P667

Comprehensive Analysis of the Editing Window of TALE Base Editors

Maria Feola[1], Sylvain Pulicani[2], Diane Tkach[1], Robert Hong[1], Alex Boyne[1], Louisa Mayer[1], Aymeric Duclert[2], Philippe Duchateau[2], Alexandre Juillerat[1]

[1] Cellectis Inc, New York, NY; [2] Cellectis SA, Paris, France

Contact: maria.feola@cellectis.com

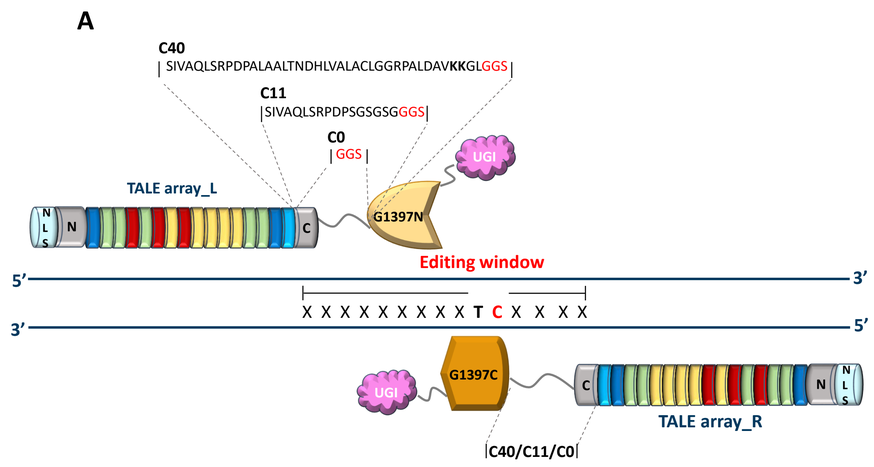
Abstract

One of the most recent advances in the genome editing field has been the addition of the so-called “C-to-T TALE base editors” (TALE-BE), an innovative platform for cell therapy that relies on the deamination of cytidines within double strand DNA, through the formation of an uracil (U) intermediate. These molecular tools are fusions of a transcription activator-like effector (TALE) domain for the binding of a specific DNA sequence, split-DddA deaminase halves that will catalyze the conversion of a cytosine (C) to a thymine (T) upon reconstitution, and an uracil glycosylase inhibitor (UGI).

Here we aimed to systematically investigate the influence of the sequence context surrounding the targeted Cytosine on TALE-BE C to T conversion efficiency. Recently we developed a strategy that allowed us to comprehensively characterize editing efficiencies as a function of the of TC position within the TALE-BE editing windows. This experimental setup provides a high throughput screening format for editing efficiency in a precisely defined genomic context in cellulo. In particular excluding or minimizing biases that would arise from microenvironmental and/or epigenetic differences, such as chromosome relaxation, that would be present among different genomic loci. This method is specifically taking advantage of the highly precise and efficient TALEN mediated ssODN knock-in in primary T cells, allowing us to focus on how target composition and spacer variations can affect TALE-BE activity/efficiency. The robustness and versatility of this screening strategy enabled us to explore the influence of the bases positioned before and after a fixed TC sequence and defined their contributions to the editing efficiency and delineate rules to determine optimal TALE-BE design. Moreover, the in cellulo high throughput screening system could be used to engineer the DddAtox deaminase and validated new mutants with defined characteristics. Overall, we believe that the knowledge obtained will allow for better design of efficient TALE-BE while improving the specificity profiles of this innovative editing platform.

