1210

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

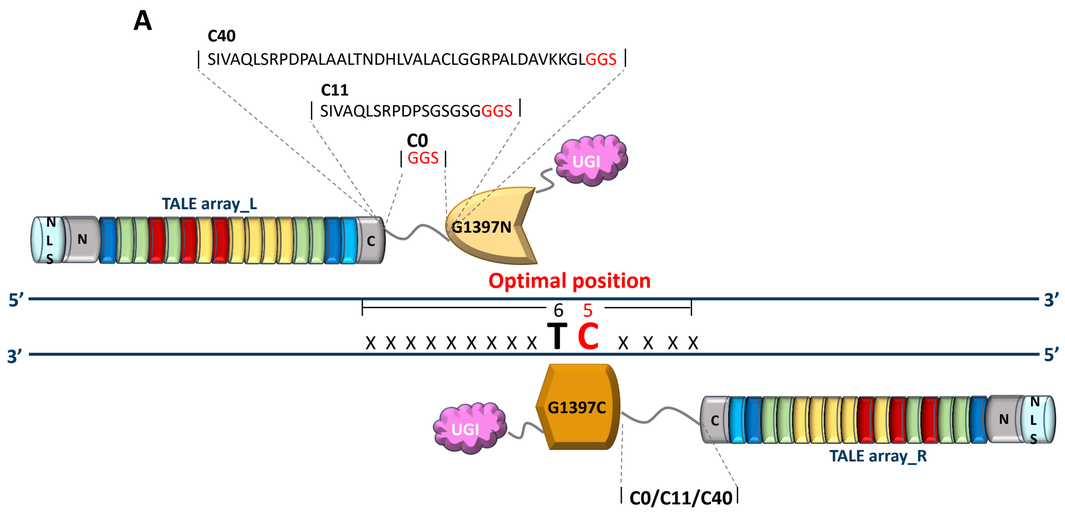
Abstract

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-mediated 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 define 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.

.

