piggyBac-system-mediated genomic integration of large variant libraries for deep mutational scanning in mammalian cells

Yanjie Zhao [1#], Yifan Li [1#], Yi Wang [1], Weijun Su [2*], Shuai Li [1*]

1. Department of Breast Cancer Pathology and Research Laboratory, Tianjin Medical University Cancer Institute and Hospital, National Clinical Research Center for Cancer; Key Laboratory of Cancer Prevention and Therapy, Tianjin; Key Laboratory of Breast Cancer Prevention and Therapy, Tianjin Medical University, Ministry of Education; Tianjin's Clinical Research Center for Cancer, Tianjin 300060, China.

2. School of Medicine, Nankai University, Tianjin 300071, China.

Contribute equally.

* To whom correspondence should be addressed. Shuai Li: Tel: +86 22 23340123; Email: shuaili@tmu.edu.cn & shuai.li2001@gmail.com. Weijun Su: Tel: +86 22 23502554; Email: suweijun@nankai.edu.cn.

Abstract

There is a need to introduce large variant libraries into mammalian cells to facilitate deep mutational scanning (DMS). In this study, we develop approaches to integrate *in-vitro* ligation dsDNA variant library into cultured mammalian cell genome via the piggyBac transposon system. T4 DNA ligase was employed to generate dsDNA ligation products containing variant library sequences flanked by piggyBac inverted terminal repeat sequences (ITRs) at both ends. The ligation products were then cotransfected with a piggyBac transposase expression vector into mammalian cells for library integration. The clone formation assay result showed that about 29,000 resistance clones were generated from one single well of 6-well-plate transfected HEK293T cells. As a proof of the principle of using the strategy for DMS, a GFP chromophore random variant library was integrated into the HEK293T cell genome. By FACS selection, we collected the GFP-positive cells, and next-generation sequencing (NGS) showed that the GFP-chromophore amino acid sequences were enriched by the selection.

Keywords

Deep mutational scanning, T4 DNA ligase, piggyBac system, double-stranded DNA

Introduction

Deep mutational scanning (DMS), also known as saturated mutagenesis screen or massively parallel genetic assay, is a powerful high-throughput screening method in protein engineering that uncovers genotype-phenotype relationships.[1-2] DMS has a wide range of applications, including antibody engineering,[3-4] vaccine design,[5-7] drug-resistant mutation identification,[8] and so on. In a typical DMS experiment, thousands of mutations are introduced into the specific genomic regions and then subjected to selection. After screening, next-generation sequencing (NGS) is employed to score each mutation, thereby connecting sequences to functions.[1-2]

Mammalian cells are the ideal platform for phenotypic screening for many proteins, such as transmembrane proteins,[9-10] proteins associated with signaling pathways,[11-12] and proteins related to cellular phenotypes.[7-8, 13] However, mammalian DMS experiments are substantially challenged by difficulties in constructing large genome-integrated variant libraries. Currently, there are two types of approaches to introduce integrated libraries into mammalian cells, both of which have their own limitations. The first type of approach is plasmid-based. Plasmids are transformed into competent cells to achieve a maximum library size of 10^9-10^{10}.[14] Then, after plasmid amplification, extraction, and cell transfection, the plasmid library is integrated into the mammalian genome via viral-,[15] transposon-,[16-17] or recombinase-[11] mediated genome integration. The library content is continuously lost in the above steps, resulting in a significant reduction in the library size (~10^4). The second type of approach is CRISPR/Cas9-mediated homology-directed repair (HDR).[3, 8, 13] CRISPR/Cas9 system is employed to create double-strand breaks (DSBs) at specific genome locations to trigger DNA damage repair mechanisms. Meanwhile, ssODNs containing bilateral

homology arms and variant coding sequences are transfected into cells as HDR templates, thereby incorporating the library into DSB sites. However, there are several technical hurdles in HDR, e.g., the need to pre-integrate the landing pad of the library, the need to optimize the CRISPR targeting efficiency, the difficulty to distinguish integrated and non-integrated cells, and the limited mutation range (<100 nt). Therefore, there is a need of developing an easy-to-perform, genomic-integrated, large variant library construction strategy to facilitate mammalian cell-based DMS.

In our previous study, we developed a mammalian cell-based nanobody library construction strategy named '*in-vitro* ligation'.[18] The random complementary determining region (CDR) sequences are first incorporated into double-stranded DNAs (dsDNAs) through PCR amplification. Then the up- and down-stream dsDNAs containing CDRs are ligated to each other by T4 DNA ligase to form an intact nanobody expression cassette containing the promoter, nanobody gene body, and polyA signal sequence. The ligation products are transfected into HEK293T cells for transient nanobody library expression. The NGS results show that our strategy could generate over 10^6 unique nanobody sequences. Based on this, here we develop novel large variant library genomic integration strategies by combining *in-vitro* ligation strategy with piggyBac transposon system. To validate the feasibility of the strategies in mammalian DMS, we enriched GFP-positive cells and restored GFP-chromophore sequences from GFP-chromophore random libraries.

Materials and Methods

Cell culture

HEK293T cell line was purchased from ATCC (American Type Culture Collection) and maintained in DMEM (Dulbecco's Modified Eagle Medium, Corning, NY, USA) supplemented with 10% FBS (Fetal bovine serum, Corning) and 1% Penicillin-Streptomycin solution (10,000 U/ml, Thermo Fisher Scientific, Waltham, MA, USA) at 37 □ with 5% CO_2.

PCR amplification, restriction digestion, and ligation

PCR template plasmids were constructed by VectorBuilder (VectorBuilder, Guangzhou, China, for plasmid sequences see supplementary materials). KOD Plus Neo DNA polymerase (Toyobo, Osaka, Japan) was used to perform PCR amplification. PCR products were purified (Universal DNA purification kit, TIANGEN, Beijing, China) after agarose electrophoresis, and the concentrations were determined by NanoDrop (Thermo Fisher Scientific). PCR primers were listed in Supplementary Table 1. The up- and down-stream PCR products were digested by BssSI (New England Biolabs, Ipswich, MA, USA). After agarose electrophoresis and gel purification, the up- and down-stream digested dsDNA were ligated to each other by T4 DNA ligase (New England Biolabs). The *in-vitro* ligation products were electrophoresed and purified for downstream transfection. To avoid UV-induced DNA damage, Blue Light Gel Imager (Sangon Biotech, Shanghai, China) was used for dsDNA visualization.

For *in-vitro*-ligation-PCR library dsDNA preparation, 1 ng or 10 ng *in-vitro* ligation dsDNAs were added into per 50 μl standard PCR reaction system, respectively. The PCR cycle numbers for each type of library were determined to ensure the amplification was in the exponential phase.

Cell transfection, resistance selection, and clone-formation assay

HEK293T cells were plated in 6-well plates the day before transfection (70~80 % confluency). *In vitro* ligation dsDNA (1 μg), PCR products (1 μg), or plasmids (1 μg) were cotransfected with a hyperactive piggyBac transposase expression vector (1 μg, VectorBuilder) into HEK293T cells using Lipofectamine 2000 (Thermo Fisher Scientific). Forty-eight hours after transfection, puromycin (10 μg/ml, Selleck Chemicals, Houston, TX, USA) was added into the cell culture medium to select integrated cells. For clone-formation assay, transfected cells were diluted 1,000 times and cultured in 6-well plates for ten days. Crystal violet staining was used to visualize the cell clones. For *in-vitro*-ligation-PCR library transfection, the library dsDNA (6 μg) and the hyperactive piggyBac transposase expression vector (6 μg) were cotransfected into a 10 cm dish plated HEK293T cells.

