Efficient Gene Editing Through an Intronic Selection Marker in Cells

Shang Wang
Sun Yat-sen University Cancer Center https://orcid.org/0000-0002-4359-2101
Yuqing Li
The Third Affiliated Hospital of Shenzhen University: Shenzhen Luohu Hospital Group Luohu People's Hospital
Li Zhong
Sun Yat-sen University Cancer Center
Kai Wu
The Third Affiliated Hospital of Shenzhen University: Shenzhen Luohu Hospital Group Luohu People's Hospital
Ruhua Zhang
Sun Yat-sen University Cancer Center
Tiebang Kang
Sun Yat-sen University Cancer Center
Song Wu
The Third Affiliated Hospital of Shenzhen University: Shenzhen Luohu Hospital Group Luohu People's Hospital
Yuanzhong Wu (wuyzh@sysucc.org.cn)
Sun Yat-sen University Cancer Center

Research Article

Keywords: Gene editing technolog, eukaryotes, homology, DNA, HDR strategy

Posted Date: July 30th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-739001/v1

Version of Record: A version of this preprint was published at Cellular and Molecular Life Sciences on January 31st, 2022. See the published version at https://doi.org/10.1007/s00018-022-04152-1.

Abstract

Background Gene editing technology has provided researchers with the ability to modify genome sequences in almost all eukaryotes. Gene-edited cell lines are being used with increasing frequency in both bench research and targeted therapy. Despite the great importance and universality of gene editing, however, precision and efficiency are hard to achieve with the prevailing editing strategies, such as homology-directed DNA repair (HDR) and the use of base editors (BEs).

Results & Discussion Our group has developed a novel gene editing technology to indicate DNA variation with an independent selection marker using an HDR strategy, which we named **G**ene **E**diting through an **I**ntronic **S**election marker (GEIS). GEIS uses a simple process to avoid nonhomologous end joining (NHEJ)-mediated false-positive effects and achieves editing efficiency as high as 91% without disturbing endogenous gene splicing and expression. We re-examined the correlation of the conversion tract and editing efficiency, and our data suggest that GEIS has the potential to edit approximately 99% of gene editing targets in human and mouse cells. The results of further comprehensive analysis suggest that the strategy may be useful for introducing multiple DNA variations in cells.

Background

Genetic mutations are strongly associated with human diseases, including cancers and heritable disorders. Therefore, gene editing technologies that can easily correct or generate mutations are critical for both bench research and clinical applications [1]. Bacteria-derived CRISPR-Cas systems mediate protospacer adjacent motif (PAM) sequence-containing target DNA cleavage under the guidance of sgRNA in the DNA-RNA base-pairing method. Cas12 and Cas9 have been widely applied for eukaryotic genome editing because of their high efficiency in generating DNA double-strand breaks (DSBs), after which the DNA is repaired. Despite the potential for precise base editing by DNA repair, such as homology-directed repair (HDR) and nonhomologous end joining (NHEJ) [2], the efficiency varies dramatically in different cell lines and is usually no more than 5% [3,4]. Recently, an alternative gene editing strategy was developed that uses dead Cas9 to target DNA via sgRNA and recruits base deaminase domains to accomplish C-to-T base conversions (with a cytosine base editor, CBE) or A-to-G base conversions (with an adenine base editor, ABE) without introducing DSBs. Although the efficiency of this strategy is as high as 60% on average, its off-target effects, inability to accomplish A-to-C or A-to-G conversion, and inaccuracy in the operation window strongly inhibit its application [5,6].

Introns are transcribed together with exons as pre-mRNA but are spliced by the spliceosome complex so that mature mRNA does not contain intronic sequences [7]. Although research has reported intronic mutations that cause genetic diseases [8,9], variations in introns do not heavily influence gene expression. Introns are frequently used as targets for HDR genome editing strategies [10,11]. To assist in the retrieval of successfully repaired clones, the target gene coding sequence (CDS) is typically tagged with a fluorescent protein or antibiotic resistance gene at the N- or C-terminus as a selectable marker [12]. However, this kind

of selection can be used only when the target sites are near the terminus, and it changes the open reading frame (ORF), which can have unpredictable negative effects on genetic regulation. Furthermore, the marker is driven by the endogenous promoter, which can be too weak to make a difference.

