

# TITLE

CRISPR-Cas9 cytidine and adenosine base editing of splice-sites mediates highly-efficient disruption of proteins in primary cells

# AUTHORS

Mitchell G. Kluesner\*, Walker S. Lahr\*, Cara-Lin Lonetree, Branden A. Smeester, Patricia N. Claudio-Vázquez, Samuel P. Pitzen, Madison J. Vignes, Samantha C. Lee, Samuel P. Bingea, Aneesha A. Andrews, Beau R. Webber† & Branden S. Moriarity†,‡

\*These authors contributed equally

†These authors contributed equally

‡Corresponding author

# ABSTRACT

Base editors allow for precise nucleotide editing without the need for genotoxic double-stranded breaks. Prior work has used base editors to knockout genes by introducing premature stop codons or by disrupting conserved splice-sites, but no direct comparison exists between these methods. Additionally, while base editor mediated disruption of splice sites has been used to shift the functional isoform pool, its utility for gene knockout requires further validation. To address these needs, we developed the program SpliceR ([z.umn.edu/spliceR](http://z.umn.edu/spliceR)) to design cytidine-deaminase base editor (CBE) and adenosine-deaminase base editor (ABE) splice-site targeting guides. We compared the splice-site targeting and premature stop codon introduction in a knockout screen against the TCR-CD3 immune synapse in primary human T-cells. Our data suggests that 1) the CBE, BE4 is more reliable than the ABE, ABE7.10 for splice-site targeting knockout and 2) for both CBEs and ABEs, splice-donor targeting is the most reliable approach for base editing induced knockout.

# INTRODUCTION

Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) systems and their CRISPR associated proteins (Cas proteins) have allowed for an unprecedented ability to manipulate the genome (1–5). Key amongst the applications of these systems is their use in gene editing for targeted gene knock-out, knock-in, and modification<sup>6</sup>. These applications are of particular interest in the fields of cellular immunotherapies, where the multiplexed knock-out of genes involved in alloreactivity (TRAC, TRBC, B2M) as well as immunosuppressive genes such as *PDCD1*, *CTLA4*, and *TGFBR2*, in tandem with the knock-in of chimeric antigen receptors (CARs) specific to tumor antigens yields promise in the development of efficacious and safe therapies to recalcitrant malignancies<sup>7–10</sup>. In the most commonly used form of CRISPR, the Cas9 nuclease from *Streptococcus pyogenes* (hereafter referred to as Cas9) is paired with a single-guide RNA (sgRNA) to induce a double stranded break (DSB) at a specific DNA site directed by the programmable complementarity of the 20nt sgRNA protospacers<sup>5</sup>. While CRISPR-Cas9 nucleases work exceptionally well for single gene editing, multiple concerns have emerged surrounding DSB induction, including large scale genomic rearrangements<sup>11,12</sup> and altered pluripotency<sup>13</sup>, leading to the potential of oncogenesis in cell based therapies. Concerns arising from DSBs are exacerbated in a multiplex setting, where multiple genes are targeted simultaneously<sup>7,14</sup>. However, as more has been elucidated about the complex genetic circuitry involved in cancer immunosurveillance<sup>15</sup>, and the hope to generate ‘off the shelf’ CAR T cells<sup>14</sup>, it is of increased interest to modulate and edit genes in a multiplex settings<sup>8</sup>.

An alternative tool to edit genes without the need for DSBs are CRISPR-Cas9 base editors. Base editors are a class of gene editing enzymes that consist of a Cas9 nickase fused to a nucleotide deaminase domain<sup>16–18</sup>. In principle, base editors localize to a target region in the genome guided by the sgRNA protospacer. Once bound, the ssDNA R-loop is displaced by the binding of the Cas9 complex. Displacement allows the tethered deaminase domain access to the ssDNA R-loop, whereby cytidine deaminase base editors (CBEs, C-to-T) deaminate C-to-U, which base pairs like T, and adenosine deaminase base editors (ABEs, A-to-G) deaminate A-to-I, which base pairs like G. Nicking of the unedited strand by the core Cas9 nickase complex then stimulates DNA repair to use the nascent base as a template thus preserving the genomic edit during subsequent cycles of DNA replication.

Previously, our group established a platform for the multiplex engineering of human lymphocytes using CRISPR-Cas9 base editors. This approach achieves high-efficiency multiplexed gene knockout without multiple DSBs and their associated complications such as chromosomal translocations and hindered proliferation<sup>14</sup>. In that work we used CBEs to KO genes by mutating CAG, CGA, and TGG codons to introduce premature stop codons (pmSTOPS, previously termed iSTOP or CRISPR-STOP)<sup>19,20</sup> and by mutating the conserved splice-site motifs to disrupt splicing of the fully functional transcript. To date, there has been substantial interest in using both CBEs and ABEs to modulate splicing which has predominantly focused on using targeted skipping of exons bearing pathogenic mutations (previously termed CRISPR-SKIP)<sup>21,22</sup> and inducing functional alternative splicing patterns<sup>23</sup>. However, there is substantial evidence for the utility of the splice-site targeting methods for functional gene knockout as opposed to strictly modulating alternative splicing (Here, distinguished as BE-splice), which has been previously demonstrated with Cas9 nuclease<sup>24,25</sup>. Despite the array of reports demonstrating the various ways base editors can be used to knockout genes and modulate splicing, there are two main gaps in the field, namely 1) a comprehensive tool for designing both CBE and ABE sgRNAs to target both splice donor (SD) and splice acceptor (SA) sites, and 2) a head-to-head comparison of the methods of base editing mediated gene knockout encompassed across, ABEs, CBEs, BE-splice methods, and pmSTOP induction.

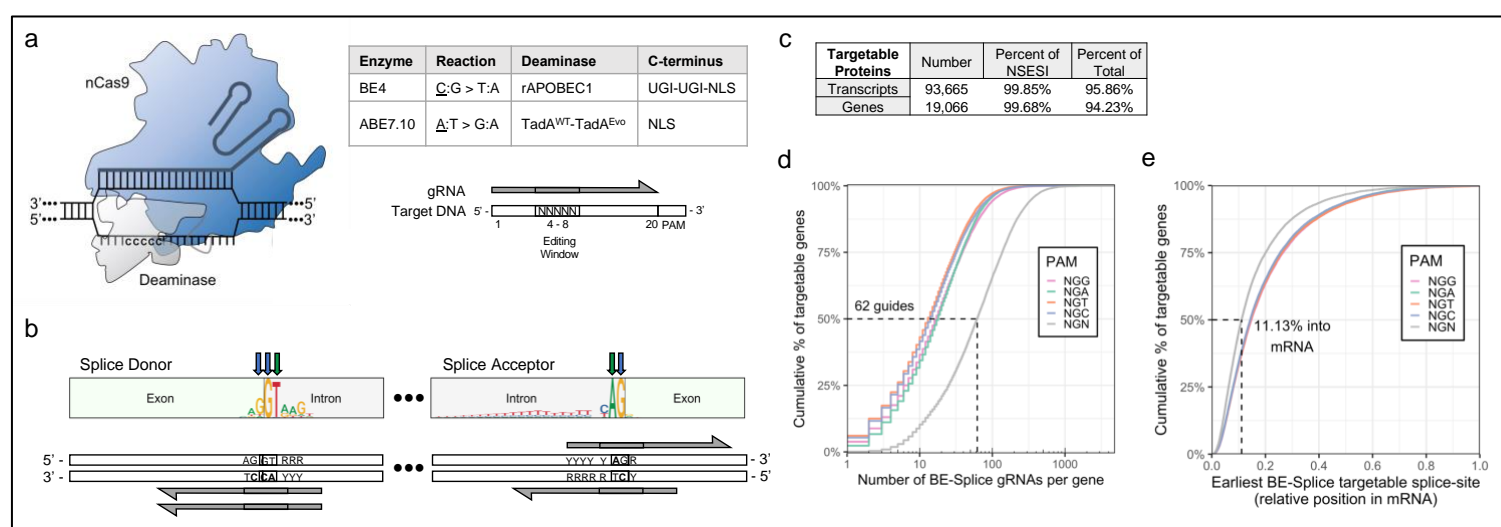
Here we present an easy-to-use webtool SpliceR ([z.umn.edu/splicer](http://z.umn.edu/splicer)), for the design of base editing sgRNAs to target splice-sites of any ensembl annotated metazoan genome using an ensembl transcript ID. We demonstrate the robustness of our approach in a focused sgRNA screen targeting the various proteins that make up the heterooctameric T cell receptor-CD3 (TCR-CD3) complex and MHC Class I immune synapse. We found that for gene editing and protein knockout 1) CBEs were more reliable than ABEs, 2) among both CBEs and ABEs targeting splice donors tended to produce more robust knockouts than targeting splice acceptors, and 3) targeting splice donor sites produced more robust knockouts than pmSTOPS. Ultimately, here we describe a highly-efficient method and program for the design of sgRNAs for gene knockout via base editing.

## RESULTS

In previous work, we and others have established the ability to target splice sites with base editing for both modulating splicing patterns and disrupting proteins<sup>14,21,23</sup>. While programs exist for designing splice acceptor targeting guides with a limited number of PAMs, there is not a comprehensive program for designing guides that target both splice donors and splice acceptors sites with CBEs or ABEs (Fig. 1a, 1b), that is compatible with any PAM identity, and that is compatible with all ensembl annotated metazoan species. To meet these needs, we developed the program SpliceR for the design of BE-splice sgRNAs ([z.umn.edu/splicer](http://z.umn.edu/splicer), github repo). SpliceR employs an interactive user interface to specify the parameters for BE-splice sgRNA design, such as Ensembl transcript ID, PAM identity, base editor type, and species of

interest (Supplementary Fig. 1a). SpliceR communicates directly with ensembl.org to query genetic information for sgRNA design, and then performs sgRNA matching and scoring.

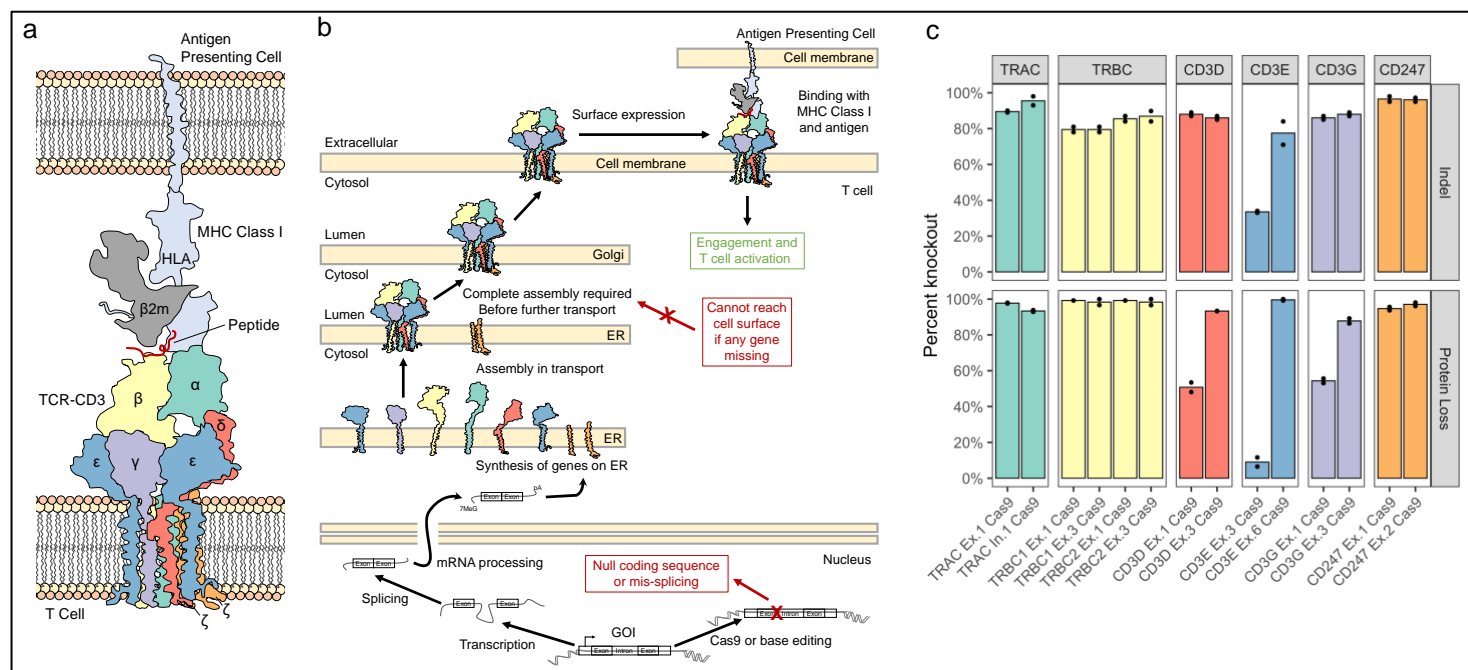
To assess the practicality of the BE-splice approach, we used SpliceR to generate sgRNAs to every protein coding gene in the human genome. When restricting the analysis to genes that undergo splicing, i.e. non-single-exon-single-isoform (NSES) genes, such that they are targetable by this approach, we found that 99.85% of transcripts and 99.68% of genes are expected to be targetable of the 94.23% of protein coding genes that are NSES genes (Fig. 1c). Furthermore, 50% of NSES genes are expected to have 62 sgRNAs or more, with the earliest sgRNA within the first 11.1% of the transcript (Fig. 1d, 1e). When broken down by targeting splice-donors or splice-acceptors, splice donors tend to have more sgRNAs (Supplementary Fig. 2), while also having the first sgRNA appear earlier in the transcript (Supplementary Fig. 3). These results suggested that the BE-splice approach is a robust method for targeting most genes, with targeting splice donors appearing more reliable given the greater availability of splice donor sgRNAs across CBEs and ABEs.