Fluorescence-activated cell sorting (FACS)

HEK293T cells were digested with trypsin. Cells were resuspended in PBS, and GFP or BFP positive cells were isolated using a BD FACSAria™ II Cell Sorter (BD Biosciences, Bedford, MA, USA). FlowJo software (FLOWJO, Ashland, OR, USA) was used to analyze the raw data.

Next-generation sequencing and data processing

Genomic DNA was extracted with a TaKaRa MiniBEST Universal Genomic DNA Extraction Kit Ver. 5.0 (TaKaRa, Kusatsu, Japan). KOD Plus Neo DNA polymerase was used to amplify the sequences containing random sequences. PCR products were subjected to sequencing library construction using the TruSeq® DNA PCR-Free Sample Preparation Kit (Illumina, San Diego, CA, USA). The libraries were loaded on a NovaSeq 6000 system (Illumina) to generate PE-150 reads. Clean reads were merged using the FLASH (Fast Length Adjustment of SHort reads) software

(v1.2.11). Inhouse Perl scripts were developed to analyze the amino acid sequences of libraries.

Quantification analysis

The chromophore amino acid sequences in each library were first normalized as RPM (read per million reads) values. Then the RPM values (> 1) from the initial and selected libraries are used to calculate the enrichment ratios for each amino acid sequence.

Results

Genomic integration of *in-vitro* ligation dsDNA via piggyBac system

In our previous study, we developed an *in-vitro* ligation strategy to introduce nanobody libraries into mammalian cells.[18] Briefly, both up- and down-stream dsDNAs were generated by PCR amplification and digested by restriction enzyme BssSI to create sticky ends. T4 DNA ligase is used to link up- and down-stream dsDNAs together to form full-length dsDNAs with the intact gene expression cassette. The obtained full-length dsDNAs were transfected into HEK293T cells to express the nanobody library. In the present study, we employ the piggyBac system to directly integrate the *in-vitro* ligation products into the mammalian genome for DMS.

The piggyBac system can efficiently insert foreign DNA into the mammalian genome via a non-viral route.[19] Here, we add piggyBac recognition sequences, the inverted terminal repeat sequences (ITRs), to the 5' or 3' end of the up- and down-stream dsDNA products, respectively (Figure 1A). T4 DNA ligase was employed to ligate BssSI-digested up- and down-stream dsDNAs to each other to form a full puromycin resistance gene expression cassette (Figure 1A). Then the *in-vitro* ligation products and a hyperactive piggyBac transposase expression vector [20] were

cotransfected into HEK293T cells. Forty-eight hours after transfection, the cells were diluted 1,000 times and cultured with puromycin. Ten days later, the cell clones were stained with crystal violet for counting (Figure 1B). As shown in Figure 1C, *in-vitro* ligation products could generate about 29,000 resistant clones in one single well of 6-well-plate transfected HEK293T cells. While the PCR product and plasmid, both of which contain puromycin resistance gene expression cassette and ITR sequences, led to the formation of 1.98×10^5 and 2.03×10^5 resistance cell clones, respectively (Figure 1C).

Screening of GFP chromophore randomized library

We next employed the piggyBac system to integrate the *in-vitro* ligation dsDNA variant library into the mammalian genome for downstream functional DMS. We randomized the codons encoding GFP chromophore via NNK degeneracy and expected to restore the GFP coding sequence by FACS. The full-length dsDNAs containing two NNK codons (corresponding to GFP T65-Y66) or three NNK codons (corresponding to GFP T65-Y66-G67) were generated by *in-vitro* ligation, respectively (Figure 2A). Then the dsDNA ligation products and piggyBac transposase expression vector were cotransfected into HEK293T cells. For each library, the transfection was performed in one single well of a 6-well-plate. Forty-eight hours after transfection, puromycin was added. After a 10-day puromycin selection, the resistant cells were subjected to FACS isolation. GFP-positive cells were collected and cultured for the subsequent round of FACS isolation. For the 2-NNK group, the initial GFP-positive rate was 8.47%, which was raised to 80.2% after one round of sorting, and 91.4% after two rounds (Figure 2B). For the 3-NNK group, the initial GFP positive rate was 0.319% and was raised to 94.3% after two rounds of sorting (Figure 2C). We

also isolated BFP-positive cells from the 2-NNK library. The initial BFP-positive rate was 0.639% (Figure 3). After the first round of FACS, the BFP rate was raised to 67.4%, and it was further raised to 90.8% after two rounds of selection.

Assessment of the chromophore sequences by NGS

The chromophore sequences of sorted GFP-positive cells were then analyzed by NGS. For the 2-NNK group, eight out of the top ten enriched amino acid sequences show a conserved "XY" pattern (Table 1). For the 3-NNK group, five out of the top ten amino acid sequences show a conserved "XYG" pattern (Table 2). The chromophore sequences of sorted BFP-positive cells were also analyzed. As shown in Supplementary Table 2, the BFP chromophore sequence SH ranks first.

The development of *in-vitro*-ligation-PCR genomic integration library strategy

The above *in-vitro* ligation dsDNA is difficult for large-scale production, which blocks the preparation of high-content genomic-integrated libraries *via* mass transfection. To facilitate the large-scale production of dsDNA, we employed PCR to amplify the dsDNA by taking *in-vitro* ligation dsDNA as PCR templates (Figure 4A). 1 ng or 10 ng *in-vitro* ligation dsDNAs were added into per 50 μl standard PCR reaction system, respectively, to generate two types of PCR products, *in-vitro*-ligation-PCR (1 ng) and *in-vitro*-ligation-PCR (10 ng). The PCR products were then integrated into HEK293T cell (cultured in a 10 cm dish) genome by the piggyBac system to generate 3NNK GFP chromophore randomized libraries (Figure 4A). The *in-vitro*-ligation-PCR (1 ng) and *in-vitro*-ligation-PCR (10 ng) library cells were subjected to FACS selections to enrich

GFP positive cells. Both libraries reach about 85% GFP-positive rate after two rounds of selections (Figure 4BC). The chromophore sequences of sorted GFP-positive cells were subjected to NGS. As shown in Table 3 and Table 4, eight XYG amino acid sequences pattern was enriched for both libraries.

Discussion

To facilitate DMS in cultured mammalian cells, we developed novel large variant library genomic integration approaches by combining the *in-vitro* ligation library construction strategy with the piggyBac system. The methods are easy to implement, only requiring PCR, restriction digestion, ligation, and cell transfection. The feasibility of the methods in mammalian DMS was confirmed by the isolation of GFP-positive cells from GFP-chromophore randomized libraries. NGS results revealed the overrepresentation of GFP-chromophore amino acid sequences in isolated GFP-positive cells.

Here we choose BssSI (also known as BauI) restriction enzyme to digest the dsDNAs. The non-palindromic characteristics of BssSI bring both advantages and disadvantages. Non-palindromic sticky ends block mistaken up-stream to up-stream or down-stream to down-stream dsDNA ligations, leading to a much higher yield of accurate ligation products compared to palindromic cutting sites. However, the number of commercially available non-palindromic endonucleases is far less than that of palindromic endonucleases. Thus, in some sequence contexts, there may be no appropriate non-palindromic restriction sites for *in-vitro* ligation.