To solve the shortcomings of the current gene editing methods, we established an efficient gene editing system based on HDR-mediated intronic fluorescent protein insertion without disruption of endogenous gene splicing and expression, which we named **G**ene **E**diting through an **I**ntronic **S**election marker (GEIS). This strategy artfully avoids donor DNA-mediated false positive cell clones and produces as many as 91% gene-edited cells. The results of further studies reveal its strong potential for use in almost 99% of exon editing applications and for multiple mutation introduction.

Results

The GEIS workflow generates RELA/p65 S276C HEK293T cells within 1 month.

p65 is a REL-associated protein involved in NF-κB heterodimer formation, nuclear translocation, and downstream gene transactivation [13]. We applied GEIS to generate the S276C mutation in *RELA*/p65. A lentiCRISPR-v2 plasmid carrying sgRNA targeting intron 8 of the *RELA* gene was used to generate DSBs. To avoid disrupting RNA splicing, we did not target the splicing element. The donor DNA template contained a cytomegalovirus (CMV) promoter-driven DsRed-expressing cassette between the left and right homology arms (HAs), while the desired S276C mutation was located on the left arm (Fig. 1A). The lentiCRISPR-v2 plasmid and donor DNA were cotransfected into HEK293T cells for 24 hours, and then puromycin selection was conducted for 72 hours to kill nontransfected cells. The surviving cells were subjected to fluorescence-activated cell sorting (FACS) to isolate the DsRed-positive cells. To increase the selection efficiency, a second round of FACS was performed. The sorted cells were seeded into 96-well plates for single-cell clone growth. We obtained positive cell clones with the S276C mutation within 1 month with this workflow (Fig. 1B, C). Reverse transcription PCR (RT-PCR) and quantitative PCR (qPCR) showed that the inclusion of the CMV-DsRed cassette in the intron neither disturbed the splicing of the two adjacent exons nor affected mRNA transcription (Fig. 1D, E).

HDR with an ssDNA template reduces production of false-positive cell clones.

In this strategy, the use of dsDNA as donor DNA produces false-positive cell clones via direct transcription and translation or via random integration into the genome through canonical NHEJ (c-NHEJ) [14]. Recent studies have demonstrated that ssDNA donors show superior performance compared to dsDNA donors in mammalian systems by reducing the probability of NHEJ [15]. To effectively obtain ssDNA sequences as large as 5000 nt, we denatured the dsDNA from PCR at 95°C and with 100 mM NaCl for 10 minutes (Fig. 2A). Our data demonstrated that a single-stranded CMV-DsRed donor led to significantly lower fluorescence intensity than a double-stranded donor (Fig. 2B, S1A). We speculated that the use of an ssDNA donor would increase the true-positive rate of FACS-enriched DsRed-expressing cells. As shown in Fig. 2C, with ssDNA, the recombination rate for *RELA* S276C reached 66.7% (8 out of 12), while with

dsDNA, it was only 16.7% (2 out of 12) after two rounds of sorting. Elevated recombination rates were also observed at the *NABP2* and *EGFR* loci, and no abnormal splicing or mRNA changes were detected (Fig. 2C, D, E).

A conversion tract longer than 490 bp is necessary for NABP2 GEIS.

Despite the efficient selection of successfully recombined cells, GEIS still exhibits a low editing efficiency when the conversion tract is too long. Because the CMV-DsRed cassette must be located in an intron to avoid disrupting endogenous gene splicing and expression, the sgRNA target site should usually be intronic, but the expected conversion site is usually exonic. The distance from the DSB to the conversion site (conversion tract) affects the gene editing efficiency [16].

To estimate the influence of conversion tract length on editing efficiency, we first evaluated the HA length required for efficient insertion of the selection cassette into the intron. Using the EGFR locus as an example, we designed a series of donors with 250, 500, 800 and 1000 bp HAs. HAs longer than 500 bp were determined to be necessary for recombination at this locus (**Fig. S2 A, B**). Next, we designed donor DNA with a left HA (800 bp) containing nucleotide variations 45, 90, 171, 281, 490, 596 and 696 bp away from the DSB site for GEIS of *NABP2* (Fig. 3A). The genomic DNA of the GEIS-processed cell group was PCR-amplified with the forward primer located outside the left HA on the genome and the reverse primer at the DsRed cassette. The PCR product was cloned into pLV-MCS-puro-Green for Sanger sequencing. A total of 624 clones were sequenced, and the conversion efficiency was calculated. The results showed that the conversion efficiency decreased as the tract became further away from the DSB. In contrast to previous research based on 80 cell clones, which reported an efficiency of only 20% when the tract was 200 bp long [16], our data showed that nearly 60% of recombinants were converted at the 490 bp site (Fig. 3B).