**Figure 1.** Overview of the BE-splice approach. (a) Diagram of base editor complex, table of the enzymes used in this study, and base editing sgRNAs with editing window. (b) Positioning of BE-splice sgRNAs within conserved splice donor and splice acceptors motif. Logo plots were generated from all human protein coding gene splice sites. Arrows indicate the base targeted by either CBEs (blue), or ABEs (green). (c) Breakdown of transcripts and genes targetable by BE-splice, showing the vast majority of non-single-exon-single-isoform (NSES) genes are targetable by this approach (99.68%). (d) Distribution of BE-splice sgRNA density across each gene. 50% of genes have 62 or more sgRNAs mapping to them when accounting for all PAM identities and both CBE and ABE approaches. (e) Distribution of the position of the first sgRNA for each gene, with 50% having their first sgRNA 11.13% way through the mRNA or earlier.

Despite the number of publications that have used base editors to disrupt or modulate genes by targeting splice sites with CBEs and ABEs, or introducing pmSTOPS with CBEs<sup>14,19,20,22,23,26–28</sup>, to our knowledge there is no comprehensive, direct comparison among these methods. Therefore, we sought to compare these methods directly and assess the predictions made from SpliceR in a mid-throughput screen. Given the interest of the TCR-CD3 and MHC class I immune synapse in the context of immunotherapies<sup>29–33</sup>, the presence of multiple NSES genes that comprise these complexes (Fig. 2a), the necessity of every gene for surface expression (Fig. 2b, reviewed in Ref. 34<sup>34</sup>) and the ease of a functional readout with flow cytometry, we saw this complex as an ideal model to compare base editing mediated knockout approaches. To validate this model, we first designed Cas9 nuclease sgRNAs to each gene in the TCR-CD3 complex and the subunit of the MHC class I complex. Primary human T cells were then transfected with sgRNAs and Cas9 nuclease, and indel formation was measured via Sanger sequencing, while protein knockout was measured via flow cytometry

(Supplementary Fig. 4). We found a high level of indel formation ( $81.7\% \pm 15.1\%$ ) and high corresponding protein knockout ( $73.4\% \pm 37.2\%$ ) across all genes in the complexes, including CD3 $\zeta$  in the TCR-CD3 complex (Fig. 2c). These results establish that any one of the genes in the TCR-CD3 complex can be edited to induce a common phenotype of functional knockout.



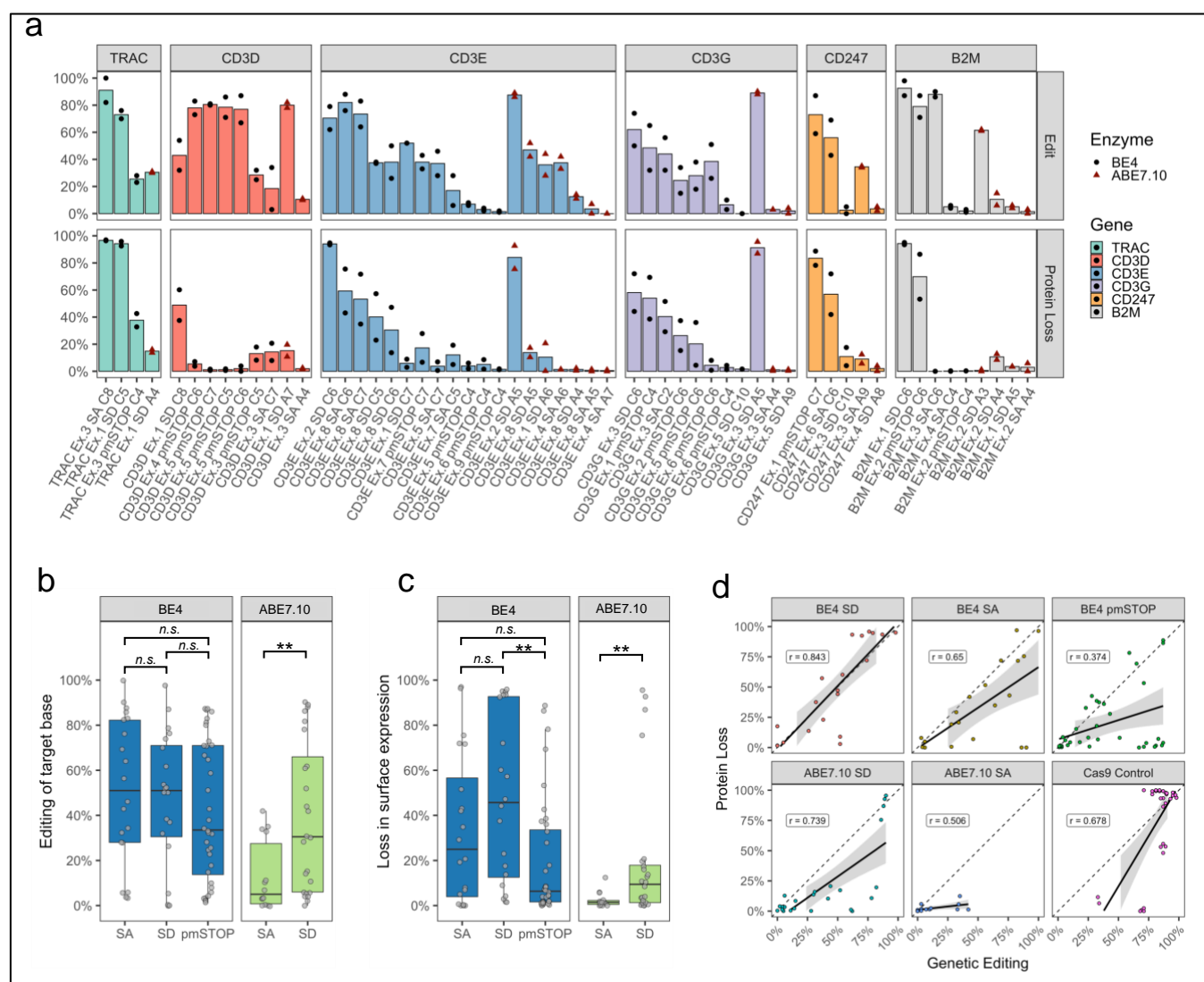
**Figure 2.** Conception and validation of the TCR-CD3-MHC Class I immune synapse as a screening model for functional knockout. (a) Diagram of the multimeric TCR-CD3 complex and MHC Class I immune synapse containing multiple NSESI genes, based on the solved structures (PDB 6JXR<sup>43</sup>, PDB 3T0E<sup>47</sup>, PDB 10GA<sup>48</sup>). (b) Diagram of the synthesis and localization of the TCR-CD3 complex and interaction with MHC Class I. All members of the CD3 complex are required before functional localization to the cell surface, where disruption of a single splice site within one gene member can prevent a surface expressed complex from forming. (c) Cas9 nuclease knockout of each individual member of TCR-CD3 complex validates the screening model. Two Cas9 nuclease sgRNAs were designed to exonic regions of each gene in the complex. All genes had at least one guide with >85% indel efficiency and loss in TCR-CD3 surface expression. N = 2 independent donors.

Next, we wanted to apply our model to directly compare 1) CBE vs. ABE knockout, 2) splice-donor vs. splice-acceptor knockout, and 3) BE-splice vs. pmSTOP knockout approaches. We used SpliceR to generate a screen of BE-splice sgRNAs targeting the TCR-CD3 complex, and the iSTOP database<sup>19</sup> to design pmSTOP sgRNAs. Using the same workflow as with Cas9, primary human T cells were transfected with sgRNAs and either the cytidine base editor BE4, or the adenosine base editor ABE7.10, expressed via codon optimized mRNA<sup>35</sup>. Samples exhibited a wide range of gene editing at the target base across CBE and ABE approaches ( $M \pm SD$ ,  $39.6\% \pm 31.6\%$ , range 0%-100%), and a wide range of protein KO ( $M \pm SD$ ,  $25.1\% \pm 31.9\%$ , range 0%-97.0%) (Fig. 3a). Both editing of the target base and protein knockout was higher among CBE than ABE treated samples ( $t = 2.87$ ,  $df = 76$ ,  $p\text{-value} = 0.0053$ ;  $t = 3.75$ ,  $df = 76$ ,  $p\text{-value} = 0.00034$ ) (Fig. 3b, 3c).

Within CBE treated samples, there was no significant difference in editing of the target base between splice donors, splice acceptors, and pmSTOPS ( $t = 0.416$ ,  $df = 38$ ,  $p\text{-value} = 0.680$ ;  $t = -0.846$ ,  $df = 54$ ,  $p\text{-value} = 0.401$ ;  $t = -1.27$ ,  $df = 54$ ,  $p\text{-value} = 0.204$ ) (Fig. 3b). Furthermore, the rate of protein knockout was not significantly different between CBE splice donors and CBE splice acceptors ( $t = -1.27$ ,  $df = 38$ ,  $p\text{-value} = 0.213$ ), however splice donors had significantly higher knockout efficiencies relative to pmSTOP sgRNAs ( $t = 3.33$ ,  $df = 54$ ,  $p\text{-value} = 0.001$ ).



1 value = 0.00157), while splice acceptors were non-significantly higher than pmSTOP sgRNAs ( $t$   
2 = 1.76,  $df$  = 54,  $p$ -value = 0.0841). Among CBEs, protein knockout was well correlated with  
3 editing of the splice donor and splice acceptor sites ( $r$  = 0.843,  $p$ -value = 3.11e-06;  $r$  = 0.65  
4  $p$ -value = 0.00192), yet poorly correlated with the introduction of premature stop codons ( $r$  =  
5 0.374,  $p$ -value = 0.0247) (Fig. 3d). In contrast, ABE splice donor sgRNAs had significantly  
6 higher editing compared to ABE splice acceptors ( $t$  = -3.25,  $df$  = 34.9,  $p$ -value = 0.00257), which  
7 translated into the ABE splice donor protein knockout efficiency being significantly higher than  
8 the ABE splice acceptor protein knockout efficiency ( $t$  = -2.81,  $df$  = 23.9,  $p$ -value = 0.00972).  
9 Furthermore, protein knockout among ABE treated samples was well correlated with target base  
10 editing in splice donor ( $r$  = 0.739,  $p$ -value = 3.73e-05), but not in splice acceptor sgRNAs ( $r$  =  
11 0.506,  $p$ -value = 0.0651). Among the well correlated BE-splice approaches; CBE splice donors,  
12 CBE splice acceptors, and ABE splice donor, the correlation was similar to that observed by  
13 Cas9 nuclease ( $r$  = 0.678,  $p$ -value = 2e-05) (Fig. 3d). Ultimately these results established that 1)  
14 CBE mediated knockout is more reliable than ABE mediated knockout, 2) splice donors are a  
15 more robust target across CBEs and ABEs, and 3) Among the genes in our screen, the BE-  
16 splice method produced more robust knockouts than pmSTOPS.

