Broad Specificity Profiling of TALE Nucleases (TALENs) Results in Engineered TALENs With Improved DNA Cleavage Specificity

*John P. Guilinger,[1] Vikram Pattanayak,[1] Deepak Reyon,[2,3] Shengdar Q. Tsai,[2,3] Jeffry D. Sander,[2,3] J. Keith Joung,[2,3] & David R. Liu[1]**

SUPPLEMENTARY INFORMATION

Supplementary Results	Additional Details of the *In Vitro* Selection
Supplementary Results	Comparison of Classifier and Purely Computational TALEN Off-Target Site Identification
Supplementary Results	Engineering N-Terminal Domains for Improved TALEN DNA Cleavage Specificity
Supplementary Results	Effects of N-Terminal Domain, C-Terminal Domain, and TALEN Concentration on Specificity
Supplementary Results	DNA Spacer-Length and Cut-Site Preferences
Supplementary Results	Details of the Classifier for TALEN Off-target Site Identification
Supplementary Results	TALE Repeats Productively Bind Base Pairs with Relative Independence
Supplementary Results	Longer TALENs Are Predicted to Induce Less Off-target Cleavage in a Genomic Context
Supplementary Discussion	Comparison of Our Method with Others for Characterizing TALEN Specificity and Identifying Genomic TALEN Off-target Sites
Supplementary Figure S1	Targeted DNA Sequences in Human *CCR5* and *ATM* Genes
Supplementary Figure S2	TALEN and Restriction Enzyme Digestion of Concatemerized On-Target or Mutant DNA Sites
Supplementary Figures S3 to S8	Specificity Profiles of All 41 TALEN Selections
Supplementary Figure S9	Modifications Induced by TALENs at On-Target And Predicted Off-Target Genomic Sites.
Supplementary Figure S10	Correspondence of Modification Induced by CCR5A TALENs at Genomic Sites Between Independent Biological Replicates
Supplementary Figure S11	Observed Versus Predicted Double-Mutant Sequence Enrichment Values
Supplementary Figure S12	Predicted Off-Target Genomic Cleavage as a Function of TALEN Length Considering Both TALEN Specificity And Off-Target Site Abundance in the Human Genome
Supplementary Figure S13	Effects of Engineered TALEN Domains and TALEN Concentration on Specificity
Supplementary Figure S14	Spacer-Length Preferences of TALENs
Supplementary Figure S15	DNA Cleavage-Site Preferences of TALENs

Supplementary Table S1	TALEN Constructs and Concentrations Used in the Selections
Supplementary Table S2	Statistics of Sequences Selected by TALEN Digestion
Supplementary Table S3	Statistics of Sequences From Pre-Selection Libraries
Supplementary Table S4	Enrichment Values of Sequences as a Function of Number of Mutations
Supplementary Table S5	Genomic Off-Target Site Abundance in the Human Genome
Supplementary Table S6	Predicted Off-Target Sites in the Human Genome
Supplementary Table S7	Cellular Modification Induced by TALENs at On-Target and Predicted Off-Target Genomic Sites
Supplementary Table S8	Cellular Modification Induced by HDAC1, PMS2 and SDHD TALENs at On-Target and Off-Target Genomic Sites
Supplementary Table S9	Cellular Modification Induced by CCR5A TALENs at On-Target and Predicted Off-Target Genomic Sites Generated by TALENoffer and Generated by This Study
Supplementary Table S10	Exponential Fitting of Enrichment Values as a Function of Mutation Number
Supplementary Table S11	Exponential Fitting and Extrapolation of Enrichment Values as a Function of Mutation Number
Supplementary Table S12	Oligonucleotides Used in This Study
Supplementary Algorithm	Computational Filtering Of Pre-Selection Sequences And Selected Sequences
Supplementary Algorithm	Identification of Indels in Genomic Sequences
Supplementary Algorithm	Computational Search For Genomic Off-Target Sites Related to the CCR5B Target Site
Supplementary Sequences	TALEN Protein Sequences

SUPPLEMENTARY RESULTS

Additional Details of the *In Vitro* Selection

TALEN concentrations in the selection were empirically determined to avoid overdigestion, since the selection isolates DNA products 1.5 target sites in length. For example, CCR5A TALEN digestion of concatemerized repeats of the CCR5 DNA target sites under the conditions used in the selection yielded the expected ladder of integral units of target sites (Supplementary Fig. S2). Under these conditions, there is little cleavage to monomeric target sites compared to digestion of the same concatemerized DNA with a restriction enzyme that yields mostly monomeric sites. TALEN concentrations used in the selection (3 nM to 40 nM) correspond to ~20 to ~200 dimeric TALEN molecules per human cell nucleus,[1] which is in the lower range of cellular protein expression[2, 3] and therefore approaches the lower limit of TALEN concentration possible in a cell.

A completely random library of 25% of each base pair at each position would not yield sufficient library coverage of potential target sites with relatively few mutations relative to the on-target site. Therefore, a partially randomized library of 79% on-target base pair at each position was used. At this randomization frequency, 0.021% of the library contains an on-target site (no mutations), based on the expected binomial distribution of 79% on-target base pairs across the 36-bp target site. 14% of the library will have six mutations, with ~100 copies on average of all possible six-mutation sequences possible in a 36-bp site. (In theory, the library will have 1.0×10^{12} molecules with exactly six mutations which is ~100-fold more than the 1.4×10^{9} different six-mutation sequences possible in a 36-bp site.) The DNA that survived the selection contained significantly fewer mean mutations in the targeted half-sites than were present in the pre-selection libraries (Fig. 2A and 2B; Supplementary Table S2 and S3). For example, the mean number of combined mutations in the 18-bp left half-site and 18-bp right half-site among DNA sequences surviving selection after treatment with TALENs was 4.06 for CCR5A and 3.18 for ATM sequences, respectively, compared to 7.54 and 6.82 mutations in the corresponding pre-selection libraries (Fig. 2A and 2B).

The entire selection process can be completed in less than one week, with the single most time-consuming step being high-throughput DNA sequencing, a step that is necessary for any method that attempts to detect low levels of TALEN-induced modification at genomic off-target sites.

Details of the Classifier for TALEN Off-target Site Identification

The "classifier" algorithm calculates the posterior probability of each nucleotide in each position of a target to occur in a sequence that was cleaved by the TALENs in opposition to sequences from the target library that were not observed to be cleaved.[4] These posterior probabilities were then used to score the likelihood that the TALEN used to train the algorithm would cleave every possible target sequence in the human genome with monomer spacing of 10 to 30 bps. Since sites containing the

longest tested DNA spacers (24 bp) were productively cleaved during the *in vitro* selection (Supplementary Fig. S14), target sites with even longer spacers were considered for potential off-target modification in the genome.

Comparison of Classifier and Purely Computational TALEN Off-target Site Identification

To better understand the likelihood of detecting genomic off-target sites using other methods, three previously reported TALENs targeting separate sites in the human HDAC1, PMS2, and SDHD genes[5] were constructed with ELD/KKR *Fok*I domains. These TALENs were chosen because their target sites are only three to five mutations away from potential off-target sites in the human genome. (Supplementary Table S6C). The HDAC1, PMS2, and SDHD putative off-target and on-target sites were amplified from genomic DNA isolated from cells expressing the appropriate TALENs. Since only three out of six of these more closely related off-target sites containing three to five mutations were significantly modified (Supplementary Table S8), it is likely that more distant off-target sites with 8 to 12 mutations would be very difficult to predict and detect, consistent with reports[6-9] that fail to identify TALEN-induced off-target site modification in cells. The challenge is compounded by the presence of more than 10,000 potential genomic off-target sites containing 8 to 12 mismatches out of 36 recognized bases revealed by both computational identification and statistical modeling (Supplementary Table S5)

To compare genomic off-target site prediction by our classifier to a purely computational approach, we used a recently developed *ab initio* genomic off-target site prediction algorithm, TALENoffer,[10] to identify the 32 best-scoring genomic off-target sites for the CCR5A TALEN. These TALENoffer-predicted off-target sites were amplified from genomic DNA purified from cells expressing the CCR5A TALEN with homodimeric FokI domains, and assayed for cellular modification (Supplementary Table S6D). Since the TALENoffer sites were only eight to 11 mutations from the on-target site, they were on average less distant from the on-target sequence than the off-target sites predicted by our classifier (9.2 versus 11.1 mean mutations in the off-target sites from the TALENoffer versus from our classifier, respectively). Therefore, from our classifier prediction of genomic off-target sites for the CCR5A TALEN, only CCR5A genomic off-target sites containing eight to 11 mutations were considered for comparison to TALENoffer sites (Supplementary Table S6). Both sets of sites were assayed for cellular modification by the CCR5A TALEN with homodimeric FokI domains. One genomic off-target site predicted by both TALENoffer and our classifier was modified in cells (see Supplementary Fig. S10 for a comparison of modification percentages between biological replicates). Of the remaining 25 amplified genomic off-target sites predicted exclusively by TALENoffer, only one was significantly modified in cells (Supplementary Table S9A). In contrast, 5 of the 25 amplified

genomic off-target sites exclusively predicted by our classifier were significantly modified in cells (Supplementary Table S9B). These results indicate that our classifier outperformed purely computational prediction for off-target CCR5A TALEN substrates. We also note that ten new, amplified genomic off-target sites were assayed for cellular modification by CCR5A TALENs (Supplementary Table S9B). Combining these ten new CCR5A sites with the 35 CCR5A and 31 ATM genomic off-target sites previously assayed yields a total of 76 assayed genomic off-target sites of which 16 demonstrated significant TALEN-induced modification at off-target sites in cells.

TALE Repeats Productively Bind Base Pairs with Relative Independence

In order to assess whether mutations at one position in the target sequence affect the ability of TALEN repeats to productively bind other positions, we generated an expected enrichment value for every possible double-mutant sequence for the L13+R13 CCR5B TALENs assuming independent contributions from the two corresponding single-mutation enrichments. In general, the predicted enrichment values closely resembled the actual observed enrichment values for each double-mutant sequence (Supplementary Fig. 11A and 11B), suggesting that component single mutations independently contributed to the overall cleavability of double-mutant sequences. The difference between the observed and predicted double-mutant enrichment values was relatively independent of the distance between the two mutations, except that two adjacent mismatches were slightly better tolerated than would be expected (Supplementary Fig. 11C). To determine the potential interdependence of more than two mutations, we evaluated the relationship between selection enrichment values and the number of mutations in the post-selection target for the L13+R13 CCR5B TALEN (Fig. 4A, black line). For 0 to 5 mutations, enrichment values closely followed a simple exponential function of the mean number of mutations (m) (Supplementary Table S10). This relationship is consistent with a model in which each successive mutation reduces the binding energy by a constant amount ($\Delta\Delta G$), resulting in an exponential decrease in TALEN binding ($K_{eq}(m)$) such that $K_{eq}(m) \sim e^{\Delta\Delta G * m}$. The observed exponential relationship therefore suggests that the mean reduction in binding energy from a typical mismatch is independent of the number of mismatches already present in the TALEN:DNA interaction. Collectively, these results indicate that TALE repeats bind their respective DNA base pairs independently beyond a slightly increased tolerance for adjacent mismatches.

Longer TALENs Are Predicted to Induce Less Off-Target Cleavage in a Genomic Context

Although longer TALENs are more tolerant of mismatched sequences (Fig. 4) than shorter TALENs, in the human genome there are far fewer closely related off-target sites for a longer target site

than for a shorter target site (Supplementary Fig. S12A). Since off-target site abundance and cleavage efficiency both contribute to the number of off-target cleavage events in a genomic context, we calculated overall genome cleavage specificity as a function of TALEN length by multiplying the extrapolated mean enrichment value of mutant sequences of a given length with the number of corresponding mutant sequences in the human genome (we note that this estimation assumes that extrapolated mean enrichments of highly mutant off-target substrates correlate with cleavage rates in cells). The decrease in potential off-target site abundance resulting from the longer target site length is large enough to outweigh the decrease in specificity per recognized base pair observed for longer TALENs (Supplementary Fig. S12B). As a result, longer TALENs are predicted to be more specific against the set of potential cleavage sites in the human genome than shorter TALENs for the tested TALEN pairs targeting a total of 20 to 32 base pairs.

Engineering N-Terminal Domains for Improved TALEN DNA Cleavage Specificity

The model of TALEN binding and specificity described in the main text predicts that reducing excess TALEN binding energy will increase TALEN DNA cleavage specificity. To further test this prediction and potentially further augment TALEN specificity, we mutated one ("N1", K150Q), two ("N2", K150Q and K153Q), or three ("N3", K150Q, K153Q, and R154Q) Lys or Arg residues to Gln in the N-terminal domain of TALENs targeting CCR5A and ATM. These N-terminal residues have been shown in previous studies to bind non-specifically to DNA, and mutations at these specific residues to neutralize the cationic charge decrease non-specific DNA binding energy.[11] We hypothesized the reduction in non-specific binding energy from these N-terminal mutations would decrease excess TALEN binding energy resulting in increased specificity. *In vitro* selections on these three TALEN variants revealed that the less cationic N-terminal TALENs indeed exhibit greater enrichment values of on-target cleavage (Supplementary Table S4).

Effects of N-Terminal Domain, C-Terminal Domain, and TALEN Concentration on Specificity

All TALEN constructs tested specifically recognize the intended base pair across both half-sites (Supplementary Fig. S3 to S8), except that some of the ATM TALENs do not specifically interact with the base pair adjacent to the spacer (targeted by the most C-terminal TALE repeat) (Supplementary Fig. S5 and S6). To compare the broad specificity profiles of canonical TALENs with those containing engineered C-terminal or N-terminal domains, the specificity scores of each target base pair from selections using CCR5A and ATM TALENs with the canonical, Q3, or Q7 C-terminal domains and N1, N2, or N3 N-terminal domains were subtracted by the corresponding specificity scores from selections on the canonical TALEN (canonical 63-aa C-terminal domain, wild-type N-terminal domain). The

results are shown in Supplementary Fig. S13. Mutations in the C-terminal domain that increase specificity did so most strongly in the middle and at the C-terminal end of each half-site. Likewise, the specificity-increasing mutations in the N-terminus tended to increase specificity most strongly at positions near the TALEN N-terminus (5' DNA end) although mutations in the N-terminus of ATM TALEN targeting the right half-site did not significantly alter specificity. These results are consistent with a local binding compensation model in which weaker binding at either terminus demands increased specificity in the TALE repeats near this terminus. To characterize the effects of TALEN concentration on specificity, the specificity scores from selections of ATM and CCR5A TALENs performed at three different concentrations ranging from 3 nM to 16 nM were each subtracted by the specificity scores of corresponding selections performed at the highest TALEN concentration assayed, 24 nM for ATM, or 32 nM for CCR5A. The results (Supplementary Fig. S13) indicate that specificity scores increase fairly uniformly across the half-sites as the concentration of TALEN is decreased.

DNA Spacer-Length and Cut-Site Preferences

To assess the spacer-length preference of various TALEN architectures (C-terminal mutations, N-terminal mutations, and *Fok*I variants) and various TALEN concentrations, the enrichment values of library members with 10- to 24- base pair spacer lengths in each of the selections with CCR5A and ATM TALEN with various combinations of the canonical, Q3, Q7, or 28-aa C-terminal domains; N1, N2, or N3 N-terminal mutations; and the EL/KK or ELD/KKR *Fok*I variants at 4 nM to 32 nM CCR5A and ATM TALEN were calculated (Supplementary Fig. S14). All of the tested concentrations, N-terminal variants, C-terminal variants, and *Fok*I variants demonstrated a broad DNA spacer-length preference ranging from 14- to 24- base pairs with three notable exceptions. First, the CCR5A 28-aa C-terminal domain exhibited a much narrower DNA spacer-length preference than the broader DNA spacer-length preference of the canonical C-terminal domain, consistent with previous reports.[12-14] Second, the CCR5A TALENs containing Q7 C-terminal domains showed an increased tolerance for 12-base spacers compared to the canonical C-terminal domain variant (Supplementary Fig. S14). This slightly broadened spacer-length preference may reflect greater conformational flexibility in the Q7 C-terminal domain, perhaps resulting from a smaller number of non-specific protein:DNA interactions along the TALEN:DNA interface. Third, the ATM TALENs with Q7 C-terminal domains and the ATM TALENs with N3 mutant N-terminal domains showed a narrowed spacer preference. We speculate that these more specific TALENs (Supplementary Table S4) with lower DNA-binding affinity may have faster off-rates that are competitive with the rate of cleavage of non-optimal DNA spacer lengths, altering the observed spacer-length preference. While previous reports have focused on the length of the TALEN C-terminal domain as a primary determinant of DNA spacer-length preference, these results suggest the net charge of the C-terminal domain as well as overall DNA-binding affinity can also affect

TALEN spacer-length preference. For many of the TALENs assayed, the spacer preferences revealed in the *in vitro* selection results are broader than a previous report by Miller *et al*[13] of a comparably narrower spacer length preference in cells. However, the broader spacer preferences revealed in the *in vitro* selection are consistent with a previous report by Mussolino *et al*.[15] This discrepancy could be explained by higher affinity TALENs, or by a higher concentration of TALENs used in our study and used in the study of Mussolino *et al*., leading to saturation of binding sites that allows cleavage of otherwise non-optimal DNA spacer lengths.

We also characterized the location of TALEN DNA cleavage within the spacer. We created histograms reporting the number of spacer DNA bases observed preceding the right half-site in each of the sequences from the selections with CCR5A and ATM TALEN with various combinations of the canonical, Q3, Q7, or 28-aa C-terminal domains; N1, N2, or N3 N-terminal mutations; and the EL/KK or ELD/KKR *Fok*I variants (Supplementary Fig. S15). The peaks in the histogram were interpreted to represent the most likely locations of DNA cleavage within the spacer. The cleavage positions are dependent on the length of the DNA spacer between the TALEN binding half-sites, as might be expected from conformational constraints imposed by the TALEN C-terminal domain and DNA spacer lengths.

SUPPLEMENTARY DISCUSSION

Comparison of Our Method with Others for Characterizing TALEN Specificity and Identifying Genomic TALEN Off-target Sites

While previous reports[16,17] have investigated TALEN specificity by identifying genomic off-target modifications for only a single TALEN pair, our study profiles the genomic specificity of two independent TALEN pairs (CCR5A and ATM) and is the first to compare different TALEN variants (canonical 63-aa vs. Q3 vs. Q7 C-terminal domains). It is difficult to directly compare our study with those that characterize different TALENs, although we note that our study has identified 16 bona fide heterodimeric off-target sites with eight to 12 mutations modified by TALENs in cells while the previous studies, using SELEX[16] or IDLV[17], collectively identified only three such sites that are distant from their corresponding on-target sequences. Furthermore, the same strategy used in this study has been previously shown to identify more off-target sites of a ZFN[6, 18] than the purely cellular IDLV study[19] or using SELEX data to predict off-target sites.[20] Of the 76 total IDLV insertion sites identified in cells treated with TALENs,[17] seven were at the target site while only three off-target sites were consistent with nuclease-mediated IDLV insertion. Only one of these three genomic off-target sites was heterodimeric (the use of TALENs containing homodimeric *Fok1* domains likely allowed for homodimeric off-target sites in which both right and left half-sites were targeted by the same TALEN). The on-target:off-target activity ratios of 170 and 1,140 for the two bona fide TALEN genomic off-targets sites out of 19 identified by SELEX[16] are similar to the 12 to 890 on:off-target activity ratios observed in this study.

While a comparison of the specificity of various programmable genome-editing technologies including ZFNs, TALENs and CRISPRs would be of interest, in order for such a comparison to be rigorous, many new lines of experiments are needed so that all three technologies are evaluated on their ability to cleave the same target sequences *in vitro* and in cells under the same conditions.

SUPPLEMENTARY FIGURES

A

CCR5 target sites

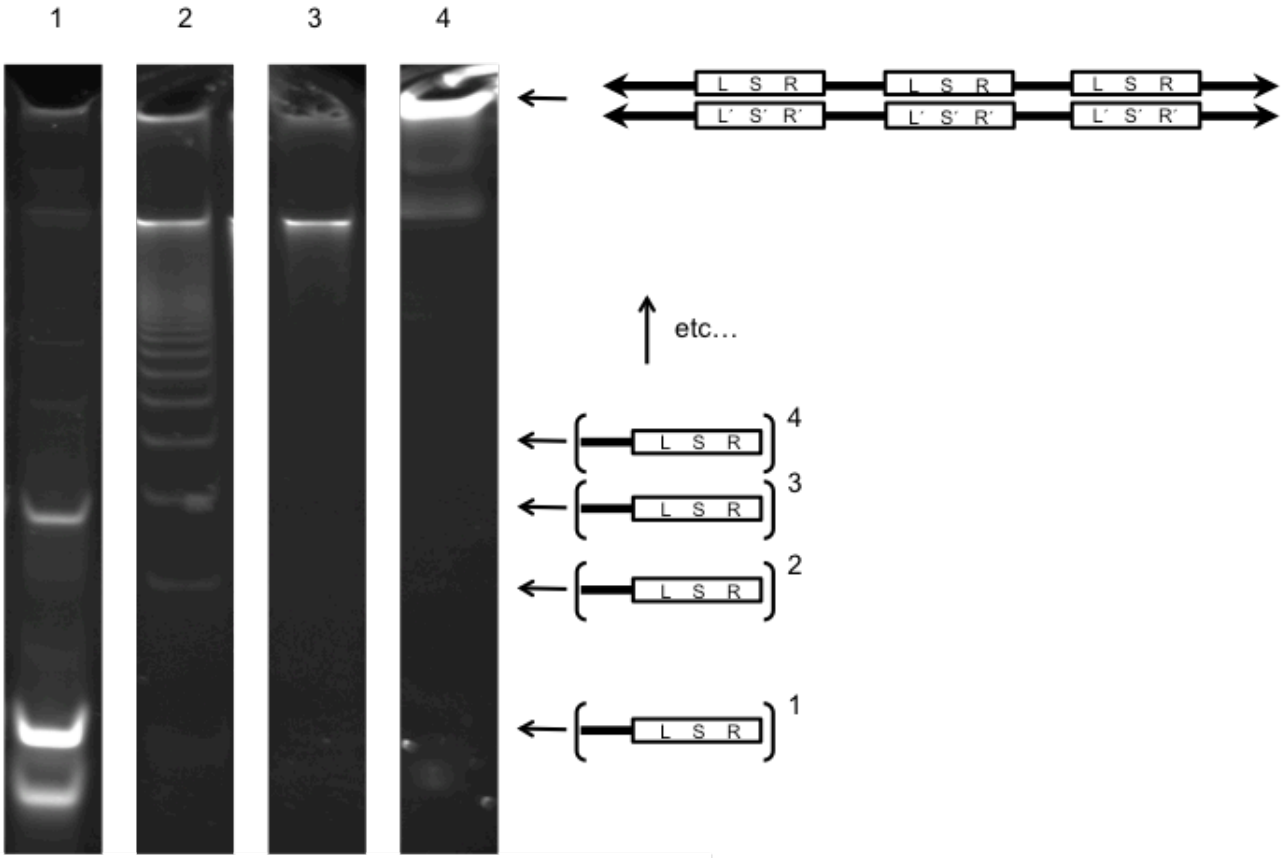
```
TALEN monomer
CCR5A L18      5′-TTCATTACACCTGCAGCT
CCR5B L16    5′-TCTTCATTACACCTGC
CCR5B L13       5′-TCATTACACCTGC
CCR5B L10          5′-TTACACCTGC

                 TCTTCATTACACCTGCAGCTCTCATTTTCCATACAGTCAGTATCAATTCTGGAAGA
                 AGAAGTAATGTGGACGTCGAGAGTAAAAGGTATGTCAGTCATAGTTAAGACCTTCT

CCR5A R18                                                  TCATAGTTAAGACCTTCT-5′
CCR5B R16                                          GTATGTCAGTCATAGT-5′
CCR5B R13                                          GTATGTCAGTCAT-5′
CCR5B R10                                          GTATGTCAGT-5′
```

B

ATM target site

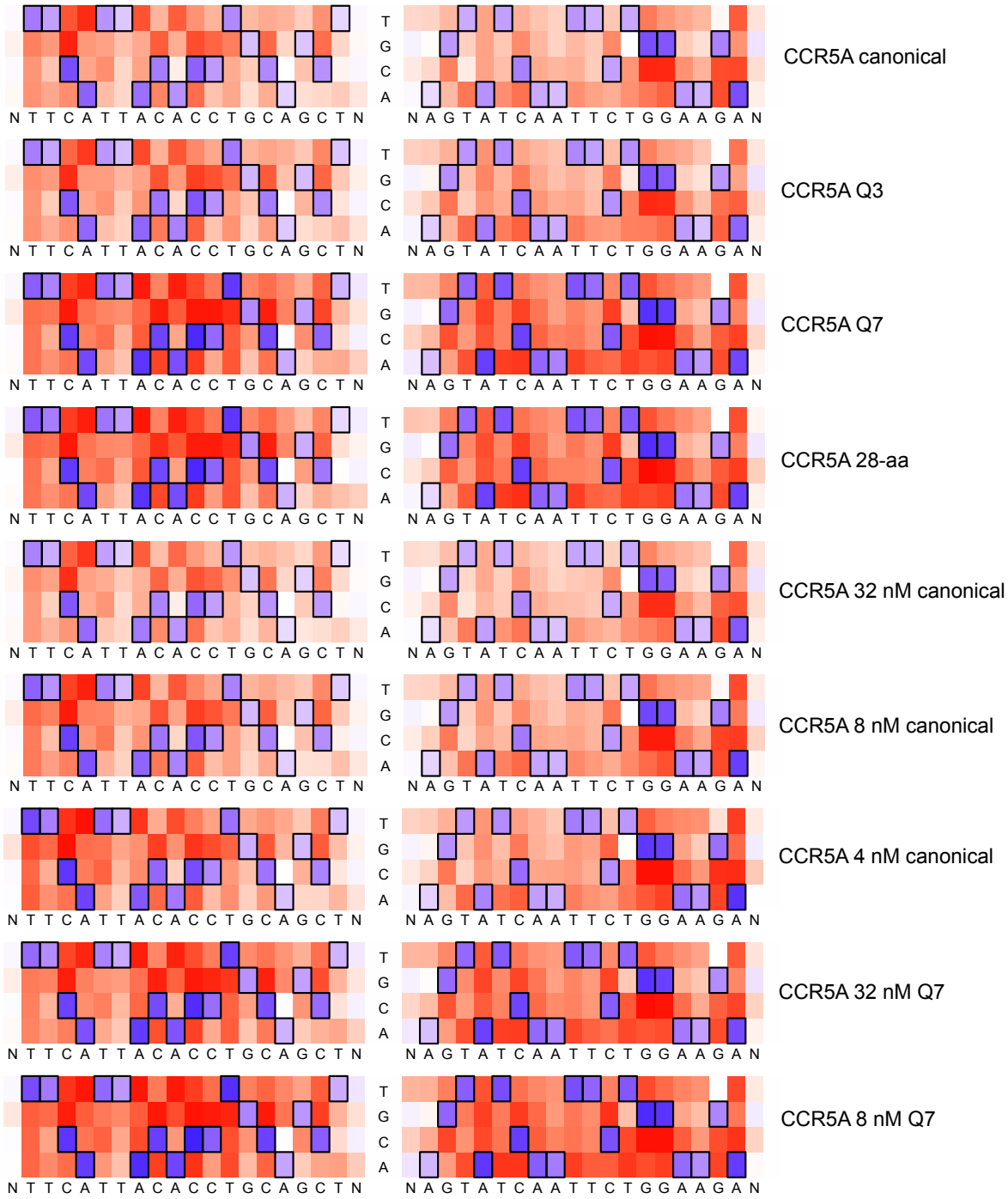
```
TALEN monomer
ATM L18   5′-TGAATTGGGATGCTGTTT

              TGAATTGGGATGCTGTTTTTAGGTATTCTATTCAAATTTATTTTACTGTCTTTA
              ACTTAACCCTACGACAAAAATCCATAAGATAAGTTTAAATAAAATGACAGAAAT

ATM R18                                            AAATAAAATGACAGAAAT-5′
```

Supplementary Figure S1. Target DNA Sequences in Human *CCR5* and *ATM* Genes. The target DNA sequences for the TALENs used in this study are shown in black. The N-terminal TALEN end recognizing the 5′ T for each half-site target is noted (5′) and TALENs are named according to number of base pairs targeted. TALENs targeting the CCR5 L18 and R18 shown are referred to as CCR5A TALENs while TALENs targeting the L10, L13, L16, R10, R13 or R16 half-sites shown are referred to as CCR5B TALENs.