Because introns adjacent to the target exon on both the left and right sides are available for GEIS DSB generation, a nearer intron can always be found for the exonic editing site for GEIS, which needs less than half of the exon length as the conversion tract. The conversion tract of 490 bp indicated that GEIS has an approximately 60% probability of generating mutations for exons as large as 980 bp in the locus (**Fig. S2C**).

To assess the applicability of GEIS in the human and mouse genomes, we analyzed the distributions of exon length in these two species from the Consensus CDS (CCDS) Project **(Fig. S2D)** [17–19]. Most exons longer than 1000 bp were the first or last exons, which contain long 5’ or 3’ untranslated regions (UTRs); however, DSBs can still be introduced by sgRNA in the first or last intron. When we excluded the UTRs and reanalyzed the distribution of exon lengths, only approximately 1% of exons had lengths greater than 1000 bp (Fig. 3C). Based on the conversion tract analysis from the *NABP2* locus, we speculate that GEIS might be able to edit 99% of gene targets with relatively high efficiency.

GEIS has the potential to introduce multiple DNA variations.

To evaluate the possibility of introducing multiple genome alterations in one GEIS reaction, we analyzed the mutation distributions in each of the 624 clones. A heatmap was created to show the percentage of alterations that occurred at the remaining sites (horizontal axis) when an alteration occurred at the indicated site (vertical axis) (Fig. 3D). According to the map, mutations at a further site largely indicated successful editing of the nearer site, and genome editing showed a high extent of linkage rather than independence, indicating that multiple genome alterations can be introduced in one GEIS reaction.

Discussion

Here, we have developed a universal and efficient HDR-based gene editing strategy in cell lines. Rather than tagging GFP to the target gene ORF for FACS selection, we chose to insert a pCMV-driven DsRed selection marker into introns so that the selection marker did not influence target gene expression or splicing when indicating the desired DNA variations for FACS selection. To eliminate the possibility of false-positive cell clone generation by pCMV-driven DsRed-containing donor DNA and random integration of the DNA into the genome via NHEJ, we used ssDNA as the donor template [20,21]. To obtain bulk ssDNA sequences as large as 5000 nt [22], we simply denatured dsDNA at a high temperature in a certain concentration of salt, which is widely applied for all kinds of PCR [23,24]. Given that an editing target that is too far away from an intron might result in failure to introduce a DNA variation, we explored the editing efficiency based on the conversion tract length, and the results are of great value for other gene editing strategies [25–27]. Considering these results and the results of our analysis of the exon length distribution in humans and mice, we speculate that GEIS can accomplish nearly 99% of exon editing tasks. In addition, GEIS's ability to introduce multiple DNA variations was also assessed in this study. Such an ability has strong potential for use in the development of novel features based on a directed molecular evolution strategy [28,29]. Overall, we have precisely and concisely introduced a new gene editing strategy and have described its risks as well as methods to avoid those risks. Our data can also be used in other gene editing applications and can support the use of novel strategies for specific editing tasks.

Indication of exonic gene mutation by intronic selection marker is accessible by simply denaturing donor dsDNA into ssDNA. This strategy is unlike intronic insertion knock-in HDR strategies, which are restricted to editing of bases near the stop or start codons for tagging of markers that must be located at the N or C-terminus of a CDS. The independent intronic marker strategy untethers the HDR gene editing strategy, allowing it to be used to solve editing problems in any exon in cells. Although similar intron targeting strategies have been reported, none of them eliminate the burden of in-frame marker tagging [11,30,31]. Notably, a series of intronic DNA variations have been reported to affect gene expression, RNA splicing, and to cause cancers and nervous system disorders [32–36]. Such risks accompany GEIS and may affect experiments. To overcome this defect, if needed, we can turn to scareless gene editing strategy [37] that employs an additional process of HDR to remove the DNA imprint in introns.