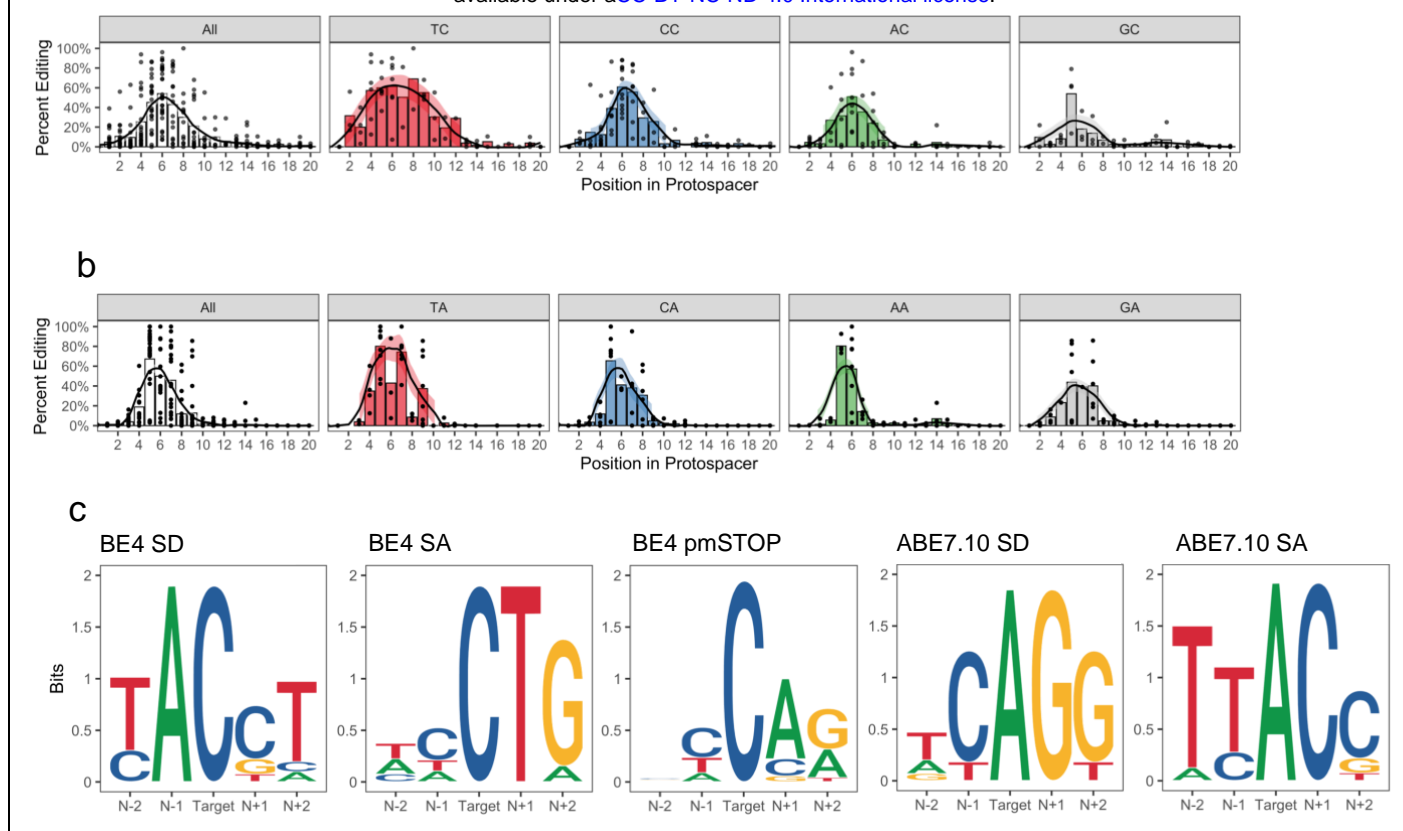


**Figure 3.** BE-splice sgRNAs mediate robust editing and disruption of TCR-CD3 MHC Class I immune synapse. (a) Editing efficiency (top) and surface protein loss (bottom) from each guide in the sgRNA screen. Results grouped by gene and enzyme used in descending order by protein loss. X-axis label indicates position of target base within sgRNA. TRBC1 and TRBC2 were omitted from the BE-splice screen due to the inability to design single BE-splice sgRNAs to target both paralogs simultaneously. (b) Base editing efficiencies grouped by enzyme and target motif analyzed with Student's t-test or Welch's t-test. (c) Protein loss efficiencies grouped by enzyme and target motif analyzed with Student's t-test or Welch's t-test. (d) Consistency of editing efficiency and protein loss across all approaches employed here. Relationship between protein loss and base editing efficiency is comparable to that observed in Cas9 control. N = 2 independent donors, performed on different days. n.s.  $P > 0.05$ , \*  $P \leq 0.05$ , \*\*  $P \leq 0.01$ , \*\*\*  $P \leq 0.001$ .

With these differences in mind, we wanted to investigate what may cause the disparities in baseline editing among the different approaches. Previous work has established that base editing efficiency is context dependent, with particular preference dictated by the nucleotide preceding the target base<sup>36–38</sup>. These works also sought to change the context dependencies of the preceding nucleotide by employing cytidine deaminase paralogs, orthologs, and engineered variants to change the context dependencies of base editors. Understanding the dinucleotide context dependencies of base editing would aid in the selection of BE-splice sgRNAs, however, to the best of our knowledge there does not exist a comprehensive analysis of the dinucleotide context dependencies by position of the commonly used base editors employed here, rAPOBEC1-BE4, and TadA<sub>WT</sub>-TadA<sub>Evo</sub>-ABE7.10.

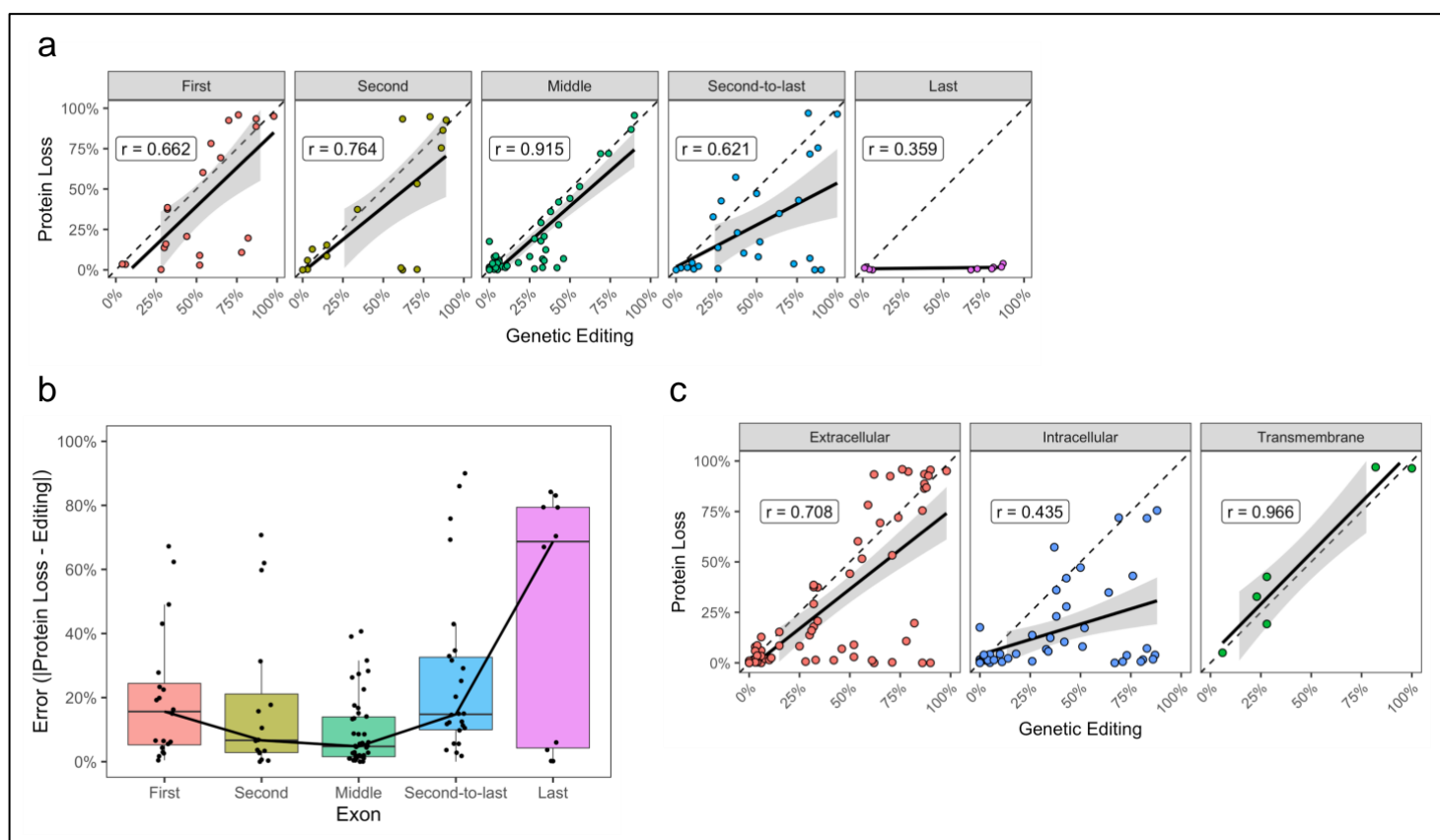
To determine these dependencies, we performed an aggregated analysis for both BE4 and ABE7.10 with data across multiple cells types, genes and delivery methods from the literature<sup>18,39–41</sup> and data generated by our group (6 papers, 102 guides, 447 edits in total). We chose to focus our analysis on editing efficiency as a function of the position in the protospacer and the identity of the preceding base. Aggregated analysis of BE4 across all nucleotide contexts produced a smooth distribution of editing activity centered about position 6 of the protospacer (Fig. 4a). Consistent with previous work, the editing window was dependent on the identity of the preceding nucleotide<sup>17,26,37,42</sup>, where TC dinucleotides exhibited the broadest editing window, while AC and CC exhibited smaller, comparable windows and GC exhibited a highly suppressed window with  $\geq 20\%$  activity only observed at positions 5-6 (Fig. 4a). When comparing TC to GC dinucleotides, the identity of the preceding nucleotide alone decreased the average editing efficiency by 3.2-fold from 28.7% to 6.9% (Supplementary Fig. 5a). In contrast, aggregated analysis of ABE7.10 yielded a narrower window with  $\geq 20\%$  editing activity between positions 4-8, along with thinner tails to the editing distribution (Fig. 4b). Interestingly, ABE7.10 exhibited similar preceding nucleotide context dependencies as BE4 with TA having the broadest and tallest window, followed by CA, AA, and then GA (Fig. 4b). When comparing TA to GA dinucleotides, the identity of the preceding nucleotide alone decreased the average editing efficiency by 2.9-fold from 24.5% to 6.3% (Supplementary Fig. 5b). Among both BE4 and ABE7.10, the post dinucleotide base did not appear to have as large of an effect on editing efficiency (Supplementary Fig. 5c-5f).

Given the distinct differences in editing among dinucleotide contexts, we hypothesized that this could partially explain the differences in baseline editing efficiency of the BE-splice sgRNAs. To make this comparison we generated a consensus pentanucleotide motif for each approach we tested for base editing mediated protein knockout (Fig. 4c). Consistent with the expectation among CBE splice-site motifs, we observed an ACN motif for splice donors, NCT motif for splice acceptors, and a NCA motif for pmSTOPs. The lack of enhancing TC motifs, and inhibitory GC motifs across these approaches likely accounts for non-significant baseline editing efficiencies observed across all CBE treated samples. In contrast, ABE splice donors exhibited a preferred TAC motif, while the acceptors exhibited a non-disfavored CAG motif, which likely contributed to the significant difference in baseline editing among ABE treated samples (Fig. 3c).



**Figure 4.** Context dependencies of base editors and BE-splice target motifs. (a) Preceding dinucleotide context dependencies of rAPOBEC1-BE4. Results normalized and aggregated across published results and our own work. Smoothed distributions generated using LOESS regression with span = 0.5. (b) Preceding dinucleotide context dependencies of TadAwt-TadAevr-ABE7.10. Results normalized and aggregated across published results and our own work. Smoothed distributions generated using LOESS regression with span = 0.45. 6 papers, 102 guides, 447 edits. (c) Logo plots of pentanucleotide motif of each enzyme and target motif combination in this work. Plots are consistent with conserved splice site sequences.

1 Lastly, to investigate how the position of the sgRNA within the transcript affects the  
2 reliability of a knockout, we binned all base editing sgRNAs into first, second, middle, second-to-  
3 last, and last exons based on which exons they targeted (Supplementary Fig. 6). Strikingly, we  
4 found the highest degree of correlation between protein loss and base editing among guides  
5 that targeted middle exons ( $r = 0.915$ ,  $p\text{-value} = 1.6e-20$ ), and no significant correlation among  
6 guides that targeted the last exon ( $r = 0.359$ ,  $p\text{-value} = 0.309$ ), as might be expected (Fig. 5a).  
7 Next, we wanted to see how the error in protein loss, defined as the absolute value of the  
8 observed protein loss minus the editing of target base, varied as a function of the exons being  
9 targeted. We found that the error was minimized when targeting inner exons, and increased in  
10 guides targeting the second-to-last and last exons (Fig. 5b). Given the extracellular and  
11 transmembrane domains of the TCR-CD3 complex are essential to interchain interactions that  
12 assemble the complex<sup>43</sup>, allow for surface localization<sup>34</sup>, and signal transduction<sup>44-46</sup>, we wanted  
13 to study how the positioning of these sgRNAs affected the reliability of protein knockout. To do  
14 this, we mapped each sgRNA within the tertiary structure of the TCR-CD3 complex (PDB  
15 6JXR<sup>43</sup>) and MHC Class I complex (PDB 3T0E<sup>47</sup>; PDB 10GA<sup>48</sup>), as determined by the solved  
16 structures (Supplementary Fig. 6-13). From these mappings, we classified each sgRNA as  
17 extracellular, intracellular, or transmembrane targeting. Consistent with the functional role of  
18 these regions, we found that guides targeting the transmembrane region of genes in the  
19 complex had the highest correlation of protein loss to editing ( $r = 0.966$ ,  $p\text{-value} = 0.0017$ , albeit  
20  $n = 6$ ), followed by extracellular guides ( $r = 0.734$ ,  $p\text{-value} = 1.18e-11$ ), and intracellular guides  
21 ( $r = 0.435$ ,  $p\text{-value} = 0.00314$ ) (Fig. 5c). Ultimately, these results suggest that the greatest  
22 reliability of base editing mediated knockout is found targeting early, or inner exons within  
23 regions known to be functionally crucial.