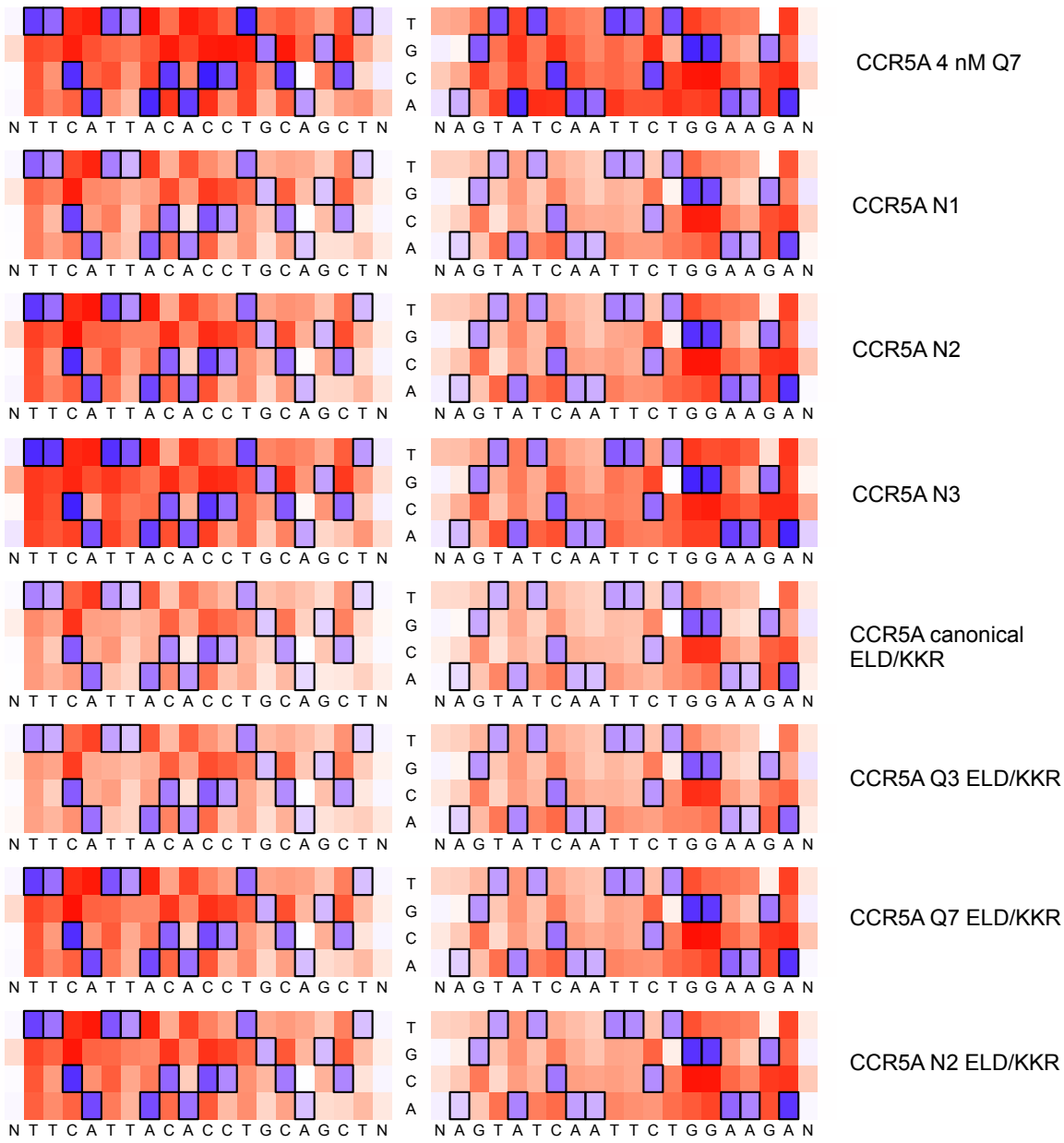


Supplementary Figure S2. TALEN and Restriction Enzyme Digestion of Concatemerized On-target or Mutant DNA sites. A single-stranded discrete DNA oligonucleotide containing a left half-site (L), spacer (S), right half-site (R), and constant region (represented by a thick black line) was circularized, then concatemerized by rolling circle amplification. The protocol for *In Vitro* Selection for DNA Cleavage (Online Methods) was preformed using CCR5Aon-Circ oligos (CCR5 on-target DNA sites) or CCR5Amut-Circ oligos (CCR5 mutant DNA sites) except after TALEN digestion samples were analyzed by PAGE gel. For lanes 1, 2, and 4 the DNA target site was the CCR5 on-target sequence but for lane 3 the DNA sequence contained a completely randomized right half-site. For lanes 1, 2, and 3 the resulting concatemerized DNAs were incubated with in vitro-translated CCR5A TALENs and the samples were purified by M-column (Qiagen) and analyzed by PAGE. Lane 1 shows digestion of concatemerized on-target DNA sites with a restriction enzyme (*Tsp*45I); lane 2 shows digestion of concatemerized on-target DNA sites with a CCR5A TALEN; lane 3 shows digestion of concatemerized mutant DNA sites with a CCR5A TALEN; and lane 4 shows the concatemerized DNA substrate before digestion or column purification.

A

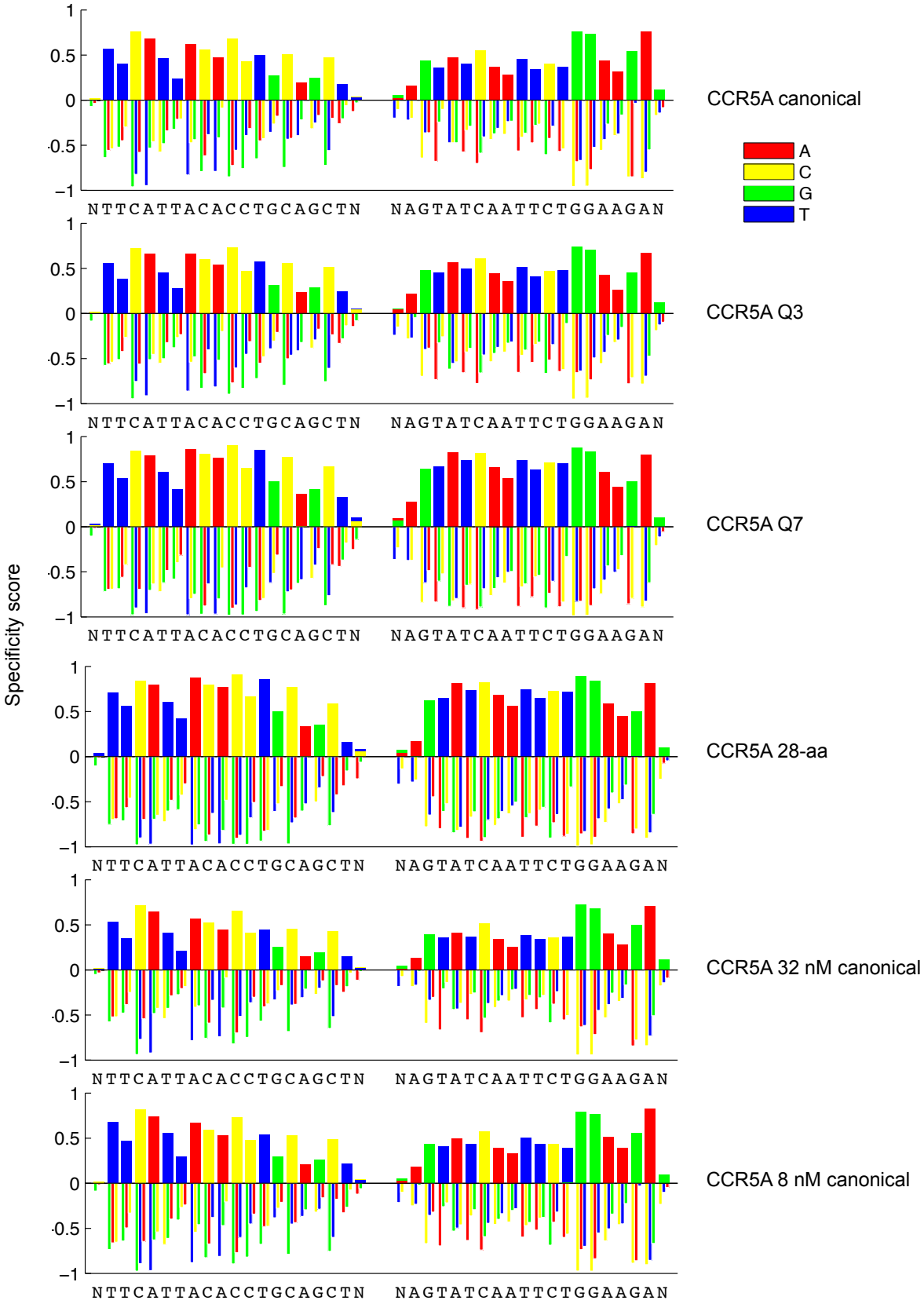

T G C A
N T T C A T T A C A C C T G C A G C T N
N A G T A T C A A T T C T G G A A G A N
CCR5A canonical
CCR5A Q3
CCR5A Q7
CCR5A 28-aa
CCR5A 32 nM canonical
CCR5A 8 nM canonical
CCR5A 4 nM canonical
CCR5A 32 nM Q7
CCR5A 8 nM Q7

B

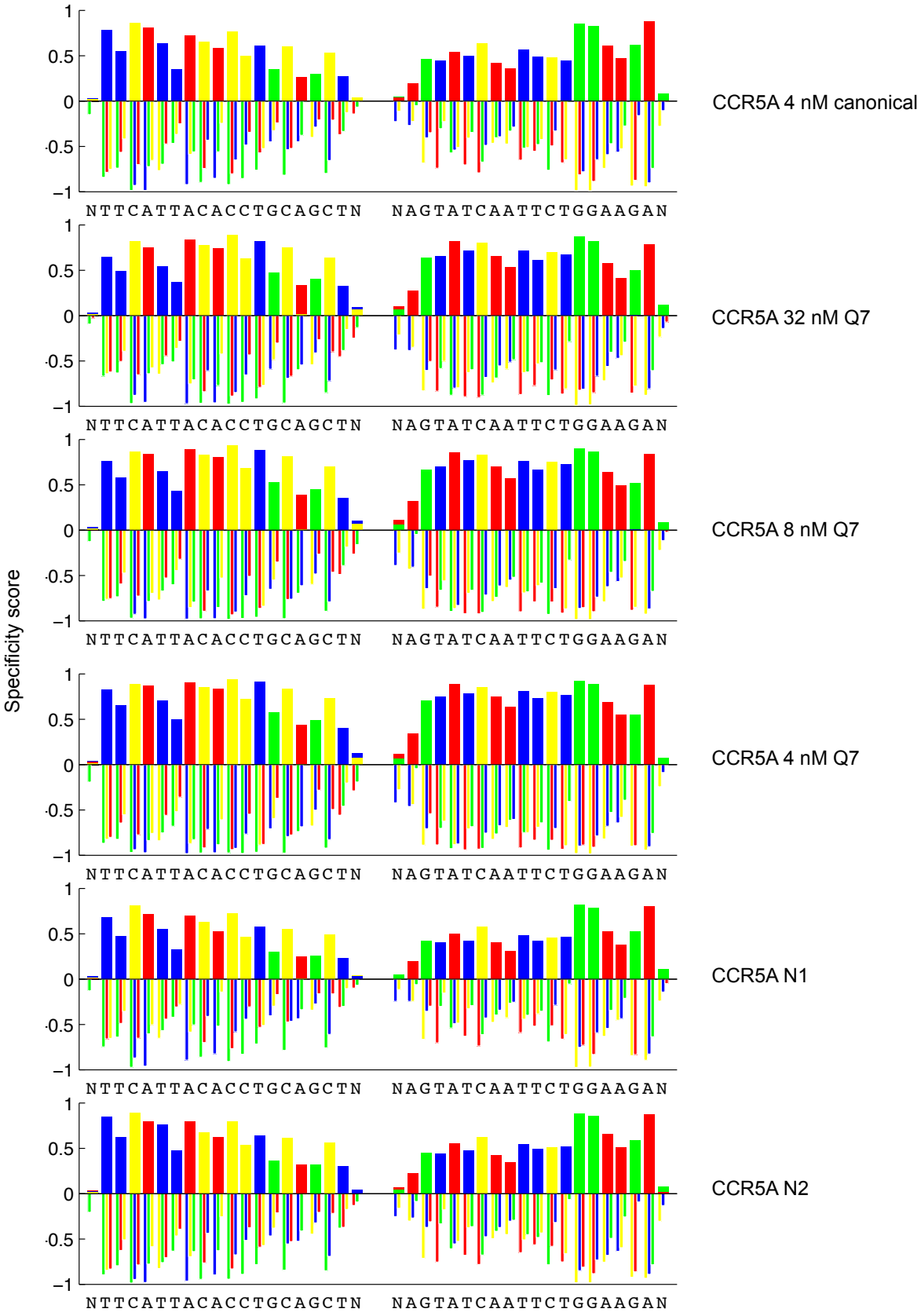


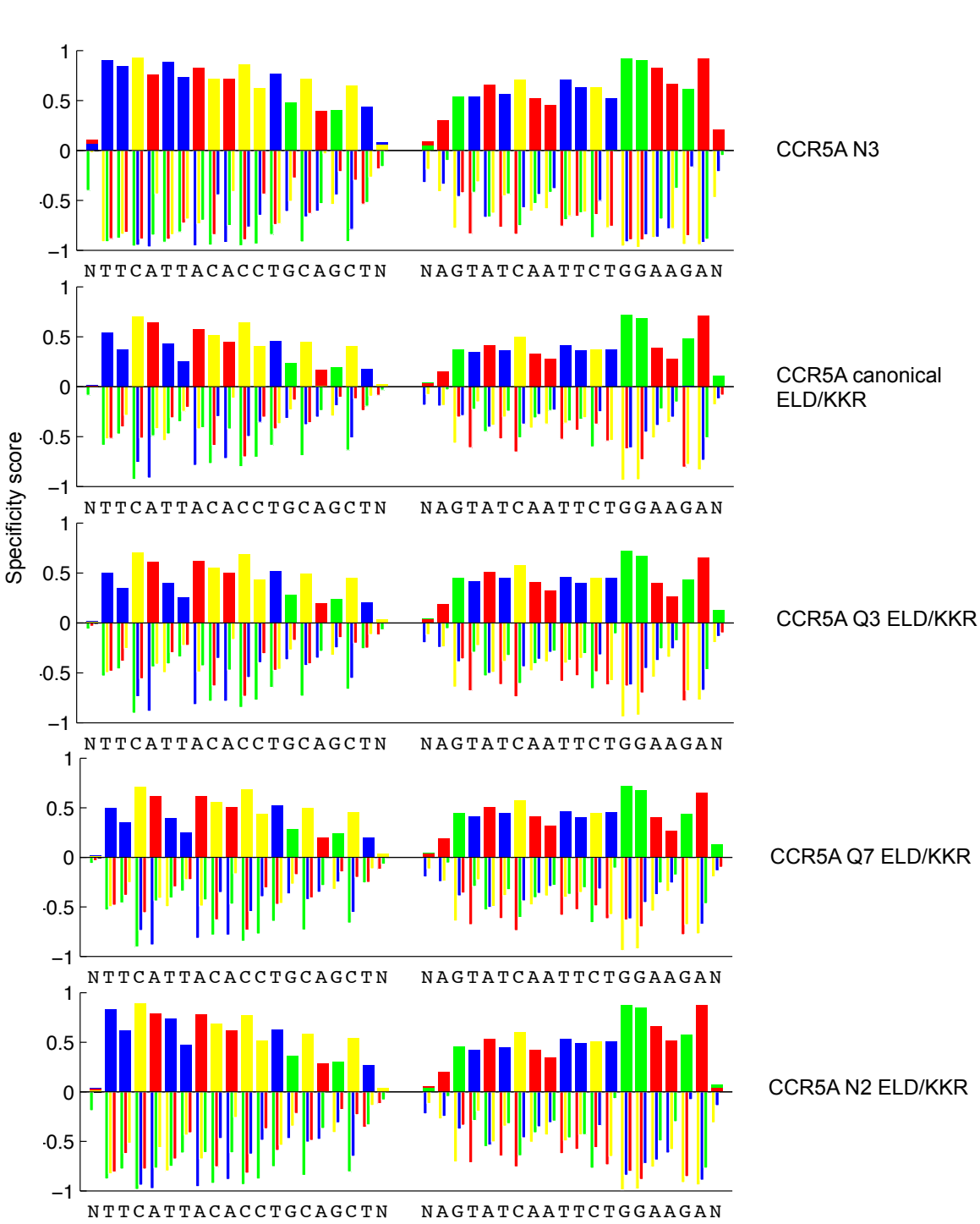
Supplementary Figure S3. Specificity Profiles from All CCR5A TALEN Selections as Heat Maps. Specificity scores for every targeted base pair in selections of CCR5A TALENs are shown. Specificity scores for the L18+R18 CCR5A TALEN at all positions in the target half-sites plus a single flanking position. The colors range from dark blue (maximum specificity score of 1.0) to white (score of 0, no specificity) to dark red (maximum negative score of -1.0); see the main text for details. Boxed bases represent the intended target base. The titles to the right indicate if the TALEN used in the selection differs from the canonical TALEN architecture, which contains a canonical C-terminal domain, wild-type N-terminal domain, and EL/KK *Fok*I variant. Selections correspond to conditions listed in Supplementary Table S1. (A) Specificity profiles of canonical, Q3, Q7, 28-aa, 32 nM canonical, 8 nM canonical, 4 nM canonical, 32 nM Q7 and 8 nM Q7 CCR5A TALEN selections. (B) Specificity profiles of 4 nM Q7, N1, N2, N3, canonical ELD/KKR, Q3 ELD/KKR, Q7 ELD/KKR and N2 ELD/KKR CCR5A TALEN selections. When not specified, TALEN concentration was 16 nM.

A


Specificity score
1
0.5
0
-0.5
−1
NTTCATTACACCTGCAGCTN NAGTATCAATTCTGGAAGAN
CCR5A canonical
A
C
G
T
CCR5A Q3
CCR5A Q7
CCR5A 28-aa
CCR5A 32 nM canonical
CCR5A 8 nM canonical

B

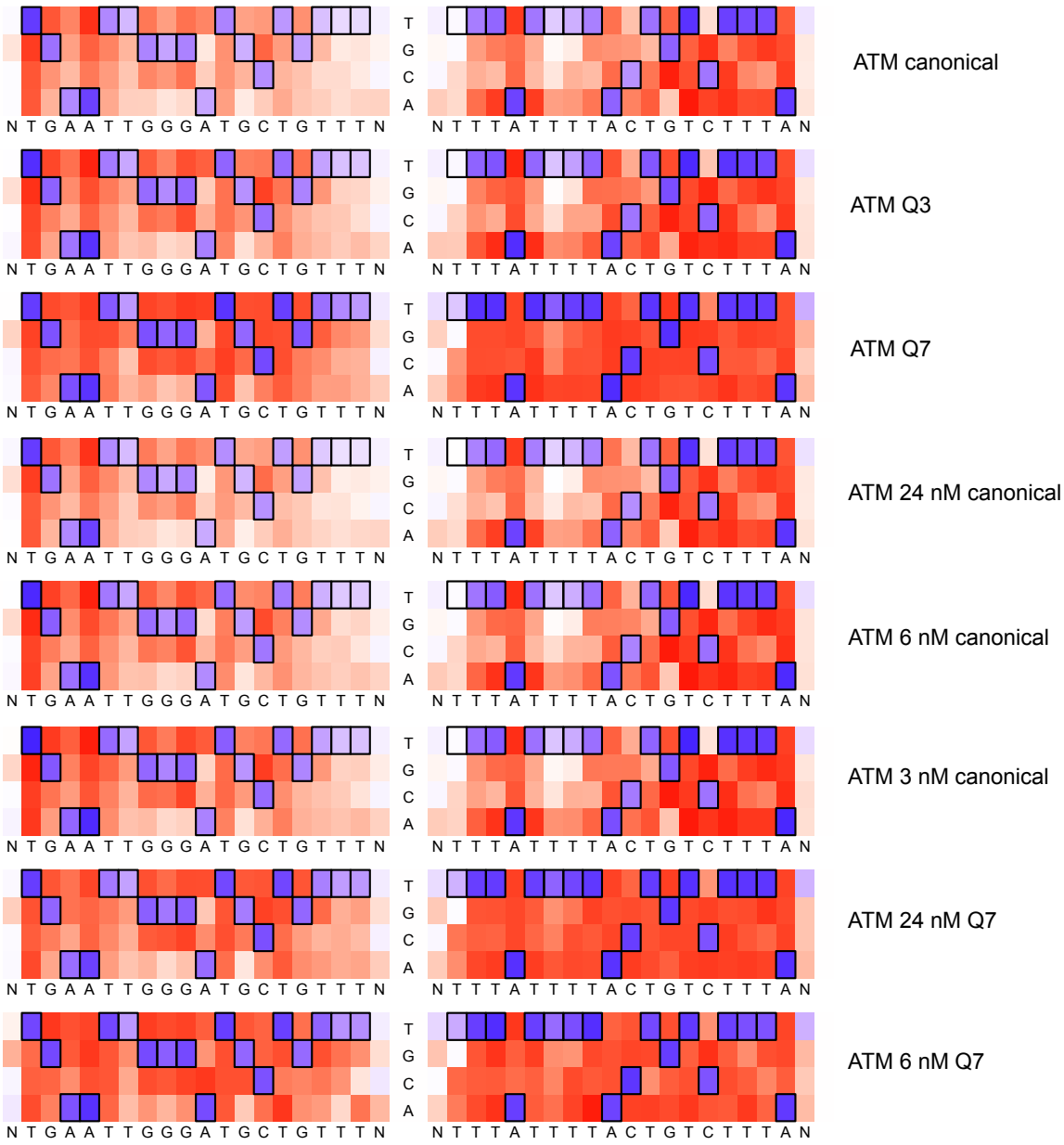

Specificity score
1
0.5
0
-0.5
−1
NTTCATTACACCTGCAGCTN NAGTATCAATTCTGGAAGAN
CCR5A 4 nM canonical
CCR5A 32 nM Q7
CCR5A 8 nM Q7
CCR5A 4 nM Q7
CCR5A N1
CCR5A N2



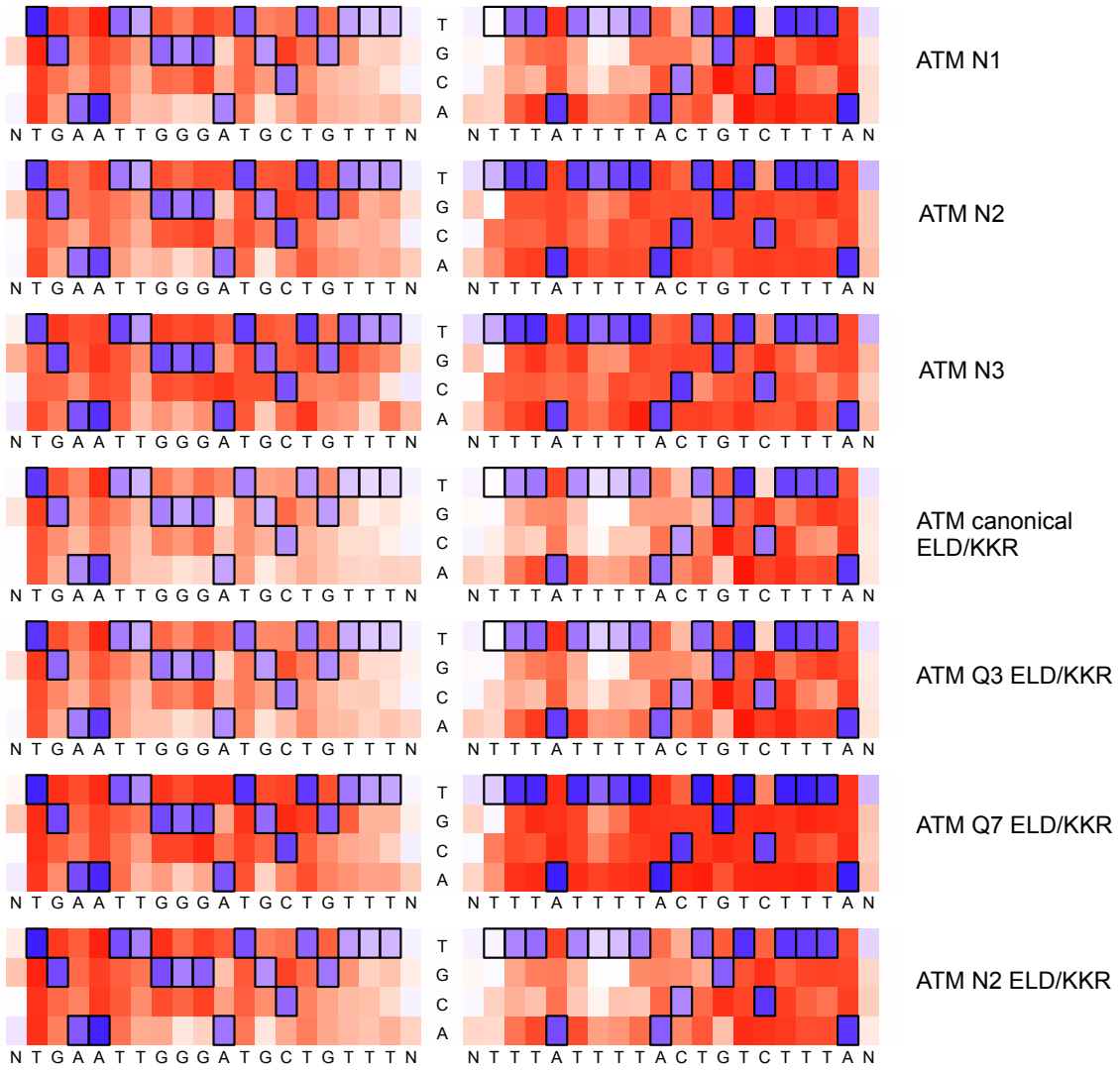
Supplementary Figure S4. Specificity Profiles from All CCR5A TALEN Selections as Bar Graphs. Specificity scores for every targeted base pair in selections of CCR5A TALENs are shown. Positive specificity scores, up to complete specificity at a specificity score of 1.0, signify enrichment of that base pair over the other possibilities at that position. Negative specificity scores, down to complete anti-specificity of –1.0, represents enrichment against that base pair. Specified positions (specificity score > 0) were plotted as stacked bars above the axis (multiple specified base pairs at the same position were plotted over each other with the shortest bar in front, and not end-to-end) while anti-specified base pairs were plotted as narrow, grouped bars. The titles to the right indicate if the TALEN used in the selection differs from the canonical TALEN architecture, which contains a canonical C-terminal

domain, wild-type N-terminal domain, and EL/KK *Fok*I variant. Selections correspond to conditions listed in Supplementary Table S1. (A) Specificity profiles of canonical, Q3, Q7, 28-aa, 32 nM canonical, and 8 nM canonical CCR5A TALEN selections. (B) Specificity profiles of 4 nM canonical, 32 nM Q7, 8 nM Q7, 4 nM Q7, N1, and N2 CCR5A TALEN selections. (C) Specificity profiles of N3, canonical ELD/KKR, Q3 ELD/KKR, Q7 ELD/KKR, and N2 ELD/KKR CCR5A TALEN selections. When not specified, TALEN concentration was 16 nM.