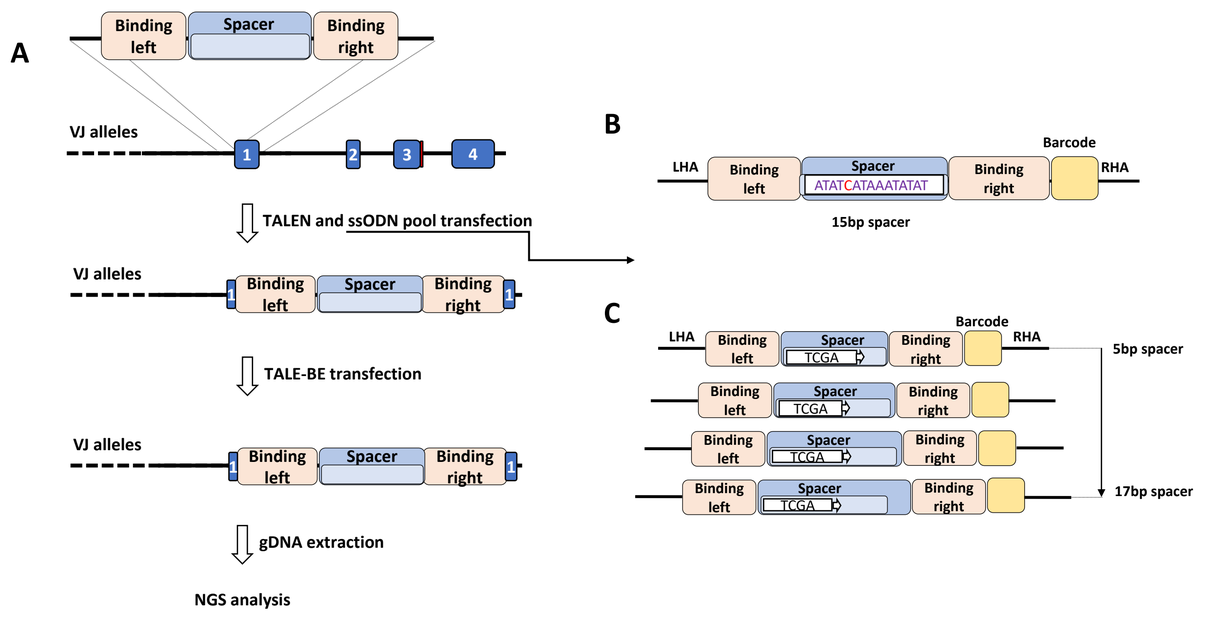
#1 Background

Base editing is a technology that leads to the introduction of point mutations (C>T transitions) in defined loci of a targeted DNA sequence. Base editors create mutations by deamination of the targeted bases (C or A), which are then converted into T or G during DNA repair process. This occurs without creating DNA double strand breaks, making it a promising therapeutic strategy for genetic diseases. The first development of TALE-BE relies on the use of a TALEN® scaffold; here, the domain linking the TALE DNA binding domain to the catalytic domain (linker) 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. It has been shown that, 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 efficiency within the editing window. To extend our understanding of key determining factors allowing efficient TALE-BE editing, 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).



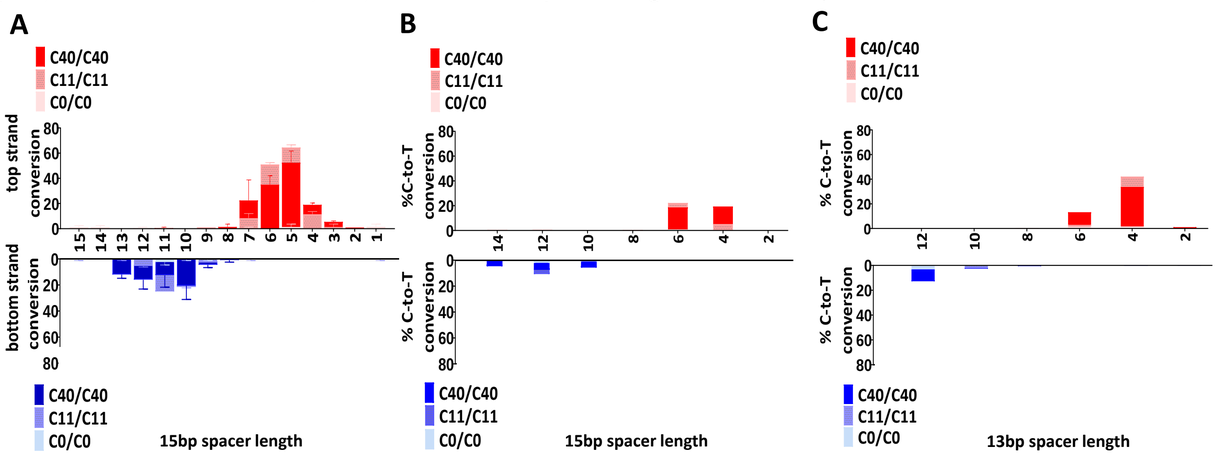
#2 Definition of the optimal TALE-BE editing window: Artificial spacer pool for high throughput testing

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 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. Screening strategy 1.** Schematic representation of the 15 ssODN pool collection **to determine the impact of the three linkers (C40, C11 and C0) on editing of targets containing a unique TC within a spacer of 15bp**, known to be optimal for the C40 TALE-BE scaffold. Additionally, to facilitate the sequence analysis, a unique barcode was added to each construct. **C. Screening strategy 2.** 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 the length of the linker and DNA spacer on editing efficiency**. A TCGA quadruplex target sequence was incorporated into the spacer at different positions from 5' to 3'. Spacer lengths were limited to odd numbers (eg 5bp, 7bp, 11bp, etc.). LHA and RHA stand for left and right homology arms, respectively.