**Figure 5.** Consistency of editing efficiency and protein loss across mRNA and protein regions. (a) Scatter plots of protein loss and editing efficiency by exon group across all base editor approaches employed. The strongest relationship is observed among middle exons, while the weakest is observed in the last exon. (b) Error in protein loss as a function of editing efficiency across each exon group. The least error is observed among middle exons, while the greatest is observed in the last exon. (c) Scatter plots of the protein loss and editing efficiency grouped by where each BE sgRNA maps to the TCR-CD3 and MHC Class I structures. sgRNAs that map to transmembrane and extracellular regions exhibit the greatest consistency between protein loss and base editing efficiency.

## DISCUSSION

In this work we studied the use of CRISPR-Cas9 base editors for highly efficient gene disruption in primary human T cells using both cytidine and adenosine deaminase base editors. In this approach, the conserved splice donor (exon|GT-intron) and splice acceptor (intron-AG|exon) sites are edited via a transition mutation (C:G>T:A, or A:T>G:C), nullifying the splice-site and disrupting the gene at the transcriptional level<sup>21,23</sup>. To improve accessibility to the BE-splice approach, we developed the program SpliceR ([z.umn.edu/splicer](http://z.umn.edu/splicer)) as an online tool for the design of BE-splice sgRNAs. Analysis of the entire human genome showed that 95.86% of all protein coding gene transcripts, and 99.85% of all protein coding gene transcripts from genes that do not consist of a single-exon-single-isoform are targetable. We assessed these predictions, and compared the BE-splice approach to pmSTOP sgRNAs<sup>19</sup> with a mid-throughput screen of sgRNAs targeting proteins involved in the TCR-CD3 MHC Class I immune synapse. From this screen we discovered three main trends, detailed below.

## Major findings



Firstly, CBE mediated knockout was more reliable than ABE mediated knockout. The higher rate of CBE knockout can be primarily attributed to the higher levels of baseline editing observed in CBE compared to ABE. The higher baseline editing of CBEs was consistent with the larger activity window of BE4 relative ABE7.10 observed in our aggregated analysis of these enzymes. Furthermore, across all dinucleotide contexts BE4 exhibited a smoother, more normally distributed window compared to ABE7.10, which may be the result of the processive activity of the CBE deaminase, *R. norvegicus* APOBEC1, while in contrast the ABE deaminase was evolved from *E. coli* TadA which acts endogenously on a single adenosine within the tRNA<sub>Arg</sub> anticodon loop<sup>49,50</sup>. In general, we found current ABEs to be less reliable editors than CBEs, underpinning the need to develop more active ABEs by employing homologs, orthologs, and evolved variants as has been widely demonstrated among CBEs<sup>17,26,37,42</sup>. As such, during the final preparations of this manuscript two independent reports emerged describing the development of newer ABEs, termed ABE8s. These works employed either large-scale mutation libraries<sup>51</sup> or phage assisted directed evolution<sup>52</sup> to produce ABEs with substantially higher on-target editing. Thus, employing these enzymes will likely reduce the disparity in baseline editing efficiency between CBEs and ABEs.

Secondly, across CBE and ABE, splice-donor sgRNAs produced more robust knockouts. In CBEs, splice donors and splice acceptors were not significantly different in editing efficiency, while in ABEs splice donors had a significantly higher editing efficiency. These trends were also mirrored in the rate of protein knockout with both enzymes, with the rate of protein loss being non-significantly higher in CBE splice donors relative to splice acceptors.

Among ABEs, this disparity could be attributed in part to the highly preferred TAC motif in splice donor guides, and the lesser preferred CAG motif in splice acceptor guides. The TAC preference motif is consistent with the preference motif of the parental enzyme *E. coli* TadA, and other adenosine deaminases<sup>49,50,53</sup>.

Additionally, previous work has demonstrated that targeting the splice donor sequence with an indel forming Cas9 nuclease causes robust protein and lncRNA knockout<sup>24,25</sup>. The reliability of disrupting the splice donor may be explained by the critical nature of the splice donor site in initiating splicing. During splicing, the pre-mRNA nucleotides contained in the splice donor site form a sequence-dependent RNA:RNA duplex with either the U1 or U12 snRNP. This Watson-Crick-Franklin interaction defines the exon boundary and initiates splicing<sup>54,55</sup>. Therefore, if this interaction is disrupted at the outset of splicing, then the ability to undergo the native splicing event may be principally inhibited.

Moreover, when considering the enhanced reliability of targeting splice donors from the splice acceptor perspective, previous work demonstrates that disrupting splice acceptors tends to favor exon skipping which may retain the reading frame, as opposed to a nonsense inducing outcome<sup>21</sup>. However, as demonstrated by the multiple instances of clinical splice site mutations (reviewed in Ref. 56<sup>56</sup>), it is important to note that splice acceptor mutations do not always result in a clean, single exon skipping outcome, as previous publications have suggested<sup>21</sup>. Rather, splice acceptor mutations are capable of inducing alternative splicing patterns via activation of cryptic splice sites, such as full intron retention, partial intron retention, and partial skipping of an exon<sup>57–59</sup>, all of which have the potential to knockout a gene by introducing a frameshift mutation, or removing a functionally critical region of the molecule.

Thirdly, among the genes in our screen, CBE-splice sgRNAs produced more frequent protein knockouts than pmSTOPS sgRNAs. At the genetic level, there was no significant difference in editing efficiency of BE-splice guides compared to pmSTOP guides, however, at the protein level pmSTOPS produced significantly less knockout compared to splice donors, and non-significantly less knockout compared to splice acceptors. These disparate results may be attributed to different levels of induction of nonsense mediated decay. Premature stop codons are known to incompletely induce a nonsense mediated decay<sup>60</sup>, or they can form relatively functional truncated variants<sup>61</sup>. In contrast, there is evidence that one-third of human genes

have alternative isoforms that are actively suppressed by nonsense mediated decay<sup>62</sup>. This suggests that if a base edited splice site induces an alternative isoform that naturally occurs due to basal splicing errors, then these isoforms may be more readily subjected to nonsense mediated decay than newly introduced premature stop codons which have not been previously encountered by the cell.

Regardless of the mechanisms at play, the average knockout efficiencies of BE-splice and pmSTOP approaches could both be improved by increasing the baseline editing rates with more active CBEs such as EvoFERNY-BE<sub>437</sub>, or hA3A-BE<sub>442</sub>. Moreover, it is important to note that in our own view and application of using base editing to knockout genes, we see both the BE-splice and pmSTOP approaches as highly useful, complementary approaches in the screening of guides for protein knockout.

### **Considerations for designing BE-splice sgRNAs and other BE-mediated knockout approaches**

When designing BE-splice sgRNAs with SpliceR it is important to use the ensembl transcript table to choose whichever transcript or transcripts are most relevant to the biological phenomenon of interest. In instances where knockout of the main function of the gene is desired, we recommend targeting transcripts with merged Ensembl, Havana and Uniprot annotation ("gold labelled"). Furthermore, our results demonstrate that targeting all exons besides the last exon can produce high efficiency knockouts, where the innermost exons produce the most proportional protein knockout to genetic editing efficiency. We recommend initially screening 3-4 guides per gene with a greater than 50% predicted editing efficiency. If additional CBE sgRNAs are desired, we also recommend including a high predicted efficiency, inner exon pmSTOP sgRNAs to a known functional region, from the iSTOP database.

Furthermore, given situations where none of the aforementioned approaches are available, we also recommend trying to model known pathogenic variants such as those documented on ClinVar, Uniprot, or COSMIC for generating functional knockouts. However, as is the case with Cas9 nuclease editing, with any base editing mediated knockout approach we recommend functionally validating the effect of a base edit at the protein level, rather than assuming high editing efficiency corresponds to a functional knockout.

Lastly, much excitement has been generated by the prospect of prime editing<sup>63</sup>, which in principle allows for the precise editing and induction of mutations 80bp or smaller in size, including transversion, transition, and indel mutations. The ability to target a wide variety of mutations lends prime editing potential superiority over base editing in the correction of pathogenic mutations for gene therapy. This advantage is mainly derived from the ability of prime editing to precisely introduce desired mutations without undesired bystander edits to adjacent bases in the editing window as is seen with base editing. However, in the context of knocking out a gene by disrupting a conserved sequence, unintended bystander edits are of less concern when the goal is to nullify a splice site. Furthermore, despite the low-to-moderate level of indels observed in prime editing, this method still needs to be evaluated in a multiplex setting where compounding indels will likely lead to translocations<sup>63</sup>. The effect of these indels will also need to be addressed in stem cells, where double strand breaks have been associated with reduced potency<sup>64,65</sup> and an increased risk of oncogenesis through inhibition of regulators the cell cycle and genomic stability<sup>12,66</sup>.

As more is understood about the deleterious consequences of Cas9 nuclease induced DSBs, it is of increasing interest to deploy gene editing methods that do not rely on DSBs. Here we show that cytidine deaminase base editors and adenosine deaminase base editors can allow for high efficiency knockout of proteins in primary T cells by disrupting conserved splice-sites. In a direct comparison of the major methods of base editing mediated knockout we validated with a sgRNA screen against the TCR-CD3 MHC Class I immune synapse. Our results demonstrate that 1) the CBE rAPOBEC1-BE4 is more reliable than ABE ABE7.10 for

editing and knockout, and 2) across CBEs and ABEs, targeting splice donors is the most reliable method for base editing mediated knockout. Furthermore, in a composite analysis of data available from the literature as well as our own, we describe the dinucleotide contexts in detail of rAPOBEC1-BE4 and TadA<sub>WT</sub>-TadA<sub>Evo</sub>-ABE7.10. Collectively, these results inform the application of base editing for protein knockout, and in selecting optimal BE-splice guides generated by our program, SpliceR. Ultimately, we believe the BE-splice knockout approach for protein knockout is a widely applicable technique for genetic engineering, and cell based therapies, holding particular promise in the context of multiplex editing.

## METHODS

### T cell isolation

Peripheral blood mononuclear cells (PBMCs) were isolated from Trima Accel leukoreduction system (LRS, Memorial Blood Center) chambers using ammonium chloride-based red blood cell lysis followed by Ficoll-Paque gradient (GE Healthcare). PBMCs were further purified for CD3+ cells by immunomagnetic negative selection using the EasySep Human T-cell Isolation Kit (STEMCELL Technologies, Cambridge, MA). T cells were frozen at  $10-20 \times 10^6$  cells per 1 mL of Cryostor CS10 (STEMCELL Technologies, Cambridge, MA) and thawed into culture for knockout experiments.

### Cell culture

Primary human T cells were cultured in OpTmizer CTS T cell Expansion SFM (ThermoFisher, Waltham, MA) containing 2.5% CTS Immune Cell SR (ThermoFisher), L-Glutamine, Penicillin/Streptomycin, N-Acetyl-L-cysteine (10 mM, Sigma-Aldrich), IL-2 (300 IU/mL, PeproTech), IL-7 (5 ng/mL, PeproTech), and IL-15 (5 ng/mL, PeproTech) at 37 °C and 5% CO<sub>2</sub>. Prior to electroporation T cells were activated with Dynabeads Human T-Activator CD3/CD28 (ThermoFisher, Waltham, MA) at a 2:1 bead:cell ratio for 48 h. Following electroporation, T cells were maintained at  $1 \times 10^6$  cells/mL in a 24-well or 12-well plate.

### Transfection

Forty-eight hours post activation, Dynabeads were magnetically removed and cells were washed one time with PBS prior to resuspension in the appropriate electroporation buffer. T cells were electroporated with 1 µg of chemically modified sgRNA (Synthego, Menlo Park, CA) and 1.5 µg codon optimized BE4 mRNA (TriLink Biotechnologies, San Diego, CA) codon optimized ABE7.10 mRNA (TriLink Biotechnologies, San Diego, CA), or codon optimized SpCas9 mRNA (TriLink Biotechnologies, San Diego, CA). Cas9 nuclease sgRNAs were designed using the Synthego knockout guide design (<https://design.synthego.com/#/>), BE-splice sgRNAs were designed using SpliceR ([z.umn.edu/splicer](http://z.umn.edu/splicer)), and pmSTOP sgRNAs were designed using the iSTOP database (<http://www.cicciolab-database.com/istop>). T cells were electroporated with the 4D-nucleofector system (Lonza, Basel, Switzerland) using a P3 16-well Nucleocuvette kit, with  $1 \times 10^6$  T cells per 20 µL cuvette using the program EO-115. T cells were allowed to recover in electroporation cuvettes for 15 minutes before resuspension in antibiotic-free medium at 37 °C, 5% CO<sub>2</sub> for 20 min following gene transfer, and were then cultured in complete CTS OpTmizer T cell Expansion SFM as described above.

### Flow cytometry

Seven days after electroporation a total of  $5 \times 10^6$  T cells were collected and stained in 100 µL with fluorophore-conjugated anti-human antibodies for CD3 (BD Biosciences #564001), beta-2-microglobulin (BioLegend #316306) and Fixable Viability Dye eFluor780 or LIVE/DEAD Fixable Aqua Dead Cell Stain (ThermoFisher #65-0865-14 1:500 dilution, #L34966 1:40 dilution) were used to assess cell viability. Events were acquired on LSR II or LSRFortessa flow cytometers

using FACSDiva software, and data was analyzed using FlowJo v10 software. See Supplementary Fig. 4 for gating strategy. The percent of positive events were normalized, by dividing over the percent positivity seen in pulse alone controls. Protein loss was then calculated as  $[1 - \% \text{ normalized sample positivity}]$ .