The piggyBac system is highly effective for integrating foreign DNA, usually plasmids, into the mammalian host genome.[19] piggyBac transposase recognizes ITR sequences on both ends of the cargo sequences and inserts them into the chromosomal TTAA sites. In this study, for the first time, we showed that *in-vitro* ligation dsDNAs and PCR products containing ITR sequences on both ends could be integrated into the mammalian genome by the piggyBac system. As revealed by the clone formation assay, the integration efficiency of PCR products (11.0%) is comparable to that of plasmids (11.3%) (Figure 1BC). However, the integration efficiency of *in-vitro* ligation dsDNAs is only 1.61%, lower than that of PCR products. The underlying reasons require further investigation.

In principle, other transposon systems, such as Sleeping Beauty [21] and Tol2,[22] can also integrate *in-vitro* ligation dsDNAs or *in-vitro*-ligation-PCR products containing the corresponding transposase recognition sequences on both ends into the mammalian genome. Moreover, instead of ssODNs or plasmids, the *in-vitro* ligation dsDNAs or *in-vitro*-ligation-PCR products could serve as homologous repair templates for CRISPR/Cas9-mediated HDR. And by pre-integrating a single landing pad per cell, the *in-vitro* ligation dsDNA or *in-vitro*-ligation-PCR library could be inserted into the genome in a single-variant-per-cell manner.[3, 11, 13]

As initial applications of these strategies, we integrated the GFP expression cassettes containing randomized chromophore coding sequences into HEK29T cells and isolated the GFP-positive cells by FACS (Figure 2A, 4A). The theoretical frequencies of the GFP chromophore amino acid

sequence for 2-NNK and 3-NNK libraries are 0.195% and 0.0122%, respectively. After two rounds of FACS selection, both the *in-vitro* ligation 2-NNK group and the 3-NNK group enriched the cell population with over 90% of GFP-positive cells (Figure 2B). While for *in-vitro*-ligation-PCR strategies, the GFP-positive rates reached about 85% (Figure 4B). NGS results rev ealed that the Y66-G67 motif was conserved among FACS-selected GFP-positive cells (Table 1-4). This is consistent with the facts that G67 is important for GFP-chromophore formation and Y66 helps to maintain GFP specific absorption and emission spectral. [23-24] Moreover, all eight enriched amino-acids-65 (A, C, G, L, M, S, T, and V) were reported to have GFP excitation spectra (Table 1, 3, 4). [23, 25] Meanwhile, several fluorescence intensities weakened (lower than wild type SYG) amino-acids-65, such as R, N, D, F, and W, were not enriched. [25] All these results demonstrated the feasibility of our strategies.

In conclusion, we developed novel piggyBac-based large variant library integration approaches, which can be applied to perform DMS in cultured mammalian cells.

Acknowledgements

This work was supported by the National Natural Science Foundation of China [31870860 to S.L., 31971388 to W.S.].

Competing Interests

The authors have declared that no competing interest exists.

Figure Legends

Figure 1. Schematics and abundance analysis of piggyBac-mediated *in-vitro* ligation dsDNAs genomic integration. (A) Using *in-vitro* ligation strategy to generate dsDNA containing a Puro-resistance gene expression cassette, which is integrated into the genome by the piggyBac system to make cells resistant to puromycin. (B) Forty-eight hours after transfection, HEK293T cells were diluted 1,000-fold and cultured with puromycin for ten days. Cell clone formation was shown by crystal violet staining. PCR Product: PCR product containing Puro-resistance cassette; Plasmid: plasmid containing Puro-resistance cassette; no-transfection: HEK293T cells that have not been transfected. (C) Statistics of cell clone numbers. The total number of clones formed in each group shown in the bar graph is calculated by the clone number in the corresponding well in (B) multiplied by 1,000.

Figure 2. piggyBac system-mediated genomic integration of *in-vitro* ligation GFP chromophore randomized library and the enrichment of GFP positive cells. (A) Schematics of piggyBac-mediated genomic integration of *in-vitro* ligation GFP chromophore random library (3-NNK). Puromycin resistance gene links to GFP coding region via a P2A element. FACS isolation of GFP positive cells from *in-vitro* ligation 2-NNK (B) and *in-vitro* ligation 3-NNK (C) library.

Figure 3. FACS isolation of BFP positive cells from the 2-NNK library.

Figure 4. piggyBac system-mediated genomic integration of *in-vitro*-ligation-PCR GFP chromophore randomized library and the enrichment of GFP positive cells. (A) Schematics of piggyBac-mediated genomic integration of *in-vitro*-ligation-PCR GFP chromophore random library (3-NNK). Puromycin resistance gene links to GFP coding region via a P2A element. PCR step is marked by the red dotted line. FACS isolation of GFP positive cells from *in-vitro*-ligation-PCR (1 ng) (B) and *in-vitro*-ligation-PCR (10 ng) (C) library.

References

1. Araya, C. L.; Fowler, D. M., Deep mutational scanning: assessing protein function on a massive scale. *Trends Biotechnol* **2011,** *29* (9), 435-42.

2. Fowler, D. M.; Fields, S., Deep mutational scanning: a new style of protein science. *Nat Methods* **2014,** *11* (8), 801-7.

3. Mason, D. M.; Weber, C. R.; Parola, C.; Meng, S. M.; Greiff, V.; Kelton, W. J.; Reddy, S. T., High-throughput antibody engineering in mammalian cells by CRISPR/Cas9-mediated homology-directed mutagenesis. *Nucleic Acids Res* **2018,** *46* (14), 7436-7449.

4. Hanning, K. R.; Minot, M.; Warrender, A. K.; Kelton, W.; Reddy, S. T., Deep mutational scanning for therapeutic antibody engineering. *Trends Pharmacol Sci* **2021**.

5. Starr, T. N.; Greaney, A. J.; Hilton, S. K.; Ellis, D.; Crawford, K. H. D.; Dingens, A. S.; Navarro, M. J.; Bowen, J. E.; Tortorici, M. A.; Walls, A. C.; King, N. P.; Veesler, D.; Bloom, J. D., Deep Mutational Scanning of SARS-CoV-2 Receptor Binding Domain Reveals Constraints on Folding and ACE2 Binding. *Cell* **2020,** *182* (5), 1295-1310 e20.

6. Burton, T. D.; Eyre, N. S., Applications of Deep Mutational Scanning in Virology. *Viruses* **2021,** *13* (6).

7. Narayanan, K. K.; Procko, E., Deep Mutational Scanning of Viral Glycoproteins and Their Host Receptors. *Front Mol Biosci* **2021,** *8*, 636660.

8. Ma, L.; Boucher, J. I.; Paulsen, J.; Matuszewski, S.; Eide, C. A.; Ou, J.; Eickelberg, G.; Press, R. D.; Zhu, L. J.; Druker, B. J.; Branford, S.; Wolfe, S. A.; Jensen, J. D.; Schiffer, C. A.; Green, M. R.; Bolon, D. N., CRISPR-Cas9-mediated saturated mutagenesis screen predicts clinical drug resistance with improved accuracy. *Proc Natl Acad Sci U S A* **2017,** *114* (44), 11751-11756.

9. Jones, E. M.; Lubock, N. B.; Venkatakrishnan, A. J.; Wang, J.; Tseng, A. M.; Paggi, J. M.; Latorraca, N. R.; Cancilla, D.; Satyadi, M.; Davis, J. E.; Babu, M. M.; Dror, R. O.; Kosuri, S., Structural and functional characterization of G protein-coupled receptors with deep mutational scanning. *Elife* **2020,** *9*.