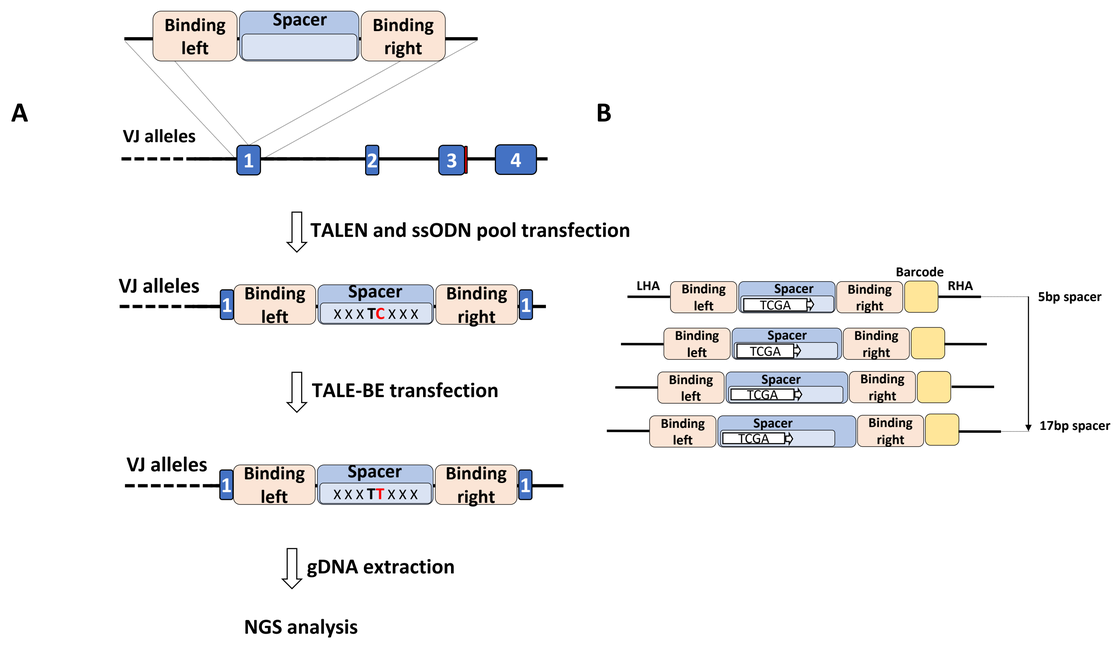
#1 Background

Base editing is a technology that leads to the introduction of point mutations (C>T transitions) in defined loci of a targeted DNA sequence. It create mutations by deamination of the targeted bases (C or A), which are then converted into T or G during DNA repair process without creating DNA double strand breaks, making it a promising therapeutic strategy for genetic diseases. The first development of TALE-BE rely on the use of a TALEN scaffold; here, the domain linking the TALE DNA binding domain to the catalytic domain is composed of the native first 40 amino acids from the C-ter domain of TALE from Xanthomonas (AvrBs3, here called C40), to which a GGS sequence was appended. In a TALEN context, the linker could impact the overall efficiency of such molecular tools. Here we describe the impact of the linker nature (length and composition), on the C-to-T conversion editing efficiency within the editing window. We compared the **C40/C40** scaffold with 2 additional scaffolds that were obtained by removing most of the C terminal end of the TALE domain (keeping only the first 11 amino acids, amino acids 887-897 plus the GGS flexible linker, called **C11/C11**), or completely eliminate it (maintaining only the GGS linker, called **C0/C0**). **A.** Schematic representation of a TALE-BE in the **C40**, **C11** and **C0** scaffolds).