## Genetic analysis

Seven days after electroporation Genomic DNA was isolated from T cells by spin column-based purification (GeneJET, ThermoFisher). Editing efficiency was analyzed on the genomic level by PCR amplification of the targeted loci (Additional Data 7 for primers and guide protospacer), Sanger sequencing of the PCR amplicons (ACGT or Eurofins Genomics). Sanger sequencing traces of base edited samples were analyzed using EditR ([z.umn.edu/editr](https://z.umn.edu/editr))<sup>67</sup>, while indel efficiency was analyzed using the Synthego ICE (<https://ice.synthego.com/#/>)<sup>68</sup>.

## SpliceR development and whole genome guide prediction

SpliceR was written in the statistical programming language R (v. 3.6.1) using RStudio (v. 1.1.383). The SpliceR web app was developed using R shiny (<https://shiny.rstudio.com/>). All files are available online (<https://github.com/MoriarityLab/SpliceR>); dependencies.R is responsible for installing the necessary libraries to run the application; global.R defines the functions needed to interact with ensembl, and generate BE-splice sgRNAs. server.R and ui.R communicate with each other to handle the user inputs, generate sgRNAs, and return the output to the user interface. SpliceR relies on the various tidyverse (<https://www.tidyverse.org/>) and Bioconductor (<https://www.bioconductor.org/>) packages. For the whole genome guide prediction, the *generate\_guides()* function from the SpliceR web app was modified to run as a command line executable. All human protein coding gene ensembl transcript IDs were pulled from the GENCODE database (<https://www.encodegenes.org/human/>). The *generate\_guides()* function was then run in parallel across all Ensembl transcript IDs using the Minnesota Supercomputing Institute (MSI).

## Aggregate analysis of CBE and ABE context

Papers employing the CBE rAPOBEC1-BE4, and the ABE TadA<sub>WT</sub>-TadA<sub>Evo</sub>-ABE7.10 were found using PubMed and Google scholar. Using a combination of main and supplementary figures the editing values for each targetable base (adenosines for ABEs, and cytidines for CBEs) within the protospacer was manually recorded. Data generated from our own work was similarly entered. All manual data entries were double entered over two independent times to check for consistency of values. To control for variable electroporation and baseline editing efficiencies across different works, each editing value was normalized to the maximum editing efficiency observed in each cell type, for each work. To account for the same guide being used in different studies, normalized percent editing was averaged across each position of each unique guide. Data was then analyzed based on the pre-dinucleotide, and post-dinucleotide context of each base. R script for reproducible analysis is available in additional file 1, with aggregate analysis data from additional file 4 and additional file 5.

## Data analysis and visualization

All statistical analyses were performed in R using RStudio. The level of significance was set at  $\alpha = 0.05$ . All statistical tests were first subjected to assumptions of homoscedasticity. For samples with equal variance, Student's two-sample, unpaired two-tailed t-test was used, while for samples with unequal variance Welch's two-sample, unpaired two-tailed t-test was used. Data were visualized in R studio employing various tidyverse (<https://www.tidyverse.org/>), Bioconductor (<https://www.bioconductor.org/>) packages, and ggseqlogo (<https://github.com/omarwagih/ggseqlogo>) packages. See R script in additional file 1 for reproducible analyses.



# FIGURE LEGENDS

1. Overview of the BE-splice approach. (a) Diagram of base editor complex, table of the enzymes used in this study, and base editing sgRNAs with editing window. (b) Positioning of BE-splice sgRNAs within conserved splice donor and splice acceptors motif. Logo plots were generated from all human protein coding gene splice sites. Arrows indicate the base targeted by either CBEs (blue), or ABEs (green). (c) Breakdown of transcripts and genes targetable by BE-splice, showing the vast majority of non-single-exon-single-isoform (NSES) genes are targetable by this approach (99.68%). (d) Distribution of BE-splice sgRNA density across each gene. 50% of genes have 62 or more sgRNAs mapping to them when accounting for all PAM identities and both CBE and ABE approaches. (e) Distribution of the position of the first sgRNA for each gene, with 50% having their first sgRNA 11.13% way through the mRNA or earlier.
2. Conception and validation of the TCR-CD3-MHC Class I immune synapse as a screening model for functional knockout. (a) Diagram of the multimeric TCR-CD3 complex and MHC Class I immune synapse containing multiple NSES genes, based on the solved structures (PDB 6JXR<sub>43</sub>, PDB 3T0E<sub>47</sub>; PDB 10GA<sub>48</sub>). (b) Diagram of the synthesis and localization of the TCR-CD3 complex and interaction with MHC Class I. All members of the CD3 complex are required before functional localization to the cell surface, where disruption of a single splice site within one gene member can prevent a surface expressed complex from forming. (c) Cas9 nuclease knockout of each individual member of TCR-CD3 complex validates the screening model. Two Cas9 nuclease sgRNAs were designed to exonic regions of each gene in the complex. All genes had at least one guide with >85% indel efficiency and loss in TCR-CD3 surface expression. N = 2 independent donors.
3. BE-splice sgRNAs mediate robust editing and disruption of TCR-CD3 MHC Class I immune synapse. (a) Editing efficiency (top) and surface protein loss (bottom) from each guide in the sgRNA screen. Results grouped by gene and enzyme used in descending order by protein loss. X-axis label indicates position of target base within sgRNA. TRBC1 and TRBC2 were omitted from the BE-splice screen due to the inability to design single BE-splice sgRNAs to target both paralogs simultaneously. (b) Base editing efficiencies grouped by enzyme and target motif analyzed with Student's t-test or Welch's t-test. (c) Protein loss efficiencies grouped by enzyme and target motif analyzed with Student's t-test or Welch's t-test. (d) Consistency of editing efficiency and protein loss across all approaches employed here. Relationship between protein loss and base editing efficiency is comparable to that observed in Cas9 control. N = 2 independent donors, performed on different days. *n.s.*  $P > 0.05$ , \*  $P \leq 0.05$ , \*\*  $P \leq 0.01$ , \*\*\*  $P \leq 0.001$ .
4. Context dependencies of base editors and BE-splice target motifs. (a) Preceding dinucleotide context dependencies of rAPOBEC1-BE4. Results normalized and aggregated across published results and our own work. Smoothed distributions generated using LOESS regression with span = 0.5. (b) Preceding dinucleotide context dependencies of TadA<sub>WT</sub>-TadA<sub>Evo</sub>-ABE7.10. Results normalized and aggregated across published results and our own work. Smoothed distributions generated using LOESS regression with span = 0.45. 6 papers, 102 guides, 447 edits. (c) Logo plots of pentanucleotide motif of each enzyme and target motif combination in this work. Plots are consistent with conserved splice site sequences.
5. Consistency of editing efficiency and protein loss across mRNA and protein regions. (a) Scatter plots of protein loss and editing efficiency by exon group across all base editor approaches employed. The strongest relationship is observed among middle exons, while the weakest is observed in the last exon. (b) Error in protein loss as a function of editing efficiency across each exon group. The least error is observed among middle

exons, while the greatest is observed in the last exon. (c) Scatter plots of the protein loss and editing efficiency grouped by where each BE sgRNA maps to the TCR-CD3 and MHC Class I structures. sgRNAs that map to transmembrane and extracellular regions exhibit the greatest consistency between protein loss and base editing efficiency.

## DATA AVAILABILITY

Data and scripts for analysis and figure reproduction are available on GitHub. Source data are available from corresponding author upon request.

## REFERENCES

1. Adli, M. The CRISPR tool kit for genome editing and beyond. *Nat Commun* **9**, 1911 (2018).
2. Mali, P. *et al.* RNA-Guided Human Genome Engineering via Cas9. *Science* **339**, 823–826 (2013).
3. Cong, L. *et al.* Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science* **339**, 819–823 (2013).
4. Jinek, M. *et al.* RNA-programmed genome editing in human cells. *eLife* **2**, e00471 (2013).
5. Jinek, M. *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816–821 (2012).
6. Ran, F. A. *et al.* Genome engineering using the CRISPR-Cas9 system. *Nat Protoc* **8**, 2281–2308 (2013).
7. Stadtmauer, E. A. *et al.* CRISPR-engineered T cells in patients with refractory cancer. *Science* **367**, eaba7365 (2020).
8. Jung, I.-Y. & Lee, J. Unleashing the Therapeutic Potential of CAR-T Cell Therapy Using Gene-Editing Technologies. *Molecules and Cells* **41**, 717–723 (2018).
9. Waseem Qasim. Molecular remission of infant B-ALL after infusion of universal TALEN gene-edited CAR T cells. (2017).
10. Ren, J. *et al.* A versatile system for rapid multiplex genome-edited CAR T cell generation. *Oncotarget* **8**, (2017).
11. Cullot, G. *et al.* CRISPR-Cas9 genome editing induces megabase-scale chromosomal truncations. *Nat Commun* **10**, 1136 (2019).
12. Kosicki, M., Tomberg, K. & Bradley, A. Repair of double-strand breaks induced by CRISPR–Cas9 leads to large deletions and complex rearrangements. *Nat Biotechnol* **36**, 765–771 (2018).
13. Ihry, R. J. *et al.* p53 inhibits CRISPR–Cas9 engineering in human pluripotent stem cells. *Nat Med* **24**, 939–946 (2018).
14. Webber, B. R. *et al.* Highly efficient multiplex human T cell engineering without double-strand breaks using Cas9 base editors. *Nat Commun* **10**, 5222 (2019).
15. Chen, J. *et al.* NR4A transcription factors limit CAR T cell function in solid tumours. *Nature* **567**, 530–534 (2019).
16. Nishida, K. *et al.* Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. *Science* **353**, 10.1126/science.aaf8729. Epub 2016 Aug 4 (2016).
17. Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A. & Liu, D. R. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* (2016).
18. Gaudelli, N. M. *et al.* Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. *Nature* **551**, 464–471 (2017).
19. Billon, P. CRISPR-Mediated Base Editing Enables Efficient Disruption of Eukaryotic Genes through Induction of STOP Codons. *Molecular Cell* 1068–1079 (2017).

20. Kuscu, C. *et al.* CRISPR-STOP: gene silencing through base-editing-induced nonsense mutations. *Nat. Methods* **14**, 710–712 (2017).
21. Gapinske, M. *et al.* CRISPR-SKIP: programmable gene splicing with single base editors. *Genome Biol* **19**, 107 (2018).
22. Winter, J. *et al.* Targeted exon skipping with AAV-mediated split adenine base editors. *Cell Discov* **5**, 41 (2019).
23. Yuan, J. *et al.* Genetic Modulation of RNA Splicing with a CRISPR-Guided Cytidine Deaminase. *Molecular Cell* **72**, 380-394.e7 (2018).
24. Liu, Y. *et al.* Genome-wide screening for functional long noncoding RNAs in human cells by Cas9 targeting of splice sites. *Nat Biotechnol* **36**, 1203–1210 (2018).
25. Garc, I. Splice donor site sgRNAs enhance CRISPR/Cas9-mediated knockout efficiency. *PLOS ONE* **14**, 19 (2019).
26. Liu, Z. *et al.* Improved base editor for efficient editing in GC contexts in rabbits with an optimized AID-Cas9 fusion. *FASEB j.* **33**, 9210–9219 (2019).
27. Xie, J. *et al.* Efficient base editing for multiple genes and loci in pigs using base editors. *Nat Commun* **10**, 2852 (2019).
28. Tanaka, S. *et al.* In vivo targeted single-nucleotide editing in zebrafish. *Sci Rep* **8**, 11423 (2018).
29. Depil, S., Duchateau, P., Grupp, S. A., Mufti, G. & Poirot, L. 'Off-the-shelf' allogeneic CAR T cells: development and challenges. *Nat Rev Drug Discov* **19**, 185–199 (2020).
30. Dustin, M. L. The Immunological Synapse. *Cancer Immunology Research* **2**, 1023–1033 (2014).
31. Zhang, J. & Wang, L. The Emerging World of TCR-T Cell Trials Against Cancer: A Systematic Review. *Technol Cancer Res Treat* **18**, 153303381983106 (2019).
32. Zhao, L. & Cao, Y. J. Engineered T Cell Therapy for Cancer in the Clinic. *Front. Immunol.* **10**, 2250 (2019).
33. Chandran, S. S. & Klebanoff, C. A. T cell receptor-based cancer immunotherapy: Emerging efficacy and pathways of resistance. *Immunol Rev* **290**, 127–147 (2019).
34. Alcover, A., Alarcón, B. & Di Bartolo, V. Cell Biology of T Cell Receptor Expression and Regulation. *Annu. Rev. Immunol.* **36**, 103–125 (2018).
35. Koblan, L. W. *et al.* Improving cytidine and adenine base editors by expression optimization and ancestral reconstruction. *Nat Biotechnol* **36**, 843–846 (2018).
36. Martin, A. St. *et al.* A panel of eGFP reporters for single base editing by APOBEC-Cas9 editosome complexes. *Sci Rep* **9**, 497 (2019).
37. Thuronyi, B. W. *et al.* Continuous evolution of base editors with expanded target compatibility and improved activity. *Nat Biotechnol* **37**, 1070–1079 (2019).
38. Yu, Y. *et al.* Next-generation cytosine base editors with minimized unguided DNA and RNA off-target events and high on-target activity.  
<http://biorxiv.org/lookup/doi/10.1101/2020.02.11.944165> (2020)  
doi:10.1101/2020.02.11.944165.
39. Kang, B.-C. *et al.* Precision genome engineering through adenine base editing in plants. *Nature Plants* **4**, 427–431 (2018).
40. Li, G. *et al.* Efficient Generation of Pathogenic A-to-G Mutations in Human Triprenuclear Embryos via ABE-Mediated Base Editing. *Molecular Therapy - Nucleic Acids* **17**, 289–296 (2019).
41. Ryu, S.-M. *et al.* Adenine base editing in mouse embryos and an adult mouse model of Duchenne muscular dystrophy. *Nat Biotechnol* **36**, 536–539 (2018).
42. Gehrke, J. M. *et al.* An APOBEC3A-Cas9 base editor with minimized bystander and off-target activities. *Nat Biotechnol* **36**, 977–982 (2018).
43. Dong, D. *et al.* Structural basis of assembly of the human T cell receptor–CD3 complex. *Nature* **573**, 546–552 (2019).