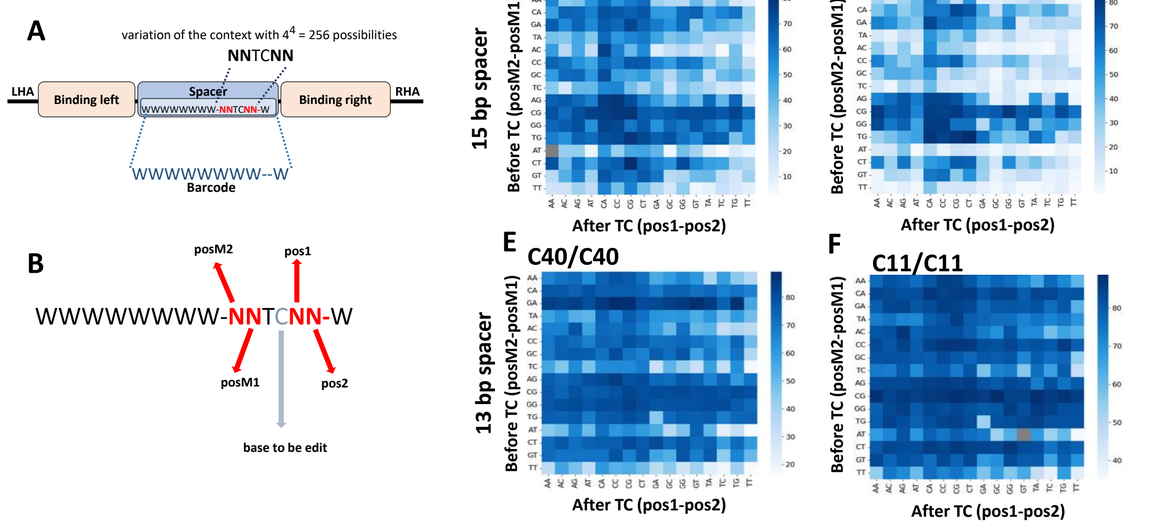
#3 Analysis of the linker nature and spacer length influence on TALEB C-to-T conversion

A. Dataset obtained from the **screening strategy 1** demonstrate that C0/C0 linker combination shows almost no activity on both top and bottom strand (max editing 3% C-to-T conversion, light pink/light blue bar); C11/C11 linker combination had stronger preference towards the C at position 5 and 6 within the 15bp spacer (50-64% top strand, C-to-T conversion, patterned pink/blue bar); the standard C40/40 arm displayed high editing frequencies on multiple positions (C at position 4,5,6 and 7; red/blue bar), lower compared to C11/C11 (C5= 53% vs 64%; C6= 35% vs 50%).
From the analysis obtained by testing different linkers on different spacer lengths, 15bp and 13bp showed to be the best size for high editing. **B-C.** Dataset on 15 and 13 bp spacer from **screening strategy 2**. results on 15 bp spacer confirm data reported in **A**, with almost no editing with C0/C0 linker combination, higher and narrower editing of C11/C11 scaffold on almost exclusively the C at position 6, whereas C40/C40 shows equal high editing on both C at position 4 and 6. Dataset on 13 bp spacer showed similar trend of activity for both C40/C40 and C11/C11 linker combinations, but overall higher editing on exclusively C at position 4 compared to 15 bp spacer (C11/C11 scaffold 43% vs 22%; C40/C40 34% vs 18%). For all figures: N=2, independent T-cells donors.