Methods

Plasmid and donor DNA

DSBs were generated by CRISPR-Cas9 technology with LentiCRISPR-V2 (Addgene, #98290) carrying the indicated intron-targeting sgRNAs (RELA: GGCUCUGUGCCGUGAGAGAG, NABP2: GGGCAAAGGGGUUUGCAAGG, EGFR: GCCAGCAUUUUCCUGACACC). **Important: sgRNAs should avoid targeting the splicing elements in introns!** The pDonor-GEIS plasmid was used as the framework for preparing the donor DNA. According to the sequence, two EcoRV sites were located adjacent to the pCMV-driving DsRed cassette for HA cloning. **Important: donor DNA should contain no terminators.** HAs were PCR amplified from HEK293T genomic DNA and inserted into the plasmid using a Gibson assembly cloning strategy. Mutations in HAs were generated by mutation- or truncation-containing primers using an overlap extension PCR strategy.

Donor DNA was generated from well-prepared pDonor-GEIS (RELA, NABP2, and EGFR) by PCR using a pair of universal primers: Donor-F: TGTGGTGGAATTCTGCAGAT and Donor-R: GCGGCCGCCACTGTGCTGGAT. PCRs were carried out in 28 cycles on an Eppendorf thermocycler with denaturation at 94°C for 15 seconds, annealing at 58°C for 15 seconds, and extension at 72°C for 3 minutes using PrimeStar (TaKaRa, Japan). PCR products were purified using an Ultra-Sep Gel Extraction Kit (Omega).

Fluorescence microscopy

Donor dsDNA and ssDNA were separately transfected into HEK293T cells. dsDNA was prepared from purified PCR product that was dissolved in ddH_2O. ssDNA was prepared by denaturation of dsDNA at 95°C for 10 minutes in 100 mM NaCl. Transfected cells were photographed 24 hours after transfection under a fluorescence microscope (NIKON).

Cell culture and FACS

HEK293T cells were obtained from the American Type Culture Collection. Cells were cultured in DMEM (Dulbecco's modified Eagle's medium) containing 10% fetal bovine serum at 37°C and 5% CO_2. Fifteen micrograms of LentiCRISPR-V2 and 2 μg of donor DNA were cotransfected into 5 X 10^7 cells with Lipofectamine 2000™ (Invitrogen) according to the manufacturer's instructions. Subsequently, transfected cells were treated with 1 μg/mL puromycin for 72 hours. Cells were washed with PBS and treated with 0.05% trypsin. The cell suspension was filtered through a 40 μm cell strainer (BD Falcon) before FACS. Flow cytometry analysis and FACS were performed using BD LSR II. Cells isolated by FACS were then cultured for one week and processed by FACS again to enhance the positive rate. Harvested cells were seeded in 96-well plates at 1/2 cell per well for single-clone growth.

Genomic DNA extraction and analysis

Genomic DNA was extracted using the TIANamp Genomic DNA Kit (#DP304-03). PCR of single clones derived from genomic DNA was processed to verify that the clone possessed the desired sequence. PCRs were performed for 35 cycles on an Eppendorf thermocycler with denaturation at 94°C for 15 seconds,

annealing at 58°C for 15 seconds, and extension at 72°C for 30 seconds using PrimeStar. Forward primers were located outside the left HA on the genome, and the universal reverse primer was located in the DsRed cassette.

The primers used were as follows:

RELA-gTest-F: GCTCATTGCCAAGGTGGGTA

NABP2-gTest-F: GGATGGACCGAGTCCCGGCT

EGFR-gTest-F: ATAAGAAGTCTGCAGAACTT

Red Uni-R: TTGGACATGACTCCACAT

Conversion tract length detection

Multiple DNA variations on left HA for NABP2 GEIS were introduced to HEK293T cells. FACS-sorted cells were collected for genomic DNA extraction. PCR was performed to amplify the successfully integrated DNA fragment from the donor DNA. PCRs were performed for 28 cycles on an Eppendorf thermocycler with denaturation at 94°C for 15 seconds, annealing at 58°C for 15 seconds, and extension at 72°C for 1 minute using PrimeStar. PCR products were cloned into the vector pLV-MCS-puro-Green (digested by EcoRI) using Gibson assembly. A total of *E. coli* 624 colonies were sequenced. The mutations identified in the colonies were mapped to the wild-type NABP2 genomic sequence, and the 7 candidate DNA alterations were recorded and calculated. The percentage of every variation was calculated as the number of mutated clones divided by 624 (total number of clones) and is illustrated in Fig. 2B.