44. Brazin, K. N. *et al.* The T Cell Antigen Receptor  $\alpha$  Transmembrane Domain Coordinates Triggering through Regulation of Bilayer Immersion and CD3 Subunit Associations. *Immunity* **49**, 829–841.e6 (2018).
45. Krshnan, L., Park, S., Im, W., Call, M. J. & Call, M. E. A conserved  $\alpha\beta$  transmembrane interface forms the core of a compact T-cell receptor–CD3 structure within the membrane. *Proc Natl Acad Sci USA* **113**, E6649–E6658 (2016).
46. Birnbaum, M. E. *et al.* Molecular architecture of the  $\alpha\beta$  T cell receptor–CD3 complex. *Proc Natl Acad Sci USA* **111**, 17576–17581 (2014).
47. Yin, Y., Wang, X. X. & Mariuzza, R. A. Crystal structure of a complete ternary complex of T-cell receptor, peptide-MHC, and CD4. *Proceedings of the National Academy of Sciences* **109**, 5405–5410 (2012).
48. Stewart-Jones, G. B. E., McMichael, A. J., Bell, J. I., Stuart, D. I. & Jones, E. Y. A structural basis for immunodominant human T cell receptor recognition. *NATURE IMMUNOLOGY* **4**, 7 (2003).
49. Wolf, J. tadA, an essential tRNA-specific adenosine deaminase from Escherichia coli. *The EMBO Journal* **21**, 3841–3851 (2002).
50. Kim, J. *et al.* Structural and Kinetic Characterization of *Escherichia coli* TadA, the Wobble-Specific tRNA Deaminase  $\dagger$ . *Biochemistry* **45**, 6407–6416 (2006).
51. Gaudelli, N. M. *et al.* Directed Evolution of Adenine Base Editors with Increased Activity and Therapeutic Application. <http://biorxiv.org/lookup/doi/10.1101/2020.03.13.990630> (2020) doi:10.1101/2020.03.13.990630.
52. Michelle F. Richter, Kevin T. Zhao, Elliot Eton, Audrone Lapinaite, Gregory A. Newby, Benjamin W. Thuronyi, Christopher Wilson, Luke W. Koblan, Jing Zeng, Daniel E. Bauer, Jennifer A. Doudna & David R. Liu. Phage-assisted evolution of an adenine base editor with improved Cas domain compatibility and activity. *Nat Biotechnol* (2020) doi:<https://doi.org/10.1038/s41587-020-0453-z>.
53. Eggington, J. M., Greene, T. & Bass, B. L. Predicting sites of ADAR editing in double-stranded RNA. *Nat Commun* **2**, 319 (2011).
54. Yuan Zhuang. A Compensatory Base Change in U1 snRNA Suppresses a 5' Splice Site Mutation. (1986).
55. Siliciano, P. G. & Guthrie, C. 5' splice site selection in yeast: genetic alterations in base-pairing with U1 reveal additional requirements. *Genes & Development* **2**, 1258–1267 (1988).
56. Anna, A. & Monika, G. Splicing mutations in human genetic disorders: examples, detection, and confirmation. *J Appl Genetics* **59**, 253–268 (2018).
57. Mora, M. *et al.* New Splice Site Acceptor Mutation in AIRE Gene in Autoimmune Polyendocrine Syndrome Type 1. *PLoS ONE* **9**, e101616 (2014).
58. Tereza Grymová. SERPING1 exon 3 splicing variants using alternative acceptor splice sites. (2019).
59. Smith, D. *et al.* A rare IL33 loss-of-function mutation reduces blood eosinophil counts and protects from asthma. *PLoS Genet* **13**, e1006659 (2017).
60. Anderson, J. L. *et al.* mRNA processing in mutant zebrafish lines generated by chemical and CRISPR-mediated mutagenesis produces unexpected transcripts that escape nonsense-mediated decay. *PLoS Genet* **13**, e1007105 (2017).
61. Savas, S., Tuzmen, S. & Ozcelik, H. Human SNPs resulting in premature stop codons and protein truncation. *Hum Genomics* **2**, 274 (2006).
62. Lewis, B. P., Green, R. E. & Brenner, S. E. Evidence for the widespread coupling of alternative splicing and nonsense-mediated mRNA decay in humans. *Proceedings of the National Academy of Sciences* **100**, 189–192 (2003).
63. Anzalone, A. V. *et al.* Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* **576**, 149–157 (2019).



64. Dever, D. P. *et al.* CRISPR/Cas9  $\beta$ -globin gene targeting in human haematopoietic stem cells. *Nature* **539**, 384–389 (2016).
65. Wu, Y. *et al.* Highly efficient therapeutic gene editing of human hematopoietic stem cells. *Nat Med* **25**, 776–783 (2019).
66. Haapaniemi, E., Botla, S., Persson, J., Schmierer, B. & Taipale, J. CRISPR–Cas9 genome editing induces a p53-mediated DNA damage response. *Nat Med* **24**, 927–930 (2018).
67. Kluesner, M. G. *et al.* EditR: A Method to Quantify Base Editing from Sanger Sequencing. *The CRISPR Journal* **1**, 239–250 (2018).
68. Hsiao, T. *et al.* *Inference of CRISPR Edits from Sanger Trace Data*. <http://biorxiv.org/lookup/doi/10.1101/251082> (2018) doi:10.1101/251082.

## ACKNOWLEDGEMENTS

We thank Dr. Matthew Johnson at the University of Minnesota for discussions surrounding flow cytometry. We thank Dr. Kevin Holden at Synthego for insight about chemically modified sgRNAs.

## AUTHOR INFORMATION

These authors contributed equally: Mitchell G. Kluesner, Walker S. Lahr  
These authors contributed equally: Beau R. Webber, Branden S. Moriarity

## Affiliations

*Department of Pediatrics, University of Minnesota, Minneapolis, MN, USA*

Mitchell G. Kluesner, Walker S. Lahr, Cara-Lin Lonetree, Branden A. Smeester, Patricia N. Claudio-Vázquez, Samuel P. Binge, Beau R. Webber & Branden S. Moriarity

*Masonic Cancer Center, University of Minnesota, Minneapolis, MN, USA*

Mitchell G. Kluesner, Walker S. Lahr, Cara-Lin Lonetree, Branden A. Smeester, Patricia N. Claudio-Vázquez, Samuel P. Binge, Beau R. Webber & Branden S. Moriarity

*Center for Genome Engineering, University of Minnesota, Minneapolis, MN, USA*

Mitchell G. Kluesner, Walker S. Lahr, Cara-Lin Lonetree, Branden A. Smeester, Patricia N. Claudio-Vázquez, Samuel P. Binge, Beau R. Webber & Branden S. Moriarity

*Stem Cell Institute, University of Minnesota, Minneapolis, MN, USA*

Mitchell G. Kluesner, Walker S. Lahr, Cara-Lin Lonetree, Branden A. Smeester, Patricia N. Claudio-Vázquez, Samuel P. Binge, Beau R. Webber & Branden S. Moriarity

*Department of Genetics, Cell Biology, and Development, University of Minnesota, Minneapolis, MN, USA*

Patricia N. Claudio-Vázquez, Samuel P. Pitzen & Branden S. Moriarity

*College of Biological Sciences, University of Minnesota, Minneapolis, MN, USA*

Madison J. Vignes, Samantha C. Lee & Aneisha A. Andrews

## Contributions

M.G.K., W.S.L., B.S.M., and B.R.W. designed experiments. M.G.K., W.S.L., C.L.L., B.A.S., P.N.C., M.J.V., S.C.L., S.P.B., and A.A.A. performed experiments. M.G.K., W.S.L., C.L.L., S.P.B., S.P.P., and S.C.L. analyzed the data. M.G.K. wrote the program. M.G.K., W.S.L., B.S.M., and B.R.W. wrote the manuscript. All authors approved and read the final manuscript.

# **Corresponding author**

Correspondence and materials request should be addressed to Dr. Branden S. Moriarity.

## **ETHICS DECLARATIONS**

### **Competing interests**

B.R.W. and B.S.M. are consultants for Beam Therapeutics. B.R.W and B.S.M. have financial interests in Beam Therapeutics. Both B.R.W. and B.S.M.'s interests were reviewed and are managed by the University of Minnesota in accordance with their conflict of interest policies. Patents have also been filed on the findings and concepts of utilizing base editors for gene knockout and gene correction.

## **SUPPLEMENTARY INFORMATION**

1. Supplementary figures
2. Script for figure reproduction of analyses and figures
3. Whole genome BE-splice sgRNA prediction data
4. Experimental data for each sample
5. CBE aggregate analysis data
6. ABE aggregate analysis data
7. SnapGene maps of genes with annotated BE-splice sgRNAs
8. Primer sequences for each sgRNA

# CRISPR-Cas9 cytidine and adenosine base editing of splice-sites mediates highly-efficient disruption of proteins in primary cells

Figures

Updated April 3<sup>rd</sup>, 2020

## AUTHORS

Mitchell G. Kluesner<sup>1,2,3,4\*</sup>, Walker S. Lahr<sup>1,2,3,4</sup>, \* Cara-Lin Lonetree<sup>1,2,3,4</sup>, Branden A. Smeester<sup>1,2,3,4</sup>, Patricia N. Claudio-Vázquez<sup>1,2,3,4,5</sup>, Samuel P. Pitzén<sup>5</sup>, Madison J. Vignes<sup>6</sup>, Samantha C. Lee<sup>6</sup>, Samuel P. Bingea<sup>1,2,3,4</sup>, Aneesha A. Andrews<sup>6</sup>, Beau R. Webber<sup>1,2,3,4,†</sup>, Branden S. Moriarity<sup>1,2,3,4,†</sup>

\*These authors contributed equally

†These authors contributed equally

## AFFILIATIONS

<sup>1</sup>Department of Pediatrics, University of Minnesota, Minneapolis, MN, USA

<sup>2</sup>Masonic Cancer Center, University of Minnesota, Minneapolis, MN, USA

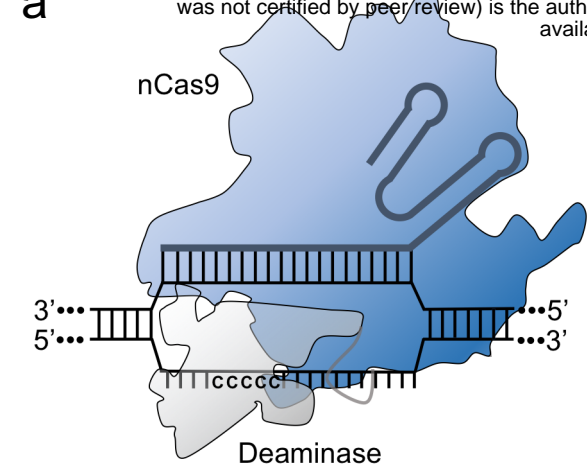
<sup>3</sup>Center for Genome Engineering, University of Minnesota, Minneapolis, MN, USA

<sup>4</sup>Stem Cell Institute, University of Minnesota, Minneapolis, MN, USA

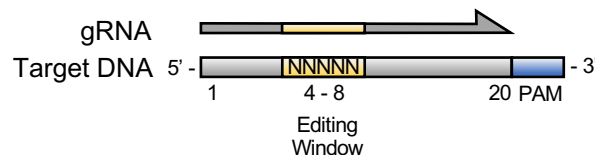
<sup>5</sup>Department of Genetics, Cell Biology, and Development, University of Minnesota, Minneapolis, MN, USA

<sup>6</sup>College of Biological Sciences, University of Minnesota, Minneapolis, MN, USA