10. Heredia, J. D.; Park, J.; Brubaker, R. J.; Szymanski, S. K.; Gill, K. S.; Procko, E., Mapping Interaction Sites on Human Chemokine Receptors by Deep Mutational Scanning. *J Immunol* **2018,** *200* (11), 3825-3839.

11. Matreyek, K. A.; Stephany, J. J.; Fowler, D. M., A platform for functional assessment of large variant libraries in mammalian cells. *Nucleic Acids Res* **2017,** *45* (11), e102.

12. Findlay, G. M.; Boyle, E. A.; Hause, R. J.; Klein, J. C.; Shendure, J., Saturation editing of genomic regions by multiplex homology-directed repair. *Nature* **2014,** *513* (7516), 120-3.

13. Erdogan, M.; Fabritius, A.; Basquin, J.; Griesbeck, O., Targeted In Situ Protein Diversification and Intra-organelle Validation in Mammalian Cells. *Cell Chem Biol* **2020,** *27* (5), 610-621 e5.

14. Galan, A.; Comor, L.; Horvatic, A.; Kules, J.; Guillemin, N.; Mrljak, V.; Bhide, M.,

Library-based display technologies: where do we stand? *Mol Biosyst* **2016,** *12* (8), 2342-58.

15. Beerli, R. R.; Bauer, M.; Buser, R. B.; Gwerder, M.; Muntwiler, S.; Maurer, P.; Saudan, P.; Bachmann, M. F., Isolation of human monoclonal antibodies by mammalian cell display. *Proc Natl Acad Sci U S A* **2008,** *105* (38), 14336-41.

16. Waldmeier, L.; Hellmann, I.; Gutknecht, C. K.; Wolter, F. I.; Cook, S. C.; Reddy, S. T.; Grawunder, U.; Beerli, R. R., Transpo-mAb display: Transposition-mediated B cell display and functional screening of full-length IgG antibody libraries. *MAbs* **2016,** *8* (4), 726-40.

17. Hellmann, I.; Waldmeier, L.; Bannwarth-Escher, M. C.; Maslova, K.; Wolter, F. I.; Grawunder, U.; Beerli, R. R., Novel Antibody Drug Conjugates Targeting Tumor-Associated Receptor Tyrosine Kinase ROR2 by Functional Screening of Fully Human Antibody Libraries Using Transpo-mAb Display on Progenitor B Cells. *Front Immunol* **2018,** *9*, 2490.

18. Zhao, Y.; Wang, Y.; Su, W.; Li, S., Construction of Synthetic Nanobody Library in Mammalian Cells by dsDNA-Based Strategies*. *Chembiochem* **2021,** *22* (20), 2957-2965.

19. Woodard, L. E.; Wilson, M. H., piggyBac-ing models and new therapeutic strategies. *Trends Biotechnol* **2015,** *33* (9), 525-33.

20. Yusa, K.; Zhou, L.; Li, M. A.; Bradley, A.; Craig, N. L., A hyperactive piggyBac transposase for mammalian applications. *Proc Natl Acad Sci U S A* **2011,** *108* (4), 1531-6.

21. Ivics, Z.; Hackett, P. B.; Plasterk, R. H.; Izsvak, Z., Molecular reconstruction of Sleeping Beauty, a Tc1-like transposon from fish, and its transposition in human cells. *Cell* **1997,** *91* (4), 501-10.

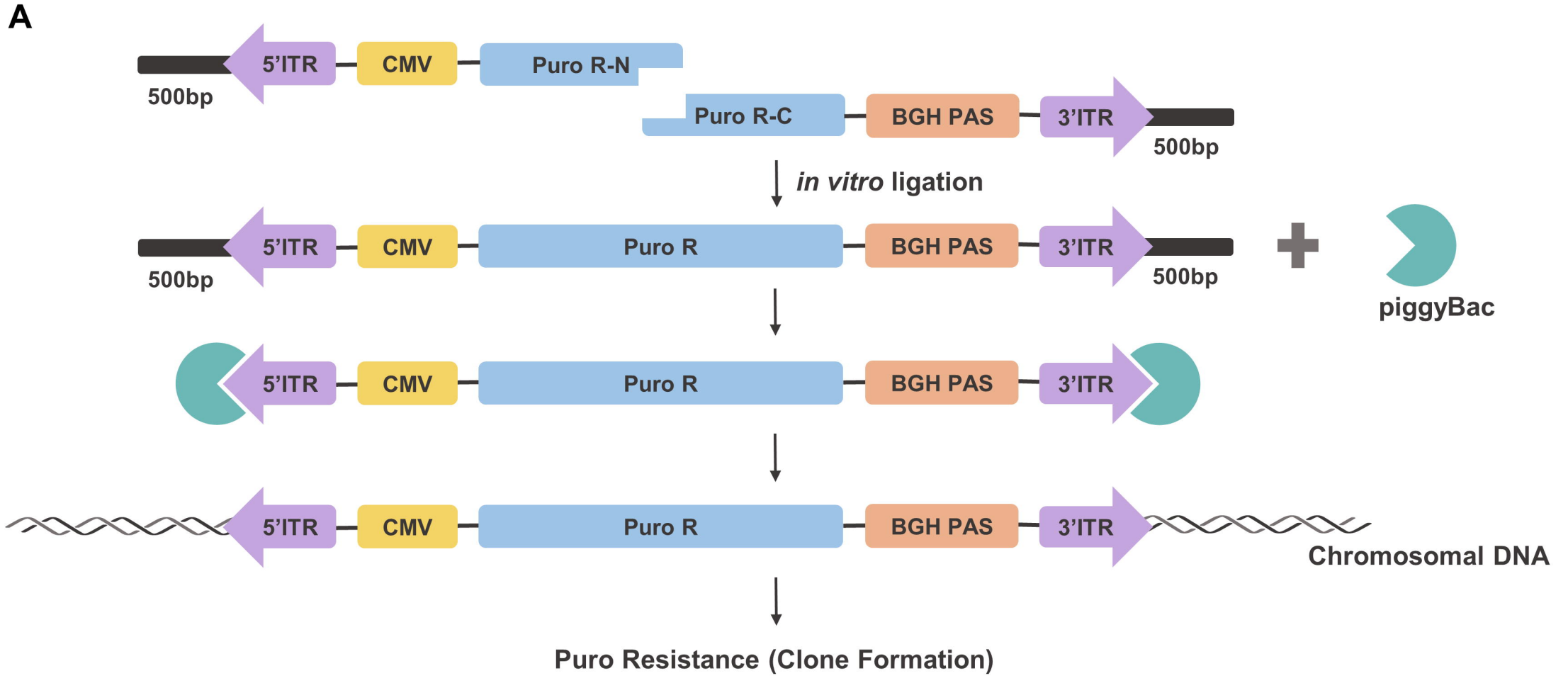
22. Kawakami, K., Tol2: a versatile gene transfer vector in vertebrates. *Genome Biol* **2007,** *8 Suppl 1*, S7.

23. Cubitt, A. B.; Heim, R.; Adams, S. R.; Boyd, A. E.; Gross, L. A.; Tsien, R. Y., Understanding, improving and using green fluorescent proteins. *Trends Biochem Sci* **1995,** *20* (11), 448-55.