The primers used were as follows:

EcoRI + NABP2-F: TTCTAGAGCTAGCGAATTGGATGGACCGAGTCCCGGCT

EcoRI + UniR: CCGATTTAAATTCGAATTTTGGACATGACTCCACAT

Human and mouse exon length distribution

Human (GRCh38, release 37) and mouse (GRCm38, release M25) genomic annotation files from GENCODE were used to evaluate the distribution of CDSs and exon lengths. In brief, each exon (UTR contained or excluded) was identified, and its length was calculated based on the end and start positions in the genome. The calculated lengths were then grouped and illustrated.

RNA extraction, RT-PCR and qPCR

Total RNA was extracted using a MolPure Cell/Tissue Total RNA Kit (Yeasen, China). RNA concentration was quantified by Nanodrop^{C} (Thermo, US). cDNA was processed with DNAase treatment and reverse transcription from 500 ng total RNA using the Hifair III 1st Strand cDNA Synthesis Kit (Yeasen). Reverse transcription was processed on thermocycler at 25°C 5min, 55°C 15min and then 85°C 5min. RT-PCR was processed for 35 cycles on an Eppendorf thermocycler with denaturation at 94°C for 15 seconds,

annealing at 58°C for 15 seconds, and extension at 72°C for 20 seconds using PrimeStar. RT-PCR primers were set on exons adjacent to the processed intron to determine whether any alternative variants were produced. qPCR was performed to check whether there were any significant differences in expression in edited cells. qPCR reactions were processed using Hieff UNICON Universal Blue qPCR SYBR Green Master Mix (Yeasen) on QuantStudio Dx (ABI), and qPCR primer sequences were derived from PrimerBank. The qPCR reaction was repeated 3 times.

The RT-PCR and qPCR primers were as follows:

RT-RELA-F: CTCGGTGGGGATGAGATCTT

RT-RELA-R: TTCTTCATGATGCTCTTGAA
RT-NABP2-F: GACAAAACAGGCAGCATCAA

RT-NABP2-R: GGGTTTGGCTCACTGAAGTT
RT-EGFR-F: GTGATGGCCAGCGTGGACAA

RT-EGFR-R: GGGATTCCGTCATATGGCTT

qGAPDH-F: GGAGCGAGATCCCTCCAAAAT

qGAPDH-R: GGCTGTTGTCATACTTCTCATGG

(product length = 197bp)

qNABP2-F: TCTGGGACGATGTTGGCAAT

qNABP2-R: GGTGCCTGCTGGGTGCTGTA

(product length = 202bp)

qRELA-F: CCCAACACTGCCGAGCTCAA

qRELA-R: CCTTTTACGTTTCTCCTCAA

(product length = 348bp)

qEGFR-F: AGGCACGAGTAACAAGCTCAC

qEGFR-R: ATGAGGACATAACCAGCCACC

(product length = 177bp)

Declarations

DATA AVAILABILITY

Python script used in this study were online: https://github.com/wukaiyeah/Genomic_CDS_stats.git.

FACS data can be downloaded from http://flowrepository.org/id/FR-FCM-Z3UN.

Sanger sequencing data for conversion tract tests were submitted to genebank as MZ399804 - MZ400408

ACCESSION NUMBERS

Sanger sequencing data for conversion tract tests were submitted to genbank as MZ399804 - MZ400408

ACKNOWLEDGEMENT

We thank Dr. Kang Wu for technical support in FACS experiments.

FUNDING

This project was supported by grants from the National Key Research and Development Program of China [2017YFA0105900 to Song Wu], the National Nature Science Foundation of China (NSFC) [81772922 to Y.W., 81802741 to Y. L.], The China Postdoctoral Science Foundation [2020M682949 to Shang Wang]. The Shenzhen Basic Research Project [JCYJ20170412155305340 to Y. L.]. The Special Support Plan for Leading talents in scientific and technological innovation in Guangdong Province [2017TX04R091 to Song Wu]

COMPETING INTERESTS

The authors declare that they have no competing interests.