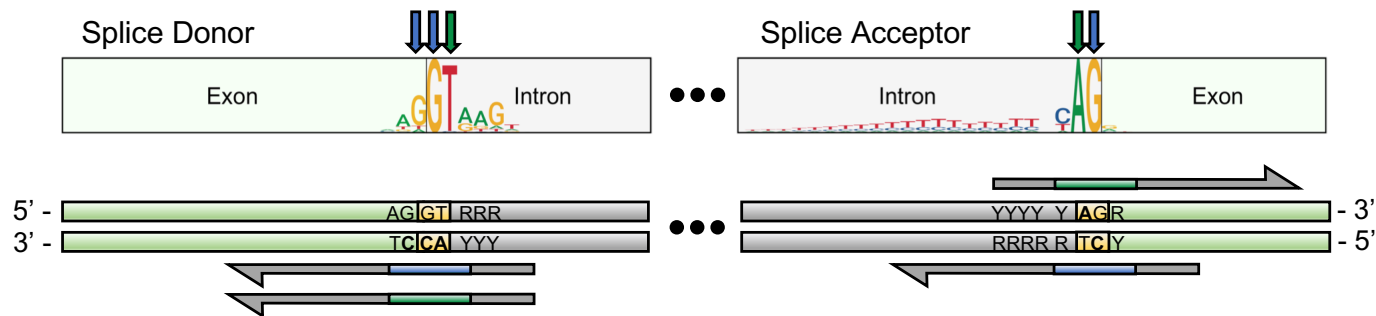
**a**



Enzyme	Reaction	Deaminase	C-terminus
BE4	C:G > T:A	rAPOBEC1	UGI-UGI-NLS
ABE7.10	A:T > G:A	TadA <sup>WT</sup> -TadA <sup>Evo</sup>	NLS



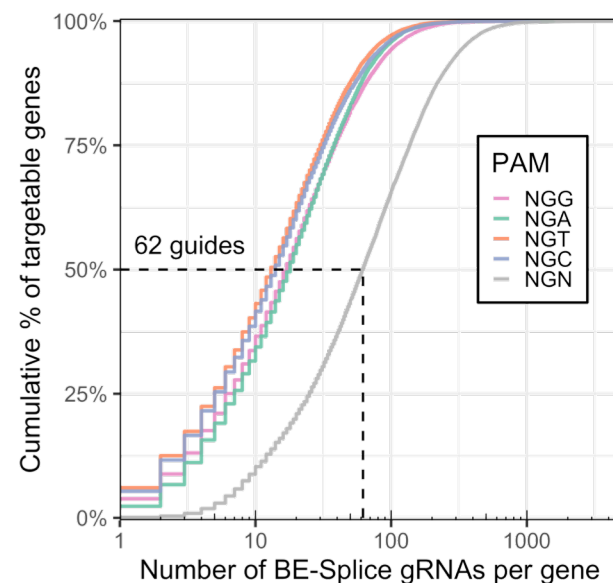
**b**



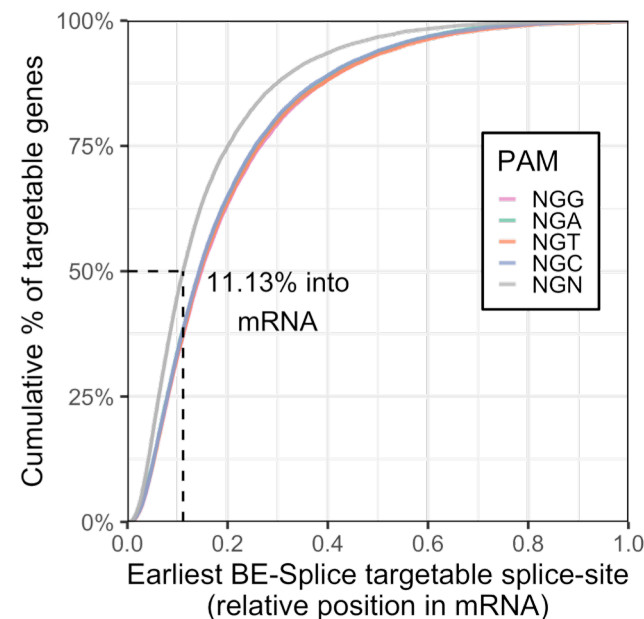
**c**

Targetable Proteins	Number	Percent of NSESI	Percent of Total
Transcripts	93,665	99.85%	95.86%
Genes	19,066	99.68%	94.23%

**d**

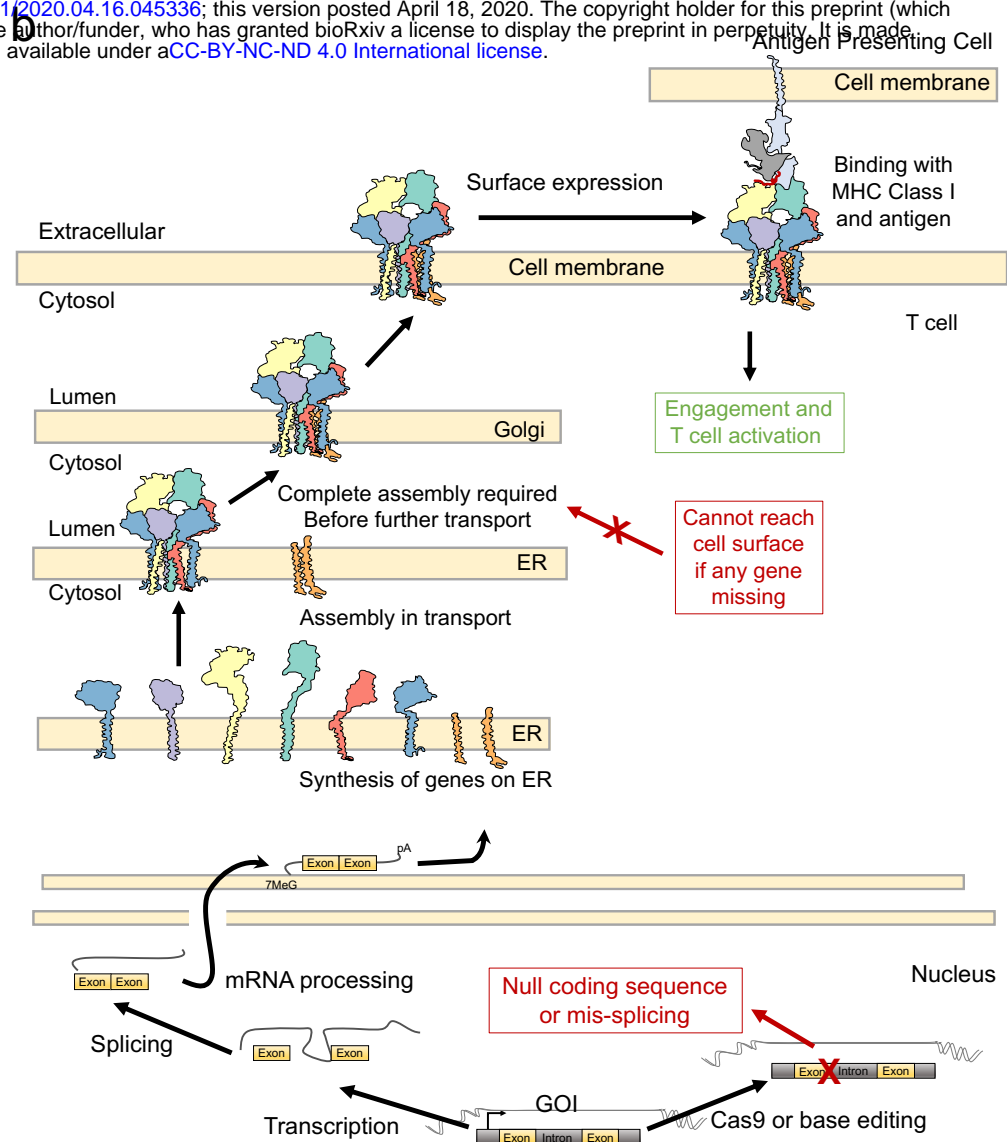
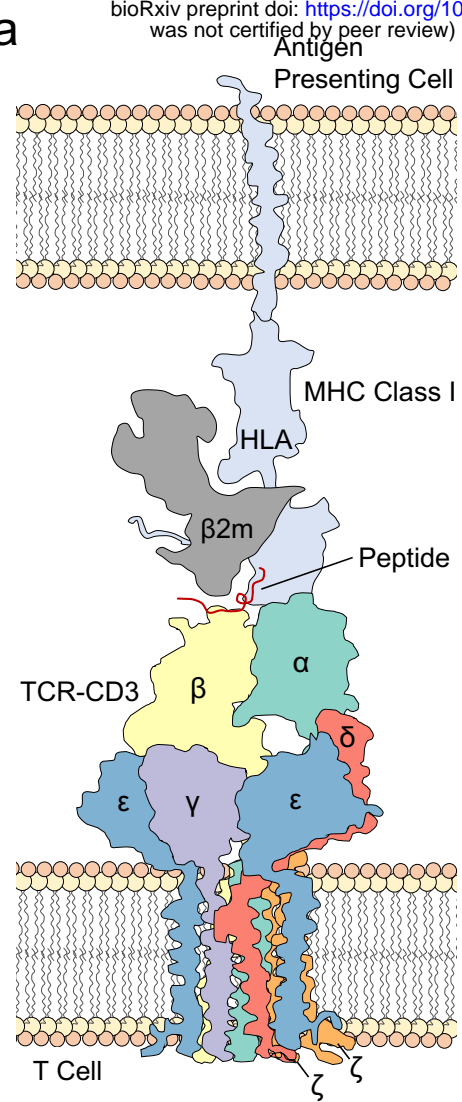