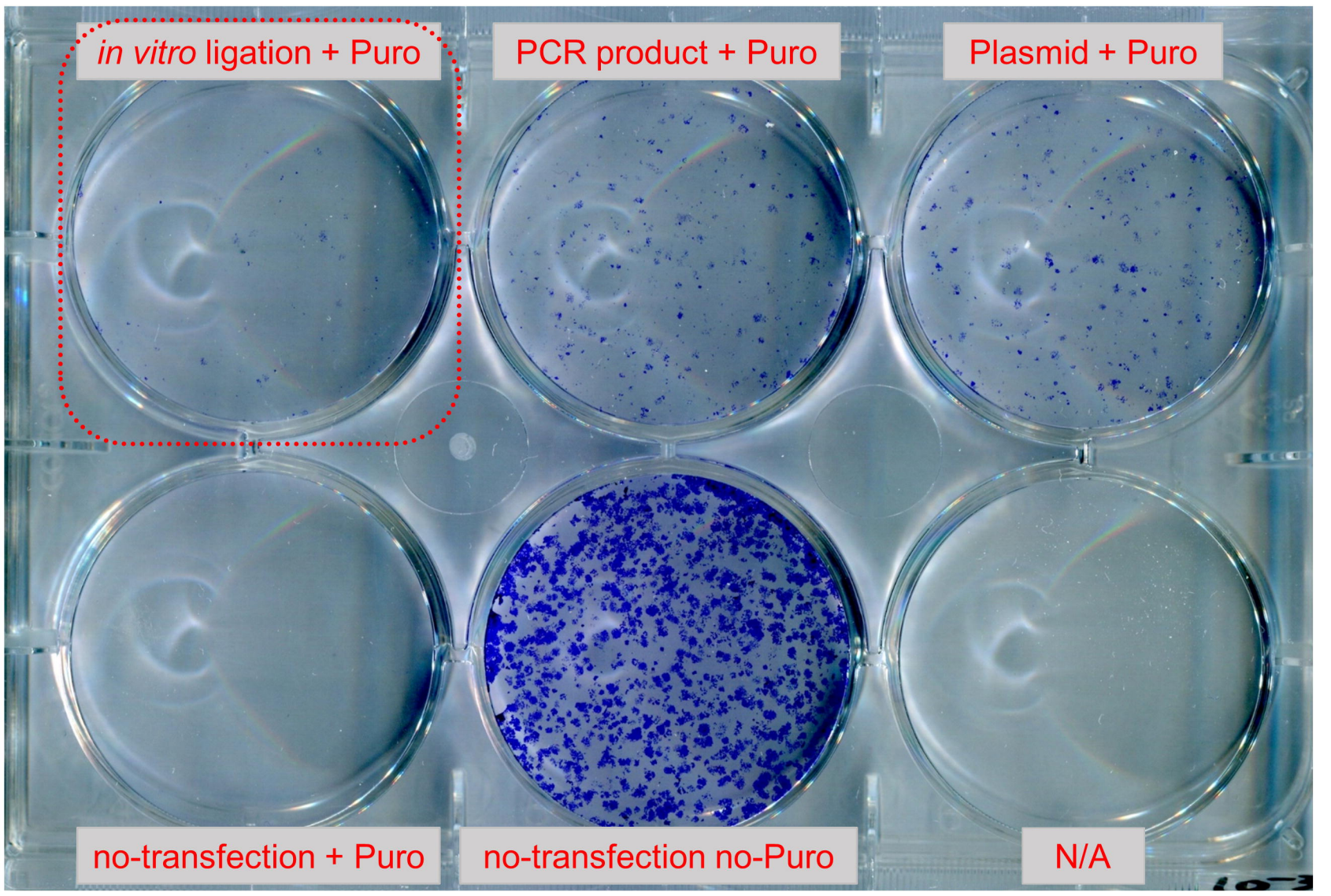
24. Day, R. N.; Davidson, M. W., The fluorescent protein palette: tools for cellular imaging. *Chem Soc Rev* **2009,** *38* (10), 2887-2921.

