The integration preference of Sleeping Beauty at non-TA site is related to the transposon end sequences

Yiting Zhou
Sun Yat-Sen University

Guangwei Ma
Sun Yat-Sen University Cancer Center Department of Radiotherapy

Jiawen Yang
Sun Yat-Sen University

Yabin Guo (✉ guoyb9@sysu.edu.cn )
Sun Yat-Sen University  https://orcid.org/0000-0001-8316-8527

Short report

Keywords: Sleeping Beauty, transposon, integration, non-TA sites, consensus sequence

Posted Date: December 18th, 2019

DOI: https://doi.org/10.21203/rs.2.19101/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Version of Record: A version of this preprint was published at Frontiers in Genetics on March 10th, 2021. See the published version at https://doi.org/10.3389/fgene.2021.639125.
The integration preference of Sleeping Beauty at non-TA site is related to the transposon end sequences

Authors: Yiting Zhou, Guangwei Ma, Jiawen Yang and Yabin Guo*

Affiliation: Guangdong Provincial Key Laboratory of Malignant Tumor Epigenetics and Gene Regulation, Medical Research Center, Sun Yat-sen Memorial Hospital, Sun Yat-sen University, Guangzhou, 510120 China.

*Correspondence: guoyb9@sysu.edu.cn Tel: +86-13922465749

Email: Yiting Zhou, zhouyt28@mail2.sysu.edu.cn
Jiawen Yang, yangjw28@mail2.sysu.edu.cn
Guangwei Ma, maguangwei77@163.com
Abstract

Background: Sleeping Beauty (SB) transposon had been thought to strictly integrate into TA dinucleotides. Recently, we found that SB also integrates into non-TA sites at a lower frequency. Here we performed further study on the non-TA integration of SB.

Results: 1) SB can integrate into non-TA sites in HEK293T cells as well as in mouse cell lines. 2) Both the hyperactive transposase SB100X and the traditional SB11 catalyze integrations at non-TA sites. 3) The consensus sequence of the non-TA target sites only occur at the opposite side of the sequenced junction between transposon end and the genomic sequences, indicating that the integrations at non-TA sites are mainly aberrant integrations. 4) The consensus sequence of the non-TA target sites is corresponding to the transposon end sequence. When the transposon end sequence is mutated, the consensus sequences changed too.

Conclusion: The interaction between the SB transposon end and genomic DNA may be involved in the target site selection of the SB integrations at non-TA sites.

Keywords: Sleeping Beauty, transposon, integration, non-TA sites, consensus sequence.
Sleeping Beauty (SB) transposon is the most widely used transposon genetic tool [1-3]. It was thought that SB strictly integrates into TA dinucleotides [4-6], but we recently proved that SB could also integrate into non-TA sites through an aberrant pathway [7]. At the same time, our study also raised several questions: 1) while integrations at non-TA sites were found in mouse cell lines, are there integration at non-TA sites in human cell lines? 2) The non-TA integrations we found were mediated by the hyperactive transposase version, SB100X [8]. Does the traditional SB11 transposase catalyze non-TA integration too? 3) Why does this consensus sequence only occur at one side of the integration site? 4) We found that the consensus sequence flanking the integration site is the same as the sequence of the transposon ends, which was speculated the result of the interaction between the transposasase and the target site, but is it possible that this phenomenon is the result of the interaction between the transposon end and the target site sequence?

To answer these questions, we performed further study on the SB integration at non-TA sites in a human cell line. We constructed a series of plasmids containing puromycin resistance gene flanked by the inverted repeat sequences of SB (IR/DR) (Fig. 1A). The plasmids were transfected into HEK293T cells with plasmids expressing SB100X or SB11. After puromycin screen, the cells were collected and genomic DNA (gDNA) samples were isolated. Then, ligation-mediated PCR (LM-PCR) and Illumina sequencing were performed to detect the integration sites. After the sequences were aligned to the human genome, non-TA sites were identified, which is similar to the observation in mouse BaF3 cells (Table 1). We found non-TA integrations in the co-transfection of both SB100X and SB11 plasmids, indicating that SB11 can mediate integrations at non-TA sites as well as SB100X.

Usually only the junctions between the SB left end and the genomic sequences were sequenced in the SB screening assays, because the left side gives better results in LM-PCR. Here, we sequenced both left and right junction of SB integrations. Non-TA integrations were detected from both sides with similar proportions (Table 1).