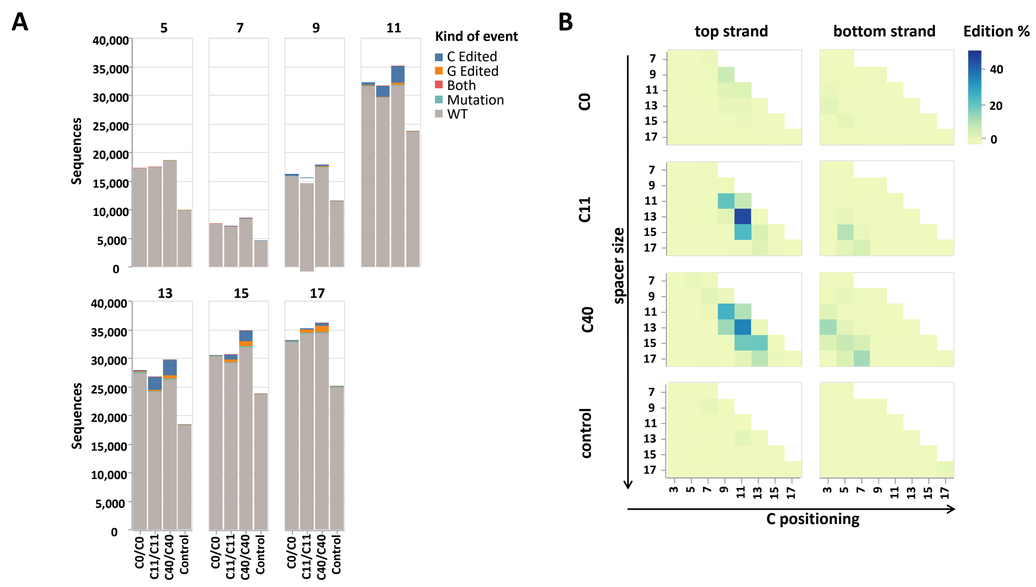
#2 Definition of the optimal TALE-BE editing window: Artificial spacer pool for high throughput testing

A. Scheme of the strategy to generate artificial base editor target sites. In the first step a pool of ssODN encoding various base editor spacer sequences flanked by a specific TALE-targeted sequences is inserted into TRAC locus. In a second step the corresponding TALE-BE is transfected. Two days post transfection the genomic DNA is collected, and the inserted sequence is analyzed by NGS. **B.** Schematic representation of the 37 unique ssODN pools characterized by spacers with varying lengths spanning from 5 to 17 bp, **to determine the impact of shortening the linker on editing efficiency as a function of the spacer length**. A TCGA quadruplex target sequence was incorporated into the spacer at varying positions from 5’ to 3’. Spacer lengths were limited to odd numbers (eg 5bp, 7bp, 11bp, etc.)



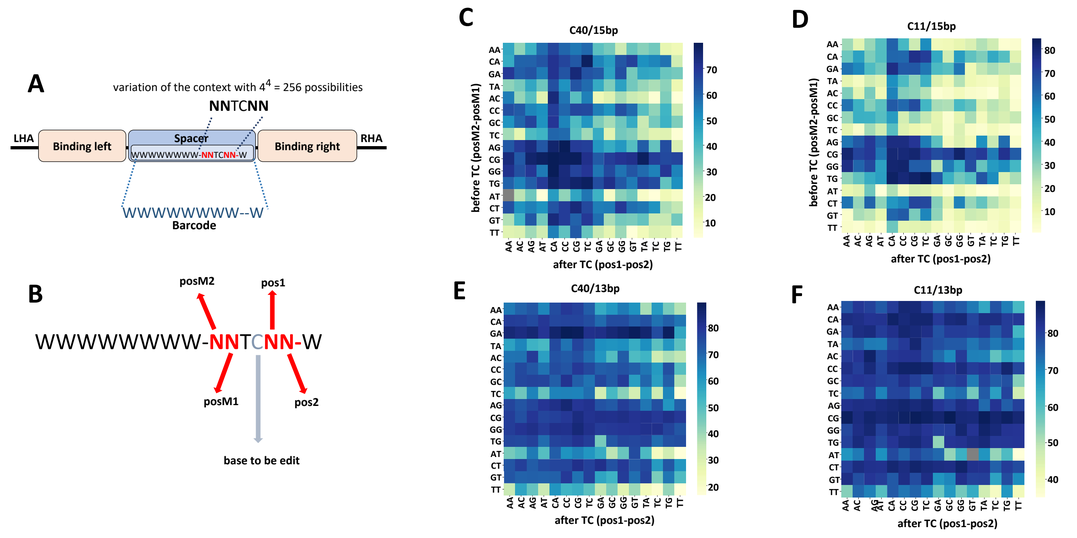
#3 Analysis of the linker nature and spacer length influence on TALE-BE C-to-T conversion