25. Heim, R.; Cubitt, A. B.; Tsien, R. Y., Improved green fluorescence. *Nature* **1995,** *373* (6516), 663-4.

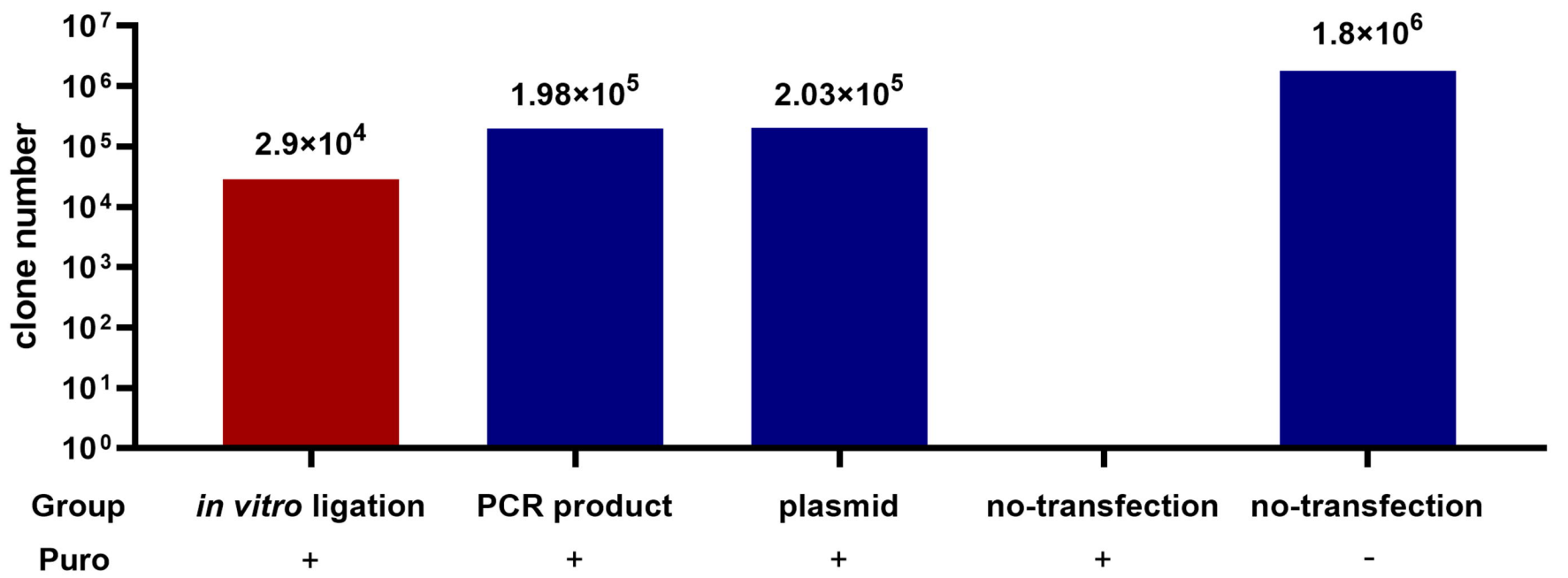
A

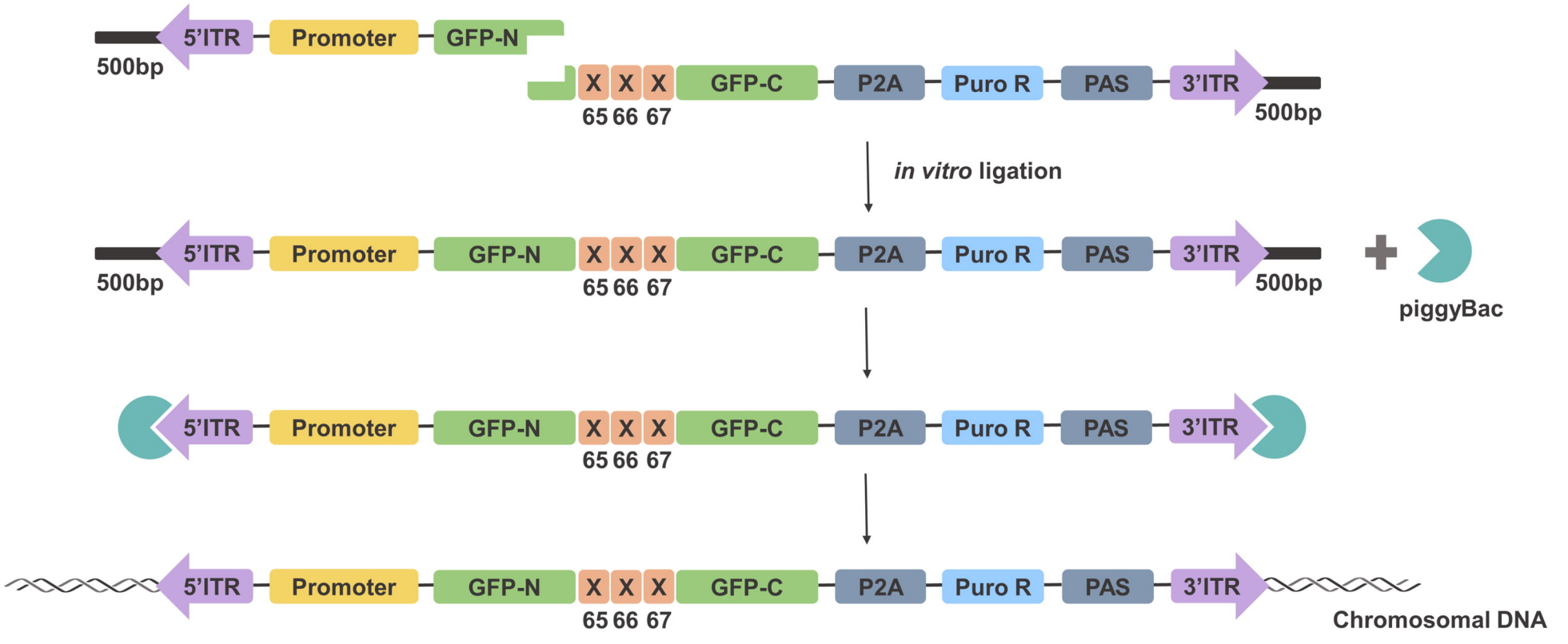

5'ITR
CMV
Puro R-N
500bp
Puro R-C
BGH PAS
3'ITR
500bp
in vitro ligation
5'ITR
CMV
Puro R
BGH PAS
3'ITR
500bp
500bp
piggyBac
5'ITR
CMV
Puro R
BGH PAS
3'ITR
5'ITR
CMV
Puro R
BGH PAS
3'ITR
Chromosomal DNA
Puro Resistance (Clone Formation)

