the phosphoramidite corresponding to the preceding nucleotide and 4 mol% of each of the other three canonical phosphoramidites. "/5Phos/" denotes a 5' phosphate group installed during synthesis.

SUPPLEMENTARY ALGORITHMS

All scripts were written in C++. Scripts are available upon request. Algorithms used in this study are as previous reported (reference) with modification.

Sequence binning

1) designate sequence pairs starting with the barcode "AACA" or "TTCA" as post-selection library members

2) for post-selection library members (with illustrated example):

example read:
AACACATGGGTCGACACAAACACAACTCGGCAGGTACTTGCAGATGTAGTCTTTCCACATGGGTCGACACAAACACAACT
CGGCAGGTATCTCGTATGCC

i) search both paired reads for the positions, pos1 and pos2, of the constant sequence "CTCGGCAGGT"

ii) keep only sequences that have identical sequences between the barcode and pos1 and preceding pos2

iii) keep the region between the two instances of the constant sequence (the region between the barcode and pos1 contains a cut half-site; the region that is between the two instances of the constant sequence contains a full site)

example: ACTTGCAGATGTAGTCTTTCCACATGGGTCGACACAAACACAA

ii) search the sequence for a selection barcode ("TGTGTTTGTGTT" for CLTA1, "AGAAGAAGAAGA" for CLTA2, "TTCTCTTTCTCT" for CLTA3, "ACACAAACACAA" for CLTA4)

example: ACTTGCAGATGTAGTCTTTCCACATGGGTCGACACAAACACAA
CLTA4

iii) the sequence before the barcode is the full post-selection library member (first four and last four nucleotides are fully randomized flanking sequence)

example: ACTT GCAGATGTAGTCTTTCCACATGG GTCG

iv) parse the quality scores for the positions corresponding to the 23 nucleotide post-selection library member

example read:
AACACATGGGTCGACACAAACACAACTCGGCAGGTACTTGCAGATGTAGTCTTTCCACATGGGTCGACACAAACACAACTCGGCAGGTATCTC
GTATGCC

CCCFFFFFHHHHHJJJJJJJJJJJJJJJJJJJJGIJJJJIJIJJJIIIHIIJJJHHHGHAEFCDDDDDDDDDDDDDDDDDDDDDDD?CDDED
D@DCCCD

v) keep sequences only if the corresponding quality score string (underlined) FASTQ quality characters for the sequence are '?' or higher in ASCII code (Phred quality score >= 30)

NHEJ sequence calling

example read:

CAATCTCCCGCATGCGCTCAGTCCTCATCTCCCTCAAGCAGGCCCCGCTGGTGCACTGAAGAGCCACCCTGTGAAACACTACATCTGCAATAT
CTTAATCCTACTCAGTGAAGCTCTTCACAGTCATTGGATTAATTATGTTGAGTTCTTTTGGACCAAACC

example quality scores:

CCBCCFFFFCCCGGGGGGGGGGHHHHHHHHHHHHHHHHHHHHHHGGGGGGGGGHHHHHHHHHHHHHHHHHHGHHHHHHHHHHHHHHHHHHHHHHGHH
HHHHHHHHHHHHHHHHFHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHGHFHHHHHF

1) identify the 20 base pairs flanking both sides of 20 base pair target site + three base pair PAM for each target site

example flanking sequences:

GCTGGTGCACTGAAGAGCCA

AATATCTTAATCCTACTCAG

2) search all sequence reads for the flanking sequences to identify the potential off-target site (the sequence between the flanking sequences)

example potential off-target site:

CCCTGTGAAACACTACATCTGC

3) if the potential off-target site contains indels (length is less than 23), keep sequence as potential off-target site if all corresponding FASTQ quality characters for the sequence are '?' or higher in ASCII code (Phred quality score >= 30)

example potential off-target site length = 22

example corresponding FASTQ quality characters: HHGHHHHHHHHHHHHHHHHHHH

4) bin and manually inspect all sequences that pass steps 2 and 3 and keep sequences as potential modified sequences if they have at least one deletion involving position 16, 17, or 18 (of 20 counting from the non-PAM end) of if they have an insertion between position 17 and 18, consistent with the most frequent modifications observed for the intended target site (Figure 3)

example potential off-target site (reverse complement, with positions labeled) with reference sequence:

```
                          11111111112222
non-PAM end 12345678901234567890123 PAM end
            GCAGATGTAGTGTTTC-ACAGGG
            GCAGATGTAGTGTTTCCACAGGG
```

4) repeat steps 1-3 for read2 and keep only if the sequence is the same

5) compare overall counts in Cas9+sgRNA treated sample to Cas9 alone sample to identify modified sites

Filter based on cleavage site (for post-selection sequences)

1) tabulate the cleavage site locations across the recognition site by identifying the first position in the full sequenced recogntion site (between the two constant sequences) that is identical to the first position in the sequencing read after the barcode (before the first constant sequence)

2) after tabulation, repeat step 1, keeping only sequences with cleavage site locations that are present in at least 5% of the sequencing reads

SUPPLEMENTARY REFERENCES

1. Jiang, W., Bikard, D., Cox, D., Zhang, F. & Marraffini, L.A. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nature biotechnology* **31**, 233-239 (2013).
2. Pattanayak, V., Ramirez, C.L., Joung, J.K. & Liu, D.R. Revealing off-target cleavage specificities of zinc-finger nucleases by in vitro selection. *Nature methods* **8**, 765-770 (2011).
3. Schneider, T.D. & Stephens, R.M. Sequence logos: a new way to display consensus sequences. *Nucleic acids research* **18**, 6097-6100 (1990).