A

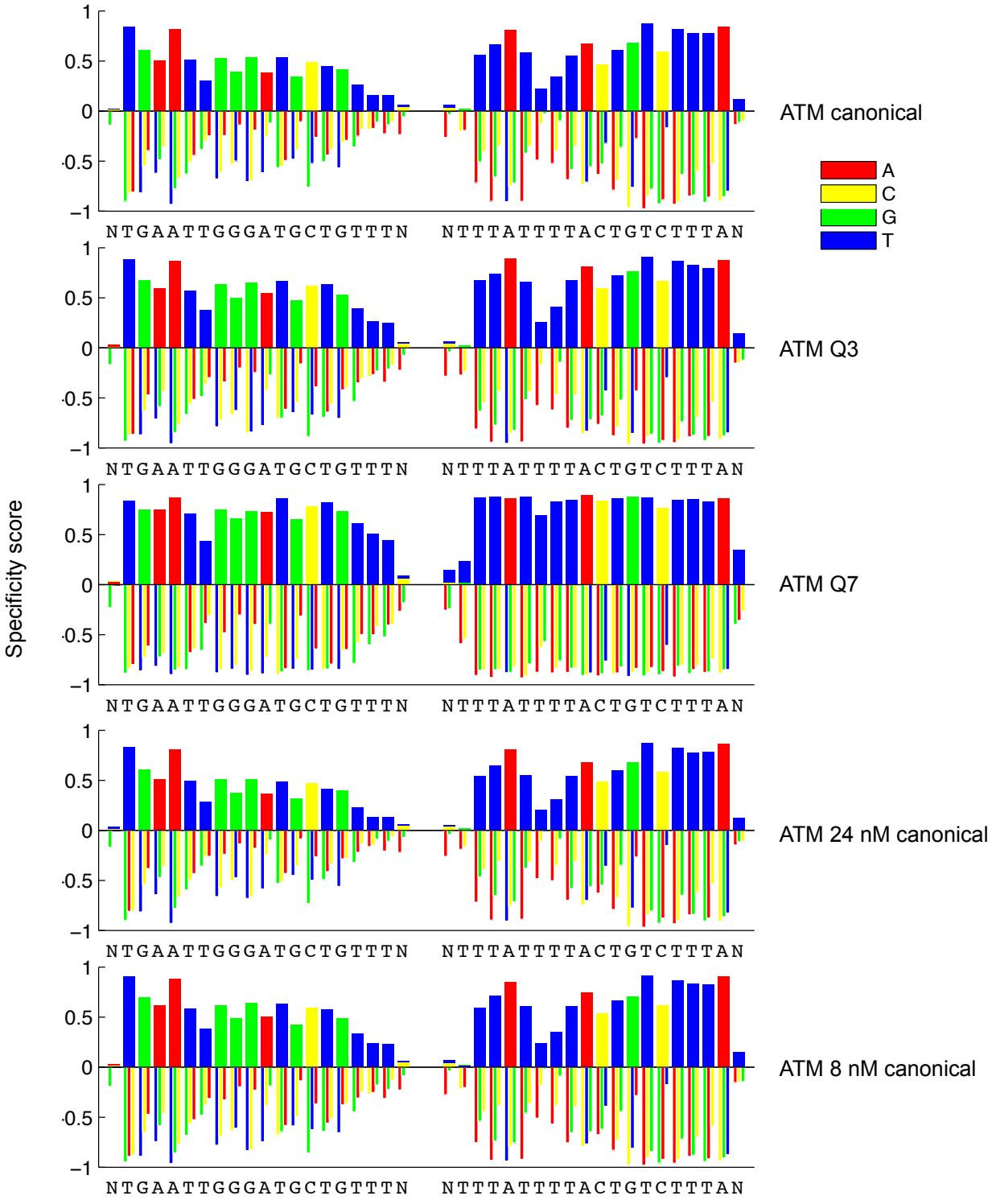

T
G
C
A
N T G A A T T G G G A T G C T G T T T N
N T T T A T T T T T A C T G T C T T T A N
ATM canonical
ATM Q3
ATM Q7
ATM 24 nM canonical
ATM 6 nM canonical
ATM 3 nM canonical
ATM 24 nM Q7
ATM 6 nM Q7

B

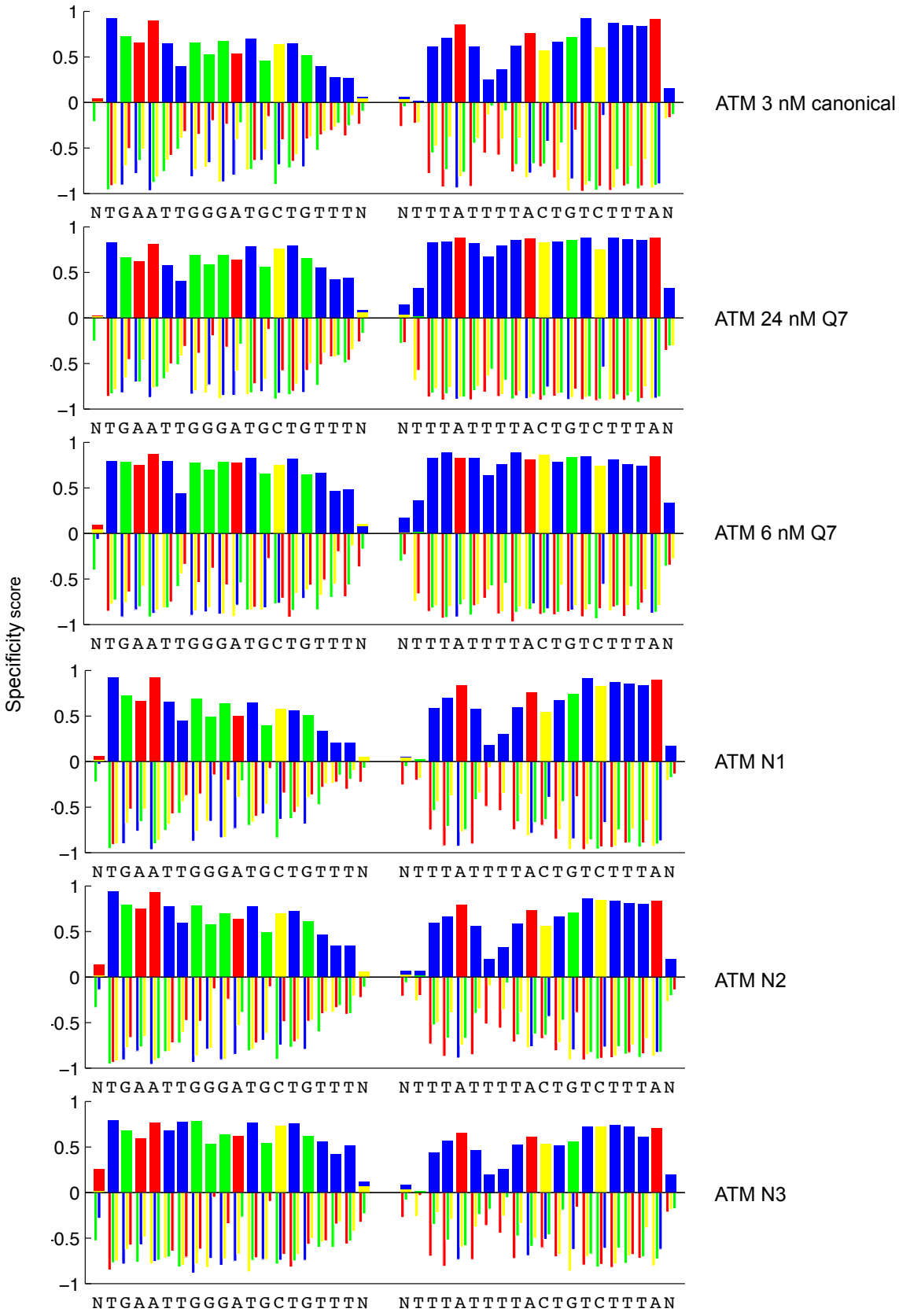


Supplementary Figure S5. Specificity Profiles from All ATM TALEN Selections as Heat Maps. Specificity scores for every targeted base pair in selections of ATM TALENs are shown. Specificity scores for the L18+R18 ATM TALEN at all positions in the target half-sites plus a single flanking position. The colors range from dark blue (maximum specificity score of 1.0) to white (score of 0, no specificity) to dark red (maximum negative score of -1.0); see the main text for details. Boxed bases represent the intended target base. The titles to the right indicate if the TALEN used in the selection differs from the canonical TALEN architecture, which contains a canonical C-terminal domain, wild-type N-terminal domain, and EL/KK *Fok*I variant. Selections correspond to conditions listed in Supplementary Table S1. (A) Specificity profiles of (12 nM) canonical, Q3, (12 nM) Q7, 24 nM canonical, 6 nM canonical, 3 nM canonical, 24 nM Q7, and 6 nM Q7 ATM TALEN selections. (B) Specificity profiles of N1, N2, N3, canonical ELD/KKR, Q3 ELD/KKR, Q7 ELD/KKR, and N2 ELD/KKR ATM TALEN selections. When not specified, TALEN concentration was 12 nM.

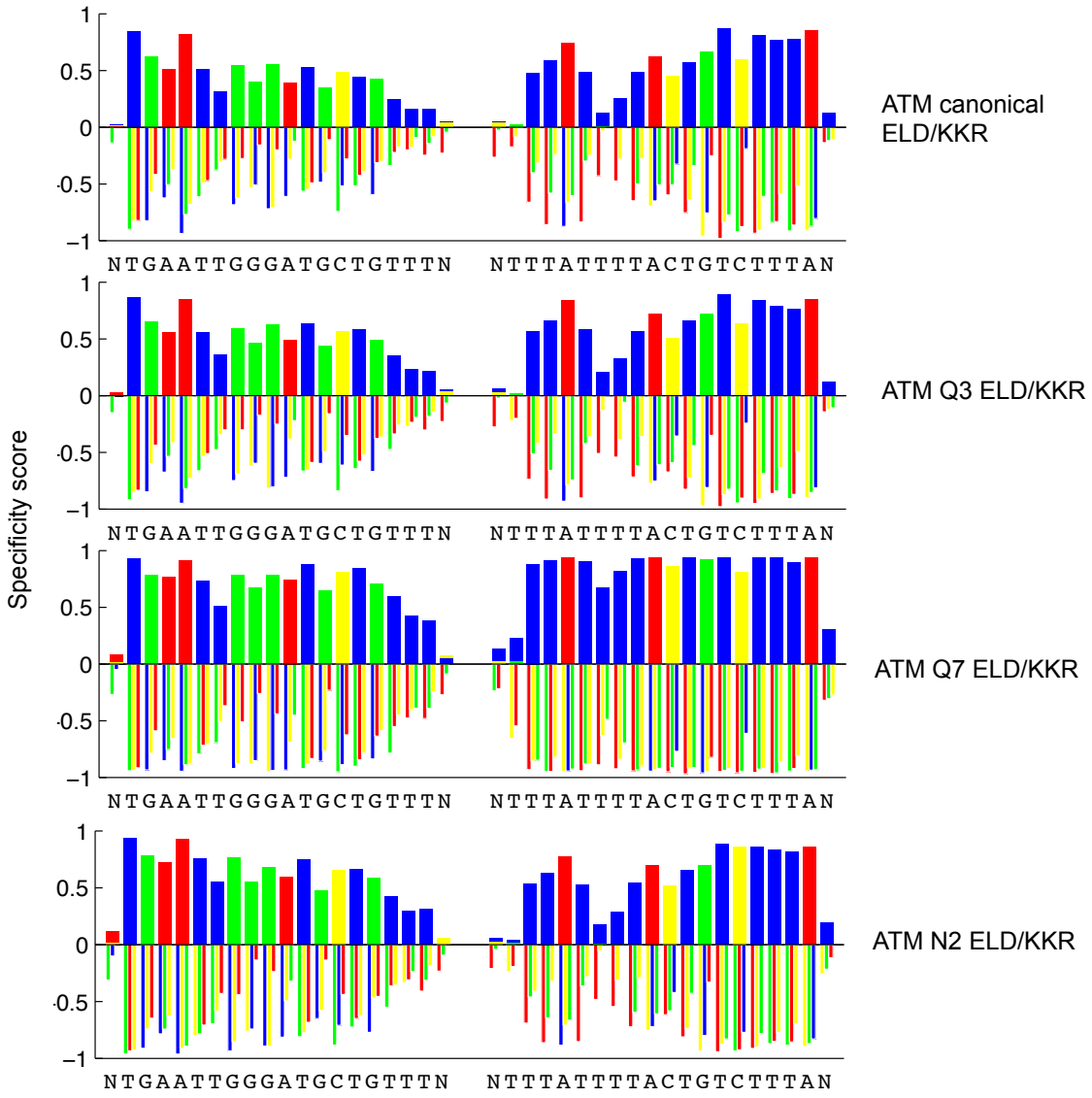
A


Specificity score
1
0.5
0
-0.5
−1
NTGAATTGGGATGCTGTTTN NTTTATTTTACTGTCTTTAN
ATM canonical
A
C
G
T
ATM Q3
ATM Q7
ATM 24 nM canonical
ATM 8 nM canonical

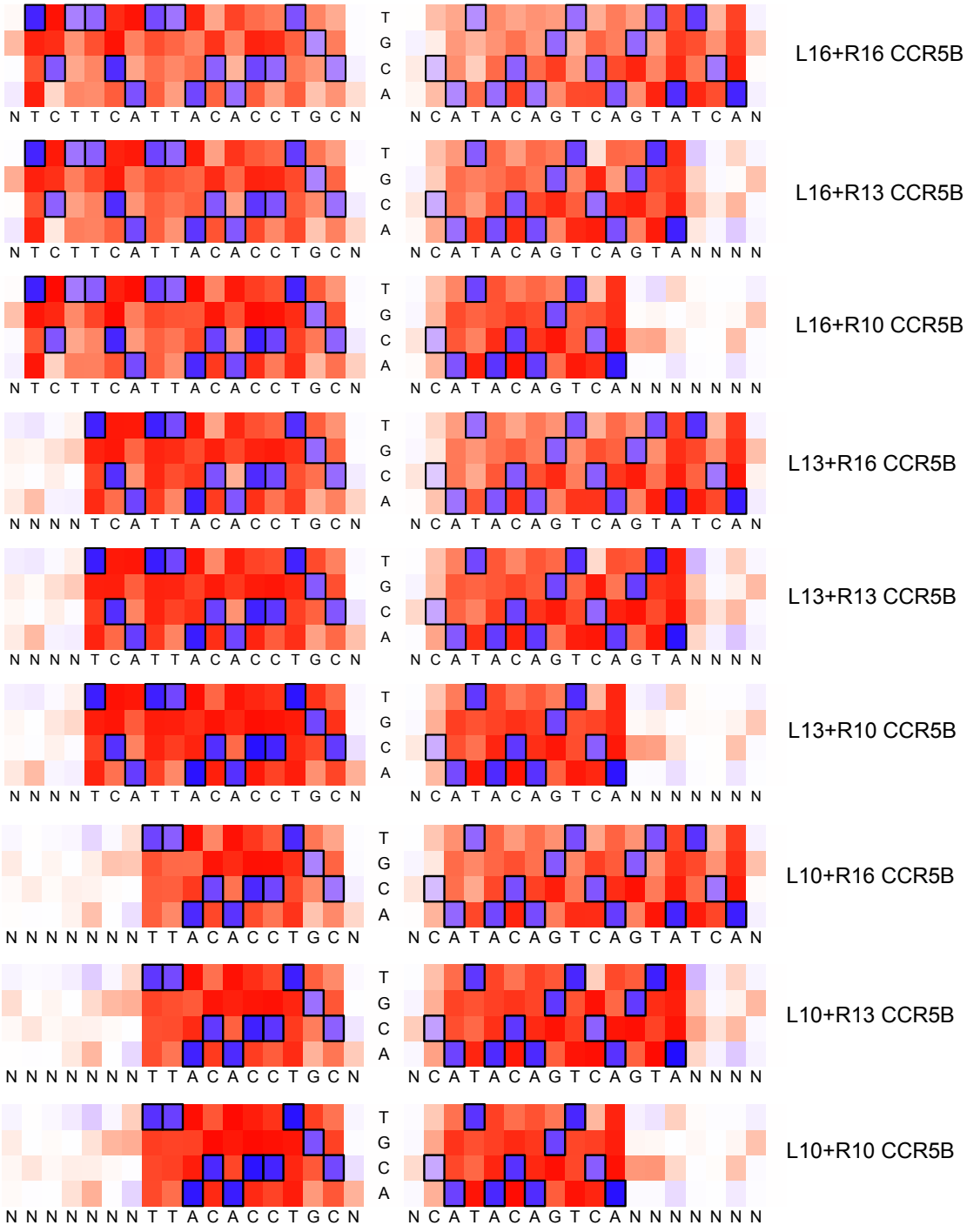
B


Specificity score
1
0.5
0
-0.5
−1
NTGAATTGGGATGCTGTTTN
NTTTATTTTACTGTCTTTAN
ATM 3 nM canonical
ATM 24 nM Q7
ATM 6 nM Q7
ATM N1
ATM N2
ATM N3

C

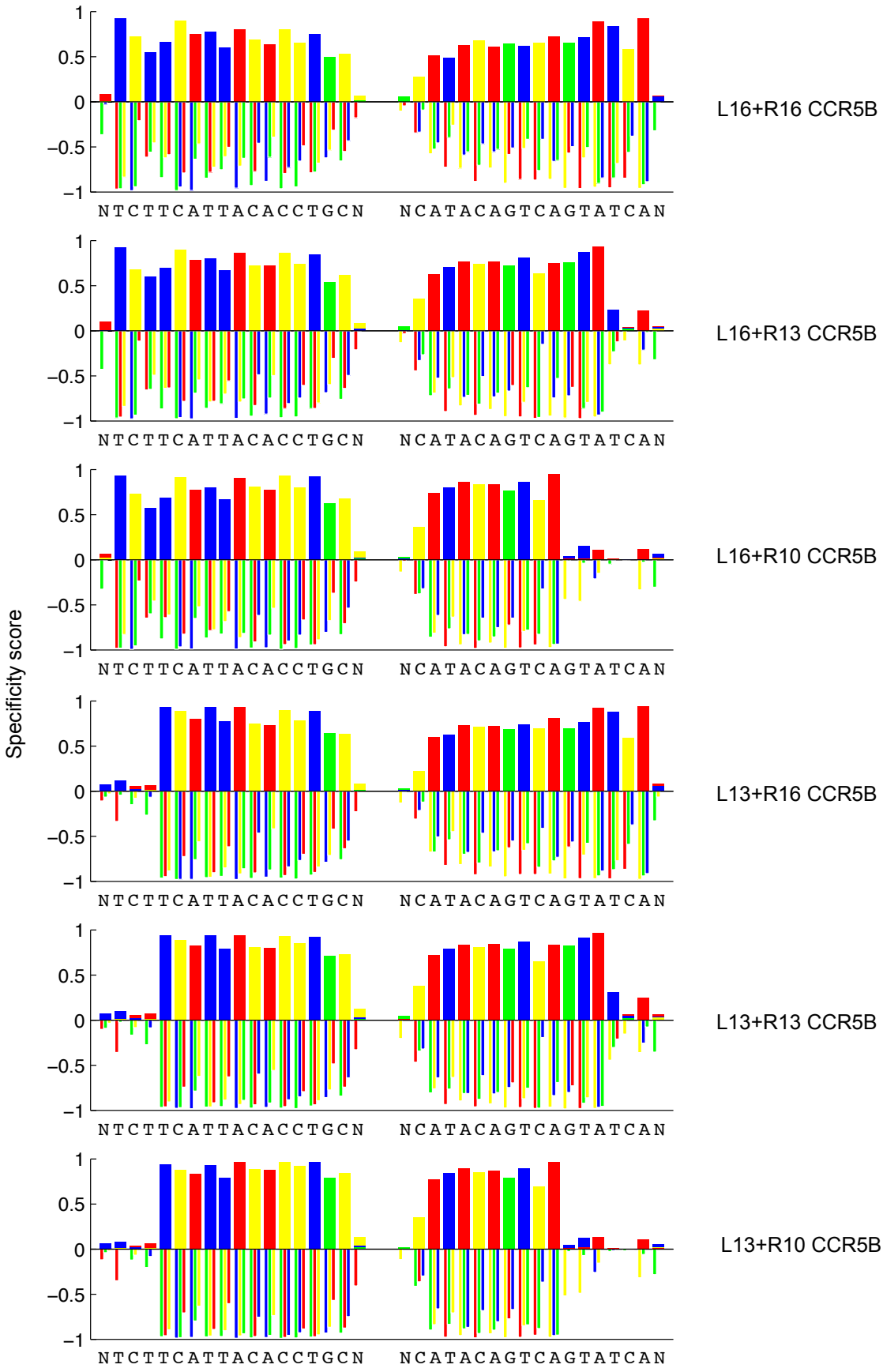


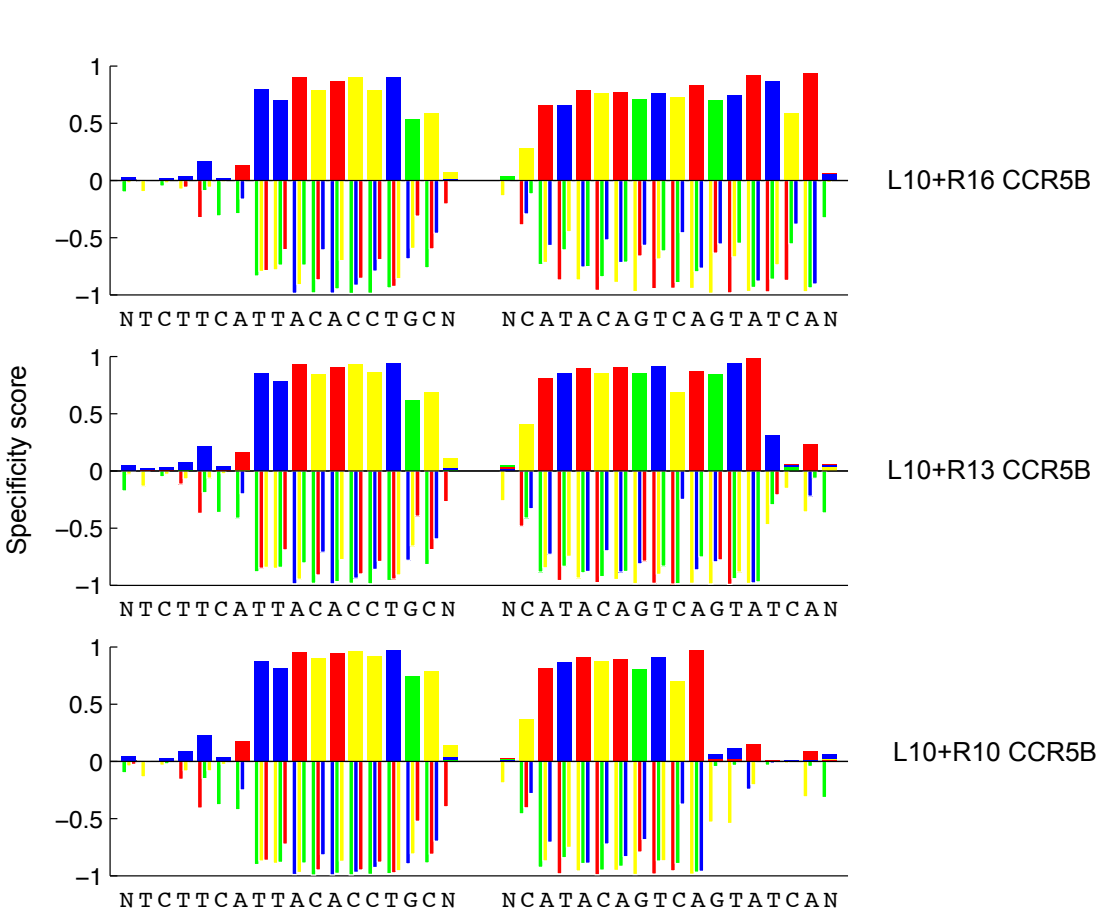
Supplementary Figure S6. Specificity Profiles from All ATM TALEN Selections as Bar Graphs. Specificity scores for every targeted base pair in selections of ATM TALENs are shown. Positive specificity scores, up to complete specificity at a specificity score of 1.0, signify enrichment of that base pair over the other possibilities at that position. Negative specificity scores, down to complete anti-specificity of –1.0, represents enrichment against that base pair. Specified positions (specificity score > 0) were plotted as stacked bars above the axis (multiple specified base pairs at the same position were plotted over each other with the shortest bar in front, and not end-to-end) while anti-specified base pairs were plotted as narrow, grouped bars. The titles to the right indicate if the TALEN used in the selection differs from the canonical TALEN architecture, which contains a canonical C-terminal domain, wild-type N-terminal domain, and EL/KK *Fok*I variant. Selections correspond to conditions listed in Supplementary Table S1. (A) Specificity profiles of canonical, Q3, Q7, 32 nM canonical, and 8 nM canonical ATM TALEN selections. (B) Specificity profiles of 3 nM canonical, 24 nM Q7, 6 nM Q7, N1, N2, and N3 ATM TALEN selections. (C) Specificity profiles of canonical ELD/KKR, Q3 ELD/KKR, Q7 ELD/KKR, and N2 ELD/KKR ATM TALEN selections. When not specified, TALEN concentration was 12 nM.



Supplementary Figure S7. Specificity Profiles from All CCR5B TALEN Selections as Heat Maps. Specificity scores for every targeted base pair in selections of CCR5B TALENs are shown. Specificity scores for CCR5B TALENs targeting all possible combinations of the left (L10, L13, L16) and right (R10, R13, R16) half-sites at all positions in the target half-sites plus a single flanking position. The colors range from dark blue (maximum specificity score of 1.0) to white (score of 0, no specificity) to dark red (maximum negative score of -1.0); see the main text for details. Boxed bases represent the intended target base. The titles to the right notes the targeted left (L) and right (R) target half-sites for the CCR5B TALEN used in the selection. Selections correspond to conditions listed in Supplementary Table S1.

A


Specificity score
1
0.5
0
−0.5
−1
NTCTTCATTACACCTGCN NCATACAGTCAGTATCAN
L16+R16 CCR5B
L16+R13 CCR5B
L16+R10 CCR5B
L13+R16 CCR5B
L13+R13 CCR5B
L13+R10 CCR5B



Supplementary Figure S8. Specificity Profiles from All CCR5B TALEN Selections as Bar Graphs. Specificity scores for every targeted base pair in selections of CCR5B TALENs are shown. Positive specificity scores, up to complete specificity at a specificity score of 1.0, signify enrichment of that base pair over the other possibilities at that position. Negative specificity scores, down to complete anti-specificity of –1.0, represents enrichment against that base pair. Specified positions (specificity score > 0) were plotted as stacked bars above the axis (multiple specified base pairs at the same position were plotted over each other with the shortest bar in front, and not end-to-end) while anti-specified base pairs were plotted as narrow, grouped bars. The titles to the right notes the targeted left (L) and right (R) target half-sites for the CCR5B TALEN used in the selection. Selections correspond to conditions listed in Supplementary Table S1.

OnCCR5A

TTCATTACACCTGCAGCTCTCATTTTCCATACAGTCAGTATCAATTCTGGAAGA (5263)
TTCATTACACCTGCAGCTCTCAT-------ACAGTCAGTATCAATTCTGGAAGA (76)
TTCATTACACCTGCAG------------------TCAGTATCAATTCTGGAAGA (63)
TTCATTACACCTG-----------------------------------GAAGA (61)

OffC-5

TCCAATACCTCTGCCACACCCAGGCATTGGCCAGGAGCAACTCTGGGAGA (16660)
TCCAATACCTCTGCCACAC-----------CCAGGAGCAACTCTGGGAGA (28)
TCCAATACCTCTG----------GCATTGGCCAGGAGCAACTCTGGGAGA (12)
TCCAATAC--------------------------------CTCTGGGAGA (10)

OffC-15

TCCATGACACAAAAGACTTCCCTGATTTCTTCTAAGGCATCACTGGTATCTATCCTGGAATA (6951)
TCCATGACACAAAAGACTTCCCTGATTTCTTCTAAGG-----CTGGTATCTATCCTGGAATA (6)

OffC-16

TTCCTTCCACCAGTGTCCACAGTCTTCACACTGATCACCAAATCCCAGCATCAATCCTGGAAGA (38524)
TTCCTTCCACCAGTGTCCACAGTC-----------CACCAAATCCCAGCATCAATCCTGGAAGA (4)

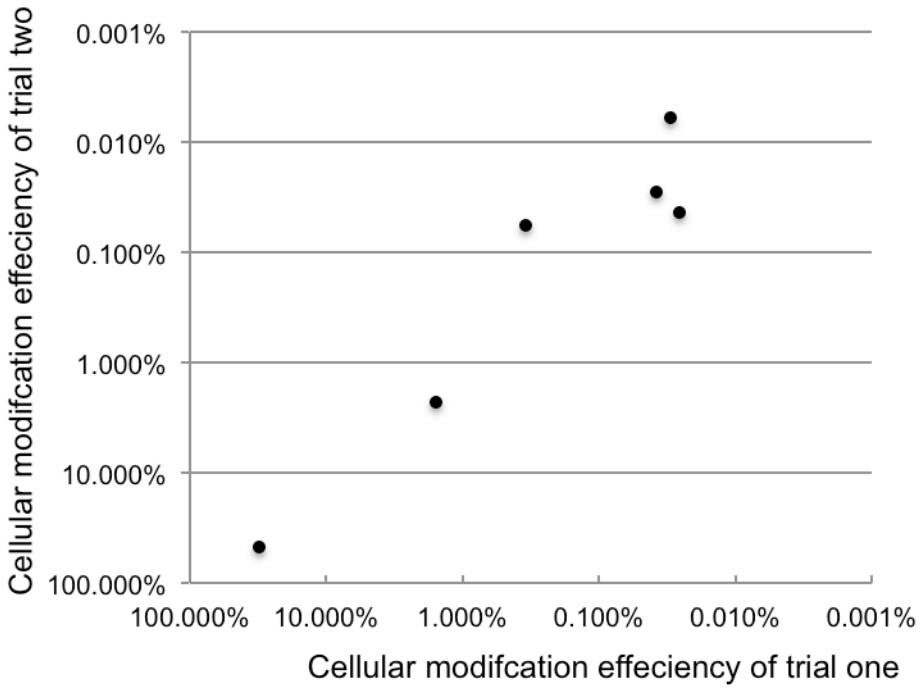
OffC-28

TTTATTACACTTCCAGATCTTTTATTTTAAGTTACCAGATATCCTTTCTGGAAGA (7367)
TTTATTACACTT----------------------CCAGATATCCTTTCTGGAAGA (3)
TTTATTACACTTCCAGATCTTTT---------------ATATCCTTTCTGGAAGA (2)
TTTATTACACTTCCAGATCTTT-----------------TATCCTTTCTGGAAGA (2)

OffC-36

CTCCTAATACCTGCAAATTATAAGGACACTATTTGACTTGATATTATTTCTGGAGGA (12442)
CTCCTAATACCTGCAAATTATAAGGACACT----GACTTGATATTATTTCTGGAGGA (11)

Supplementary Figure S9. Modifications induced by TALENs at on-target and predicted off-target genomic sites. Examples of modified sequences at the on-target site and off-target sites for cells treated with CCR5A TALENs containing the ELD/KKR *Fok*I domains. For each example shown, the unmodified genomic site is the first sequence, followed by the top three sequences containing deletions. The numbers in parentheses indicate sequencing counts and the half-sites are underlined and bolded.

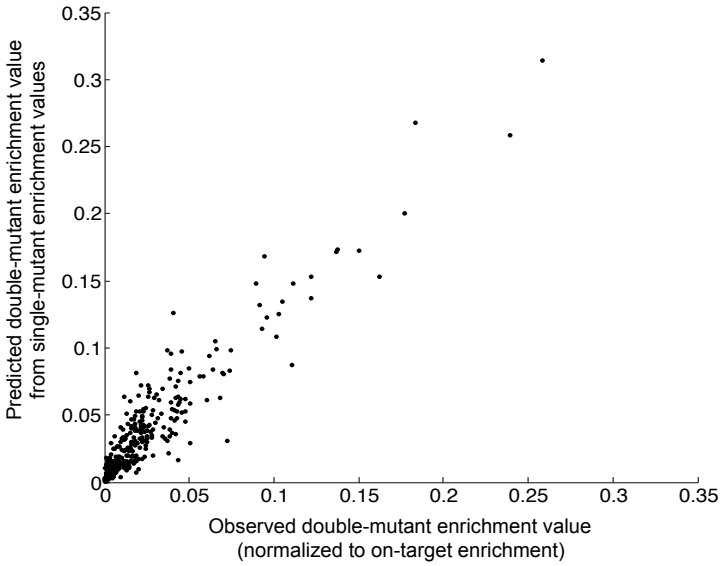


Supplementary Figure S10. Correspondence of Modification Induced by CCR5A TALENs at Genomic Sites Between Independent Biological Replicates. The CCR5A TALEN-induced cellular modification efficiencies of OnCCR5A, OffC5, OffC15, OffC16, OffC28 and OffC36 genomic sites from two independent biological replicates are plotted in the graph. The modification efficiency was determined from DNA sequencing CCR5A on-target and each genomic off-target site amplified from genomic DNA isolated from human cells treated with CCR5A TALENs containing canonical C-terminal domains and ELD/KKR heterodimeric *Fok*I domains. Trial one data is shown in Supplementary Table S9C and trial two data is shown in Supplementary Table S7A.