A. Quality control graphs. Number of sequences per base editor transfection and spacer size. The total number of sequences per scaffold are figured as bar plots. The subplots show the counts by spacer size. The color encodes the events, with "WT" meaning no event, "C Edited" meaning a sequence with a TG instead of the CG, "G Edited" meaning a CA instead of the CG, "Both" meaning a TA instead of the CG and "Mutation" meaning that the C (respectively the G) is mutated into something else than a T (resp. an A). No *indels* were found in the spacer sequences. **B.** Dataset demonstrate absent or very low editing (max editing values: C40/C40: C_4 1.3%; C11/C11: C_6 3.3%; C0/C0: C_2 1%), on both top and bottom strand, for spacer length below 9 bp for any of the three linker pairs. The C40/C40 and C11/C11 linker combinations demonstrated editing to some extent on the 9 and 17 bp spacer (top strand max editing value: C40/C40: C_4 3.2%; C11/C11: C_6 %; C0/C0: C_4 7%; bottom strand max editing value: C40/C40: C_4 1.7%; C11/C11: C_6 0.3%; C0/C0: C_2 0.6%). On the spacer of 11bp, these two linker combinations showed similar editing levels almost exclusively on the top strand (max editing: C40/C40: C_4 25%; C11/C11: C_4 20%), while the highest editing rates were obtained on 13 bp and 15bp spacers (max editing: C40/C40: C_4 34%; C11/C11: C_4 44% on 13bp; C40/C40: C_4 20%; C11/C11: C_4 22% on 13bp). Shortening the spacer from 15 to 13bp favored editing closer to the right end of the spacer (moving from position 6 to position 4). For all figures: N=2, independent T-cells donors.



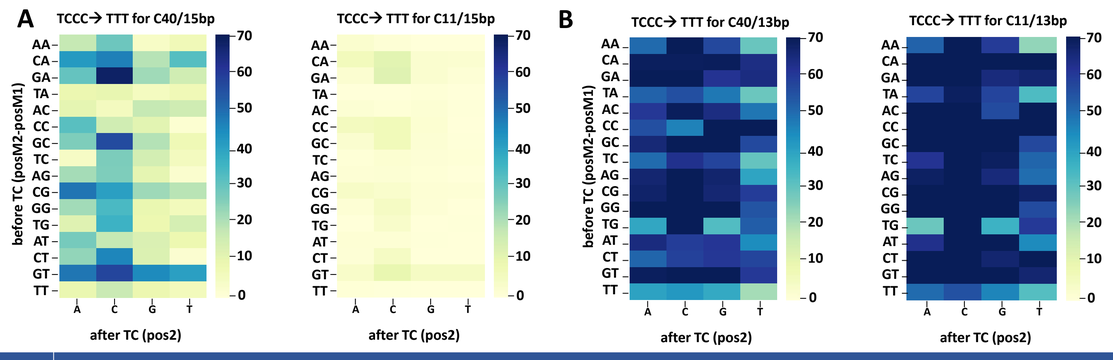
#4 Influence of the context around TC: 15bp and 13bp spacer length