References

1. Anzalone, A. V., Koblan, L. W. & Liu, D. R. Genome editing with CRISPR-Cas nucleases, base editors, transposases and prime editors. *Nat Biotechnol* **38**, 824-844, doi:10.1038/s41587-020-0561-9 (2020).
2. Yao, X. *et al.* Homology-mediated end joining-based targeted integration using CRISPR/Cas9. *Cell Res* **27**, 801-814, doi:10.1038/cr.2017.76 (2017).
3. Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819-823, doi:10.1126/science.1231143 (2013).
4. Doman, J. L., Raguram, A., Newby, G. A. & Liu, D. R. Evaluation and minimization of Cas9-independent off-target DNA editing by cytosine base editors. *Nat Biotechnol* **38**, 620-628, doi:10.1038/s41587-020-0414-6 (2020).
5. Wrighton, K. H. Cytosine base editors go off-target. *Nat Rev Genet* **20**, 254-255, doi:10.1038/s41576-019-0110-x (2019).

6. Zhou, S. *et al.* Programmable Base Editing of the Sheep Genome Revealed No Genome-Wide Off-Target Mutations. *Front Genet* **10**, 215, doi:10.3389/fgene.2019.00215 (2019).
7. Wilkinson, M. E., Charenton, C. & Nagai, K. RNA Splicing by the Spliceosome. *Annu Rev Biochem* **89**, 359-388, doi:10.1146/annurev-biochem-091719-064225 (2020).
8. Jaganathan, K. *et al.* Predicting Splicing from Primary Sequence with Deep Learning. *Cell* **176**, 535-548 e524, doi:10.1016/j.cell.2018.12.015 (2019).
9. Sanchez-Roige, S. *et al.* Genome-wide association study of delay discounting in 23,217 adult research participants of European ancestry. *Nat Neurosci* **21**, 16-18, doi:10.1038/s41593-017-0032-x (2018).
10. Kocher, T. *et al.* Cut and Paste: Efficient Homology-Directed Repair of a Dominant Negative KRT14 Mutation via CRISPR/Cas9 Nickases. *Mol Ther* **25**, 2585-2598, doi:10.1016/j.ymthe.2017.08.015 (2017).
11. Yoshimi, K. *et al.* Combi-CRISPR: combination of NHEJ and HDR provides efficient and precise plasmid-based knock-ins in mice and rats. *Hum Genet* **140**, 277-287, doi:10.1007/s00439-020-02198-4 (2021).
12. Doudna, J. A. & Charpentier, E. Genome editing. The new frontier of genome engineering with CRISPR-Cas9. *Science* **346**, 1258096, doi:10.1126/science.1258096 (2014).
13. Nowak, D. E. *et al.* RelA Ser276 phosphorylation is required for activation of a subset of NF-kappaB-dependent genes by recruiting cyclin-dependent kinase 9/cyclin T1 complexes. *Mol Cell Biol* **28**, 3623-3638, doi:10.1128/MCB.01152-07 (2008).
14. Zhang, X., Li, T., Ou, J., Huang, J. & Liang, P. Homology-based repair induced by CRISPR-Cas nucleases in mammalian embryo genome editing. *Protein Cell*, doi:10.1007/s13238-021-00838-7 (2021).
15. Yoshimi, K. *et al.* ssODN-mediated knock-in with CRISPR-Cas for large genomic regions in zygotes. *Nat Commun* **7**, 10431, doi:10.1038/ncomms10431 (2016).
16. Elliott, B., Richardson, C., Winderbaum, J., Nickoloff, J. A. & Jasin, M. Gene conversion tracts from double-strand break repair in mammalian cells. *Mol Cell Biol* **18**, 93-101, doi:10.1128/mcb.18.1.93 (1998).
17. Pujar, S. *et al.* Consensus coding sequence (CCDS) database: a standardized set of human and mouse protein-coding regions supported by expert curation. *Nucleic Acids Res* **46**, D221-D228, doi:10.1093/nar/gkx1031 (2018).
18. Pruitt, K. D. *et al.* The consensus coding sequence (CCDS) project: Identifying a common protein-coding gene set for the human and mouse genomes. *Genome Res* **19**, 1316-1323, doi:10.1101/gr.080531.108 (2009).
19. Harte, R. A. *et al.* Tracking and coordinating an international curation effort for the CCDS Project. *Database (Oxford)* **2012**, bas008, doi:10.1093/database/bas008 (2012).
20. Maurissen, T. L. & Woltjen, K. Synergistic gene editing in human iPS cells via cell cycle and DNA repair modulation. *Nat Commun* **11**, 2876, doi:10.1038/s41467-020-16643-5 (2020).