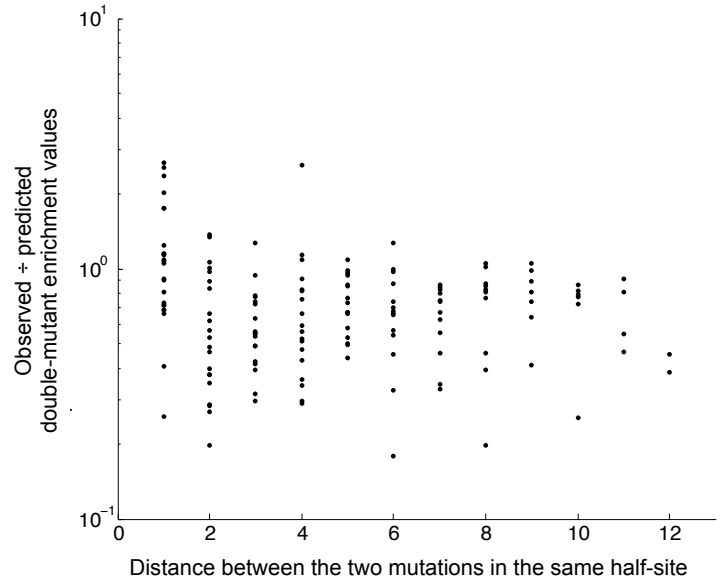
A

Sequence	Observed enrichment value from selection	Predicted enrichment value assuming independence
TCATaACACCTGC	0.46	
TCATTACACCTGt	0.47	
TCATaACACCTGt	0.25	0.46 x 0.47 = 0.22

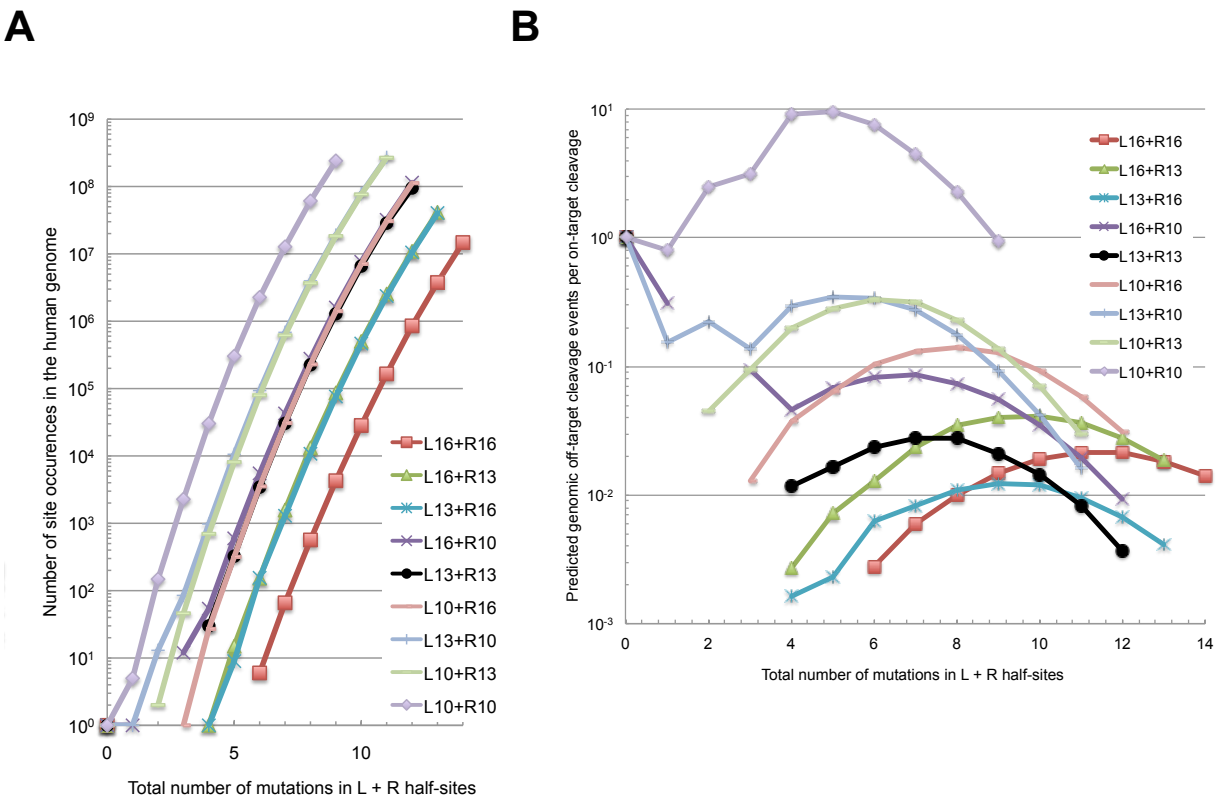
B



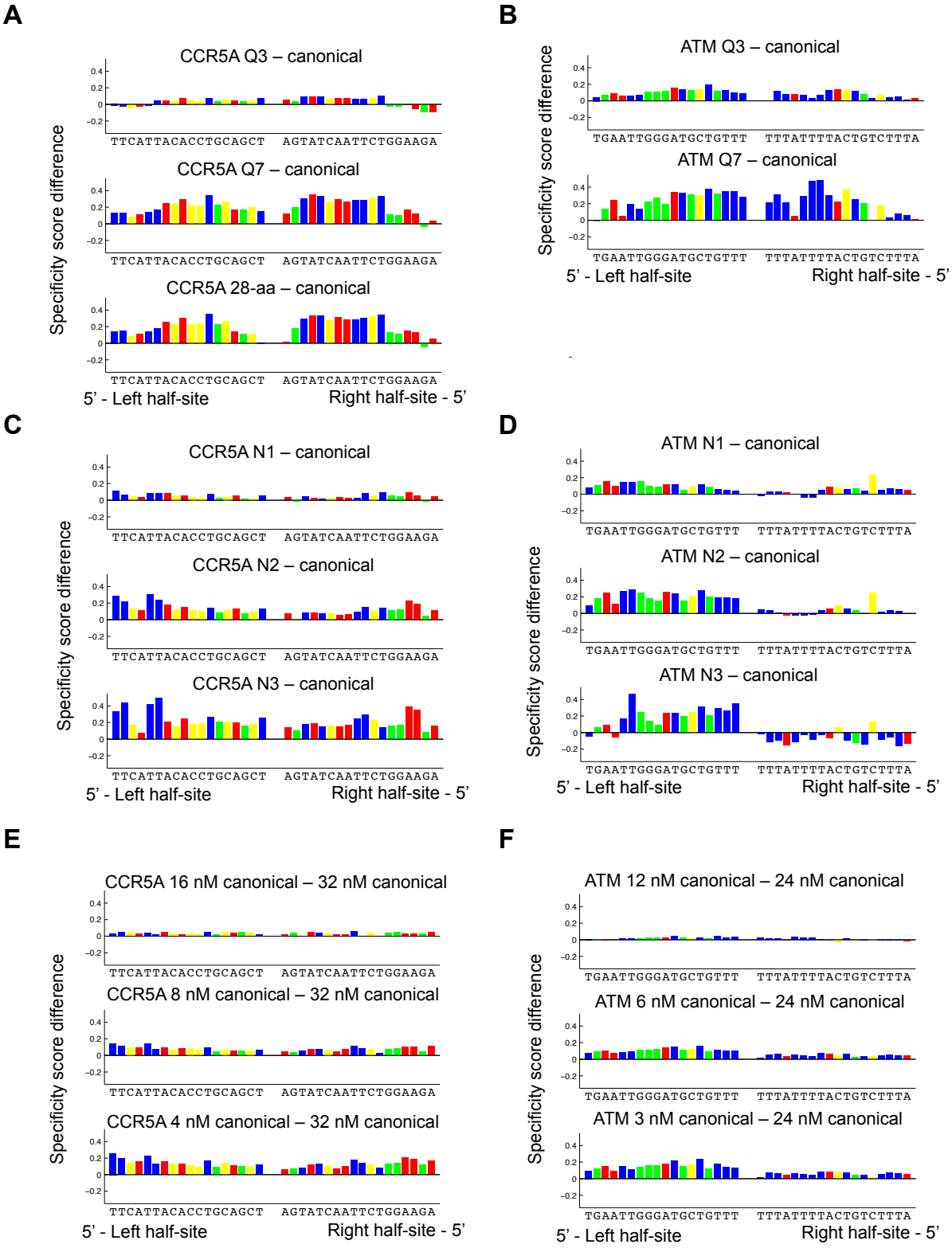
C



Supplementary Figure S11. Observed Versus Predicted Double-Mutant Sequence Enrichment Values. (A) An example calculation of a predicted double-mutant enrichment value from the L13+R13 CCR5A TALEN selection. All enrichment values were normalized to the on-target enrichment value (= 1.0 by definition). Observed enrichment values of each single-mutant are shown and were multiplied together to calculate the predicted enrichment value of the corresponding double-mutant sequence. This predicted enrichment value can then be compared to the observed enrichment value for the same double-mutant sequence. (B) For the L13+R13 CCR5A TALEN selection, the observed double-mutant enrichment values of individual sequences (post-selection sequence abundance ÷ pre-selection sequence abundance) were normalized to the on-target enrichment value (= 1.0 by definition) and plotted against the corresponding predicted double-mutant enrichment values calculated by multiplying the enrichment value of the component single-mutants normalized to the on-target enrichment. The predicted double-mutant enrichment values assume independent contributions from each single mutation to the double-mutant's enrichment value. (C) The observed double-mutant sequence enrichment divided by the predicted double-mutant sequence enrichment plotted as a function of the distance (in base pairs) between the two mutations. Only sequences with two mutations in the same half-site were considered.



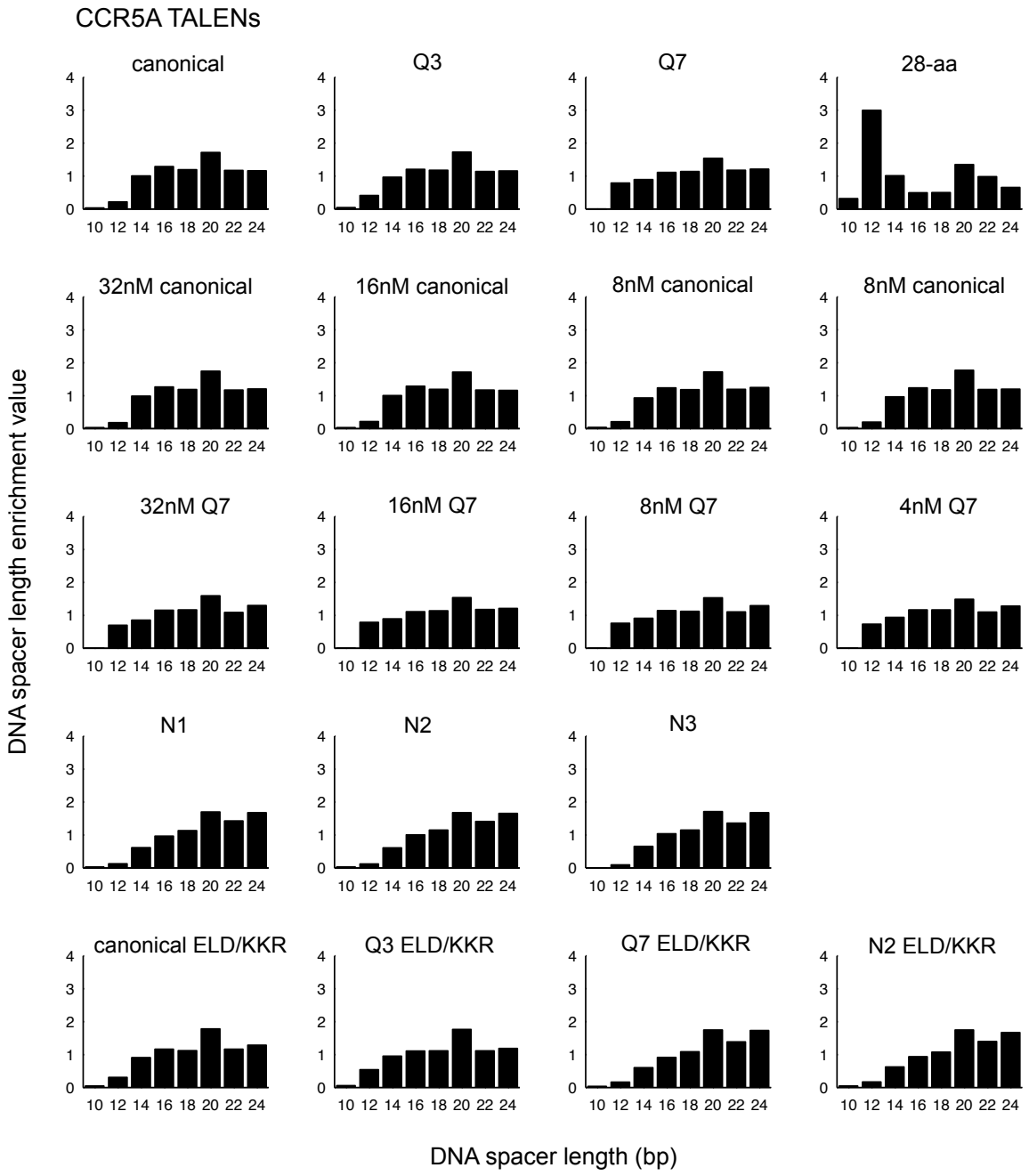
Supplementary Figure S12. Predicted off-target genomic cleavage as a function of TALEN length considering both TALEN specificity and off-target site abundance in the human genome. (A) Number of sites in the human genome related to each of the nine CCR5B on-target sequences (L10, L13, or L16 combined with R10, R13, or R16), allowing for a spacer length from 12 to 25 bps between the two half-sites (Supplementary Algorithm). The TALENs targeted DNA sites of 32 bp (L16+R16), 29 bp (L16+R13 or L13+R16), 26 bp (L16+R10 or L13+R13 or L10+R16), 23 bp (L13+R10 or L10+R13) or 20 bp (L10+R10) in length (B) For all nine CCR5B TALENs, overall genomic off-target cleavage frequency was predicted by multiplying the number of sites in the human genome containing a certain number of mutations by the enrichment value of off-target sequences containing that same number of mutations shown in (Fig. 4). Because enrichment values level off at high mutation numbers likely due to the limit of sensitivity of the selection, it was necessary to extrapolate high-mutation enrichment values by fitting low-mutation enrichment values as function of mutation number (Supplementary Table S11). The overall predicted genomic cleavage was calculated only for mutation numbers with sites observed to occur more than once in the human genome. For L16+R10 there are no genomic sequences with two mutations, causing the break in the corresponding line.

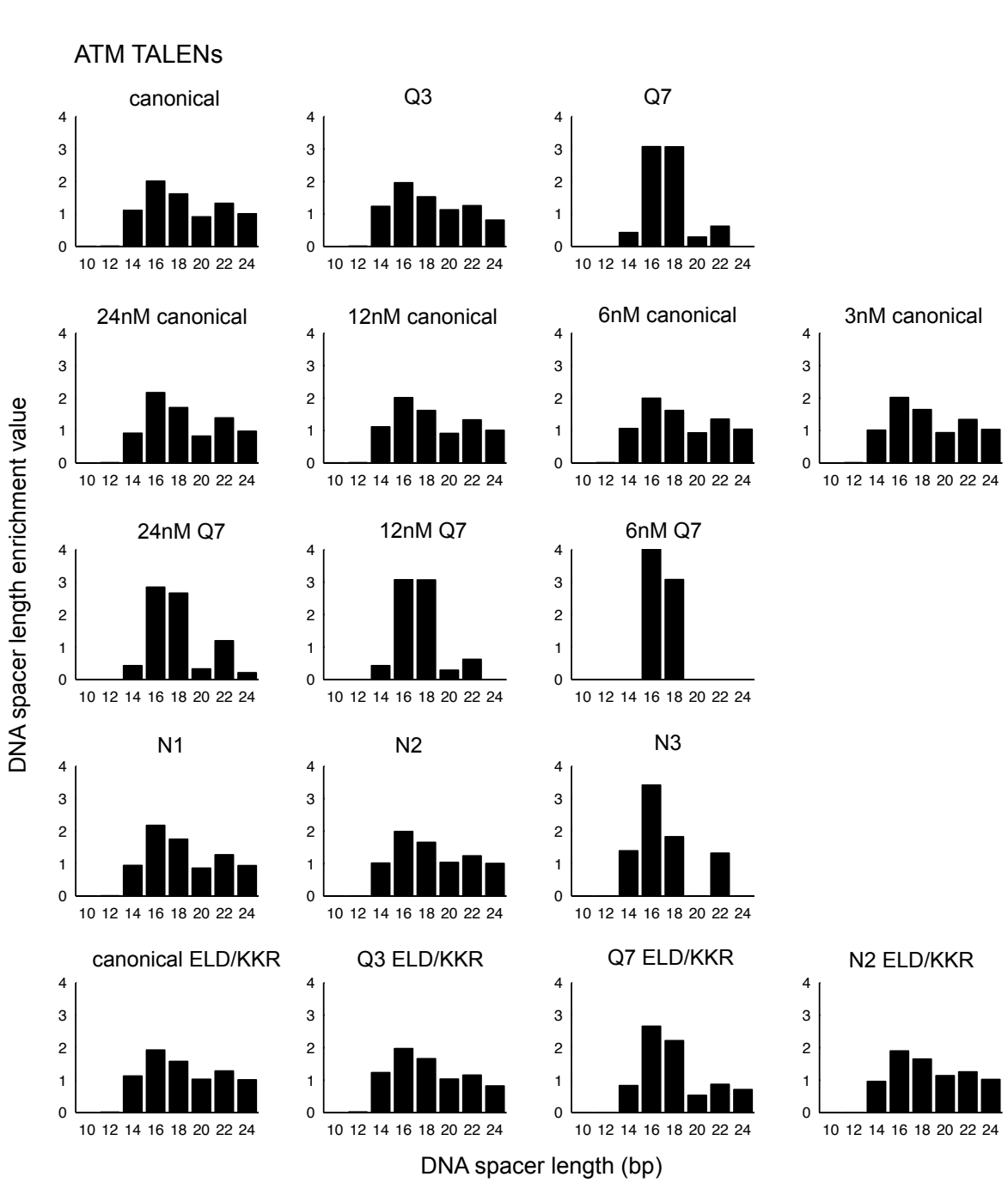


Supplementary Figure S13. Effects of Engineered TALEN Domains and TALEN Concentration on Specificity. (A) The specificity score of the targeted base pair at each position of the CCR5A site was calculated for CCR5A TALENs containing the canonical, Q3, Q7, or 28-aa C-terminal domains. The specificity scores of the Q3, Q7, or 28-aa C-terminal domain TALENs subtracted by the specificity scores of the TALEN with the canonical C-terminal domain are shown. (B) Same as (A) but for

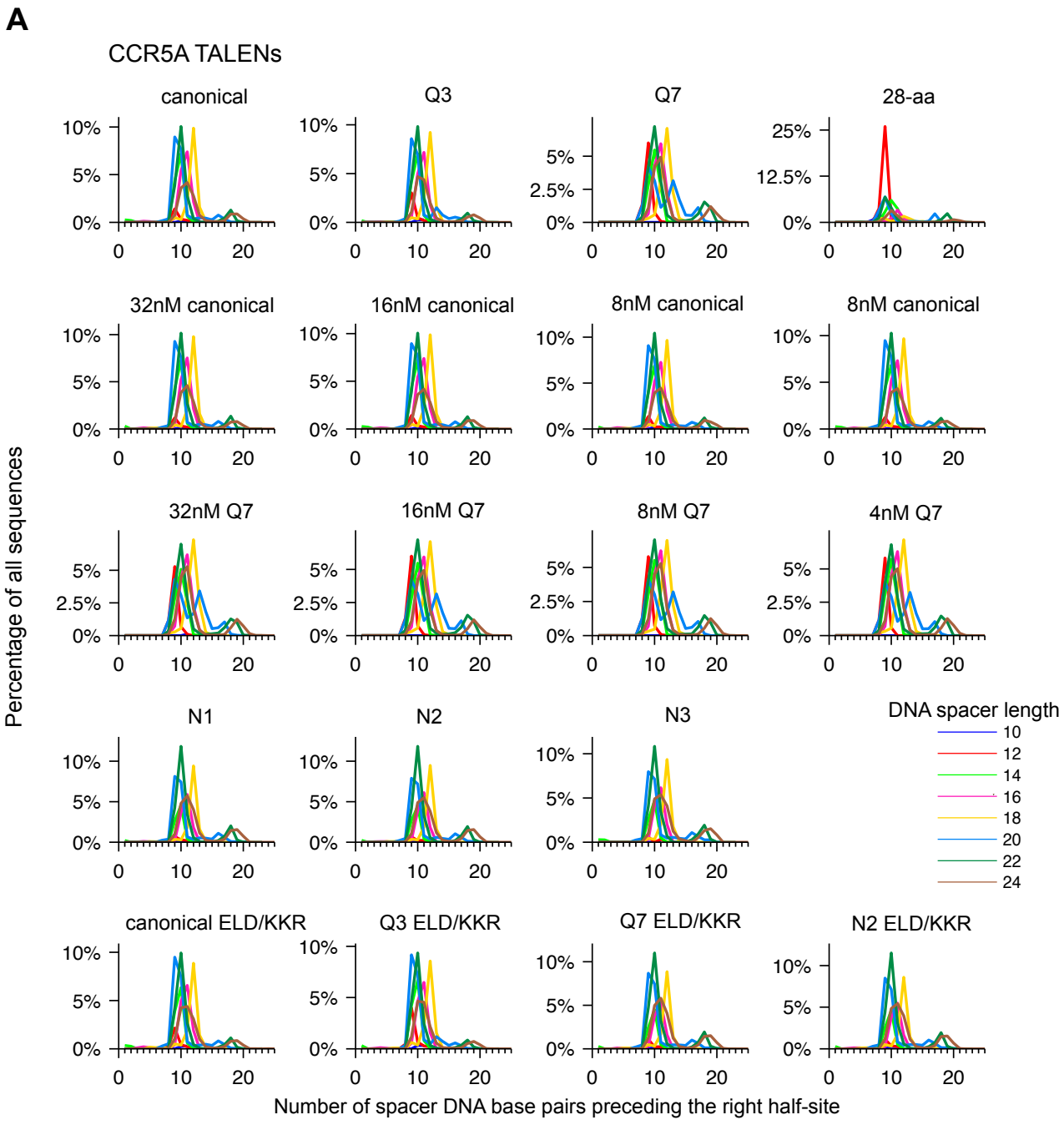
CCR5A TALENs containing engineered N-terminal domains N1, N2, or N3. (C) Same as (A) but comparing specificity scores differences of the canonical CCR5A TALEN assayed at 16 nM, 8 nM or 4 nM subtracted by the specificity scores of canonical CCR5A TALENs assayed at 32 nM. (D-F) Same as (A-C) but for ATM TALENs. Selections correspond to conditions listed in Supplementary Table S1.

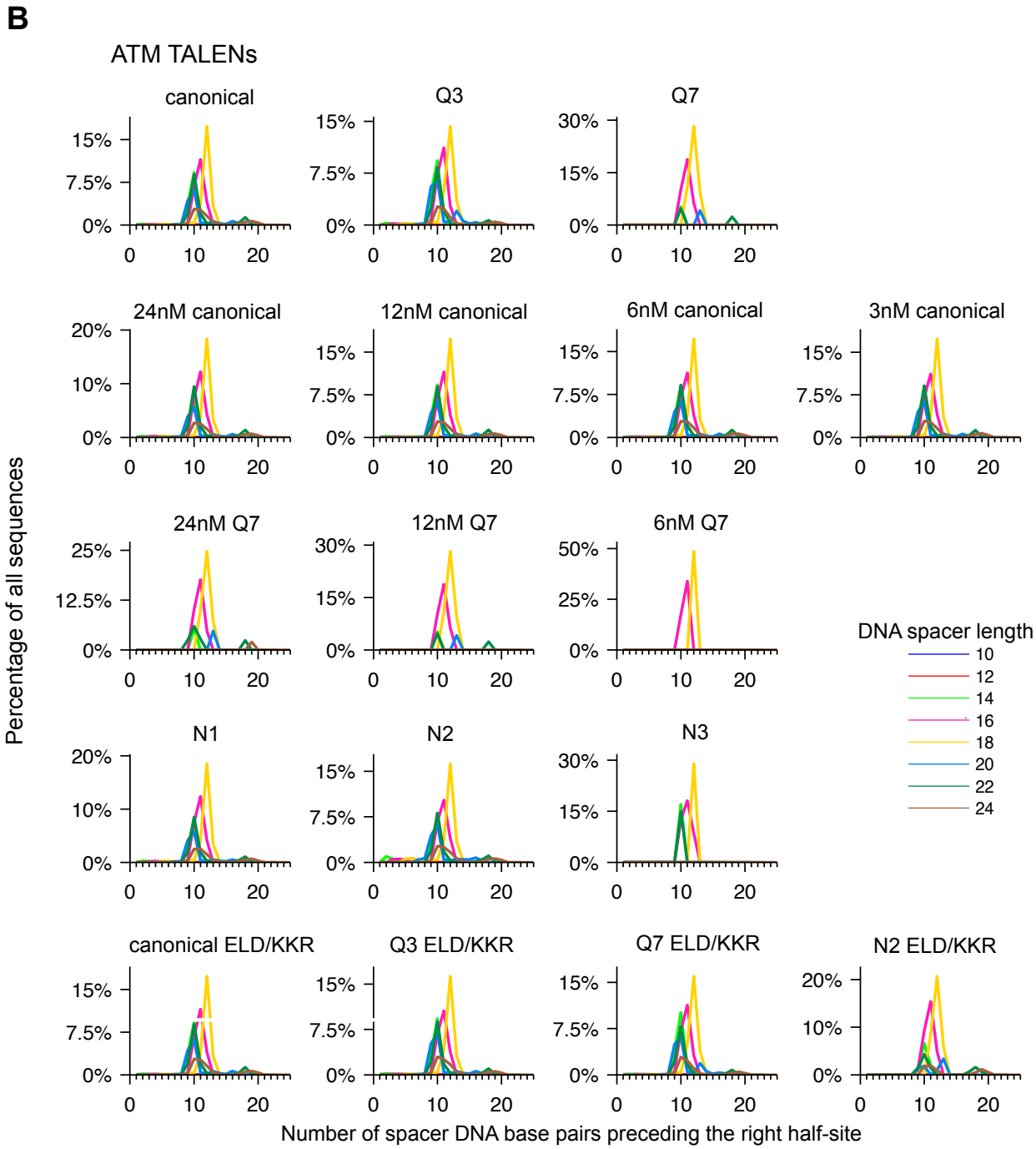
A


CCR5A TALENs
canonical
Q3
Q7
28-aa
32nM canonical
16nM canonical
8nM canonical
8nM canonical
32nM Q7
16nM Q7
8nM Q7
4nM Q7
N1
N2
N3
canonical ELD/KKR
Q3 ELD/KKR
Q7 ELD/KKR
N2 ELD/KKR
DNA spacer length enrichment value
DNA spacer length (bp)



Supplementary Figure S14. Spacer-Length Preferences of TALENs. (A) For each selection with CCR5A TALENs containing various combinations of the canonical, Q3, Q7, or 28-aa C-terminal domains; N1, N2, or N3 N-terminal mutations; and the EL/KK or ELD/KKR *Fok*I variants and at 4, 8, 16, or 32 nM, the DNA spacer-length enrichment values were calculated by dividing the abundance of DNA spacer lengths in post-selection sequences by the abundance of DNA spacer lengths in the pre-selection library sequences. (B) Same as (A) but for ATM TALENs.


A
CCR5A TALENs
canonical
Q3
Q7
28-aa
32nM canonical
16nM canonical
8nM canonical
8nM canonical
32nM Q7
16nM Q7
8nM Q7
4nM Q7
N1
N2
N3
DNA spacer length
10
12
14
16
18
20
22
24
canonical ELD/KKR
Q3 ELD/KKR
Q7 ELD/KKR
N2 ELD/KKR
Percentage of all sequences
Number of spacer DNA base pairs preceding the right half-site



Supplementary Figure S15. DNA Cleavage-Site Preferences of TALENs. (A) For each selection with CCR5A TALENs with various combinations of canonical, Q3, Q7, or 28-aa C-terminal domains; N1, N2, or N3 N-terminal mutations; and the EL/KK or ELD/KKR *Fok*I variants and at 4, 8, 16, or 32 nM, histograms of the number of spacer DNA base pairs preceding the right half-site for each possible DNA spacer length, normalized to the total sequence counts of the entire selection, are shown. (B) Same as (A) for ATM TALENs.