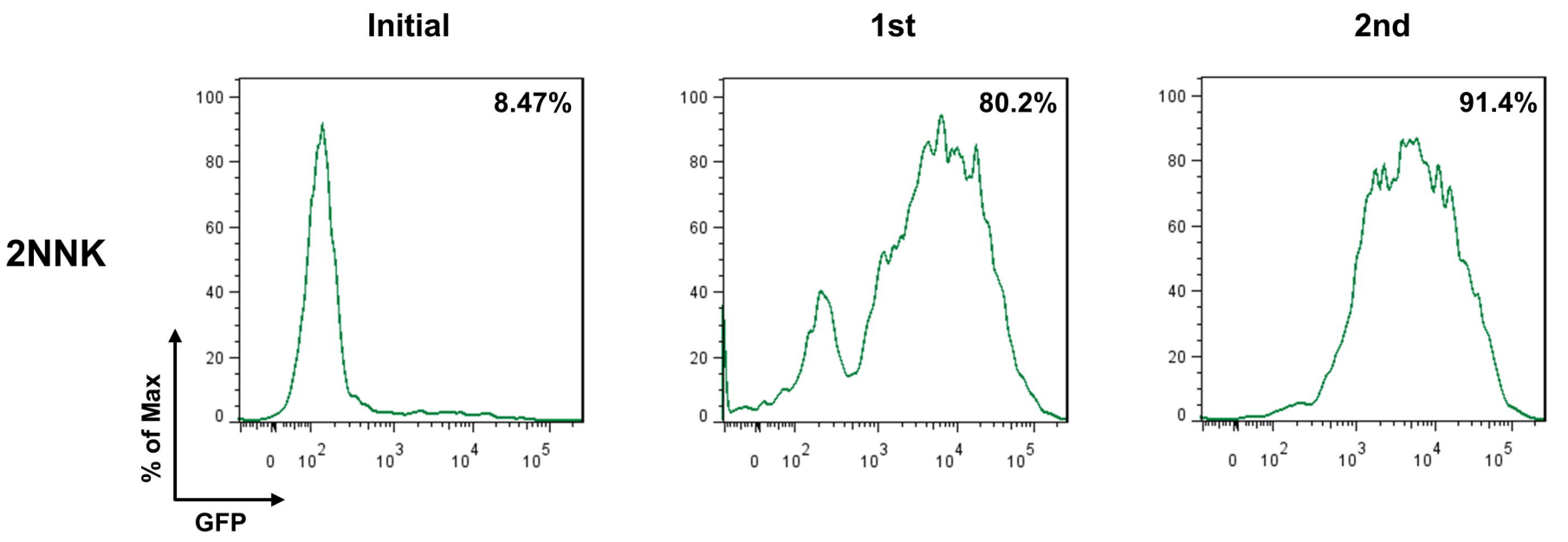
B

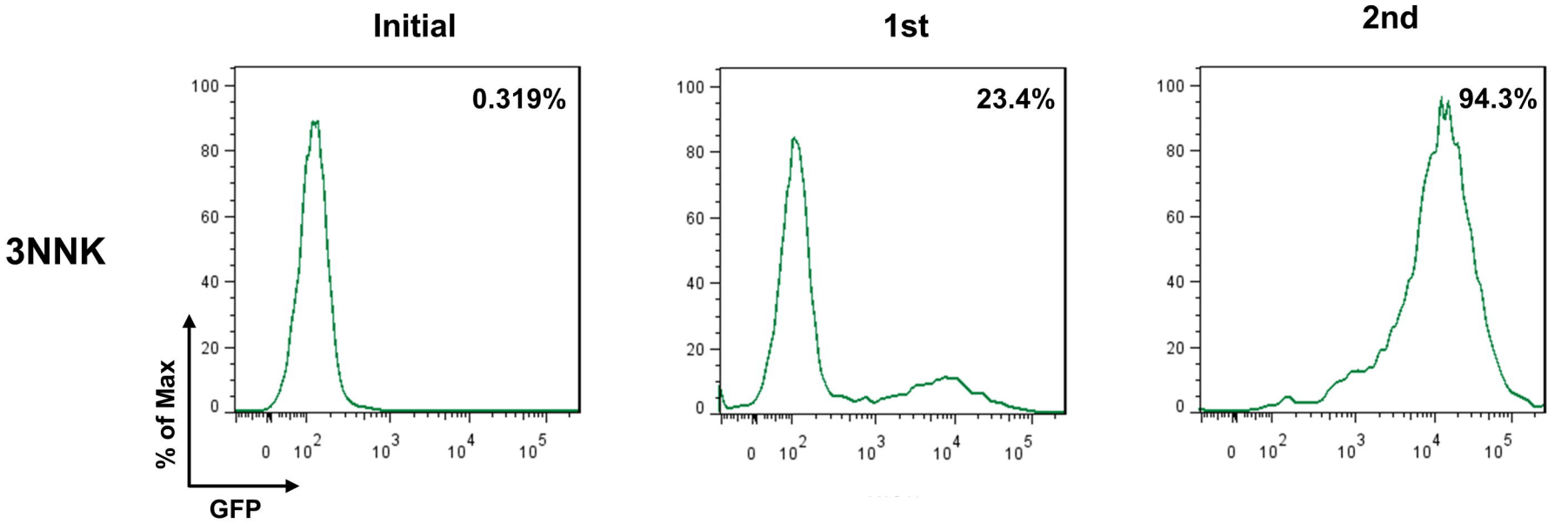

1000 X dilution
in vitro ligation + Puro
PCR product + Puro
Plasmid + Puro
no-transfection + Puro
no-transfection no-Puro
N/A

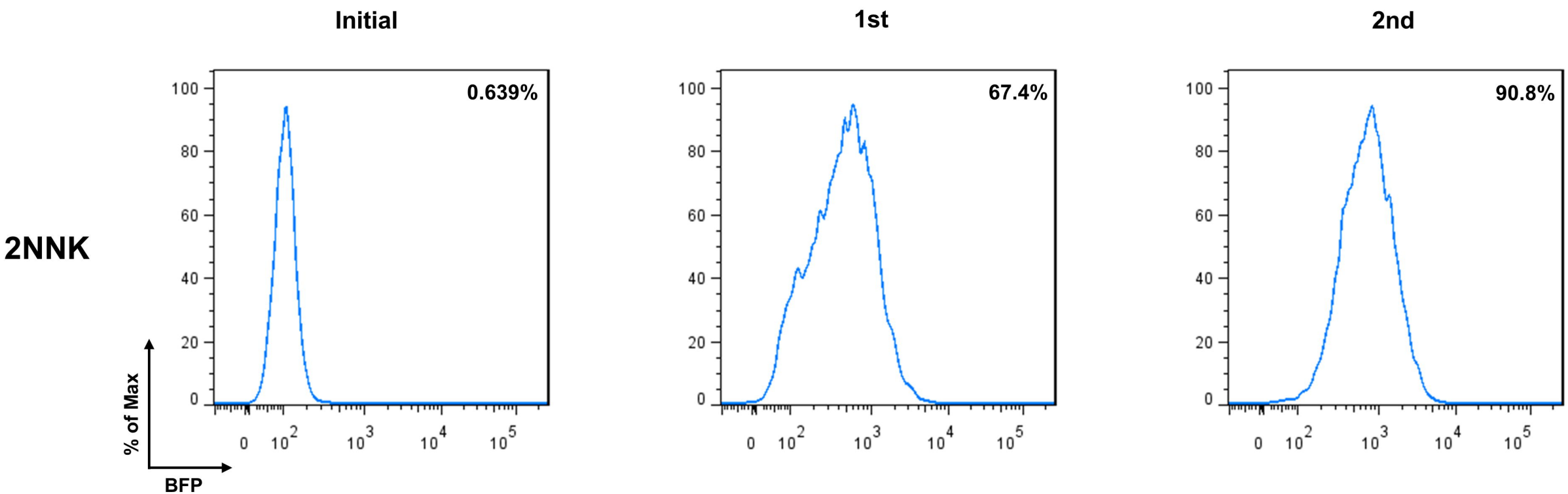
C

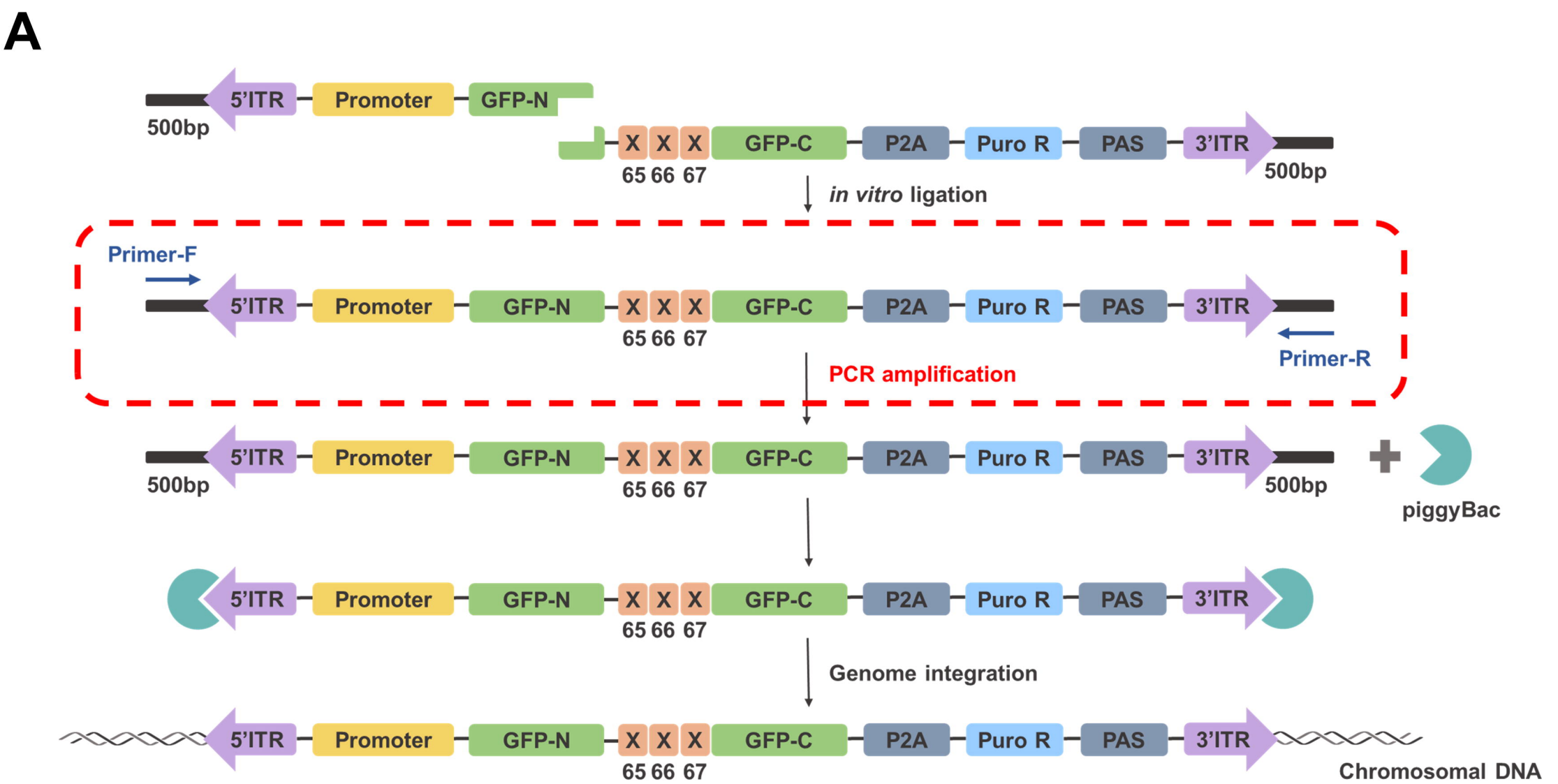

clone number
10^7
10^6
10^5
10^4
10^3
10^2
10^1
10^0
2.9×10^4
1.98×10^5
2.03×10^5
1.8×10^6
Group
in vitro ligation
PCR product
plasmid
no-transfection
no-transfection
Puro
+
+
+
+
-