In our last study, we found a consensus sequence at the non-TA target sites which is identical to the SB IR/DR end sequences. Here we performed the same analysis
with the integration data of this study. Fig. 1B showed the similar pattern to what was found in our last study. The strong CA is corresponding to the CA/TG of SB ends. When we looked at the consensus sequence at the non-TA sites identified by sequencing the right end of SB, the consensus sequence occurred at the left side of the logo (Fig. 1C). Interestingly, the consensus sequence is not fixed to the left or right side, but always occurs at the opposite side of the sequencing primers, which indicates that integrations at non-TA sites are mainly aberrant ones. The non-TA dinucleotides only occur at one side, whereas, those at the other side are still TA dinucleotides, thus were treated as canonical integrations when sequenced from the sides with TA dinucleotides.

To test whether the consensus sequence flanking the non-TA integration sites is related to the IR/DR sequences, we constructed plasmids with mutated IR/DR ends (Fig. 2A). It is previously reported that the two nucleotides at the very end of the IR/DR are critical for SB transposition; mutation at the IR/DR ends almost abolish the transposition [9]. Therefore, we kept the first nucleotide unchanged and mutated the second and the third nucleotides from AG/CT to GA/TC (Fig. 2A). The transposition efficiencies of SB with these mutated ends are similar to that of WT transposon in 293T cells. Non-TA integrations were identified as well as in the integrations with native transposon ends (Table S1) and it seems that the proportions of non-TA integrations of the transposons with mutated ends are higher than those with native ends.

The genomic sequences flanking integration sites were extracted and aligned. Surprisingly, the consensus sequences were all changed according to the changes of the transposon end sequences (Fig. S1). Since the number of total sites identified in this assay is small, to get a better view for the consensus sequence, the target sequences from both left and right sides were aligned together by the mutated ends (Fig. 2B). Obviously, the consensus sequence (--ATCG) perfectly reproduced the mutated transposon end. We also sequenced the left junction of the integrations of pYT22, which only has mutation at the right end. Fig. 2C showed that the consensus sequence still reproduced the canonical transposon end (--ACTG) as the previous
In our last study, following the suggestion of the reviews, we hypothesized that the consensus sequence is the result of the interaction between the transposase and the target DNA [7]. However, the current study seems indicate that the consensus sequence is due to the interaction between the transposon end DNA and the target DNA. Therefore, the integration preference of transposons is not only determined by the transposase, but it also can be influenced by the transposon end sequences. Our results might bring new ideas to the mechanism study on the target site determination of transposons.
**Brief Methods**

A puromycin resistance gene with promoter and polyA site were inserted between the IR/DRs of SB transposon, and this cassette was cloned into pUC19 backbone between HindIII and EcoRI restriction sites. pYT11 is the plasmid with classical SB ends. pYT22 and pYT23 have mutations at the IR/DR ends as described in the main text and Fig. 2. The plasmids were transfected into 293T cells together with the transposase expression plasmids, SB100X or SB11. After puromycin selection, cells were collected and genomic DNA samples were isolated. Then, ligation-mediated PCR assays were performed [3] and the amplicons were submitted for Illumina sequencing. The sequencing data were analyzed as previously described [7]. The sequence logos were generated using DNAlogo (https://www.biorxiv.org/content/10.1101/096933v2).

**Abbreviations:** SB, Sleeping Beauty; IRDR, inverted repeat direct repeat.
Declarations

Ethics approval and consent to participate
Not applicable. No patients or animals were involved in this study.

Consent for publication
Not applicable

Availability of data and materials
The Illumina sequencing data will be uploaded to the SRA database upon acceptance.

Competing interest: The authors declare that they have no competing interests.

Funding: This work was supported by National Natural Science Foundation of China (81872295 to Y. G.); Guangdong Natural Science Foundation (2018A030313819 to Y. G.); Guangdong Science and Technology Department (2017B030314026).

Authors' contributions
YG conceived the idea for the project; YZ and YG designed the experiments; YZ and GM performed the experiments; YZ, JY and YG analyzed the data; YZ and YG wrote the manuscript.
References

1. Dupuy AJ, Akagi K, Largaespada DA, Copeland NG, Jenkins NA: Mammalian mutagenesis using a highly mobile somatic Sleeping Beauty transposon system. *Nature* 2005, 436(7048):221-226.

2. Starr TK, Allaei R, Silverstein KA, Staggs RA, Sarver AL, Bergemann TL, Gupta M, O'Sullivan MG, Matise I, Dupuy AJ et al: A transposon-based genetic screen in mice identifies genes altered in colorectal cancer. *Science* 2009, 323(5922):1747-1750.