SUPPLEMENTARY TABLES

A

Selection name	Target site	Left+Right half-site	Site length	N-terminal domain	C-terminal domain	*Fok*I domain	TALEN conc. (nM)
CCR5A 32 nM canonical	CCR5A	L18+R18	36	canonical	Canonical	EL/KK	32
CCR5A 16 nM canonical (or CCR5A canonical)	CCR5A	L18+R18	36	canonical	Canonical	EL/KK	16
CCR5A 8 nM canonical	CCR5A	L18+R18	36	canonical	Canonical	EL/KK	8
CCR5A 4 nM canonical	CCR5A	L18+R18	36	canonical	Canonical	EL/KK	4
CCR5A Q3	CCR5A	L18+R18	36	canonical	Q3	EL/KK	16
CCR5A 32 nM Q7	CCR5A	L18+R18	36	canonical	Q7	EL/KK	32
CCR5A 16 nM Q7 (or CCR5A Q7)	CCR5A	L18+R18	36	canonical	Q7	EL/KK	16
CCR5A 8 nM Q7	CCR5A	L18+R18	36	canonical	Q7	EL/KK	8
CCR5A 4 nM Q7	CCR5A	L18+R18	36	canonical	Q7	EL/KK	4
CCR5A 28-aa	CCR5A	L18+R18	36	canonical	28-aa	EL/KK	16
CCR5A N1	CCR5A	L18+R18	36	N1	Canonical	EL/KK	16
CCR5A N2	CCR5A	L18+R18	36	N2	Canonical	EL/KK	16
CCR5A N3	CCR5A	L18+R18	36	N3	Canonical	EL/KK	16
CCR5A canonical ELD/KKR	CCR5A	L18+R18	36	canonical	Canonical	ELD/KKR	16
CCR5A Q3 ELD/KKR	CCR5A	L18+R18	36	canonical	Q3	ELD/KKR	16
CCR5A Q7 ELD/KKR	CCR5A	L18+R18	36	canonical	Q7	ELD/KKR	16
CCR5A N2 ELD/KKR	CCR5A	L18+R18	36	N2	Canonical	ELD/KKR	16

B

Selection name	Target site	Left + Right half-site	Site length	N-terminal domain	C-terminal domain	*Fok*I domain	TALEN conc. (nM)
ATM 24 nM canonical	ATM	L18+R18	36	canonical	Canonical	EL/KK	24
ATM 12 nM canonical (or ATM canonical)	ATM	L18+R18	36	canonical	Canonical	EL/KK	12
ATM 6 nM canonical	ATM	L18+R18	36	canonical	Canonical	EL/KK	6
ATM 3 nM canonical	ATM	L18+R18	36	canonical	Canonical	EL/KK	3
ATM Q3	ATM	L18+R18	36	canonical	Q3	EL/KK	12
ATM 24 nM Q7	ATM	L18+R18	36	canonical	Q7	EL/KK	24
ATM 12 nM Q7 (or ATM Q7)	ATM	L18+R18	36	canonical	Q7	EL/KK	12
ATM 6 nM Q7	ATM	L18+R18	36	canonical	Q7	EL/KK	6
ATM N1	ATM	L18+R18	36	N1	Canonical	EL/KK	12
ATM N2	ATM	L18+R18	36	N2	Canonical	EL/KK	12
ATM N3	ATM	L18+R18	36	N3	Canonical	EL/KK	12
ATM canonical ELD/KKR	ATM	L18+R18	36	canonical	Canonical	ELD/KKR	12
ATM Q3 ELD/KKR	ATM	L18+R18	36	canonical	Q3	ELD/KKR	12
ATM Q7 ELD/KKR	ATM	L18+R18	36	canonical	Q7	ELD/KKR	12
ATM N2 ELD/KKR	ATM	L18+R18	36	N2	Canonical	ELD/KKR	12

C

Selection name	Target site	Left + Right half-site	Site length	N-terminal domain	C-terminal domain	*Fok*I domain	TALEN conc. (nM)
L16+R16 CCR5B	CCR5B	L16+R16	32	canonical	Canonical	EL/KK	10
L16+R13 CCR5B	CCR5B	L16+R13	29	canonical	Canonical	EL/KK	10
L16+R10 CCR5B	CCR5B	L16+R10	26	canonical	Canonical	EL/KK	10
L13+R16 CCR5B	CCR5B	L13+R16	29	canonical	Canonical	EL/KK	10
L13+R13 CCR5B	CCR5B	L13+R13	26	canonical	Canonical	EL/KK	10
L13+R10 CCR5B	CCR5B	L13+R10	23	canonical	Canonical	EL/KK	10
L10+R16 CCR5B	CCR5B	L10+R16	26	canonical	Canonical	EL/KK	10
L10+R13 CCR5B	CCR5B	L10+R13	23	canonical	Canonical	EL/KK	10
L10+R10 CCR5B	CCR5B	L10+R10	20	canonical	Canonical	EL/KK	10

Supplementary Table S1. TALEN Constructs and Concentrations Used in the Selections. For each selection using TALENs targeting the CCR5A target sequence (A), ATM target sequence (B) and CCR5B target sequence (C), the selection name, the target DNA site, the TALEN N-terminal domain, the TALEN C-terminal domain, the TALEN *Fok*I domain, and the TALEN concentration (conc.) are shown.

A

Selection name	Seq. count	Mean mut.	Stdev mut.	Mut./bp	P-value vs. library	P-value vs. other TALENs
CCR5A 32 nM canonical	53883	4.327	1.483	0.120	3.3E-10	vs. CCR5A canonical ELD/KKR = 0.26
CCR5A 16 nM canonical	28940	4.061	1.438	0.113	5.4E-10	vs. CCR5A Q3 ELD/KKR = 0.028
CCR5A 8 nM canonical	29568	3.751	1.394	0.104	3.3E-10	
CCR5A 4 nM canonical	34355	3.347	1.355	0.093	1.5E-10	
CCR5A Q3	51694	3.841	1.380	0.107	1.7E-10	
CCR5A 32 nM Q7	48473	2.718	1.197	0.076	4.4E-11	
CCR5A 16 nM Q7	56593	2.559	1.154	0.071	3.1E-11	
CCR5A 8 nM Q7	43895	2.303	1.157	0.064	3.0E-11	
CCR5A 4 nM Q7	43737	2.018	1.234	0.056	2.1E-11	
CCR5A 28-aa	47395	2.614	1.203	0.073	4.0E-11	
CCR5A N1	64257	3.721	1.379	0.103	1.1E-10	vs. CCR5A 8 nM canonical =0.039
CCR5A N2	45467	3.148	1.306	0.087	8.2E-11	
CCR5A N3	24064	2.474	1.493	0.069	8.1E-11	
CCR5A canonical ELD/KKR	46998	4.336	1.491	0.120	4.0E-10	
CCR5A Q3 ELD/KKR	56978	4.098	1.415	0.114	2.2E-10	
CCR5A Q7 ELD/KKR	54903	3.234	1.330	0.090	7.3E-11	
CCR5A N2 ELD/KKR	79632	3.286	1.341	0.091	5.2E-11	

B

Selection name	Seq. count	Mean mut.	Stdev mut.	Mut./bp	P-value vs. library	P-value vs. other TALENs
ATM 24 nM canonical	89571	3.262	1.360	0.091	6.54E-11	vs. ATM canonical ELD/KKR =0.046
ATM 12 nM canonical (or ATM canonical)	96703	3.181	1.307	0.088	5.36E-11	
ATM 6 nM canonical	78852	2.736	1.259	0.076	3.63E-11	
ATM 3 nM canonical	82527	2.552	1.258	0.071	2.71E-11	
ATM Q3	96582	2.551	1.248	0.071	2.31E-11	
ATM 24 nM Q7	10166	1.885	2.125	0.052	2.06E-10	
ATM 12 nM Q7 (or ATM Q7)	4662	1.626	2.083	0.045	5.31E-10	vs. ATM 6 nM Q7 =0.069
ATM 6 nM Q7	1290	1.700	2.376	0.047	7.16E-09	
ATM N1	84402	2.627	1.318	0.073	2.92E-11	vs. ATM N3 =0.039
ATM N2	62470	2.317	1.516	0.064	2.69E-11	
ATM N3	1605	2.720	2.363	0.076	2.69E-08	vs. ATM 6 nM canonical =0.36 vs. Q3 ELD/KKR =0.017
ATM canonical ELD/KKR	107970	3.279	1.329	0.091	5.48E-11	
ATM Q3 ELD/KKR	104099	2.846	1.244	0.079	3.15E-11	

ATM Q7 ELD/KKR	21108	1.444	1.56	0.040	3.02E-11
ATM N2 ELD/KKR	70185	2.45	1.444	0.06805	2.82E-11

C

Selection name	Seq. count	Mean mut.	Stdev mut.	Mut./bp	P-value vs. library	P-value vs. other TALENs
L16+R16 CCR5B	34904	2.134	1.168	0.067	4.7E-11	
L16+R13 CCR5B	38229	1.581	1.142	0.055	2.7E-11	
L16+R10 CCR5B	37801	1.187	0.949	0.046	2.2E-11	
L13+R16 CCR5B	46608	1.505	1.090	0.052	1.7E-11	
L13+R13 CCR5B	53973	0.996	1.025	0.038	8.8E-12	
L13+R10 CCR5B	60550	0.737	0.884	0.032	7.4E-12	
L10+R16 CCR5B	36927	1.387	0.971	0.053	3.0E-11	
L10+R13 CCR5B	58170	0.839	0.882	0.036	9.1E-12	
L10+R10 CCR5B	57331	0.646	0.779	0.032	1.0E-11	

Supplementary Table S2. Statistics of Sequences Selected by TALEN Digestion. Statistics are shown for each TALEN selection on the CCR5A target sequence (A), ATM target sequence (B), and CCR5B target sequences (C). Seq. counts: total counts of high-throughput sequenced and computationally filtered selection sequences. Mean mut.: mean mutations in selected sequences. Stdev. mut.: standard deviation of mutations in selected sequences. Mut./bp: mean mutation normalized to target site length (bp). P-value vs. library: P-values between the TALEN selection sequence distributions to the corresponding pre-selection library sequence distributions (Supplementary Table S3) were determined as previously reported[6] using a (left) one-sided *t*-test. P-value vs. other TALENs: all pairwise comparisons between each distribution of TALEN-digested sequences were calculated (e.g., comparing CCR5A 16 nM canonical and CCR5A 32 nM canonical). All P-values *above* the lower P-value limits are shown. Lower P-values limits were based on multiple comparison correction and thus any P-value below the lower value is considered significant. For CCR5A and ATM lower P-value limits were 0.0125 and for CCR5B the lower P-value limit was 0.0055 based on multiple comparisons correction using the Bonferroni method. All post-selection sequences were assumed to be binomially distributed (Fig. 2A and 2B). Note that for the 3 nM Q7 ATM and the 28-aa ATM selection not enough sequences were obtained to interpret, although these selections were performed.

Library name	Target site	Left + Right half-site	Site length	Seq. count	Mean mut.	Stdev mut.	Mut./bp
CCR5A Library	CCR5A	L18+R18	36	158643	7.539	2.475	0.209
ATM Library	ATM	L18+R18	36	212661	6.820	2.327	0.189
CCR5B Library	CCR5B	L16+R16	32	280223	6.500	2.441	0.203
CCR5B Library	CCR5B	L16+R13	29	280223	5.914	2.336	0.204
CCR5B Library	CCR5B	L16+R10	26	280223	5.273	2.218	0.203
CCR5B Library	CCR5B	L13+R16	29	280223	5.969	2.340	0.206
CCR5B Library	CCR5B	L13+R13	26	280223	5.383	2.230	0.207
CCR5B Library	CCR5B	L13+R10	23	280223	4.742	2.106	0.206
CCR5B Library	CCR5B	L10+R16	26	280223	5.396	2.217	0.208
CCR5B Library	CCR5B	L10+R13	23	280223	4.810	2.100	0.209
CCR5B Library	CCR5B	L10+R10	20	280223	4.169	1.971	0.208

Supplementary Table S3. Statistics of Sequences From Pre-selection Libraries. For each pre-selection library containing a distribution of mutant sequences of the CCR5A target sequence, ATM target sequence and CCR5B target sequences. Seq. counts: total counts of high-throughput sequenced and the computationally filtered selection sequences. Mean mut.: mean mutations of sequences. Stdev. mut.: standard deviation of sequences. All pre-selection sequences were assumed to be binomially distributed (Figure 2A and 2B). Mut./bp: mean mutation normalized to target site length (bp).

A

Selection	Enrichment value								
	0 Mut.	1 Mut.	2 Mut.	3 Mut.	4 Mut.	5 Mut.	6 Mut.	7 Mut.	8 Mut.
CCR5A 32 nM canonical	9.879	9.191	8.335	6.149	4.205	2.269	1.005	0.325	0.085
CCR5A 16 nM canonical	12.182	13.200	10.322	7.195	4.442	2.127	0.748	0.216	0.052
CCR5A 8 nM canonical	19.673	17.935	13.731	8.505	4.512	1.756	0.531	0.116	0.028
CCR5A 4 nM canonical	36.737	29.407	19.224	9.958	4.047	1.242	0.302	0.058	0.014
CCR5A Q3	18.550	16.466	12.024	8.070	4.632	1.938	0.572	0.126	0.026
CCR5A 32 nM Q7	60.583	54.117	31.082	11.031	2.640	0.469	0.073	0.013	0.006
CCR5A 16 nM Q7	62.294	64.689	35.036	10.538	2.183	0.322	0.046	0.010	0.006
CCR5A 8 nM Q7	97.020	91.633	38.634	8.974	1.485	0.189	0.029	0.010	0.007
CCR5A 4 nM Q7	197.239	130.497	38.361	6.535	0.896	0.120	0.025	0.019	0.017
CCR5A 28-aa	70.441	62.213	33.481	10.488	2.317	0.402	0.064	0.012	0.006
CCR5A N1	19.038	18.052	13.858	8.788	4.546	1.697	0.499	0.115	0.025
CCR5A N2	41.715	35.752	22.638	10.424	3.777	0.989	0.194	0.038	0.007
CCR5A N3	173.897	88.392	31.503	8.770	1.853	0.350	0.089	0.036	0.027
CCR5A canonical ELD/KKR	8.101	10.012	8.220	6.147	4.119	2.291	1.019	0.330	0.083
CCR5A Q3 ELD/KKR	14.664	12.975	9.409	6.819	4.544	2.235	0.797	0.198	0.041
CCR5A Q7 ELD/KKR	37.435	32.922	21.033	10.397	3.867	1.087	0.238	0.046	0.010
CCR5A N2 ELD/KKR	35.860	31.469	20.135	10.189	3.983	1.155	0.260	0.050	0.013

B

Selection	Enrichment value								
	0 Mut.	1 Mut.	2 Mut.	3 Mut.	4 Mut.	5 Mut.	6 Mut.	7 Mut.	8 Mut.
ATM 24 nM canonical	19.900	16.881	12.162	6.318	2.629	0.884	0.228	0.057	0.015
ATM 12 nM canonical	20.472	17.645	12.724	6.549	2.606	0.803	0.189	0.039	0.007
ATM 6 nM canonical	41.141	29.522	17.153	6.551	1.872	0.431	0.082	0.017	0.006
ATM 3 nM canonical	56.152	37.152	18.530	6.196	1.562	0.308	0.058	0.015	0.008
ATM Q3	50.403	36.687	19.031	6.245	1.513	0.294	0.057	0.016	0.010
ATM 24 nM Q7	353.148	90.350	13.475	1.531	0.186	0.128	0.116	0.118	0.103
ATM 12 nM Q7	513.385	89.962	11.310	0.860	0.190	0.093	0.115	0.092	0.111
ATM 6 nM Q7	644.427	82.074	7.650	0.677	0.170	0.205	0.163	0.164	0.071
ATM N1	57.218	35.388	17.808	6.124	1.644	0.383	0.076	0.023	0.011
ATM N2	119.240	53.618	18.977	4.742	0.992	0.233	0.076	0.044	0.037
ATM N3	201.158	55.468	15.244	3.187	0.764	0.307	0.154	0.173	0.287
ATM canonical ELD/KKR	19.356	15.692	11.855	6.403	2.706	0.899	0.224	0.054	0.011
ATM Q3 ELD/KKR	32.816	25.151	16.172	6.727	2.095	0.506	0.095	0.018	0.004

ATM Q7 ELD/KKR	447.509	93.166	13.505	1.543	0.170	0.053	0.049	0.045	0.045
ATM N2 ELD/KKR	90.625	45.525	18.683	5.369	1.267	0.274	0.076	0.035	0.027

C

Selection	Enrichment value								
	0 Mut.	1 Mut.	2 Mut.	3 Mut.	4 Mut.	5 Mut.	6 Mut.	7 Mut.	8 Mut.
L16+R16 CCR5B	59.422	35.499	13.719	3.770	0.737	0.132	0.024	0.011	0.008
L16+R13 CCR5B	80.852	31.434	7.754	1.380	0.218	0.040	0.022	0.016	0.017
L16+R10 CCR5B	64.944	20.056	3.867	0.515	0.056	0.010	0.006	0.006	0.007
L13+R16 CCR5B	101.929	34.255	8.131	1.299	0.167	0.033	0.016	0.011	0.014
L13+R13 CCR5B	113.102	22.582	3.037	0.315	0.044	0.022	0.017	0.017	0.016
L13+R10 CCR5B	74.085	11.483	1.270	0.121	0.022	0.013	0.011	0.013	0.008
L10+R16 CCR5B	60.186	22.393	5.286	0.777	0.084	0.012	0.006	0.006	0.008
L10+R13 CCR5B	74.204	13.696	1.673	0.152	0.021	0.011	0.010	0.009	0.010
L10+R10 CCR5B	43.983	7.018	0.740	0.061	0.013	0.007	0.007	0.008	0.005

Supplementary Table S4. Enrichment Values of Sequences as a Function of Number of Mutations. For each TALEN selection on the *CCR5A* target sequence (A), *ATM* target sequence (B) and CCR5B target sequence (C), enrichment values calculated by dividing the fractional abundance of post-selection sequences from a TALEN digestion by the fractional abundance of pre-selection sequences as a function of total mutations (Mut.) in the half-sites.

Mutations in site	Off-target sites to CCR5A	Statistically expected
0	1	1
1	0	0.0
2	0	0.0
3	0	0.0
4	0	0.0
5	0	0.0
6	0	0.0
7	0	0.3
8	8	3.6
9	70	34.1
10	634	275.9
11	4338	1956.3
12	27114	12226.7
13	149005	67716.9
14	648230	333747.3
15	2657598	1468488.3
16	9783617	5782172.6

S**upplementary Table S5. Genomic Off-target Site Abundance in the Human Genome.** Column 2 shows the number of sites in the human genome related to the CCR5A on-target sequence, allowing for a spacer length from 12 to 25 bps between the two half-sites (Supplementary Algorithm). Column 3 shows the statistically expected average numbers of off-target sites that would be found in a human-sized genome for a 36-bp target sequence assuming a random distribution of an equal ratio of A:C:G:T across the genome utilizing the binomial distribution of mutant sites multiplied by the number of spacer lengths allowed (14). Number of sites with m mutations = genome size ($2 \times 3.1 \times 10^9$) x spacer length allowance (14) x binomial distribution of mutations in a 36-bp target site with a 0.75 probability of a mutation occurring $[(36! / (m! \times (36 - m)!)) \times (0.75)^m \times (0.25)^{36-m}]$.

A

CCR5A Site	Score	Mut.	Left half-site	Spacer length	Right half-site
OnCCR5A	0.008	0	TTCATTACACCTGCAGCT	18	AGTATCAATTCTGGAAGA
OffC-1	0.747	9	TaCATcACAtaTGCAaaT	29	tGTATCAtTTCTGGgAGA
OffC-2	0.747	9	TaCATcACAtaTGCAaaT	29	tGTATCAtTTCTGGgAGA
OffC-3	0.747	9	TaCATcACAtaTGCAaaT	29	tGTATCAtTTCTGGgAGA
OffC-4	0.747	11	TcCATaACACaTctttCT	10	tGcATCAtTcCTGGAAGA
OffC-5	0.804	11	TcCAaTACctCTGCcaCa	14	AGgAgCAAcTCTGGgAGA
OffC-6	0.818	10	TTCAgTcCAtCTGaAaac	16	gGTATCAtTTCTGGAgGA
OffC-7	0.834	14	TaCAaaACcCtTGCcaaa	27	taTATCAATTtgGGgAGA
OffC-8	0.837	12	TcCAagACACCTGCttac	26	tcTATCAATTtgGGgAGA
OffC-9	0.874	10	TTCATaACAtCTtaAaaT	27	AaTAcCAAcTCTGGAtGA
OffC-10	0.89	12	TcCAaaACAtCTGaAaaT	25	tGgATCAAaTtgGGAAGA
OffC-11	0.896	12	TTCAgaACACaTGactac	21	tGTATCAgTTaTGGAtGA
OffC-12	0.904	13	TcCATaAtAtCTtCctCT	28	gGgATtAATTtgGGAgGA
OffC-13	0.905	11	TgCAaTAtACCTGttGaT	16	ctcATCAATTCTGGgtGA
OffC-14	0.906	12	TTCATaACACtccacctT	16	gGTATCAAaTCTGGggGA
OffC-15	0.906	12	TcCATgACACaaaagaCT	26	gGTATCtATcCTGGAAtA
OffC-16	0.906	9	TTCcTTcCACCaGtgtCc	28	AGcATCAATcCTGGAAGA
OffC-17	0.907	10	TTaATaACAtCTcCAaCT	24	gGcAcCAAaTCTGGAtGA
OffC-18	0.909	13	TcCATcACcCCTcCctCc	10	gGTgcCAgcTCTGGAgGA
OffC-19	0.909	8	TTCATTACtCCTcCttCT	30	ctTATCAcTTtTGGAAGA
OffC-20	0.912	10	TgCATTACACaTtatGtg	17	AGcAgCAcTTCTGGAAGA
OffC-21	0.913	11	TTCAaaACACaTaCAtCT	28	AacAaCAtTcCTGtAAGA
OffC-22	0.913	10	TcCATTACcaCTGCAGaT	25	gacATCAgTTaTGGAtGA
OffC-23	0.925	13	TTCcagACcCCTtCctCa	13	gacATCAAaTCTGGgAGA
OffC-24	0.927	12	TTCcaaACACCcGCttCc	26	taTATCctTTCTGGAAtA
OffC-25	0.93	12	TgaAaTACACCTGCctaT	13	gGccTCAAggCTGGAtGA
OffC-26	0.93	12	TgCcaaACctCTGtcaCc	22	AGgATCAcTTCTGGAAGA
OffC-27	0.931	12	TgCcaaACctCTGtcaCc	22	AGgATCAcTTCTGGAAGA
OffC-28	0.931	8	TTtATTACACtTcCAGaT	19	gaTATCctTTCTGGAAGA
OffC-29	0.932	13	TaCAaaAaACtTtCtGag	27	tGTATCAATTtgGGgAGA
OffC-30	0.932	11	TcCAaaACACCcaCAGac	19	gGTATagATTgTGGAAGA
OffC-31	0.934	13	TTCATTcCACaTcCccac	25	gtTATCAAcatgGGAAGA
OffC-32	0.934	11	TTCAaTAtgCCaaCAGCT	11	AGctTCAATctgGGAgGA
OffC-33	0.934	12	TTCAaTACACtTGtctaT	12	tGTgTCAtTTCTGGgttA
OffC-34	0.935	11	TTCAacACACCTtCAaaa	12	tGTgTCAtTaaTGGAAGA
OffC-35	0.935	10	TTCAaaACAtCTGacatT	10	AaTAgaAATTCTGGAAGA
OffC-36	0.935	11	cTCcTaAtACCTGCAaaT	21	gaTATtAtTTCTGGAgGA
OffC-38	0.939	10	TTCATaACAaCTtaAaCa	16	tGgATCAATTgTGGAAtA
OffC-39	0.941	9	TTaATTAtAttTttAaCa	25	AGTAcCAATTCTGGAAGA
OffC-40	0.941	9	TTaATTAtAttTttAaCa	25	AGTAcCAATTCTGGAAGA
OffC-42	0.941	9	TcaATTAtACCTaCAtaT	24	AGTtTCAATTtaGGAAGA

OffC-45	0.943	10	aTCATTgCcCCTGCAGag	23	gGTATCttTcCTGGAtGA
OffC-49	0.946	9	TTCAaTACACCTaCAtCa	15	AGTcaCAtTTCTGGAAtt
OffC-56	0.951	9	TgCATTACAaCTGaAGac	19	AGcATtAcTTCaGGAAGA
OffC-65	0.956	9	TaCATTcCACCTaCttCc	22	AGTtTgAcTTCTGGAAGA
OffC-69	0.958	9	TTCATTACACCcaCtGCT	23	caTtTCtAgTCTGGAttA
OffC-76	0.96	9	TcCATaAtcCCatCAGCT	19	AGTgTCAtTTCTGGAAGg
OffC-137	0.974	9	TTCATTAtgCCTGCAGta	14	AGcATCAATTCatGAtGA
OffC-150	0.975	9	TTCATTgCACCTGataCT	16	gGaAcCtATTCTGGAAGt

B

ATM Site	Score	Mut.	Left half-site	Spacer length	Right half-site
OnATM	0.000	0	TGAATTGGGATGCTGTTT	18	TTTATTTTACTGTCTTTA
OffA-1	0.595	7	TGAATaGGaAataTaTTT	20	TTTATTTTACTGTtTTTA
OffA-2	0.697	9	TGgATTcaGATaCTcTTT	10	TTTATTTTttTaTtTTTA
OffA-3	0.697	9	TGgATTcaGATaCTcTTT	10	TTTATTTTttTaTtTTTA
OffA-4	0.697	9	TGgATTcaGATaCTcTTT	10	TTTATTTTttTaTtTTTA
OffA-5	0.697	9	TGgATTcaGATaCTcTTT	10	TTTATTTTttTaTtTTTA
OffA-6	0.697	9	TGgATTcaGATaCTcTTT	10	TTTATTTTttTaTtTTTA
OffA-7	0.697	9	TGgATTcaGATaCTcTTT	10	TTTATTTTttTaTtTTTA
OffA-8	0.7	8	TGcATaGGaATGCTaaTT	10	TTTATTTTACTaTtTaTA
OffA-9	0.708	10	TGAATTaaaATcCTGcTT	19	gTTATaTgACTaTtTTTA
OffA-10	0.711	10	TccATTaaaATaCTaTTT	18	TTTATTTTAtTaTtTTTA
OffA-11	0.715	10	TGAATTGaGAgaagcaTT	16	TTTATTTTAtTaTtTTTA
OffA-12	0.725	10	TGAAgTGGGATaCTGTTa	29	ggTATaTTAtaaTtTTTA
OffA-13	0.729	9	TGAATTatGAaGCTacTT	17	TTTATTgTAaTaTtTTTA
OffA-14	0.731	9	TGAATaaGGATGCTaTTa	25	TTTATTTattTaTtTTTA
OffA-15	0.744	10	TGAATgGGGAcaCaGcca	29	TTTATTTTAtTaTtTTTA
OffA-16	0.752	9	TaAATgGaaATGCTGTTc	24	aTTATTTTAtTGTtTTTt
OffA-17	0.761	9	gGAAaTGGGATaCTGagT	15	TTTATgTTACTaTtTcTA
OffA-18	0.781	11	TGgATcGaagTGaTtaTT	23	TTTATTTTAtTaTtTTTA
OffA-19	0.792	11	TGAATTGaGATtCacagc	23	TTTATTTTttTaTtTTTA
OffA-20	0.803	8	TGAATTaGGAatCTGaTT	10	TTTATTTTAtTaTtaTTA
OffA-21	0.807	12	TaAATTaaaATaCTccag	23	aTTATTTTAaTGTtTTTA
OffA-22	0.811	10	TGAATaGGaATatTcTTT	12	TTTATTTattTaTtTTTA
OffA-23	0.811	9	TagATTGaaATGCTGTTT	15	TTTtTaTTAtTaTtTTTA
OffA-24	0.816	10	TGAcTaGaaATGaTGaTT	25	TTTATTTTctTaTtTTTA
OffA-25	0.817	12	TGAATTtaaAaaaTGTcc	13	aTTATTTTAtTaTtTTTA
OffA-26	0.817	12	TGAATTtaaAaaaTGTcc	13	aTTATTTTAtTaTtTTTA
OffA-27	0.817	10	TGgATccaGATaCTcTTT	10	TTTATTTTttTaTtTTTA
OffA-28	0.819	7	TGgAgTGaGATcCTGTTT	21	TTTATTTTAtTGTtaTTA
OffA-29	0.824	8	TGAAcTtGGATGaTaTaT	24	TTTATTTgAtTaTCTTTA
OffA-30	0.832	9	TGtATTGGGATaCcaTTT	26	TcTATTTTAtTaTtTTTt

OffA-31	0.833	9	TcAATTGGGATGaTcaTa	23	TTTATTcTAtTtTtTTTA
OffA-32	0.835	9	TGAAagGGaAaGtTGgaT	23	TTTATTTTACTaTtTTTA
OffA-33	0.841	9	TGgtTTGGGATcCTGTgT	27	TTTATgTTttTaTtTTTA
OffA-34	0.841	9	TGAAaTGGGATGagcTTg	28	TTTATTTTAtTaTtTTaA
OffA-35	0.844	10	TGAATTGGGATaCTGTag	29	cTTAaaTaAaTaTtTTTA
OffA-36	0.844	10	TGAATTGtGgTatTGccT	18	TTTATggTttTGTCTTTA

C

Site	Mut.	Left half-site	Spacer length	Right half-site
OnHDAC	0	TGGCGCAGACGCAGGGC	17	TTACTACTACGACGGTGA
OffHDAC-1	3	TGaCGCAGACaCAGGGC	17	TTACTACTACGACGGgGA
OnPMS	0	TCGGGTGTTGCATCCATG	18	AGGTGAGCGGGGCTCGCA
OffPMS-1	4	TCcGGTGTTGCATCCtTG	18	AGGTGAGCtGGGCTCGCg
OffPMS-2	4	TCcGGTGTTGCATCCtTG	18	AGGTGAGCtGGGCTCGCg
OnSDHD	0	TCAGGAACGAGATGGCGG	17	GCCGTTTGCGGTGCCCTA
OffSDHD-1	1	TCAGGAACGAGATGGCGG	17	GCCGTTTGCGGTGCCCaA
OffSDHD-3	3	TCAGGAACGAGATGGCGG	17	GCCcTTTGCaGTGCCCaA
OffSDHD-4	3	TCAGGAACGAGATGGCGG	17	GCCcTTTaCGGTGCCCaA
OffSDHD-5	5	TCAGGAAtGAGATGGCGG	17	aCCcTTcGCGGTGCCCaA