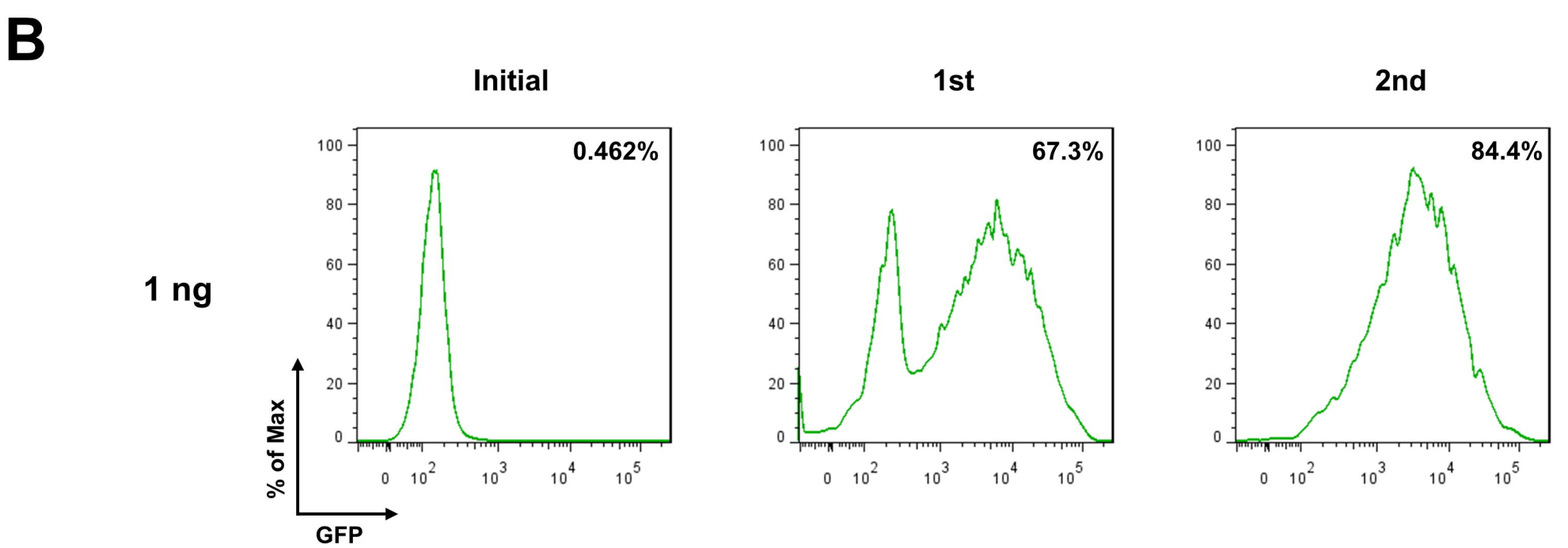

A
5'ITR
Promoter
GFP-N
500bp
X X X
65 66 67
GFP-C
P2A
Puro R
PAS
3'ITR
500bp
in vitro ligation
piggyBac
Chromosomal DNA

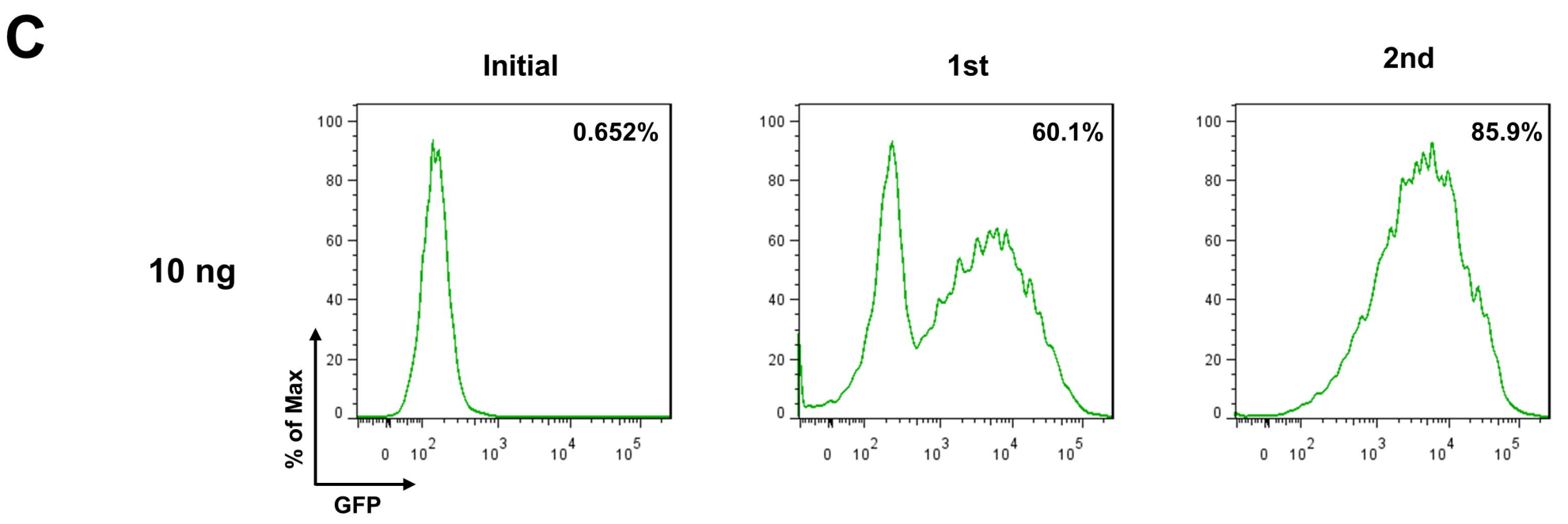

B
Initial
1st
2nd
2NNK
8.47%
80.2%
91.4%
% of Max
GFP


C
Initial
1st
2nd
3NNK
0.319%
23.4%
94.3%
% of Max
GFP


Initial
1st
2nd
2NNK
0.639%
67.4%
90.8%
% of Max
BFP
100
80
60
40
20
0
0
10^2
10^3
10^4
10^5


A
5'ITR
Promoter
GFP-N
500bp
X X X
65 66 67
GFP-C
P2A
Puro R
PAS
3'ITR
500bp
in vitro ligation
Primer-F
Primer-R
PCR amplification
piggyBac
Genome integration
Chromosomal DNA


B
Initial
1st
2nd
1 ng
0.462%
67.3%
84.4%
% of Max
GFP


C
Initial
1st
2nd
10 ng
0.652%
60.1%
85.9%
% of Max
GFP