A. Schematic representation of the ssODN pool collection containing the targeted 5’TC at the optimal positioning (cytosine at position 5), within optimal spacer lengths (13 and 15 bp, as show in Figure #3B) in which the nucleotides in posM2, posM3 and pos1, pos2 compared to the targeted 5’TC are variable for a total (of 256 different combinations). **B.** Schematic representation of nucleotide nomenclature. **C-D.** Heatmap of C-to-T conversion for C40/C40 and C11/C11 scaffold on 15bp spacer. Analysis shows similar nucleotides preferences (posM2: A = T << G < C; posM1: T = C < A << G; pos1: T < G < A << C; pos2: T < C < G < A), with C11/C11 editing being more stringent against a suboptimal context compared to C40/C40. **E-F.** Heatmap of C-to-T conversion for C40/C40 and C11/C11 scaffold on 13bp spacer. Data analysis still shows similar context preference to **C** and **D** (posM2: T < A < G < C; posM1: T < C < A < G; pos1: T < G < A < C; pos2: T < A = C < G), but context seems to have less importance for editing efficiency, which is higher overall (range : 88% to 100%, mean = 96%). For all panels: N=2, independent T-cells donors.



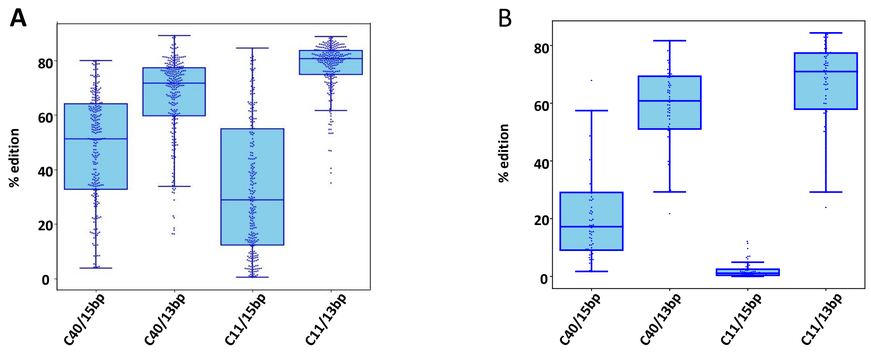
#5 Careful choice of TALE-BE design is minimizing editing in stretches of Cs

A. Heatmap of C-to-T conversion of the C in **pos1** for C40/C40 and C11/C11 scaffolds on 15bp spacer; comparison of the editing results showed a clear difference in mutation rate on the pos1 C, with the C40/C40 linker combination showing more permissive activity overall, seemingly less dependent on context requirements than C11/C11 scaffold. **B.** Heatmap of C-to-T conversion of the C in **pos1** for C40/C40 and C11/C11 scaffolds on 13bp spacer; comparison of the editing results showed a less clear difference in mutation rates on the pos1 C, with both C40/C40 and C11/C11 linker combinations showing more permissive activity, almost independent of context requirements, compared to the editing results obtained for the 15bp spacer.



#6 Editing is impacted by spacer length, TALE-BE architecture and target composition

Analysis of C-to-T conversions on the 256 **(A)** and 285 **(B)** possible targets for the 13 and 15bp spacer using the C11/C11 and C40/C40 TALE-BE. Median (middle line), interquartile (box) and 1.5 times the interquartile (whiskers) are represented, as well as the points corresponding to individual targets. **A.** Analysis showed that the overall editing on the 13 bp collection (Median: 72% for C40, 81% for C11) was found higher when compared to the 15 bp collection (median: 51% for C40, 29% for C11) which was expected from the slightly more favorable positioning of the TC (position C_4) within the former spacer length. **B.** Analysis of the editing results showed clear differences in the C-to-T conversion rates on the pos2 between the four conditions that the comparison



#6 Conclusions

We show that we have:

- Demonstrated the possibility to narrow editing on the top strand using a C11/C11 combination,
- Determined optimal spacer length (13/15 bp) for highly efficient TALE-BE for both C40/C40 and C11/C11 scaffolds
- Determined optimal common sequence context for high editing rates
- Determined that editing efficiency of the C11/C11 combination is highly dependent on Cytosine position requirements, resulting in more stringent activity in a context of 15 bp spacer size and decreasing the effects of bystander editing

Together these data demonstrate that a combination of architecture, sequence composition, and spacer length enable us to identify the best TALE-BE candidate for each specific locus