D

Site	Score	Mut.	Left half-site	Spacer length	Right half-site	Classifier
CP_CCRoff-1	-1.503	9	aTaATTACACCTGCAaCT	24	AGTgTCAgTTtgGtAAGt	
CP_CCRoff-2	-1.524	9	TTCATTACAaaTaCAGaa	23	AGaATttATTCTGtAAGA	
CP_CCRoff-3	-1.53	10	TTCtTTAaACCTaaAaCT	21	AtTATtAtTTtTGGAAGt	
CP_CCRoff-4	-1.534	8	TTCATTACtCCTcCttCT	30	ctTATCAcTTtTGGAAGA	OffC-19
CP_CCRoff-5	-1.537	8	TTttTTACtCCTGtAaCa	21	AGTcTCAATTtTGGAAGA	
CP_CCRoff-6	-1.571	9	TTCATTcaACCTtCAaCT	21	AcagTgAATTtTGGAAGA	
CP_CCRoff-7	-1.583	9	TcaATTAtACCTaCAtaT	24	AGTtTCAATTtaGGAAGA	OffC-42
CP_CCRoff-8	-1.583	9	gcCAcTgCAaCTcCAGCc	30	AGTATCAtTTtTGGAAGA	
CP_CCRoff-9	-1.594	9	TaCATcACAtaTGCAaaT	29	tGTATCAtTTCTGGgAGA	OffC-1
CP_CCRoff-10	-1.594	9	TaCATcACAtaTGCAaaT	29	tGTATCAtTTCTGGgAGA	OffC-2
CP_CCRoff-11	-1.594	9	TaCATcACAtaTGCAaaT	29	tGTATCAtTTCTGGgAGA	OffC-3
CP_CCRoff-12	-1.594	9	TaCATTcCACCTaCttCc	22	AGTtTgAcTTCTGGAAGA	
CP_CCRoff-13	-1.606	11	TcCcTTcCACCcaCAaCT	26	AGTATtAgTTCTGGggtA	
CP_CCRoff-14	-1.606	9	TTCATTtCAtCTcaAaCT	13	AaTgTCAtTTCTGtAAGA	

CP_CCRoff-15	-1.611	9	TcCccTACctCTGCAGCT	28	AGTgTCAcTTCTGGgAGg	
CP_CCRoff-16	-1.611	8	TTCATcttcCCTGCAGCg	27	AGaATCAAaTCTGtAAGA	
CP_CCRoff-17	-1.617	9	TTCATTACAaaTGCAGta	29	tGTATgAATTtgaGAAGA	
CP_CCRoff-18	-1.62	8	TTtATTACACtTcCAGaT	19	gaTATCctTTCTGGAAGA	OffC-28
CP_CCRoff-19	-1.621	10	TTCtaTAaAaCTcCAaCT	11	AGTATgAATTCTttAAtA	
CP_CCRoff-20	-1.622	9	TTCtTTACACCTcCAGCT	22	AaaATCAAagtTtGAtGA	
CP_CCRoff-21	-1.635	10	TTCATTAtcaCTcCAaCT	10	AtTAggAgTcCTGGAAGA	
CP_CCRoff-22	-1.637	10	TTCtcTACACCTGaAaCc	12	AGcATtgtgTCTGGAAGA	
CP_CCRoff-23	-1.638	8	TTaATTAaACCTGCAGtT	24	tacATtAATTCTGGAAaA	
CP_CCRoff-24	-1.65	11	TTCATTACAaCcaaAatT	29	AGTAatAATgtTGGgAGA	
CP_CCRoff-25	-1.651	10	TTCcTTACcCCTGCAtCT	21	tGTtTCAtTTtTttgAGA	
CP_CCRoff-26	-1.653	8	TggAcTgCACCTGCAGCT	23	AGctTtgATTCTGGAAGA	
CP_CCRoff-27	-1.653	11	TTCATTtCACaTaCAcCc	20	AtTATgAAgTtTtGAtGA	
CP_CCRoff-28	-1.654	10	TTCATTAaACaTGaAaCT	16	AGTATgAgcTCatGgAGA	
CP_CCRoff-29	-1.655	9	TTCAcTACACCTGCAGCc	26	AtaATgAcTTCTGGctGg	
CP_CCRoff-30	-1.655	10	TTaATTcCACCTGCAGCT	29	gGTcaCAggTtTtGgAGA	
CP_CCRoff-31	-1.659	8	TTCtaaACACCTGtAGCT	27	AGgAaCAATaCTGGAtGA	
CP_CCRoff-32	-1.662	10	TTCATTtgACCTcCcaCT	17	tGTgggAATTCTGGgAGA	

Supplementary Table S6. Predicted Off-Target Sites in the Human Genome. (A) Using a machine-learning "classifier" algorithm trained on the output of the *in vitro* CCR5A TALEN selection,[18] mutant sequences of the target site allowing for spacer lengths of 10 to 30 base pairs were scored. The resulting 36 predicted off-targets sites with the best scores for the CCR5A TALENs and the next 12 best-scoring off-target sites with nine or ten mutations are shown with their respective classifier scores, mutation numbers, left and right half-site sequences (mutations from on-target in lower case), and the length of the spacer between half-sites in base pairs. (B) Same as (A) for ATM TALENs (without the next 12 best-scoring off-target sites with nine or ten mutations). (C) List of genomic on-target and off-targets sites with one to five mutations from the on-target site of the HDAC1, PMS2, and SDHD TALENs. Sites were identified by a simple computational search for off-target sites closely related to the target site in the human genome. Sites are shown with mutation numbers, left and right half-site sequences (with mutations from on-target in lower case), and the length of the spacer between half-sites in base pairs. (D) Computationally predicted genomic off-target sites from TALENoffer program[10] allowing for spacer lengths of 10 to 30 base pairs and only searching for heterodimeric off-target sites. The resulting 32 predicted off-targets sites with the best scores for the CCR5A TALENs are shown with their respective classifier scores, mutation numbers, left and right half-site sequences (with mutations from on-target in lower case), and the length of the spacer between half-sites in base pairs.

A

C-terminal domain	No TALEN	Q7	Q7	Q3	Q3	Canonical	Canonical	Canonical
*Fok*I domain	No TALEN	EL/KK	ELD/KKR	EL/KK	ELD/KKR	EL/KK	ELD/KKR	Homo
CCR5A Sites								
OnC								
Indels	1	147	705	1430	3731	841	2004	3943
Total	42042	7192	12667	16843	15381	8546	7267	8422
% Modified	<0.006%	2.044%	5.566%	8.490%	24.257%	9.841%	27.577%	46.818%
P-value		(5.5E-122)	(<1.0E-250)	(<1.0E-250)	(<1.0E-250)	<1.0E-250	<1.0E-250	<1.0E-250
Specificity								
OffC-1								
Indels	0	0	1	1	0	0	1	1
Total	51248	38975	79857	35490	77804	34227	87496	42497
% Modified	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%
P-value								
Specificity								
OffC-2								
Indels	6	1	6	2	1	0	11	3
Total	124356	96280	157387	93337	159817	85603	163322	114663
Modified	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	0.006%	<0.006%
P-value								
Specificity								
OffC-3								
Indels	6	1	6	2	1	0	11	3
Total	124356	96280	157387	93337	159817	85603	163322	114663
Modified	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%
P-value								
Specificity								
OffC-4								
Indels	0	1	0	0	0	0	0	0
Total	45377	44674	52876	35133	53909	26034	42284	40452
Modified	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%
P-value								
Specificity								
OffC-5								
Indels	0	0	0	3	22	134	385	395
Total	27009	28172	26036	22432	25800	25273	17045	17077
Modified	<0.004%*	<0.004%*	<0.004%*	0.013%	0.085%	0.530%	2.259%	2.313%

P-value					(1.4E-07)	4.1E-43	1.2E-160	1.2E-164
Specificity		>576	>1450	635	284	19	12	20
OffC-6								
Indels	0	0	0	0	0	0	0	0
Total	10766	12309	10886	9240	10558	10500	5943	6560
Modified	<0.009%	<0.008%	<0.009%	<0.011%	<0.009%	<0.010%	<0.017%	<0.015%
P-value								
Specificity								
OffC-7								
Total	0	0	0	0	0	0	0	0
Modified	15626	28825	22138	31742	19577	11902	33200	15400
P-value	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	<0.008%	<0.006%	<0.006%
Specificity								
OffC-9								
Indels	0	0	0	1	0	0	0	0
Total	40603	39765	47974	51595	44002	34520	25211	30771
Modified	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%
P-value								
Specificity								
OffC-10								
Indels	0	0	0	0	0	0	0	0
Total	4142	9591	5187	1413	7975	4378	2215	3779
Modified	<0.024%	<0.010%	<0.019%	<0.071%	<0.013%	<0.023%	<0.045%	<0.026%
P-value								
Specificity								
OffC-11								
Indels	0	0	0	0	0	0	0	0
Total	71180	55455	65015	44847	70907	50967	65257	60191
Modified	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%
P-value								
Specificity								
OffC-12								
Indels	0	0	0	0	0	0	0	0
Total	3242	1784	30274	14006	4897	19830	9747	12910
Modified	<0.031%	<0.056%	<0.006%	<0.007%	<0.020%	<0.006%	<0.010%	<0.008%
P-value								
Specificity								
OffC-13								

Indels	0	0	0	0	0	0	0	0
Total	65518	52459	53413	38156	61600	47922	57211	78546
Modified	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%
P-value								
Specificity								

OffC-14

Indels	0	0	0	0	0	0	2	0
Total	34607	7217	26301	8339	29845	1081	9471	19026
Modified	<0.006%	<0.014%	<0.006%	<0.012%	<0.006%	<0.093%	0.021%	<0.006%
P-value								
Specificity								

OffC-15

Indels	0	0	0	0	0	0	16	2
Total	4989	4880	6026	9370	9156	7371	6967	4662
Modified	<0.020%	<0.020%	<0.017%	<0.011%	<0.011%	<0.014%	0.230%	0.043%
P-value							1.9E-04	
Specificity		>100	>335	>796	>2221	>725	120	1091

OffC-16

Indels	0	1	1	1	14	1	12	0
Total	36228	34728	34403	34866	44362	38384	38536	32636
Modified	<0.006%	<0.006%	<0.006%	<0.006%	0.032%	<0.006%	0.031%	<0.006%
P-value					(2.5E-04)		5.2E-04	
Specificity		>307	>835	>1274	769	>1476	886	>7023

OffC-17

Indels	0	0	0	0	0	0	0	0
Total	32112	23901	31273	33968	27437	29670	27133	31299
Modified	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%
P-value								
Specificity								

OffC-18

Indels	0	0	0	0	0	0	0	0
Total	9437	9661	13505	14900	13848	12720	6624	12804
Modified	<0.011%	<0.010%	<0.007%	<0.007%	<0.007%	<0.008%	<0.015%	<0.008%
P-value								
Specificity								

OffC-19

Indels	1	1	1	2	2	2	1	0
Total	22868	11479	22702	15258	20733	17449	14638	28478
Modified	<0.006%	0.009%	<0.006%	0.013%	0.010%	0.011%	0.007%	<0.006%

P-value								
Specificity								
OffC-20								
Indels	0	0	0	0	0	1	0	0
Total	23335	26164	30782	15261	20231	21184	14144	18972
Modified	<0.006%	<0.006%	<0.006%	<0.007%	<0.006%	<0.006%	<0.007%	<0.006%
P-value								
Specificity								
OffC-21								
Indels	0	0	0	0	0	0	0	0
Total	34302	27573	31694	24451	25826	27192	18110	21161
Modified	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%
P-value								
Specificity								
OffC-22								
Indels	1	0	0	0	0	0	0	0
Total	81037	86687	74274	79004	93477	92089	75359	104857
Modified	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%
P-value								
Specificity								
OffC-23								
Indels	0	0	0	0	0	0	0	0
Total	18812	19337	23034	25603	25023	28615	17172	21033
Modified	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%
P-value								
Specificity								
OffC-24								
Indels	0	0	1	0	0	0	0	1
Total	23538	21673	24594	27687	18343	29113	21709	26610
Modified	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%
P-value								
Specificity								
OffC-25								
Indels	0	0	0	0	0	0	0	0
Total	28941	25326	25871	10641	21422	20171	18946	18711
Modified	<0.006%	<0.006%	<0.006%	<0.009%	<0.006%	<0.006%	<0.006%	<0.006%
P-value								
Specificity								

OffC-26

Indels	0	0	1	0	0	0	0	0
Total	71831	48494	62650	45801	60175	65137	28795	64632
Modified	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%
P-value								
Specificity								

OffC-27

Indels	0	0	0	0	0	0	0	0
Total	12181	2423	11258	7188	5126	4003	2116	4603
% Modified	<0.008%	<0.041%	<0.009%	<0.014%	<0.020%	<0.025%	<0.047%	<0.022%
P-value								
Specificity								

OffC-28

Indels	0	0	0	0	6	1	12	5
Total	10651	6410	16179	13980	13022	7232	7379	8998
% Modified	<0.009%	<0.016%	<0.006%	<0.007%	0.046%	0.014%	0.163%	0.056%
P-value							2.2E-05	
Specificity		>131	>835	>1187	526	712	170	843

OffC-29

Indels	0	0	0	0	0	0	0	0
Total	4262	3766	4228	6960	3234	1516	2466	1810
% Modified	<0.023%	<0.027%	<0.024%	<0.014%	<0.031%	<0.066%	<0.041%	<0.055%
P-value								
Specificity								

OffC-30

Indels	0	0	0	0	0	0	0	0
Total	11840	12257	9617	34097	20507	5029	22248	6285
% Modified	<0.008%	<0.008%	<0.010%	<0.006%	<0.006%	<0.020%	<0.006%	<0.016%
P-value								
Specificity								

OffC-31

Indels	0	0	0	0	0	0	0	0
Total	64522	67791	50085	50056	56241	48287	72230	100410
% Modified	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%
P-value								
Specificity								

OffC-32

Indels	0	0	0	0	0	0	0	0
Total	1944	6888	9330	3207	4591	6699	13607	19115

% Modified	<0.051%	<0.015%	<0.011%	<0.031%	<0.022%	<0.015%	<0.007%	<0.006%
P-value								
Specificity								
OffC-33								
Indels	0	0	0	0	0	0	0	0
Total	34475	27039	18547	33467	15745	17075	4	18844
% Modified	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	<25.000%	<0.006%
P-value								
Specificity								
OffC-34								
Indels	0	0	0	0	0	0	0	0
Total	9052	18858	13647	11796	6945	6114	4979	9072
% Modified	<0.011%	<0.006%	<0.007%	<0.008%	<0.014%	<0.016%	<0.020%	<0.011%
P-value								
Specificity								
OffC-35								
Indels	0	0	0	0	0	0	0	0
Total	23839	22290	25133	24190	10	10459	22554	11897
% Modified	<0.006%	<0.006%	<0.006%	<0.006%	<10.000%	<0.010%	<0.006%	<0.008%
P-value								
Specificity								
OffC-36								
Indels	1	0	0	1	2	1	19	5
Total	23412	24394	23427	24132	19723	28369	12461	18052
Modified	<0.006%	<0.006%	<0.006%	<0.006%	0.010%	<0.006%	0.152%	0.028%
P-value							2.5E-08	
Specificity		>307	>835	>1274	2392	>1476	181	1690

B

C-term. Domain:	No TALEN	Q7	Q7	Q3	Q3	Canonical	Canonical	Canonical
*Fok*I Domain:	No TALEN	EL/KK	ELD/KKR	EL/KK	ELD/KKR	EL/KK	ELD/KKR	Homo
ATM Sites								
On-A								
Indels	1	0	46	104	309	1289	410	909
Total	15116	1869	2520	1198	1808	19025	2533	5003
Modified	0.007%	<0.006%	1.825%	8.681%	17.091%	6.775%	16.186%	18.169%
P-value	0		(4.4E-27)	(1.4E-93)	(5.6E-243)	2.6E-214	<1.0E-250	<1.0E-250
Specificity								
OffA-1								

Indels	0	0	1	0	1	0	13	34
Total	52490	45383	34195	32325	47589	39704	50349	44056
Modified	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	0.026%	0.077%
P-value							9.3E-05	2.6E-12
Specificity		>0	>274	>1302	>2564	>1016	627	235

OffA-2

Indels	0	0	0	0	0	0	0	0
Total	8777	11846	11362	12273	20704	3776	5650	5025
Modified	<0.011%	<0.008%	<0.009%	<0.008%	<0.006%	<0.026%	<0.018%	<0.020%
P-value								
Specificity								

OffA-3

Indels	0	0	0	0	1	0	0	0
Total	47338	14352	21253	17777	26512	19483	43728	29469
Modified	<0.006%	<0.007%	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%
P-value								
Specificity								

OffA-4

Indels	0	0	0	0	0	0	0	0
Total	12292	532	1383	2597	861	2598	1356	3573
Modified	<0.008%	<0.188%	<0.072%	<0.039%	<0.116%	<0.038%	<0.074%	<0.028%
P-value								
Specificity								

OffA-5

Indels	0	0	0	0	0	0	0	0
Total	60859	22846	25573	19054	25315	31754	66622	60925
Modified	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%
P-value								
Specificity								

OffA-6

Indels	0	0	0	0	0	0	0	0
Total	60859	22846	25573	19054	25315	31754	66622	60925
Modified	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%
P-value								
Specificity								

OffA-7

Indels	0	0	0	0	0	0	0	0
Total	60859	22846	25573	19054	25315	31754	66622	60925
Modified	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%

P-value								
Specificity								
OffA-8								
Indels	0	0	0	0	0	0	0	0
Total	9170	1614	5934	3215	2450	12750	10120	13003
Modified	<0.011%	<0.062%	<0.017%	<0.031%	<0.041%	<0.008%	<0.010%	<0.008%
P-value								
Specificity								
OffA-9								
Indels	0	0	0	0	0	0	0	3
Total	8753	12766	9504	10114	11086	10676	9013	11110
Modified	<0.011%	<0.008%	<0.011%	<0.010%	<0.009%	<0.009%	<0.011%	0.027%
P-value								
Specificity								
OffA-10								
Indels	1	0	0	2	2	3	5	7
Total	8151	16888	8804	7061	8891	32138	14889	40120
Modified	0.012%	<0.006%	<0.011%	0.028%	0.022%	0.009%	0.034%	0.017%
P-value								
Specificity								
OffA-11								
Indels	0	0	1	0	0	0	9	76
Total	41343	32352	28834	28709	26188	32519	24894	19586
Modified	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	0.036%	0.388%
P-value							1.5E-04	3.1E-38
Specificity		>0	>274	>1302	>2564	>1016	448	47
OffA-12								
Indels	0	0	0	0	0	0	0	0
Total	13186	2326	13981	12911	21134	9220	7792	8068
Modified	<0.008%	<0.043%	<0.007%	<0.008%	<0.006%	<0.011%	<0.013%	<0.012%
P-value								
Specificity								
OffA-13								
Indels	0	0	0	0	0	2	9	0
Total	32704	32015	12312	23645	26315	24078	36111	22364
Modified	<0.006%	<0.006%	<0.008%	<0.006%	<0.006%	0.008%	0.025%	<0.006%
P-value							4.3E-03	
Specificity		>0	>225	>1302	>2564	816	649	>2725

OffA-15								
Indels	0	0	0	0	1	0	0	0
Total	14654	15934	12313	6581	13053	18996	10916	21519
Modified	<0.007%	<0.006%	<0.008%	<0.015%	0.008%	<0.006%	<0.009%	<0.006%
P-value								
Specificity								
OffA-16								
Indels	1	0	0	0	0	0	0	12
Total	65190	35639	37252	30378	31489	22590	13594	20922
Modified	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	<0.007%	0.057%
P-value								4.3E-07
Specificity		>0	>274	>1302	>2564	>1016	>2200	317
OffA-17								
Indels	0	0	0	0	0	0	0	6
Total	1972	606	1439	2113	2862	728	597	636
Modified	<0.051%	<0.165%	<0.069%	<0.047%	<0.035%	<0.137%	<0.168%	0.943%
P-value								2.0E-04
Specificity		>0	>26	>183	>489	>49	>97	19
OffA-18								
Indels	0	0	0	0	0	0	0	0
Total	5425	995	1453	1831	3132	1934	1534	5816
Modified	<0.018%	<0.101%	<0.069%	<0.055%	<0.032%	<0.052%	<0.065%	<0.017%
P-value								
Specificity								
OffA-19								
Indels	1	2	0	1	1	1	1	3
Total	31094	41252	33213	29518	32337	25904	27575	38711
Modified	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	0.008%
P-value								
Specificity								
OffA-21								
Indels	0	0	0	0	0	0	0	0
Total	15297	9710	16719	12119	15483	21692	16558	16418
Modified	<0.007%	<0.010%	<0.006%	<0.008%	<0.006%	<0.006%	<0.006%	<0.006%
P-value								
Specificity								
OffA-22								
Indels	27	41	38	46	32	50	55	57
Total	9406	11150	11516	10269	13814	14057	11685	14291

Modified	0.287%	0.368%	0.330%	0.448%	0.232%	0.356%	0.471%	0.399%
P-value								
Specificity								
OffA-23								
Indels	1	0	0	0	0	0	10	20
Total	5671	9363	2203	7011	7078	12068	3484	8619
Modified	0.018%	<0.011%	<0.045%	<0.014%	<0.014%	<0.008%	0.287%	0.232%
P-value							4.5E-04	4.9E-04
Specificity		>0	>40	>609	>1210	>818	56	78
OffA-24								
Indels	4	0	0	1	0	1	0	2
Total	17288	7909	14261	29936	6943	6333	14973	19953
Modified	0.023%	<0.013%	<0.007%	<0.006%	<0.014%	0.016%	<0.007%	0.010%
P-value								
Specificity								
OffA-25								
Indels	0	0	0	0	0	0	0	0
Total	20089	45320	50758	108581	11574	20948	123827	74151
Modified	<0.006%	<0.006%	<0.006%	<0.006%	<0.009%	<0.006%	<0.006%	<0.006%
P-value								
Specificity								
OffA-27								
Indels	0	0	0	0	1	0	0	0
Total	47338	14352	21253	17777	26512	19483	43728	29469
Modified	<0.006%	<0.007%	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%
P-value								
Specificity								
OffA-29								
Indels	0	0	0	0	0	0	0	0
Total	5174	12618	36909	18063	16486	17934	9999	35072
Modified	<0.019%	<0.008%	<0.006%	<0.006%	<0.006%	<0.006%	<0.010%	<0.006%
P-value								
Specificity								
OffA-30								
Indels	4	4	0	7	4	4	0	3
Total	45082	56631	36333	88651	69652	20362	29180	21350
Modified	0.009%	0.007%	<0.006%	0.008%	<0.006%	0.020%	<0.006%	0.014%
P-value								
Specificity								

OffA-32								
Indels	0	0	0	0	0	0	0	0
Total	13405	6721	14013	7513	14136	22376	6407	13720
Modified	<0.007%	<0.015%	<0.007%	<0.013%	<0.007%	<0.006%	<0.016%	<0.007%
P-value								
Specificity								
OffA-33								
Indels	0	0	0	0	1	1	0	4
Total	108222	46866	157329	48611	92559	152094	201408	225805
Modified	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%
P-value								
Specificity								
OffA-34								
Indels	0	0	0	0	0	0	0	2
Total	3889	3158	2903	2235	2112	3022	2322	2481
Modified	<0.026%	<0.032%	<0.034%	<0.045%	<0.047%	<0.033%	<0.043%	0.081%
P-value								
Specificity								
OffA-35								
Indels	0	0	0	1	0	0	0	33
Total	48482	37431	38043	31033	44803	37257	41073	47273
Modified	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	0.070%
P-value								7.6E-11
Specificity		>0	>274	>1302	>2564	>1016	>2428	260
OffA-36								
Indels	0	0	2	0	0	0	0	0
Total	27115	17075	45425	35059	22298	19610	12620	27170
Modified	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	<0.006%	<0.008%	<0.006%
P-value								
Specificity								

Supplementary Table S7. Cellular Modification Induced by CCR5A and ATM TALENs at On-Target and Predicted Off-Target Genomic Sites. (A) Results from sequencing CCR5A on-target and each predicted genomic off-target site that amplified from 50ng genomic DNA isolated from human cells treated with either no TALEN or TALENs containing canonical, Q3 or Q7 C-terminal domains, and either EL/KK heterodimeric, ELD/KKR heterodimeric, or homodimeric (Homo) *Fok*I domains. Indels: the number of observed sequences containing insertions or deletions consistent with TALEN-

induced cleavage. Total: total number of sequence counts. Modified: number of indels divided by total number of sequences as percentages. Upper limits of potential modification were calculated for sites with no observed indels by assuming there is less than one indel then dividing by the total sequence count to arrive at an upper limit modification percentage, or taking the theoretical limit of detection (1/16,400), whichever value was larger. For the OffC-5 site, the limit of detection was 1/32,800 (see Online Methods). P-values: calculated as previously reported[18] using a (right) one-sided Fisher's exact test between each canonical C-terminal domain TALEN-treated sample and the untreated control sample. P-values of < 0.00025 were considered significant and shown. Cut off at 0.00025 was based on multiple comparison correction from the Benjamini-Hochberg method.[21, 22] For sites with a significant P-value in a canonical C-terminal domain TALEN-treated sample, significant P-values for other TALEN-treated samples are shown in parenthesis. Specificity is the ratio of on-target to off-target genomic modification frequency for each site. (B) Same as (A) for the *ATM* target sites; P-values < 0.005 were considered significant and are shown.

C-term. Domain:	No TALEN	Q7	Q3	Canonical
FokI Domain:	No TALEN	ELD/KKR	ELD/KKR	ELD/KKR
HDAC1 Sites				
OnHDAC				
Indels	0	697	7774	4595
Total	55902	49232	127610	88068
Modified	< 0.001%	1.416%	6.092%	5.218%
P-value		2.98E-229	< 1.00E-250	< 1.00E-250
Specificity				
OffHDAC-1				
Indels	0	2	30	169
Total	261086	292020	239664	322463
Modified	< 0.001%	< 0.001%	0.013%	0.052%
P-value			2.51E-10	4.50E-44
Specificity		> 1160	487	100
PMS2 Sites				
OnPMS				
Indels	1	591	4364	6941
Total	17023	20623	29496	34092
Modified	0.006%	2.866%	14.795%	20.360%
P-value		4.31E-151	< 1.00E-250	< 1.00E-250
Specificity				
OffPMS-1				
Indels	2	10	485	3740
Total	259211	260985	270713	259935
Modified	< 0.001%	0.004%	0.179%	1.439%
P-value		3.86E-02	1.24E-137	< 1.00E-250
Specificity		748	83	14
OffPMS-2				
Indels	1	6	334	3000
Total	63593	67027	81473	76343
Modified	0.002%	0.009%	0.410%	3.930%
P-value			8.22E-82	< 1.00E-250
Specificity		320	36	5
SDHD sites				
OnSDHD				
Indels	0	263	16217	24188
Total	94384	93203	57748	73085
Modified	< 0.001%	0.282%	28.082%	33.096%

P-value		1.55E-80	< 1.00E-250	< 1.00E-250
Specificity				
OffSDHD-1				
Indels	0	0	31	191
Total	65095	70344	79721	77706
Modified	< 0.001%	< 0.001%	0.039%	0.246%
P-value			9.88E-09	4.48E-51
Specificity		> 231	722	135
OffSDHD-2				
Indels	0	0	0	0
Total	146373	142097	153028	153595
Modified	< 0.001%	< 0.001%	< 0.001%	< 0.001%
P-value				
Specificity				
OffSDHD-3				
Indels	0	0	1	2
Total	86283	85294	114095	87973
Modified	< 0.001%	< 0.001%	< 0.001%	0.002%
P-value				
Specificity				
OffSDHD-4				
Indels	0	0	0	0
Total	53497	57541	65379	62535
Modified	< 0.001%	< 0.001%	< 0.001%	< 0.001%
P-value				
Specificity				

Supplementary Table S8. Cellular Modification Induced by HDAC1, PMS2, and SDHD TALENs at On-Target and Off-Target Genomic Sites. (A) Results from sequencing HDAC1, PMS2, and SDHD on-target and genomic off-target sites amplified from 250 ng of genomic DNA isolated from human cells treated with either no TALEN, or with corresponding TALENs containing canonical, Q3, or Q7 C-terminal domains with ELD/KKR heterodimeric *Fok*I domains. Indels: the number of observed sequences containing insertions or deletions consistent with TALEN-induced cleavage. Total: total number of sequence counts. Modified: number of indels divided by total number of sequences as percentages. Upper limits of potential modification were calculated for sites with no observed indels by assuming there is less than one indel then dividing by the total sequence count to arrive at an upper limit modification percentage, or taking the theoretical limit of detection (1/82,000), whichever value was larger. P-values: calculated as previously reported[18] using a (right) one-sided Fisher's exact test between each canonical C-terminal domain TALEN-treated sample and the untreated control sample. P-values of < 0.025 were considered significant and shown. Specificity is the ratio of on-target to off-target genomic modification frequency for each site.