21. Wang, Y. *et al.* Systematic evaluation of CRISPR-Cas systems reveals design principles for genome editing in human cells. *Genome Biol* **19**, 62, doi:10.1186/s13059-018-1445-x (2018).
22. Inoue, Y. U. *et al.* An Optimized Preparation Method for Long ssDNA Donors to Facilitate Quick Knock-In Mouse Generation. *Cells* **10**, doi:10.3390/cells10051076 (2021).
23. Ho, S. N., Hunt, H. D., Horton, R. M., Pullen, J. K. & Pease, L. R. Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene* **77**, 51-59, doi:10.1016/0378-1119(89)90358-2 (1989).
24. Liang, P. & Pardee, A. B. Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. *Science* **257**, 967-971, doi:10.1126/science.1354393 (1992).
25. Huang, T. K., Armstrong, B., Schindele, P. & Puchta, H. Efficient gene targeting in Nicotiana tabacum using CRISPR/SaCas9 and temperature tolerant LbCas12a. *Plant Biotechnol J*, doi:10.1111/pbi.13546 (2021).
26. Keith, N. *et al.* High mutational rates of large-scale duplication and deletion in Daphnia pulex. *Genome Res* **26**, 60-69, doi:10.1101/gr.191338.115 (2016).
27. Vernekar, D. V. *et al.* The Pif1 helicase is actively inhibited during meiotic recombination which restrains gene conversion tract length. *Nucleic Acids Res* **49**, 4522-4533, doi:10.1093/nar/gkab232 (2021).
28. Li, C. *et al.* Targeted, random mutagenesis of plant genes with dual cytosine and adenine base editors. *Nat Biotechnol* **38**, 875-882, doi:10.1038/s41587-019-0393-7 (2020).
29. Li, C. *et al.* Direct Modification of Multiple Gene Homoeologs in Brassica oleracea and Brassica napus Using Doubled Haploid Inducer-Mediated Genome Editing System. *Plant Biotechnol J*, doi:10.1111/pbi.13632 (2021).
30. Serebrenik, Y. V., Sansbury, S. E., Kumar, S. S., Henao-Mejia, J. & Shalem, O. Efficient and flexible tagging of endogenous genes by homology-independent intron targeting. *Genome Res* **29**, 1322-1328, doi:10.1101/gr.246413.118 (2019).
31. Suzuki, K. *et al.* Precise in vivo genome editing via single homology arm donor mediated intron-targeting gene integration for genetic disease correction. *Cell Res* **29**, 804-819, doi:10.1038/s41422-019-0213-0 (2019).
32. de Wolf, B. *et al.* Chromosomal instability by mutations in the novel minor spliceosome component CENATAC. *EMBO J*, e106536, doi:10.15252/embj.2020106536 (2021).
33. Wang, T. Y. *et al.* A pan-cancer transcriptome analysis of exitron splicing identifies novel cancer driver genes and neoepitopes. *Mol Cell*, doi:10.1016/j.molcel.2021.03.028 (2021).
34. Ishiura, H. *et al.* Expansions of intronic TTTCA and TTTTA repeats in benign adult familial myoclonic epilepsy. *Nat Genet* **50**, 581-590, doi:10.1038/s41588-018-0067-2 (2018).
35. Shi, Y. *et al.* Haploinsufficiency leads to neurodegeneration in C9ORF72 ALS/FTD human induced motor neurons. *Nat Med* **24**, 313-325, doi:10.1038/nm.4490 (2018).

36. Kim, J. H. *et al.* Real-world utility of next-generation sequencing for targeted gene analysis and its application to treatment in lung adenocarcinoma. *Cancer Med* **10**, 3197-3204, doi:10.1002/cam4.3874 (2021).
37. Ikeda, K. *et al.* Efficient scarless genome editing in human pluripotent stem cells. *Nat Methods* **15**, 1045-1047, doi:10.1038/s41592-018-0212-y (2018).