3. Guo Y, Updegraff BL, Park S, Durakoglugil D, Cruz VH, Maddux S, Hwang TH, O'Donnell KA: Comprehensive Ex Vivo Transposon Mutagenesis Identifies Genes That Promote Growth Factor Independence and Leukemogenesis. *Cancer research* 2016, 76(4):773-786.

4. Ivics Z, Hackett PB, Plasterk RH, Izsvak Z: Molecular reconstruction of Sleeping Beauty, a Tc1-like transposon from fish, and its transposition in human cells. *Cell* 1997, 91(4):501-510.

5. Plasterk RHA, Izsvák Z, Ivics Z: Resident aliens: the Tc1/mariner superfamiliy of transposable elements. *Trends in Genetics* 1999, 15(8):326-332.

6. Yant SR, Wu X, Huang Y, Garrison B, Burgess SM, Kay MA: High-resolution genome-wide mapping of transposon integration in mammals. *Mol Cell Biol* 2005, 25(6):2085-2094.

7. Guo Y, Zhang Y, Hu K: Sleeping Beauty transposon integrates into non-TA dinucleotides. *Mobile DNA* 2018, 9:8.

8. Mátés L, Chuah MKL, Belay E, Jerchow B, Manoj N, Acosta-Sanchez A, Grzela DP, Schmitt A, Becker K, Matrai J et al: Molecular evolution of a novel hyperactive Sleeping Beauty transposase enables robust stable gene transfer in vertebrates. *Nature genetics* 2009, 41(6):753-761.

9. Zayed H, Izsvák Z, Walisko O, Ivics Z: Development of Hyperactive Sleeping Beauty Transposon Vectors by Mutational Analysis. *Molecular Therapy* 2004, 9(2):292-304.
Table 1. Number of integrations at different dinucleotides.

| Dinucleotide | SB100X | SB11 |
|--------------|--------|------|
|              | Left   | Right | Left | Right |
| TA           | 29748  | 27740 | 3731 | 460   |
| CA           | 10     | 3     | 8    | 0     |
| TG           | 7      | 5     | 5    | 2     |
| TT           | 4      | 8     | 1    | 2     |
| AA           | 5      | 9     | 0    | 0     |
| GA           | 4      | 9     | 0    | 0     |
| TC           | 8      | 1     | 3    | 0     |
| AG           | 2      | 2     | 1    | 0     |
| CT           | 7      | 3     | 3    | 0     |
| GG           | 4      | 4     | 0    | 1     |
| CC           | 4      | 5     | 3    | 2     |
| AT           | 5      | 6     | 0    | 0     |
| GT           | 2      | 2     | 1    | 1     |
| AC           | 1      | 0     | 0    | 0     |
| GC           | 7      | 1     | 3    | 1     |
| CG           | 0      | 0     | 0    | 0     |
| Total        | 29818  | 27798 | 3759 | 469   |
| Proportion of non-TA | 0.235% | 0.209% | 0.745% | 1.919% |
Figure Legends

**Fig. 1.** SB transposon integrates at non-TA sites in HEK293T cells. A, the structure of plasmid pYT11, puromycin resistance gene flanked by the IR/DRs of SB. B, The consensus sequences of the SB non-TA integrations sequenced from the left end. C, The consensus sequences of the SB non-TA integrations sequenced from the right end.

**Fig. 2.** The target site consensus sequences of the non-TA integrations of the transposons with mutated IR/DR ends. A, the design of mutations at the IR/DR ends. B, the consensus sequences of non-TA target sites of pYT23. The sequences from both left and right were combined. The position corresponding to the mutations were labeled with pink asterisks. C, the consensus sequences of non-TA target sites of pYT22 sequenced from the left side (native end).
Fig. 1

A

IR/DR-L  PuroR  IR/DR-R
HindIII  AmpR  ori  EcoRI

pYT11

B

C

Bits

-10 -5  0  5  10

0  0.5  1.0

L  R

Bits

-10 -5  0  5  10

0  0.5  1.0

L  R
Fig. 2

A

pYT11: CAGTTGAA----------------------------------------------------TTCAACTG
pYT22: CAGTTGAA----------------------------------------------------TTCAACTCG
pYT23: CGATTGAA----------------------------------------------------TTCAATCG

IR/DR-L  PuroR  IR/DR-R

B

C

Bits

0 0.5 1 1.5 2

-10 -5 0 5 10

0 1 2

0 0.5 1 1.5 2

-10 -5 0 5 10
Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryMaterials.pdf