A

C-terminal domain:		No TALEN				Canonical		
FokI domain:		No TALEN				Homo		
TALENoffer site	Classifier site	Mut.	Indels	Total	Indels	Total	% Modified	P-value
OnCCR5A		0	0	9997	3006	9773	30.758%	< 1.0E-250
CP_CCRoff-1		9	0	9135	2	17468	0.011%	
CP_CCRoff-2		9	0	63390	0	36666	< 0.006%	
CP_CCRoff-3		10	14	11594	7	10460	0.067%	
CP_CCRoff-4	OffC-19	8	0	28732	0	26314	< 0.006%	
CP_CCRoff-5		8	0	21413	0	21538	< 0.006%	
CP_CCRoff-6		9	0	22764	0	28279	< 0.006%	
CP_CCRoff-8		9	2	24054	0	25702	< 0.006%	
CP_CCRoff-9	OffC-1	9	1	68283	6	59642	0.010%	
CP_CCRoff-10	OffC-2	9	0	68421	4	59558	0.007%	
CP_CCRoff-11	OffC-3	9	0	68421	4	59558	0.007%	
CP_CCRoff-12		9	0	14364	0	13806	< 0.007%	
CP_CCRoff-13		9	0	15016	0	24585	< 0.006%	
CP_CCRoff-14		9	0	28025	1	27546	< 0.006%	
CP_CCRoff-15		9	0	16105	8	13019	0.061%	1.59E-03
CP_CCRoff-16		11	0	26453	1	29619	< 0.006%	
CP_CCRoff-17		9	0	21155	1	28839	< 0.006%	
CP_CCRoff-18	OffC-28	9	0	28111	52	12591	0.413%	2.34E-19
CP_CCRoff-19		8	0	35891	0	33962	< 0.006%	
CP_CCRoff-20		9	0	64345	1	118954	< 0.006%	
CP_CCRoff-21		8	0	14857	0	11150	< 0.009%	
CP_CCRoff-22		10	0	14368	0	37008	< 0.006%	
CP_CCRoff-23		9	3	22876	7	20671	0.034%	
CP_CCRoff-24		10	4	129051	1	50695	< 0.006%	
CP_CCRoff-25		10	0	0	1	39845	< 0.006%	
CP_CCRoff-26		8	0	21677	0	24695	< 0.006%	
CP_CCRoff-27		11	0	24710	0	43452	< 0.006%	
CP_CCRoff-28		10	0	10269	0	11496	< 0.009%	
CP_CCRoff-29		8	3	121960	2	100659	< 0.006%	
CP_CCRoff-30		11	0	18320	0	12945	< 0.008%	
CP_CCRoff-31		10	0	74541	1	89994	< 0.006%	
CP_CCRoff-32		9	0	29550	1	102087	< 0.006%	

B

C-terminal	No TALEN	Canonical

domain:

FokI domain:		No TALEN		Homo		
Classifier site	Mut.	Indels	Total	Indels	Total	P-value
OffC-1	9	0	51248	1	87496	
OffC-2	9	6	124356	0	163322	
OffC-3	9	6	124356	0	163322	
OffC-4	11	0	45377	0	40452	
OffC-5	11	0	27009	395	17077	1.20E-164
OffC-6	10	0	10766	0	6560	
OffC-9	10	0	40603	0	30771	
OffC-13	11	0	65518	0	78546	
OffC-16	9	0	36228	0	32636	
OffC-17	10	0	32112	0	31299	
OffC-19	8	1	22868	0	28478	
OffC-20	10	0	23335	0	18972	
OffC-21	11	0	34302	0	21161	
OffC-22	10	1	81037	0	104857	
OffC-28	8	0	28111	52	12591	2.34E-19
OffC-30	11	0	11840	0	6285	
OffC-32	11	0	1944	0	19115	
OffC-34	11	0	9052	0	9072	
OffC-35	10	0	23839	0	11897	
OffC-36	11	1	23412	5	18052	
OffC-38	10	0	16396	9	13351	7.38E-04
OffC-39	9	0	51962	0	13562	
OffC-40	9	0	3910	0	24711	
OffC-49	9	0	26333	27	24497	2.74E-09
OffC-56	9	0	25357	0	26999	
OffC-65	9	0	14364	0	13806	
OffC-69	9	0	10407	25	27950	4.90E-04
OffC-76	9	0	61760	59	39617	8.19E-25
OffC-137	9	0	26470	0	23318	
OffC-150	9	0	22058	5	20952	

C

C-terminal domain	No TALEN	Q7	Q3	Canonical
FokI domain	No TALEN	Homo	Homo	Homo
CCR5A Sites				
OnC				
Indels	0	1023	3301	3006
Total	9997	9674	9768	9773
% Modified	< 0.001%	10.575%	33.794%	30.758%
P-value		< 1.00E-250	< 1.00E-250	< 1.00E-250

Specificity				
OffC-5				
Indels	0	0	49	487
Total	30499	31224	25581	30839
% Modified	< 0.001%	< 0.001%	0.192%	1.579%
P-value			2.03E-17	3.11E-145
Specificity		> 8670	176	19
OffC-15				
Indels	1	0	0	3
Total	14048	14607	10045	11712
% Modified	0.007%	< 0.001%	< 0.001%	0.026%
P-value				
Specificity				
OffC-16				
Indels	0	0	1	26
Total	97355	91183	75927	87049
% Modified	< 0.001%	< 0.001%	0.001%	0.030%
P-value				3.35E-09
Specificity		>8670	> 27700	1030
OffC-28				
Indels	0	1	16	94
Total	39024	44118	27961	27241
% Modified	< 0.001%	0.002%	0.057%	0.345%
P-value			8.52E-07	5.65E-37
Specificity		4665	591	89
OffC-36				
Indels	1	3	4	10
Total	28685	32915	24938	26273
% Modified	0.003%	0.009%	0.016%	0.038%
P-value				4.66E-03
Specificity		1160	2107	808

Supplementary Table S9. Cellular Modification Induced by CCR5A TALENs at On-Target and Predicted Off-target Genomic Sites Generated by TALENoffer and Generated by This Study. (A) Results from sequencing CCR5A on-target and each genomic off-target site predicted by TALENoffer that amplified from 50 ng genomic DNA isolated from human cells treated with either no TALEN or TALENs containing canonical C-terminal domains and homodimeric (Homo) *Fok*I domains. Indels: the number of observed sequences containing insertions or deletions consistent with TALEN-induced cleavage. Total: total number of sequence counts. Modified: number of indels divided by total number of sequences, expressed as percentages. Upper limits of potential modification were calculated for sites with no observed indels by assuming there is less than one indel then dividing by the total sequence

count to arrive at an upper limit modification percentage, or taking the theoretical limit of detection (1/16,400), whichever value was larger. P-values: calculated as previously reported[18] using a (right) one-sided Fisher's exact test between each canonical C-terminal domain TALEN-treated sample and the untreated control sample. P-values of < 0.00161 were considered significant and are shown. The significance cut off of 0.00161 was based on the multiple comparison correction from the Benjamini-Hochberg method.[21, 22] Specificity is the ratio of on-target to off-target genomic modification frequency for each site. (B) Same as (A) for the next 10 best-scoring amplified off-target sites with nine or ten mutations (OffC-38 to OffC-150) after the first 36 top-scoring genomic off-target site predicted by the classifer; P-values < 0.0040 were considered significant and are shown.. Results from sequencing CCR5A on-target and each genomic off-target site with eight to 11 mutations predicted by the classifier were taken from Supplementary Table S7A. (C) Results from sequencing CCR5A on-target and significantly modified genomic off-target site from Figure 3. 250ng of genomic DNA isolated from human cells treated with either no TALEN or TALENs containing canonical, Q3 or Q7 C-terminal domains, with homodimeric (Homo) *Fok*I domains. The theoretical limit of detection was (1/82,000) and P-values < 0.025 were considered significant and are shown. (C) is a biologically independent sample from the sample used to generate the results in Supplementary Table S7. The sites assayed in (A), (B) and (C) are all from the same biological sample (with the exception noted in (B)).

TALEN selection	a	b	R^2
L13+R10 CCR5B	1.00	-1.88	0.999937
L10+R10 CCR5B	1.00	-1.85	0.999901
L10+R13 CCR5B	1.00	-1.71	0.999822
L13+R13 CCR5B	1.00	-1.64	0.999771
L13+R16 CCR5B	1.00	-1.15	0.998286
L16+R10 CCR5B	1.00	-1.24	0.998252
L10+R16 CCR5B	1.01	-1.08	0.996343
L16+R13 CCR5B	1.01	-1.04	0.995844
L16+R16 CCR5B	1.03	-0.70	0.977880
L18+R18 ATM	1.08	-0.36	0.913087
L18+R18 CCR5A	1.13	-0.21	0.798923

Supplementary Table S10. Exponential Fitting of Enrichment Values as Function of Mutation Number. Enrichment values of post-selection sequences as function of mutation were normalized relative to on-target enrichment (= 1.0 by definition). Normalized enrichment values of sequences with zero to four mutations were fit to an exponential function, $a*e^b$, with R^2 reported using the non-linear least squares method.

TALEN selection	Range	a	b	R^2
L16+R16 CCR5B	3-5	1.00	-1.638	0.99998
L16+R13 CCR5B	2-4	1.00	-1.733	0.99998
L16+R10 CCR5B	2-4	1.00	-2.023	0.99999
L13+R16 CCR5B	2-4	1.00	-1.844	0.99997
L13+R13 CCR5B	1-3	1.00	-2.014	0.99998
L13+R10 CCR5B	1-3	1.00	-2.205	0.99999
L10+R16 CCR5B	2-4	1.00	-1.929	0.99995
L10+R13 CCR5B	1-3	1.00	-2.110	0.99998
L10+R10 CCR5B	1-3	1.00	-2.254	0.99999

Supplementary Table S11. Exponential Fitting and Extrapolation of Enrichment Values as Function of Mutation Number. Enrichment values of all sequences from all nine of the CCR5B selections as function of mutation number were normalized relative to enrichment values of sequences with the lowest mutation number in the range shown (= 1.0 by definition). Normalized enrichment values of sequences from the range of mutations specified were fit to an exponential function, $a*e^b$, with R^2 reported utilizing the non-linear least squares method. These exponential decrease, b, were used to extrapolate all mean enrichment values beyond five mutations.

A

oligonucleotide name	oligonucleotide sequence (5'->3')
CCR5Aon-Circ	CACCACTNTTCATTACACCTGCAGCTNNNNNNNNNNNNAGTATCAATTCTGGAAGANCGTCACGCT
CCR5Amut-Circ	CACCACTNTTCATTACACCTGCAGCTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNCGTCACGCT
TAL-Nrev	5Phos/CAGCAGCTGCCCGGT
TAL-N1fwd	5Phos/cAGATCGCGAAGAGAGGGGGAGTAACAGCGGTAG
TAL-N2fwd	5Phos/cAGATCGCGcAGAGAGGGGGAGTAACAGCGGTAG
TAL-N3fwd	5Phos/cAGATCGCGcAGcagGGGGGAGTAACAGCGGTAG
TAL-Cifwd	ATC GTA GCC CAA TTG TCC A
TAL-Cirev	GTTGGTTCTTTGGATCAATGCG
TAL-Q3	AAGTTCTCTCGGGAATCCGTTGGTTGGTTCTTTGGATCA
TAL-Q7	GAAGTTCTCTCGGGAATTTGTTGGTTGGTTTGTTGGATCAATGCGGGAGCATGAGGCAGACCTTGTTGGACTGCATC
TAL-Ciirev	CTTTTGACTAGTTGGGATCCCGCGACTTGATGGGAAGTTCTCTCGGGAAT
CCR5A Library10	5Phos/CACCACTNT%T%C%A%T%T%A%C%A%C%C%T%G%C%A%G%C%T%NNNNNNNNNNA%G%T%A%T%C%A%A%T%T%C%T%G%G%A%A%G%A%NCGTCACGCT
CCR5A Library12	5Phos/CACCACTNT%T%C%A%T%T%A%C%A%C%C%T%G%C%A%G%C%T%NNNNNNNNNNNNA%G%T%A%T%C%A%A%T%T%C%T%G%G%A%A%G%A%NCGTCACGCT
CCR5A Library14	5Phos/CACCACTNT%T%C%A%T%T%A%C%A%C%C%T%G%C%A%G%C%T%NNNNNNNNNNNNNNA%G%T%A%T%C%A%A%T%T%C%T%G%G%A%A%G%A%NCGTCACGCT
CCR5A Library16	5Phos/CACCACTNT%T%C%A%T%T%A%C%A%C%C%T%G%C%A%G%C%T%NNNNNNNNNNNNNNNA%G%T%A%T%C%A%A%T%T%C%T%G%G%A%A%G%A%NCGTCACGCT
CCR5A Library18	5Phos/CACCACTNT%T%C%A%T%T%A%C%A%C%C%T%G%C%A%G%C%T%NNNNNNNNNNNNNNNNA%G%T%A%T%C%A%A%T%T%C%T%G%G%A%A%G%A%NCGTCACGCT
CCR5A Library20	5Phos/CACCACTNT%T%C%A%T%T%A%C%A%C%C%T%G%C%A%G%C%T%NNNNNNNNNNNNNNNNNA%G%T%A%T%C%A%A%T%T%C%T%G%G%A%A%G%A%NCGTCACGCT
CCR5A Library22	5Phos/CACCACTNT%T%C%A%T%T%A%C%A%C%C%T%G%C%A%G%C%T%NNNNNNNNNNNNNNNNNNA%G%T%A%T%C%A%A%T%T%C%T%G%G%A%A%G%A%NCGTCACGCT
CCR5A Library24	5Phos/CACCACTNT%T%C%A%T%T%A%C%A%C%C%T%G%C%A%G%C%T%NNNNNNNNNNNNNNNNNNNNNA%G%T%A%T%C%A%A%T%T%C%T%G%G%A%A%G%A%NCGTCACGCT
CCR5B Library10	5Phos/CCACGCTNT%C%T%T%C%A%T%T%A%C%A%C%C%T%G%C%NNNNNNNNNNC%A%T%A%C%A%G%T%C%A%G%T%A%T%C%A%NCCTCGGGACT
CCR5B Library12	5Phos/CCACGCTNT%C%T%T%C%A%T%T%A%C%A%C%C%T%G%C%NNNNNNNNNNNNC%A%T%A%C%A%G%T%C%A%G%T%A%T%C%A%NCCTCGGGACT
CCR5B Library14	5Phos/CCACGCTNT%C%T%T%C%A%T%T%A%C%A%C%C%T%G%C%NNNNNNNNNNNNNNC%A%T%A%C%A%G%T%C%A%G%T%A%T%C%A%NCCTCGGGACT
CCR5B Library16	5Phos/CCACGCTNT%C%T%T%C%A%T%T%A%C%A%C%C%T%G%C%NNNNNNNNNNNNNNNNC%A%T%A%C%A%G%T%C%A%G%T%A%T%C%A%NCCTCGGGACT
CCR5B Library18	5Phos/CCACGCTNT%C%T%T%C%A%T%T%A%C%A%C%C%T%G%C%NNNNNNNNNNNNNNNNNC%A%T%A%C%A%G%T%C%A%G%T%A%T%C%A%NCCTCGGGACT
CCR5B Library20	5Phos/CCACGCTNT%C%T%T%C%A%T%T%A%C%A%C%C%T%G%C%NNNNNNNNNNNNNNNNNNC%A%T%A%C%A%G%T%C%A%G%T%A%T%C%A%NCCTCGGGACT
CCR5B Library22	5Phos/CCACGCTNT%C%T%T%C%A%T%T%A%C%A%C%C%T%G%C%NNNNNNNNNNNNNNNNNNNC%A%T%A%C%A%G%T%C%A%G%T%A%T%C%A%NCCTCGGGACT
CCR5B Library24	5Phos/CCACGCTNT%C%T%T%C%A%T%T%A%C%A%C%C%T%G%C%NNNNNNNNNNNNNNNNNNNNNC%A%T%A%C%A%G%T%C%A%G%T%A%T%C%A%NCCTCGGGACT
ATM Library10	Phos/CTCCGCGTNT%G%A%A%T%T%G%G%G%A%T%G%C%T%G%T%T%T%NNNNNNNNNNT%T%T%A%T%T%T%T%A%C%T%G%T%C%T%T%T%A%GGTACCCCA
ATM Library12	5Phos/CTCCGCGTNT%G%A%A%T%T%G%G%G%A%T%G%C%T%G%T%T%T%NNNNNNNNNNNNT%T%T%A%T%T%T%T%A%C%T%G%T%C%T%T%T%A%GGTACCCCA
ATM Library14	5Phos/CTCCGCGTNT%G%A%A%T%T%G%G%G%A%T%G%C%T%G%T%T%T%NNNNNNNNNNNNNNT%T%T%A%T%T%T%T%A%C%T%G%T%C%T%T%T%A%GGTACCCCA
ATM Library16	5Phos/CTCCGCGTNT%G%A%A%T%T%G%G%G%A%T%G%C%T%G%T%T%T%NNNNNNNNNNNNNNNNT%T%T%A%T%T%T%T%A%C%T%G%T%C%T%T%T%A%GGTACCCCA
ATM Library18	5Phos/CTCCGCGTNT%G%A%A%T%T%G%G%G%A%T%G%C%T%G%T%T%T%NNNNNNNNNNNNNNNNNT%T%T%A%T%T%T%T%A%C%T%G%T%C%T%T%T%A%GGTACCCCA
ATM Library20	5Phos/CTCCGCGTNT%G%A%A%T%T%G%G%G%A%T%G%C%T%G%T%T%T%NNNNNNNNNNNNNNNNNNT%T%T%A%T%T%T%T%A%C%T%G%T%C%T%T%T%A%GGTACCCCA

ATM Library22	5Phos/CTCCGCGTNT%G%A%A%T%T%G%G%G%A%T%G%C%T%G%T%T%T%NNNNNNNNNNNNNNNNNNNNNNT%T%T%A%T%T%T%T%A%C%T%G%T%C%T%T%T%A%GGTACCCCA
ATM Library24	5Phos/CTCCGCGTNT%G%A%A%T%T%G%G%G%A%T%G%C%T%G%T%T%T%NNNNNNNNNNNNNNNNNNNNNNT%T%T%A%T%T%T%T%A%C%T%G%T%C%T%T%T%A%GGTACCCCA
#1 adapter-fwd**1	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTACTGT
#1 adapter-rev**1	ACAGTAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGG
#1 adapter-fwd**2	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTCTGAA
#1 adapter-rev**2	TTCAGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGG
#1 adapter-fwd**3	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTTGCAA
#1 adapter-rev**3	TTGCAAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGG
#1 adapter-fwd**4	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTTGACT
#1 adapter-rev**4	AGTCAAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGG
#1 adapter-fwd**5	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTGCATT
#1 adapter-rev**5	AATGCAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGG
#1 adapter-fwd**6	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTCATGA
#1 adapter-rev**6	TCATGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGG
#1 adapter-fwd**7	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTATGCT
#1 adapter-rev**7	AGCATAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGG
#1 adapter-fwd**8	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTCTAGT
#1 adapter-rev**8	ACTAGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGG
#1 adapter-fwd**9	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTGCTAA
#1 adapter-rev**10	TTAGCAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGG
#1 adapter-fwd**10	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTCAGTA
#1 adapter-rev**11	TACTGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGG
#1 adapter-fwd**11	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTGTACT
#1 adapter-rev**12	AGTACAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGG
#1 adapter-fwd**12	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTACTGT
#1 adapter-rev**13	ACAGTAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGG
#1 adapter-fwd**13	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTGCTAA
#1 adapter-rev**14	TTAGCAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGG
#1 adapter-fwd**14	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTCAGTA
#1 adapter-rev**14	TACTGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGG
#1 adapter-fwd**15	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTGTACT
#1 adapter-rev**15	AGTACAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGG
#1 adapter-fwd**16	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTACTGT
#1 adapter-rev**16	ACAGTAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGG
#2A primer-fwd	AATGATACGGCGACCAC
#2A primer-rev*CCR5A	GTTCAGACGTGTGCTCTTCCGATCTNNNNAGTGGTGAGCGTGACG
#2A primer-rev*ATM	GTTCAGACGTGTGCTCTTCCGATCTNNNNACGCGGAGTGGGGTACC
#2A primer-rev*CCR5B	CAGACGTGTGCTCTTCCGATCNNNNAGCGTGGAGTCCCGAGG
#2B primer-fwd	AATGATACGGCGACCAC
#2B primer-rev**1	CAAGCAGAAGACGGCATACGAGATTGTTGACTGTGACTGGAGTTCAGACGTGTGCTCTTC
#2B primer-rev**2	CAAGCAGAAGACGGCATACGAGATACGGAACTGTGACTGGAGTTCAGACGTGTGCTCTTC
#2B primer-rev**3	CAAGCAGAAGACGGCATACGAGATTCTAACATGTGACTGGAGTTCAGACGTGTGCTCTTC
#2B primer-rev**4	CAAGCAGAAGACGGCATACGAGATCGGGACGGGTGACTGGAGTTCAGACGTGTGCTCTTC
	CAAGCAGAAGACGGCATACGAGATCGTGATGTGACTGGAGTTCAGACGTGTGCTCTTCCG
#1 Lib. adapter-fwd*CCR5A	GTACCCAGATCGGAAGAGCACACGTCTGAACTCCAGTCACACAGTGATCTCGTATGCCGTCTTCTGCTTG
#1 Lib. adapter-rev*CCR5A	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTG
#1 Lib. adapter-fwd*ATM	GTACGATGCGATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGCCGTCTTCTGCTTG
#1 Lib. adapter-rev*ATM	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCGCATC
#1 Lib. adapter-fwd*CCR5B	TCGGGAACGTGATCGGAAGAGCACACGTCTGAACTCCAGTCACCGTCTAATCTCGTATGCCGTCTTCTGCTTG
#1 Lib. adapter-rev*CCR5B	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCACGTT

#2A Lib. primer-rev	CAAGCAGAAGACGGCATACGA
#2A Lib. primer-fwd*CCR5A	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCGTCACGCTCACCACT
#2A Lib. primer-fwd*ATM	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGGTACCCCACTCCGCGT
#2A Lib. primer-fwd*CCR5B	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCCTCGGGACTCCACGCT
#2B Lib. primer-rev	CAAGCAGAAGACGGCATACGA
#2B Lib. primer-fwd	AATGATACGGCGACCAC
G adapter-fwd	ACACTCTTTCCCTACACGACGCTCTTCCGATCT
G adapter-rev	/5Phos/GATCGGAAGAGCACACGTCTGAACTCCA
G-B primer-fwd	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGAC
G-B primer-rev**1	CAAGCAGAAGACGGCATACGAGATGTGCGGACGTGACTGGAGTTCAGACGTGTGCT
G-B primer-rev**2	CAAGCAGAAGACGGCATACGAGATCGTTTCACGTGACTGGAGTTCAGACGTGTGCT
G-B primer-rev**3	CAAGCAGAAGACGGCATACGAGATAAGGCCACGTGACTGGAGTTCAGACGTGTGCT
G-B primer-rev**4	CAAGCAGAAGACGGCATACGAGATTCCGAAACGTGACTGGAGTTCAGACGTGTGCT
G-B primer-rev**5	CAAGCAGAAGACGGCATACGAGATTACGTACGGTGACTGGAGTTCAGACGTGTGCT
G-B primer-rev**6	CAAGCAGAAGACGGCATACGAGATATCCACTCGTGACTGGAGTTCAGACGTGTGCT
G-B primer-rev**7	CAAGCAGAAGACGGCATACGAGATAAAGGAATGTGACTGGAGTTCAGACGTGTGCT
G-B primer-rev**8	CAAGCAGAAGACGGCATACGAGATATATCAGTGTGACTGGAGTTCAGACGTGTGCT
CCR5AonCfwd	CGACGGTCTAGAGTCTTCATTACACCTGCAGCTCTCATTTTCCATACAGT
CCR5Amut1fwd	CGACGGTCTAGAGTCTTCATTACAtCTGCAcCTCTCATTTTCCATACAGT
CCR5Amut2fwd	CGACGGTCTAGAGTCTTCAaTACACCTGtAGCTCTCATTTTCCATACAGT
CCR5Amut3fwd	CGACGGTCTAGAGTCTTCgTTACACCTGCAtCTCTCATTTTCCATACAGT
CCR5Amut4fwd	CGACGGTCTAGAGTCTTaATTgCACCTGCAGCTCTCATTTTCCATACAGT
CCR5AonCrev	CCGACGAAGCTTTTCTTCCAGAATTGATACTGACTGTATGGAAAATGA
CCR5Amut1rev	CCGACGAAGCTTTTCTTaCAGAATTcATACTGACTGTATGGAAAATGA
CCR5Amut2rev	CCGACGAAGCTTTTCcTCCAGAgTTGATACTGACTGTATGGAAAATGA
CCR5Amut3rev	CCGACGAAGCTTTTCTTCCtGAATTGATAaTGACTGTATGGAAAATGA
CCR5Amut4rev	CCGACGAAGCTTTTCTTCCAGcATTGtTACTGACTGTATGGAAAATGA
ATMonAfwd	CGACGGTCTAGATTTGAATTGGGATGCTGTTTTTAGGTATTCTATTCAAATT
ATMmut1fwd	CGACGGTCTAGATTTGAATTGGGtTGCTGTTTTTAGGTATTCTATTCAAATT
ATMmut2fwd	CGACGGTCTAGATTTGAATTGcGATGCTGTTTTTAGGTATTCTATTCAAATT
ATMmut3wd	CGACGGTCTAGATTTGAgTTGGGATGCTGTTTTTAGGTATTCTATTCAAATT
ATMmut4fwd	CGACGGTCTAGATTTGAATTGGGATGCTGaTTTTAGGTATTCTATTCAAATT
ATMonArev	CCGACGAAGCTTAATAAAGACAGTAAAATAAATTTGAATAGAATACCTAAAA
ATMmut1rev	CCGACGAAGCTTAATAAAGACAGTgAAATAAATTTGAATAGAATACCTAAAA
ATMmut2rev	CCGACGAAGCTTAATAAAGAtAGTAAAATAAATTTGAATAGAATACCTAAAA
ATMmut3rev	CCGACGAAGCTTAATAAAGACAGTAAgATAAATTTGAATAGAATACCTAAAA
ATMmut4rev	CCGACGAAGCTTAATAAcGACAGTAAAATAAATTTGAATAGAATACCTAAAA
CCR5BonBfwd	CGACGGTCTAGAAAGGTCTTCATTACACCTGCAGCTCTCATTTTCCATACAGTCA
CCR5Bmut1fwd	CGACGGTCTAGAGTCTTCATTACACCTGtAGCTCTCATTTTC
CCR5Bmut2fwd	CGACGGTCTAGAGTCTTCATaACACCTGCAGCTCTCATTTTC
CCR5Bmut3fwd	CGACGGTCTAGAGTCTTCATTACACCcGCAGCTCTCATTTTC
CCR5Bmut4fwd	CGACGGTCTAGAGTCTTCATaACACCTGtAGCTCTCATTTTC
CCR5Bmut5fwd	CGACGGTCTAGAGTCTTCATTAtACCTaCAGCTCTCATTTTC
CCR5Bmut6fwd	CGACGGTCTAGAGTCTTCATTgCACCcGCAGCTCTCATTTTC
CCR5BonBrev	CCGACGAAGCTTTCTTCCAGAATTGATACTGACTGTATGGAAAATGAGAGCT
CCR5Bmut1rev	CCGACGAAGCTTTCTTCCAGAATTGATACTaACTGTATGGAAAATGAGAGCT
CCR5Bmut2rev	CCGACGAAGCTTTCTTCCAGAATTGATACTGACTGTATcGAAAATGAGAGCT
CCR5Bmut3rev	CCGACGAAGCTTTCTTCCAGAATTGATACTGACTGaATGGAAAATGAGAGCT
CCR5Bmut4rev	CCGACGAAGCTTTCTTCCAGAATTGATACcGACTGTATGGAAAATGAGAGCT
CCR5Bmut5rev	CCGACGAAGCTTTCTTCCAGAATTGATACTaACTGTATcGAAAATGAGAGCT
CCR5Bmut6rev	CCGACGAAGCTTTCTTCCAGAATTGATACTGAaTGTgTGGAAAATGAGAGCT
CCR5Bmut7rev	CCGACGAAGCTTTCTTCCAGAATTGATACTGAtaGTATGGAAAATGAGAGCT
pUC19Ofwd	GCGACACGGAAATGTTGAATACTCAT
pUC19Orev	CAGCGAGTCAGTGAGCGA

B

DNA substrate name	Oligonucleotide Combination
A1	ATMmut1fwd + ATMonArev
A2	ATMmut2fwd + ATMonArev
A3	ATMonAfwd +ATMmut1rev
A4	ATMonAfwd +ATMmut2rev
A5	ATMmut2fwd +ATMmut2rev
A6	ATMmut3fwd +ATMmut3rev
A7	ATMmut1fwd +ATMmut1rev
A8	ATMmut4fwd +ATMmut4rev
C1	CCR5Amut1fwd + CCR5AonCrev
C2	CCR5Amut2fwd + CCR5AonCrev
C3	CCR5Amut3fwd + CCR5AonCrev
C4	CCR5Amut4fwd + CCR5AonCrev
C5	CCR5AonAfwd + CCR5Amut1rev
C6	CCR5AonAfwd + CCR5Amut2rev
C7	CCR5AonAfwd + CCR5Amut3rev
C8	CCR5AonAfwd + CCR5Amut4rev
B1	CCR5Bmut1fwd + CCR5BonBrev
B2	CCR5Bmut2fwd + CCR5BonBrev
B3	CCR5Bmut3fwd + CCR5BonBrev
B4	CCR5BonBfwd + CCR5Bmut1rev
B5	CCR5BonBfwd + CCR5Bmut2rev
B6	CCR5BonBfwd + CCR5Bmut3rev
B7	CCR5BonBfwd + CCR5Bmut4rev
B8	CCR5Bmut4fwd + CCR5BonBrev
B9	CCR5Bmut5fwd + CCR5BonBrev
B10	CCR5Bmu6fwd + CCR5BonBrev
B11	CCR5BonBfwd + CCR5Bmut5rev
B12	CCR5BonBfwd + CCR5Bmut6rev
B13	CCR5BonBfwd + CCR5Bmut7rev
B14	CCR5Bmut1fwd + CCR5Bmut1rev
B15	CCR5Bmut2fwd + CCR5Bmut2rev
B16	CCR5Bmut1fwd + CCR5Bmut3rev

C

Site	Fwd primer	Rev primer	PCR
OnCCR5A	TCACTTGGGTGGTGGCTGTG	GACCATGACAAGCAGCGGCA	
OffC-1	AGTCCAAGACCAGCCTGGGG	AAGAACCTGTTGTCTAATCCAGCA	
OffC-2	GAACCTGTTGTCTAATCCAGCGTC	CTGCAAAGAAGGCCAGGCA	
OffC-3	GAACCTGTTGTCTAATCCAGCGTC	CTGCAAAGAAGGCCAGGCA	
OffC-4	TGACCTGTTTGTTCAGGTCTTCC	CCATATGGTCCCTGTCGCAA	
OffC-5	TCCAGTTGCTGTCCCTTCAGA	ACAGGGAGAGCCACCAATGC	
OffC-6	GCCCGGCCTGTCCTGTATTT	CACCCACACATGCACTTCCC	
OffC-7	TGGCTATTCTAGTTCTTTTGCAT	CCATGCCCTAGGGATTTGTGGA	
OffC-8	CGCTGAAGGCTGTCACCCTAA	TGGACCTAAGAGTCCTGCCCAT	ND
OffC-9	CCACCACCACACAACTTCACA	CAGCTGGCGAGAACTGCAAA	
OffC-10	TTCCAGGTCCTTTGCACAAATA	GCAAGGTCGTTGGATAGAAGTTGA	
OffC-11	CACCGAAAGCAACCCATTCC	TGATCTGCCCACCCCAGACT	
OffC-12	TTCATTCTCACCATCTGGAATTGG	TCTGGCTGGACTGCTCTGGTT	
OffC-13	TGGCATGTGGATCAGTACCCA	TAGAACATGCCCGCGAACAG	
OffC-14	CTGACGTCCATGTCAACGGG	TTTGAATTCCCCCTCCCCAT	
OffC-15	GCTCCTTTCTGAGAAGCACCCAT	GGCAGATGGTGGCAGGTCTT	