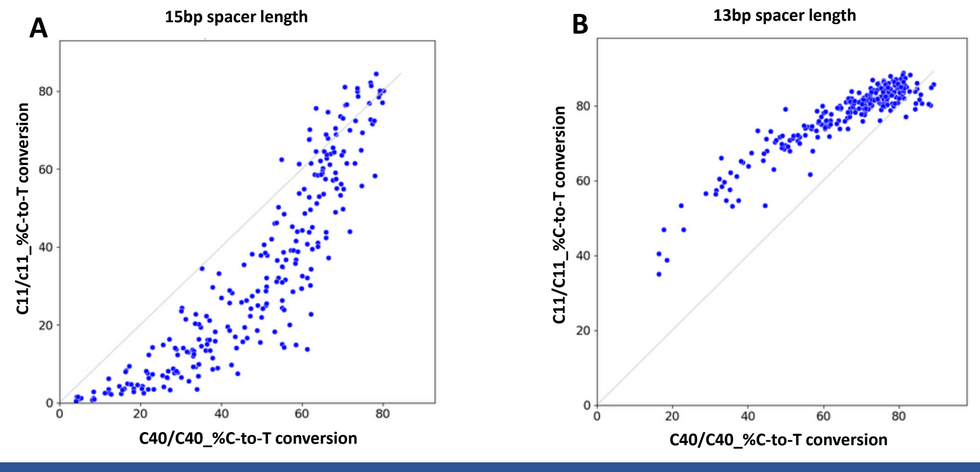
#4 Influence of the context around TC: 15bp and 13bp spacer length

A. Schematic representation of the ssODN pool collection containing the targeted 5'TC at the optimal position (cytosine at position 5), within optimal spacer lengths (13 and 15 bp, as show in Figure #3) 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 the context seems to have less importance for editing efficiency, which is overall higher than the one obtained with 15bp spacer (range : 88% to 100%, mean = 96%). For all panels: N=2, independent T-cells donors.



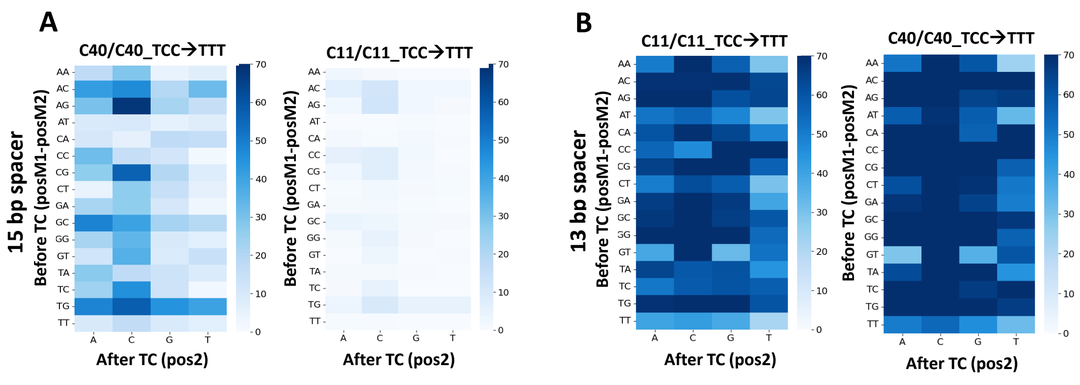
#5 Editing efficiency comparison between scaffolds on 15bp and 13bp spacer length

A. 15bp spacer: comparison of editing percentage between the two scaffolds C40/C40 and C11/C11 showed a non-linear correlation, suggesting that for contexts unfavorable to editing, activity with C11/C11 is lower than for the C40/C40 scaffold; for contexts that are the most favorable to editing, the activity is similar in C11/C11 and C40/C40, and for both scaffolds can reach upwards to 80% C-to-T conversion.
B. 13bp spacer: comparison of editing percentage between the two scaffolds C40/C40 and C11 /C11 showed a non-linear correlation, with C11/C11 scaffold being more tolerant in unfavorable contexts (reverse trend of what observed on 15bp spacer); however, both scaffolds reach high editing values (range : 88% to 100%, mean = 96%).



#6 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 from context requirements, compared to the editing results obtained for the 15bp spacer.



Conclusions

We show that we have:

- Demonstrated the possibility to narrow editing on the top strand using a C11/C11 scaffold,
- 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 scaffold 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 shades light on how TALE-BE architecture, sequence composition, and spacer length can impact editing and help educated design for specific locus