**e**

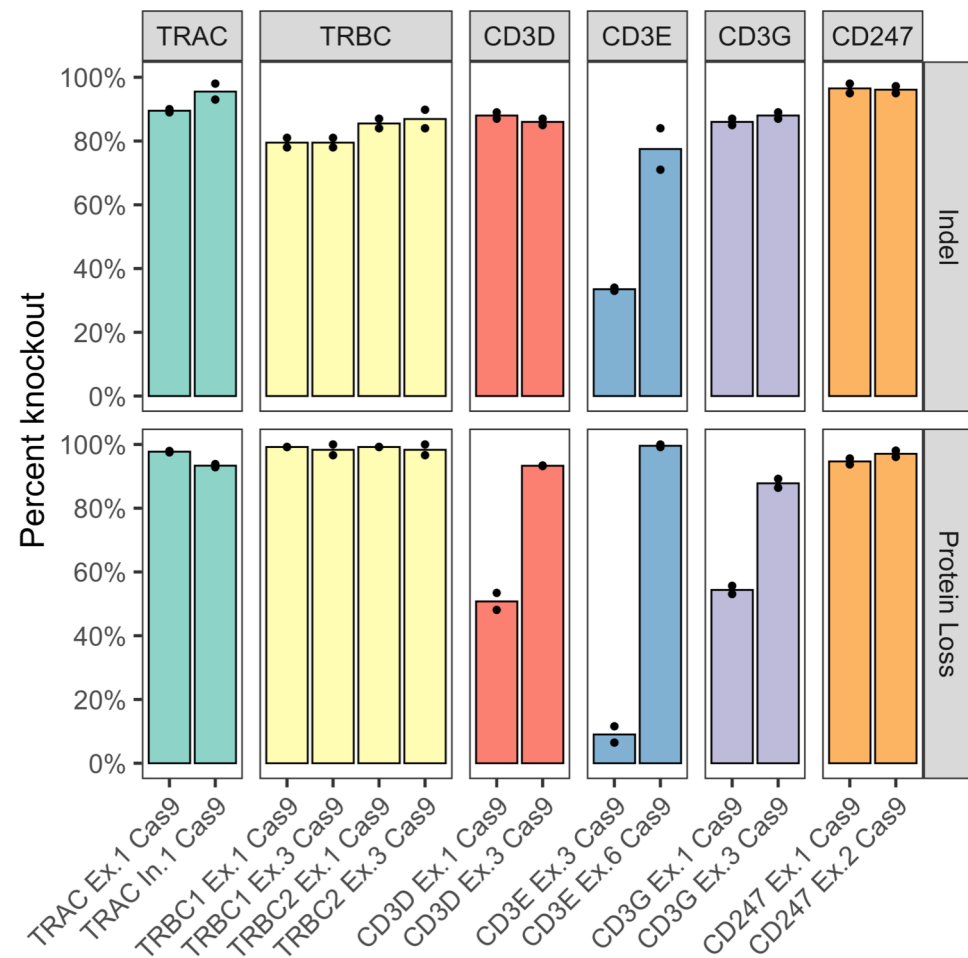


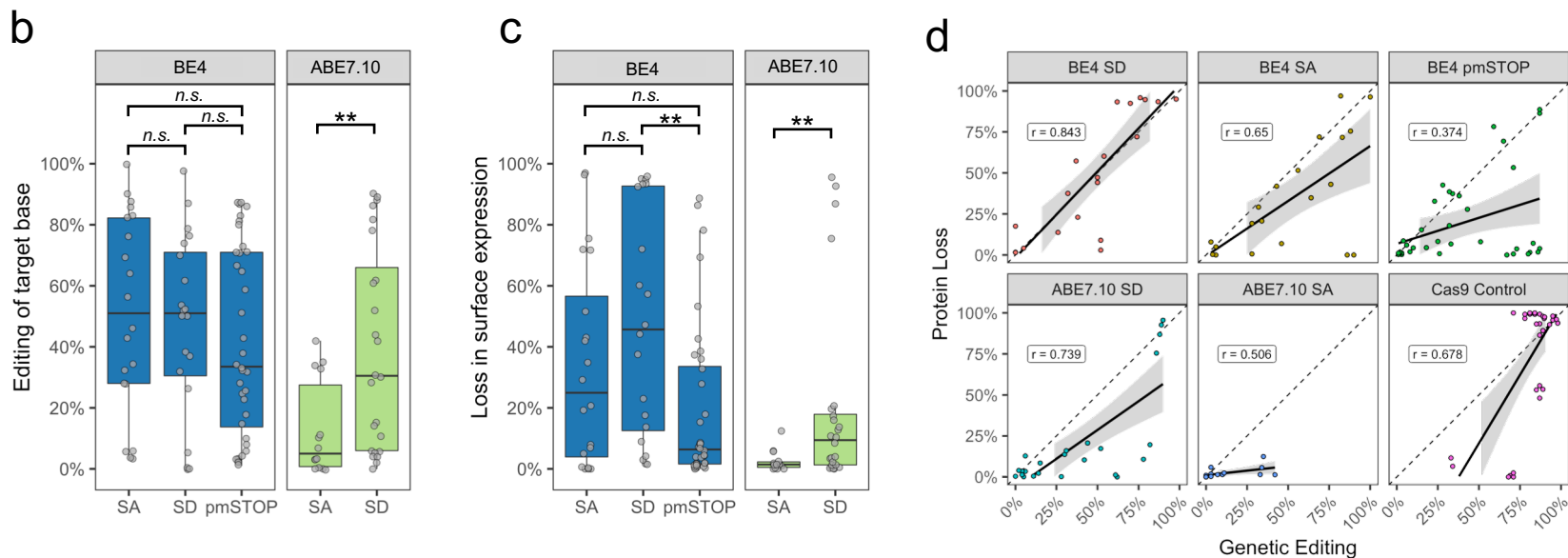
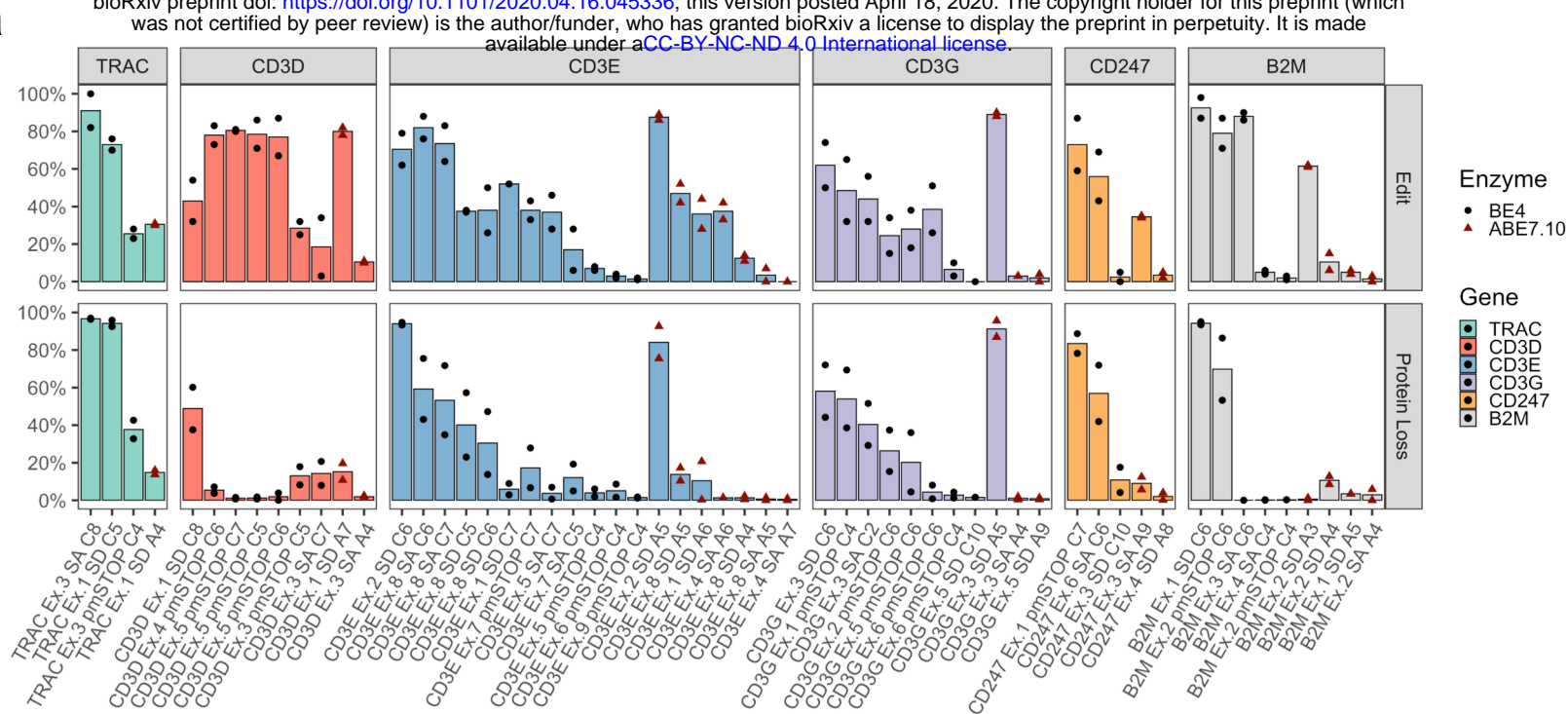


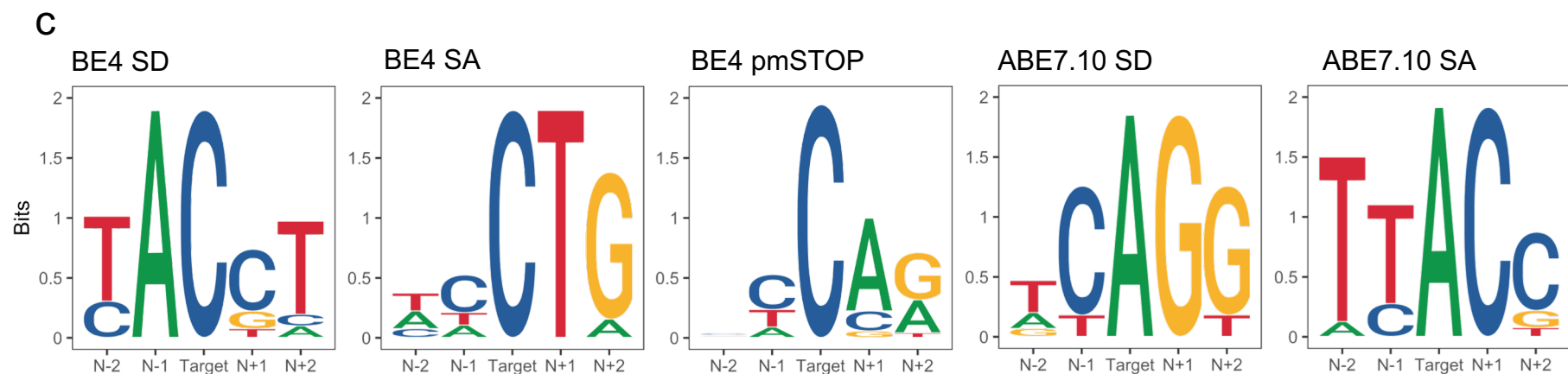
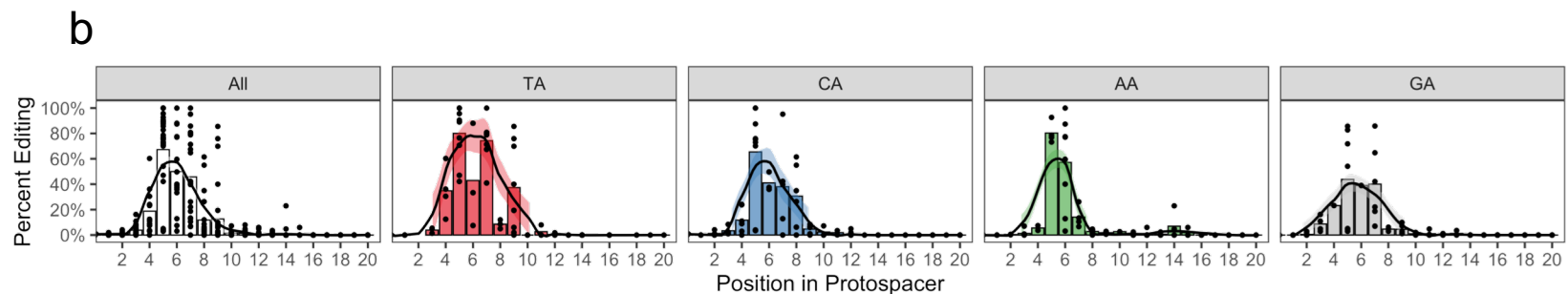
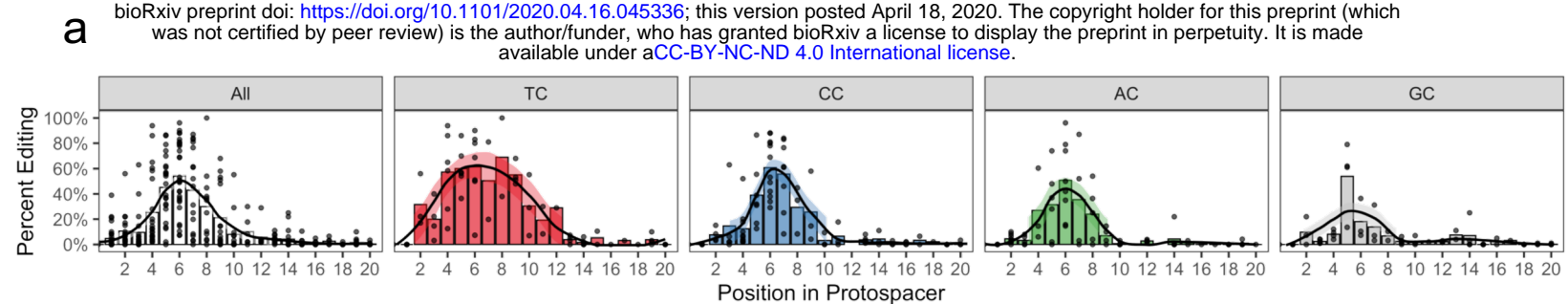
a



c

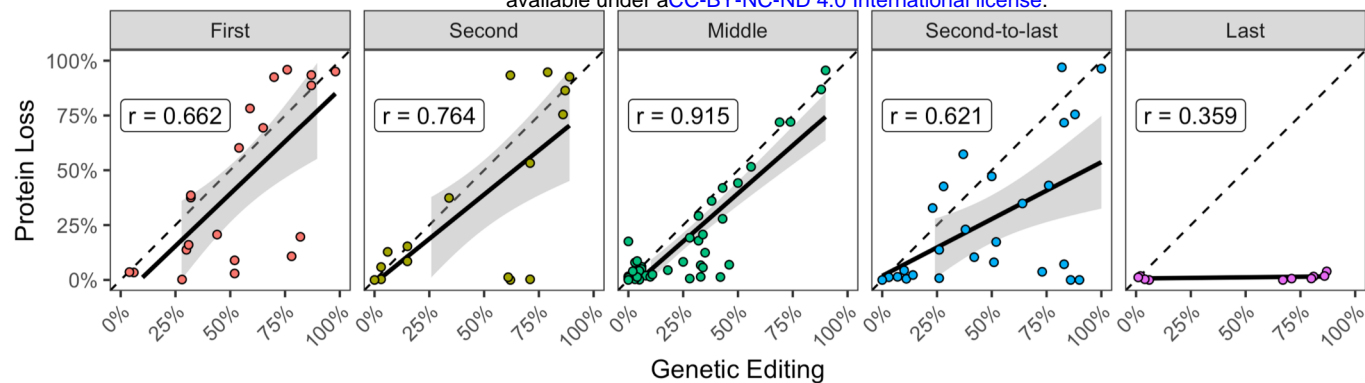
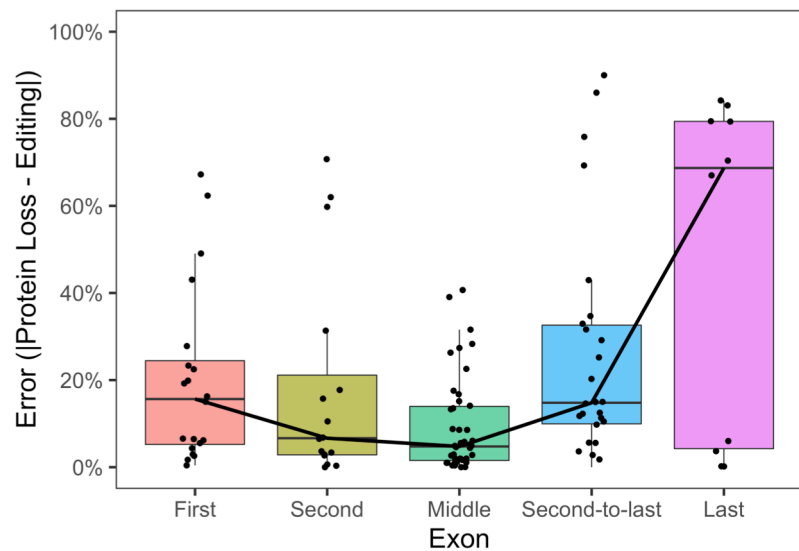






**a**

bioRxiv preprint doi: <https://doi.org/10.1101/2020.04.16.045336>; this version posted April 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

**b****c**