OffC-16	ATGAGGGCTTGGATTGGCTG	CCACCTCCCCCACTGCAATA	
OffC-17	GGAGGCCTTCATTGTGTCACG	AACTCCACCTGGGTGCCCTA	
OffC-18	CGTGGTCCCCCAGAAATCAC	GGAGCAGGAGTTGGTGGCAT	
OffC-19	GATTGCATAGGTTAGCATTGCC	GCCCCTGTTGGTTGACTCCC	
OffC-20	TTCCAGCGAATGGAAAGTGCT	AAGCCCAGGAATAAGGGCCA	
OffC-21	AAGCATGCTCACACTGTGGTGTA	TTGCTTGAGGCGGAAGTTGC	
OffC-22	TGACCCTCCAGCAAAGGTGA	CCCCAGGGACTGAGCATGAG	
OffC-23	GCTTTGCTTGCACTGTGCCTT	GGGGACAGACTGTGAGGGCT	
OffC-24	TCAAAAGGATGTGATCTGCCACA	GGCCTCTTTGAGGGCCAGTT	
OffC-25	CCAGGGCTCAATTCTTAGACCG	AAAAGAGCAGGGCTGCCATC	
OffC-26	TGTTCATGCCTGCACAGTGG	TGGATGTGCCCTCTACCACA	
OffC-27	TTTGGCAAGGAATTCACAGTTC	TCATGCCTGCACAGTGGTTG	
OffC-28	GGAGGATGTCTTTGTGGTAGGGG	CGCTGCCAAGCAAACTCAAA	
OffC-29	TCCCCCAACTTCACTGTTTTT	GCAATGAGCATGTGGACACCA	
OffC-30	TTCTCTGTTTCCAGTGATTTCAGA	GTCGCAAAACAGCCAGTTGC	+DMSO
OffC-31	TGGCTTGGTTAATGGACAATGG	CCTGCAAGGAGCAAGGCTTC	+DMSO
OffC-32	TGGGCTTCGTTGACTTAAAGAG	GGACAAGAGGGCCAGGGTTT	
OffC-33	TCTTAAACATGTGGAACCCAGTCAT	TGAAAACCCACAGAGTGGGAGA	
OffC-34	GCAGATTCATTAGCGTTTGTGGC	TGCATGGGTGTAAATGTAGCAGAAA	
OffC-35	CCAAGGATCAATACCTTTGGAGGA	GCCCTCCCTTGAATCAGGCT	
OffC-36	TTCCCCTAACCAGGGGCAGT	GTGGTGAGTGGGTGTGGCAG	+DMSO
OffC-38	CGCCCATGAGAAAGAGTCCA	CTACACCCCCTCCCCAAAGG	
OffC-39	GCTGTCATTTCCAAAGCCGC	GGCCTCTAGAGTGAGGGGGTTG	
OffC-40	GCTGTCATTTCCAAAGCGGC	GGCCAACAAGATGCCAAGGT	
OffC-42	AGTTCTCAGCAACTAAAGTATTGA	CTGGGATTACAGGCGTGAGC	ND
OffC-45	CCATCATTGGCATCATGGGA	TGGAATGGAGTGGCCAACCT	ND
OffC-49	GTCAGGACCACACAAGAAAATAAAA	GGGAATGCCAGTCTTTGCCA	
OffC-56	CCCAGCCGATGACCAGAAAT	GCCAGCTGAGCTCTTTGCTGTA	
OffC-65	ACCCATTGATAAGGCACATTCT	AGTCAAACTCATCTAACACTCCAG	
OffC-69	TCTCCCACACCCCACATTCA	TGCGATAAGGTTGCAATGACA	
OffC-76	TAGGGCGCCTCAGATCCACT	GAGCTGCAGGCCTTGAATGG	
OffC-137	AGCCACAAGGGCCTGCTGTA	ACTGCGCCCAGCCTATCACT	
OffC-150	GCTTTTCTTTTCAGCCAAATGAAAGTT	ACGTGCGTCCCTGTCACTCA	
OnATM	AGCGCCTGATTCGAGATCCT	ATGCCAAATTCATATGCAAGGC	
OffA-1	CCTGCCATTGAATTCCAGCCT	TGTCTGCCTTTCCTGTCCCC	
OffA-2	GACTGCCACTGCACTCCCAC	GGATACCCTTGCCTCCCCAC	
OffA-3	CCTCCCATTTTCCTTCCTCCA	CTGGGAGACACAGGTGGCAG	
OffA-4	TCCTCCAATTTTCCTTCCTCCA	CTGGGAGACACAGGTGGCAG	
OffA-5	CTGGGAGACACAGGTGGCAG	AGGACCAATGGGGCCAATCT	
OffA-6	CTGGGAGACACAGGTGGCAG	AGGACCAATGGGGCCAATCT	
OffA-7	CTGGGAGACACAGGTGGCAG	AGGACCAATGGGGCCAATCT	
OffA-8	GCATGCCAAAGAAATTGTAGGC	TTCCCCCTGTCATGGTCTTCA	
OffA-9	GCATCTCTGCATTCCTCAGAAGTGG	AGAAACTGAGCAAGCCTCAGTCAA	
OffA-10	GGGATACCAAAGAGCTTTTGTTTTGTT	CAGAGGCTGCATGATGCCTAATA	
OffA-11	TGCAGCTACGGATGAAAACCAT	TCAGAATACCTCCCCGCCAG	
OffA-12	GCATAAAGCACAGGATGGGAGA	TCCCTCTTTAACGGTTATGTTGGC	
OffA-13	TGGGTTAAGTAATTTCGAAAGGAGAA	ATGTGCCCCACACATTGCC	+DMSO
OffA-14	GAGTGAGCCACTGCACCCAG	CGTGTGGTGGTGGCACAAG	ND
OffA-15	CCTCCCTCTGGCTCCCTCCC	ACCAGGGCCTGTTGGGGGTT	
OffA-16	TGCTCCCTGACCTTCCTGAGA	CCATTGGAATGAGAACCTTCTGG	

OffA-17	GGTGGAACAATCCACCTGTATTAGC	GAATGTGACACCACCACCGC	
OffA-18	GGCTTTGCAAACATAAACACTCA	CCTTCTGAGCAGCTGGGACAA	
OffA-19	CACTGGAACCCAGGAGGTGG	CCTCCCATTGGAGCCTTGGT	ND
OffA-20	CAGCCTGCCTGGGTGACAG	CATCTGAGCTCAAAACTGCTGC	+DMSO
OffA-21	GCCACTGCATTGCATTTTCC	TGAGGGCAGGTCTGTTTCCTG	
OffA-22	GGGAGGATCTCTCGAGTCCAGG	CCTTGCCTGACTTGCCCTGT	
OffA-23	TGTTTAGTAATTAAGACCCTGGCTTTC	GCGACAGGTACAAAGCAGTCCAT	
OffA-24	GCCCTTTGATTTCATCTGTTTCCC	CATTGCTGCCATTGCACTCC	
OffA-25	AAACTGGCACATGTACTCCT	ACATGATTTGATTTTTCATGTGTTT	
OffA-26	GGGTGGAAGGTGAGAGGAGATT	CGCAGATGGGCATGTTATTG	ND
OffA-27	CCTCCCATTTTCCTTCCTCCA	GACTGCCACTGCACTCCCAC	
OffA-28	AGCCAAGATTGCACCATTGC	GTCCCTGACGGAGGCTGAGA	ND
OffA-29	TGGTTGGATTTTGGCTCTGTCAC	TGTCAATATCAATACCCTGCTTTCCTC	
OffA-30	TGGTTACTTTTAAAGGGTCATGATGGA	AAAAATGGATGCAAAGCCAAA	+DMSO
OffA-31	GGGACACAGAGCCAAACCGT	TGTGCACATGTACCCTAAAACT	ND
OffA-32	CAGTCATTGTTTCTAGGTAGGGGA	TTGGCAATTTGGGTGCAACA	
OffA-33	TGGATAACCTGCAGATTTGTTTCTG	TGAGCCCAGGAGTTTCAGGC	
OffA-34	TCGTGTGTGTGTGTTTGCTTCA	CAGTGGTTCGGGAAACAGCA	
OffA-35	TGGGAATGTAAATCTGACTGGCTG	CTGGAACTCTGGGCATGGCT	
OffA-36	GCTGCAATTGCTTTTTGGCA	TGGACCCCTCCCTTACACC	
CP_CCRoff-1	TTGTTCCAACCAGCTTCATGATAA	CCTCCTTAAAGCTTCTTGCCA	
CP_CCRoff-2	TGTGACACAAACCATTGCATTC	CTATGAAGAGCTATTGATGCAGAA	
CP_CCRoff-3	TCCAGAATAATCACTGTGGCTGC	ACCAGGGGAAACTAGTGGGAGG	
CP_CCRoff-4	AACTTACTATCTGCTGGACACATTG	CTAGAGCCCCTGTTGGTTGACT	
CP_CCRoff-5	GCTGGGCTAGACCACAATTTTT	CAAGGTTCAGTTTCCCTGCTCT	
CP_CCRoff-6	TGTAATAGCAAGGCTTCAGGAC	AGAAAAGTTTCAGTGAAGAAAAACG	
CP_CCRoff-7	AGTTCTCAGCAACTAAAGTATTGA	CTGGGATTACAGGCGTGAGC	ND
CP_CCRoff-8	GGAGGCTGAGGTGAGAGGTT	ATCACCCGTCTTCTGCATCG	
CP_CCRoff-9	ACCTGTTGTCTAATCCAGCGTC	GAGACCAGGAGTCCGAGACC	
CP_CCRoff-10	ACCTGTTGTCTAATCCAGCGTC	GAGACCAGGAGTCCGAGACC	
CP_CCRoff-11	CAGGAGTCCAAGACCAGCCT	AACCTGTTGTCTAATCCAGCA	
CP_CCRoff-12	ACCCATTGATAAGGCACATTCT	AGTCAAACTCATCTAACACTCCAG	
CP_CCRoff-13	TAGGGTCTCACCCACTTGCTC	GCCTTGGGTCAGTCCTGGAG	
CP_CCRoff-14	TTGGAGGGAATGAGTTGGCTG	AGTATTTGGCACAGTGATGGG	
CP_CCRoff-15	AGAAGCAGGTACCTTCCCACC	GCATGTTAAGCCATAGAAAGGGC	
CP_CCRoff-16	AAAAAGCTCCAGCCCAGTCTC	TTTGGAATTGATCCATTGAATTTGA	
CP_CCRoff-17	TCTGTCCCTCCTACTGAGGC	TGGAAGGCTCATTTTGTTGTTTCC	
CP_CCRoff-18	GGATGTCTTTGTGGTAGGGGT	TGCTGTCACTTGGGAAAAGAA	
CP_CCRoff-19	TGTCTCTAGAAGCTAATCACTTTTT	AGAATAGTCTGCAGCTCTTTCAA	
CP_CCRoff-20	CCATGTGCATCTGTGCCGC	AACCCTTGAGAGATGAGAAGAGA	
CP_CCRoff-21	GCATTTCAGGTGGTGCTGGA	AAGAGAAGAGATTCCTTGGGGG	
CP_CCRoff-22	AAAGCACCGACAAGCTCCG	AAAGACACCCCCAGTCTGCT	
CP_CCRoff-23	TGATTTTGAGAGCATTGATTTTCAT	AGCACAGCGTCAGTGATCT	
CP_CCRoff-24	AGGAAAAACAAAGTTGAGTGAGA	TAGCTAGAATCAGTTAAGTTCCTGT	
CP_CCRoff-25	GATTGTGCCACTGCACTCCA	CCCATCCCTCCTCCACCCTA	
CP_CCRoff-26	CCTGTGAAGTGTGCCGAAGA	TTTCTTCAGGAGGGGTGCCT	
CP_CCRoff-27	TGTGATCCCCACAGAAGCCA	ACCAAGGGTCTCCTCTGTAACC	
CP_CCRoff-28	AGGGTCATCTGGGAATATAGCTC	TCTGGCTGGAAATTCAAAATGA	
CP_CCRoff-29	CACTAATCATCAGGGAAATGCAA	TGAAACTGTTTTCCATAGACGTAGT	
CP_CCRoff-30	ACCATCTGTGGTAGGCAGAAT	GCTCACCTGATAACCCAGGC	

CP_CCRoff-31	CTGTCCATGGTTGATGGATGCT	CCCCCTTCTCTTTTCACTCCCT	
CP_CCRoff-32	AGGAAGCAAAAGCAGAAATCCC	TGAAACGCAAGGACATGAGG	
OnHDAC	GTCGGACGCTGAGCGGA	AGCCTCAGCCTCCAGGTTC	
OffHDAC-1	ACTCTTCACTTACCATTCCCAT	TAGCTGGGCAGCAAATTGTGAG	
OnPMS	GTCGAAAGCAGCCAATGGGA	GGAATGCCGTGGGTCTCAAA	
OffPMS-1	AAAGGGTGGAGCACAACGTC	CCCCCATTTCCAGGGAGGT	
OffPMS-2	GGAGCACAACGTCGAAAGC	CCTCGGCCATGTTCCCCC	
OnSDHD	ACCCAGCATTTCCTCTTCCC	TTCCATCCCCTTCCCTGGC	
OffSDHD-1	GACTGAACCTGTGGCACCC	CTGGGATAGGTCGGTCCTGA	
OffSDHD-3	TGGAAAGTAATCCAAGTGTCCAT	GGTCGGTCCTGAAGAAATGC	
OffSDHD-4	AAGTGTTGAATAGCTCAAGTCAT	TGCTGCACTCCACACCATTC	
OffSDHD-5	TGCAGTTAAGGTTTCAAAAGCGA	TCCACACCATTCGGGGGTAG	

Supplementary Table S12. **Oligonucleotides Used in This Study**. (A) All oligonucleotides were purchased from Integrated DNA Technologies. '/5Phos/' indicates 5' phosphorylated oligonucleotides. A % symbol indicates that the preceding nucleotide was incorporated as a mixture of phosphoramidites consisting of 79 mol% of the phosphoramidite corresponding to the preceding nucleotide and 7 mol% of each of the other three canonical phosphoramidites. An (*) indicates that the oligonucleotide primer was specific to a selection sequence (either CCR5A, ATM or CCR5B). An (**) indicates that the oligonucleotide adapter or primer had a unique sequence identifier to distinguish between different samples (selection conditions or cellular TALEN treatment). (B) Combinations of oligonucleotides used to construct discrete DNA substrates used in TALEN digestion assays. (C) Primer pairs for PCR amplifying on-target and off-target genomic sites. +DMSO: DMSO was used in the PCR; ND: no correct DNA product was detected from the PCR reaction.

SUPPLEMENTARY ALGORITHMS

All scripts were written in bash or MATLAB. Scripts are available upon request.

Computational Filtering of Pre-selection Sequences and Selected Sequences

For Pre-selection Sequences

1) Search for 16 bp constant sequence (CCR5A = CGTCACGCTCACCACT, CCR5B = CCTCGGGACTCCACGCT, ATM = GGTACCCCACTCCGCGT) immediately after first 4 bases read (random bases), accepting only sequences with the 16bp constant sequence allowing for one mutation.

2) Search for 9 bp final sequence at a position at least the minimum possible full site length away and up to the max full site length away from constant sequence to confirm the presence of a full site, accept only sequences with this 9 bp final sequence. (Final sequence: CCR5A = CGTCACGCT, CCR5B = CCTCGGGAC, ATM = GGTACGTGC)

3) Search for instances of each half-site with the least amount of mutations from the target half-site in the full site, accept any sequences with proper left and right half-site order of left then right.

4) Determine DNA spacer sequence between the two half sites, the single flanking nucleotide to left of the left half-site and single flanking nucleotide to right of the right half-site (sequence between half sites and constant sequences).

5) Filter by sequencing read quality scores, accepting sequences with quality scores of 'A' or better across three fourths of the half site positions.

For Selected Sequences

1) Output to separate files all sequence reads and position quality scores of all sequences starting with correct 5 bp barcodes corresponding to different selection conditions.

2) Search for the initial 16 bp sequence immediately after the 5 bp barcode repeated at a psotion at least the minimum possible full site length away and up to the max full site length away from initial sequence to confirm the presence of a full site with repeated sequence, accept only sequences with a 16bp repeat allowing for 1 mutation.

3) Search for 16 bp constant sequence within the full site, accept only sequences with a constant sequence allowing for one mutation. Parse sequence to start with constant sequence plus 5' sequence to second instance of repeated sequence then initial sequence after barcode to constant sequence resulting in constant sequences sandwiching the equivalent of one full site:

CONSTANT – LFLANK – LHS – SPACER – RHS – RFLANK – CONSTANT

LFLANK = Left Flank Sequence (designed as a single random base)

LHS = Left Half Site Sequence

RHS = Right Half Site Sequence

RFLANK = Right Flank Sequence (designed as a single random base)

CONSTANT = Constant Sequence (CCR5A = CGTCACGCTCACCACT, CCR5B = CCTCGGGACTCCACGCT, ATM = GGTACCCCACTCCGCGT)

4) Search for instances of each half-site with the fewest number of mutations from the target half-site in the full site, accept any sequences with proper left and right half-site order of left then right.

5) With half site positions determine corresponding spacer (sequence between the two half sites), left flank and right flank sequences (sequence between half sites and constant sequences).

6) Determine sequence end by taking sequence from the start of read after the 5 bp barcode sequence to the beginning of the constant sequence.

SEQUENCESTART – RHS – RFLANK – CONSTANT

7) Filter by sequencing read quality scores, accepting sequences with quality scores of A or better across three fourths of the half site positions.

8) Selected sequences were filtered by sequence end, by accepting only sequences with sequence ends in the spacer that were 2.5-fold more abundant than the amount of sequence end background

calculated as the mean of the number of sequences with ends zero to five base pairs into each half-site from the spacer side (sequence end background number was calculated for both half sites with the closest half site to the sequence end utilized as sequence end background for comparison).

Computational Search for Genomic Off-Target Sites Related to the CCR5B Target Site

1) The Patmatch program[23] was used to search the human genome (GRCh37/hg19 build) for pattern sequences as follows: CCR5B left half-site sequence (L16, L13 or L10) NNNNNNNNN… CCR5B right half-site sequence (R16, R13 or R10)[M,0,0] where number of Ns varied from 12 to 25 and M (indicating mutations allowed) varied from 0 to 14.

2) The number of output off-target sites were deconvoluted since the program outputs all sequences with X or fewer mutations, resulting in the number of off-target sites in the human genome that are a specific number of mutations away from the target site.

Identification of Indels in Sequences of Genomic Sites

1) For each sequence the primer sequence was used to identify the genomic site.

2) Sequences containing the reference genomic sequence corresponding to 8 bp to the left of the target site and reference genomic sequence 6 to 10 bp (or 6 bp for genomic sites at the very end of sequencing reads and 10bp for genomic sites with low complexity or highly repetitive regions) to the right of the full target site were considered target site sequences.

3) Any target site sequences corresponding to the same size as the reference genomic site were considered unmodified and any sequences not the reference size were aligned with ClustalW[24] to the reference genomic site.

4) Aligned sequences with more than two insertions or two deletions in the DNA spacer sequence between the two half-site sequences were considered indels. Since high-throughput sequencing can result in insertions or deletions of one or two base pairs (mis-phasing) at a low but relevant rate - we only considering indels of three bp that are more likely to arise from TALEN induced modifications.

SUPPLEMENTARY SEQUENCES

TALEN Protein Domain Sequences
Text colors correspond to those in Figure 1A.

N-terminus: canonical
VDLRTLGYSQQQQEKIKPKVRSTVAQHHEALVGHGFTHAHIVALSQHPAALGTVAVKYQDMI
AALPEATHEAIVGVGKQWSGARALEALLTVAGELRGPPLQLDTGQLLKIAKRGGVTAVEAVH
AWRNALTGAPLN

N-terminus: N1
VDLRTLGYSQQQQEKIKPKVRSTVAQHHEALVGHGFTHAHIVALSQHPAALGTVAVKYQDMI
AALPEATHEAIVGVGKQWSGARALEALLTVAGELRGPPLQLDTGQLL**Q**IAKRGGVTAVEAVH
AWRNALTGAPLN

N-terminus: N2
VDLRTLGYSQQQQEKIKPKVRSTVAQHHEALVGHGFTHAHIVALSQHPAALGTVAVKYQDMI
AALPEATHEAIVGVGKQWSGARALEALLTVAGELRGPPLQLDTGQLL**Q**IA**Q**RGGVTAVEAVH
AWRNALTGAPLN

N-terminus: N3
VDLRTLGYSQQQQEKIKPKVRSTVAQHHEALVGHGFTHAHIVALSQHPAALGTVAVKYQDMI
AALPEATHEAIVGVGKQWSGARALEALLTVAGELRGPPLQLDTGQLL**Q**IA**QQ**GGVTAVEAVH
AWRNALTGAPLN

TALE array of repeats: L18 CCR5A
MTPDQVVAIAS**NG**GGKQALETVQRLLPVLCQDH
GLTPEQVVAIAS**HD**GGKQALETVQRLLPVLCQAH
GLTPDQVVAIAS**NI**GGKQALETVQRLLPVLCQAH
GLTPAQVVAIAS**NG**GGKQALETVQRLLPVLCQDH
GLTPDQVVAIAS**NG**GGKQALETVQRLLPVLCQDH
GLTPEQVVAIAS**NI**GGKQALETVQRLLPVLCQAH
GLTPDQVVAIAS**HD**GGKQALETVQRLLPVLCQAH
GLTPAQVVAIAS**NI**GGKQALETVQRLLPVLCQDH
GLTPDQVVAIAS**HD**GGKQALETVQRLLPVLCQDH
GLTPEQVVAIAS**HD**GGKQALETVQRLLPVLCQAH
GLTPDQVVAIAS**NG**GGKQALETVQRLLPVLCQAH
GLTPAQVVAIAN**NN**GGKQALETVQRLLPVLCQDH
GLTPDQVVAIAS**HD**GGKQALETVQRLLPVLCQDH
GLTPEQVVAIAS**NI**GGKQALETVQRLLPVLCQAH
GLTPDQVVAIAN**NN**GGKQALETVQRLLPVLCQAH
GLTPAQVVAIAS**HD**GGKQALETVQRLLPVLCQDH
GLTPEQVVAIAS**NG**GGRPALE

C-terminal domain: 63-aa canonical
SIVAQLSRPDPALAALTNDHLVALACLGGRPALDAVKKGLPHAPALIKRTNRRIPERTSHRVA

C-terminal domain: Q3
SIVAQLSRPDPALAALTNDHLVALACLGGRPALDAVKKGLPHAPALI**Q**RTN**Q**RIPERTSH**Q**VA

C-terminal domain: Q7
SIVAQLSRPDPALAALTNDHLVALACLGGRPALDAV**QQ**GLPHAPALI**QQ**TN**QQ**IPERTSH**Q**VA

C-terminal domain: 28-aa
SIVAQLSRPDPALAALTNDHLVALACLG

FokI:homodimeric
GSQLVKSELEEKKSELRHKLKYVPHEYIELIEIARNSTQDRILEMKVMEFFMKVYGYRGKHLGG
SRKPDGAIYTVGSPIDYGVIVDTKAYSGGYNLPIGQADEMQRYVEENQTRNKHINPNEWWKVY
PSSVTEFKFLFVSGHFKGNYKAQLTRLNHITNCNGAVLSVEELLIGGEMIKAGTLTLEEVRRKFN
NGEINF*

FokI:EL
GSQLVKSELEEKKSELRHKLKYVPHEYIELIEIARNSTQDRILEMKVMEFFMKVYGYRGKHLGG
SRKPDGAIYTVGSPIDYGVIVDTKAYSGGYNLPIGQADEM**E**RYVEENQTRNKH**L**NPNEWWKV
YPSSVTEFKFLFVSGHFKGNYKAQLTRLNHITNCNGAVLSVEELLIGGEMIKAGTLTLEEVRRKF
NNGEINF*

*Fok*I:KK
GSQLVKSELEEKKSELRHKLKYVPHEYIELIEIARNSTQDRILEMKVMEFFMKVYGYRGKHLGG
SRKPDGAIYTVGSPIDYGVIVDTKAYSGGYNLPIGQADEMQRYV**K**ENQTRNKHINPNEWWKV
YPSSVTEFKFLFVSGHFKGNYKAQLTRLNH**K**TNCNGAVLSVEELLIGGEMIKAGTLTLEEVRRK
FNNGEINF*

*Fok*I:ELD
GSQLVKSELEEKKSELRHKLKYVPHEYIELIEIARNSTQDRILEMKVMEFFMKVYGYRGKHLGG
SRKPDGAIYTVGSPIDYGVIVDTKAYSGGYNLPIGQADEM**E**RYVEENQTR**D**KH**L**NPNEWWKV
YPSSVTEFKFLFVSGHFKGNYKAQLTRLNHITNCNGAVLSVEELLIGGEMIKAGTLTLEEVRRKF
NNGEINF*

*Fok*I:KKR
GSQLVKSELEEKKSELRHKLKYVPHEYIELIEIARNSTQDRILEMKVMEFFMKVYGYRGKHLGG
SRKPDGAIYTVGSPIDYGVIVDTKAYSGGYNLPIGQADEMQRYV**K**ENQTRNKHINPNEWWKV
YPSSVTEFKFLFVSGHFKGNYKAQLTRLN**RK**TNCNGAVLSVEELLIGGEMIKAGTLTLEEVRRK
FNNGEINF*

SUPPLEMENTARY REFERENCES

1. Maul, G.G. & Deaven, L. Quantitative determination of nuclear pore complexes in cycling cells with differing DNA content. *J Cell Biol* **73**, 748-760 (1977).
2. Huang, B. et al. Counting low-copy number proteins in a single cell. *Science* **315**, 81-84 (2007).
3. Beck, M. et al. The quantitative proteome of a human cell line. *Mol Syst Biol* **7**, 549 (2011).
4. Witten, I.H. & Frank, E. Data mining: practical machine learning tools and techniques, Edn. 2nd. (Morgan Kaufman, San Francisco; 2005).
5. Reyon, D. et al. FLASH assembly of TALENs for high-throughput genome editing. *Nat Biotechnol* **30**, 460-465 (2012).
6. Pattanayak, V., Ramirez, C.L., Joung, J.K. & Liu, D.R. Revealing off-target cleavage specificities of zinc-finger nucleases by in vitro selection. *Nat Methods* **8**, 765-770 (2011).
7. Ding, Q. et al. A TALEN Genome-Editing System for Generating Human Stem Cell-Based Disease Models. *Cell Stem Cell* (2012).
8. Lei, Y. et al. Efficient targeted gene disruption in Xenopus embryos using engineered transcription activator-like effector nucleases (TALENs). *Proc Natl Acad Sci U S A* **109**, 17484-17489 (2012).
9. Kim, Y. et al. A library of TAL effector nucleases spanning the human genome. *Nat Biotechnol* **31**, 251-258 (2013).
10. Grau, J., Boch, J. & Posch, S. TALENoffer: genome-wide TALEN off-target prediction. *Bioinformatics* **29**, 2931-2932 (2013).
11. Gao, H., Wu, X., Chai, J. & Han, Z. Crystal structure of a TALE protein reveals an extended N-terminal DNA binding region. *Cell Res* **22**, 1716-1720 (2012).
12. Li, T. et al. Modularly assembled designer TAL effector nucleases for targeted gene knockout and gene replacement in eukaryotes. *Nucleic Acids Res* **39**, 6315-6325 (2011).
13. Miller, J.C. et al. A TALE nuclease architecture for efficient genome editing. *Nat Biotechnol* **29**, 143-148 (2011).
14. Mahfouz, M.M. et al. De novo-engineered transcription activator-like effector (TALE) hybrid nuclease with novel DNA binding specificity creates double-strand breaks. *Proc Natl Acad Sci U S A* **108**, 2623-2628 (2011).
15. Mussolino, C. et al. A novel TALE nuclease scaffold enables high genome editing activity in combination with low toxicity. *Nucleic Acids Res* **39**, 9283-9293 (2011).
16. Hockemeyer, D. et al. Genetic engineering of human pluripotent cells using TALE nucleases. *Nat Biotechnol* **29**, 731-734 (2011).
17. Osborn, M.J. et al. TALEN-based Gene Correction for Epidermolysis Bullosa. *Molecular Therapy* (2013).
18. Sander, J.D. et al. In silico abstraction of zinc finger nuclease cleavage profiles reveals an expanded landscape of off-target sites. *Nucleic Acid Research* **41** (2013).
19. Gabriel, R. et al. An unbiased genome-wide analysis of zinc-finger nuclease specificity. *Nat Biotechnol* **29**, 816-823 (2011).
20. Perez, E.E. et al. Establishment of HIV-1 resistance in CD4+ T cells by genome editing using zinc-finger nucleases. *Nature biotechnology* **26**, 808-816 (2008).
21. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society Series B-Methodological* **57**, 289-300 (1995).
22. Noble, W.S. How does multiple testing correction work? *Nature Biotechnology* **27**, 1135-1137 (2009).
23. Yan, T. et al. PatMatch: a program for finding patterns in peptide and nucleotide sequences. *Nucleic Acids Res* **33**, W262-266 (2005).
24. Larkin, M.A. et al. Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947-2948 (2007).