Figures

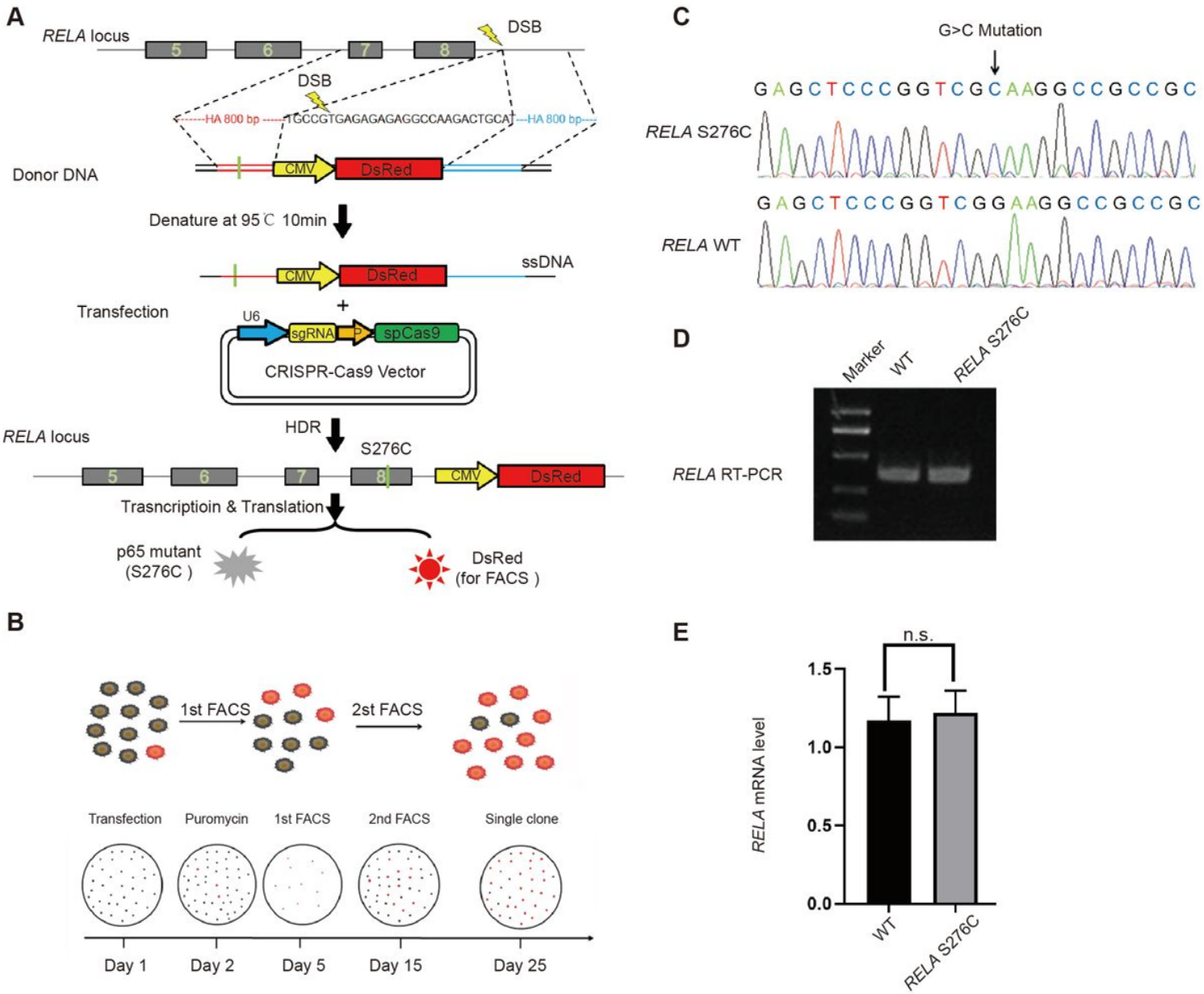

Figure 1

A. Schematic view of GEIS to introduce the S276C mutation into the RELA locus in HEK293T cells. CRISPR technology is used to generate DSB in intron 8 and donor ssDNA as a template to introduce mutations by HDR. B. Workflow of GEIS, with two rounds of FACS to enrich and fractionate individual

DsRed-positive cells. C. Sanger sequencing of RELA genomic DNA sequences derived from WT- and GEIS-treated RELA S276C gene-edited HEK293T cells. D. RT-PCR of cDNA from WT and S276C HEK293T cells. No alternative variants were found. E. Relative expression of RELA in WT and S276C HEK293T cells. No significant (n. s.) change in RELA expression was detected.

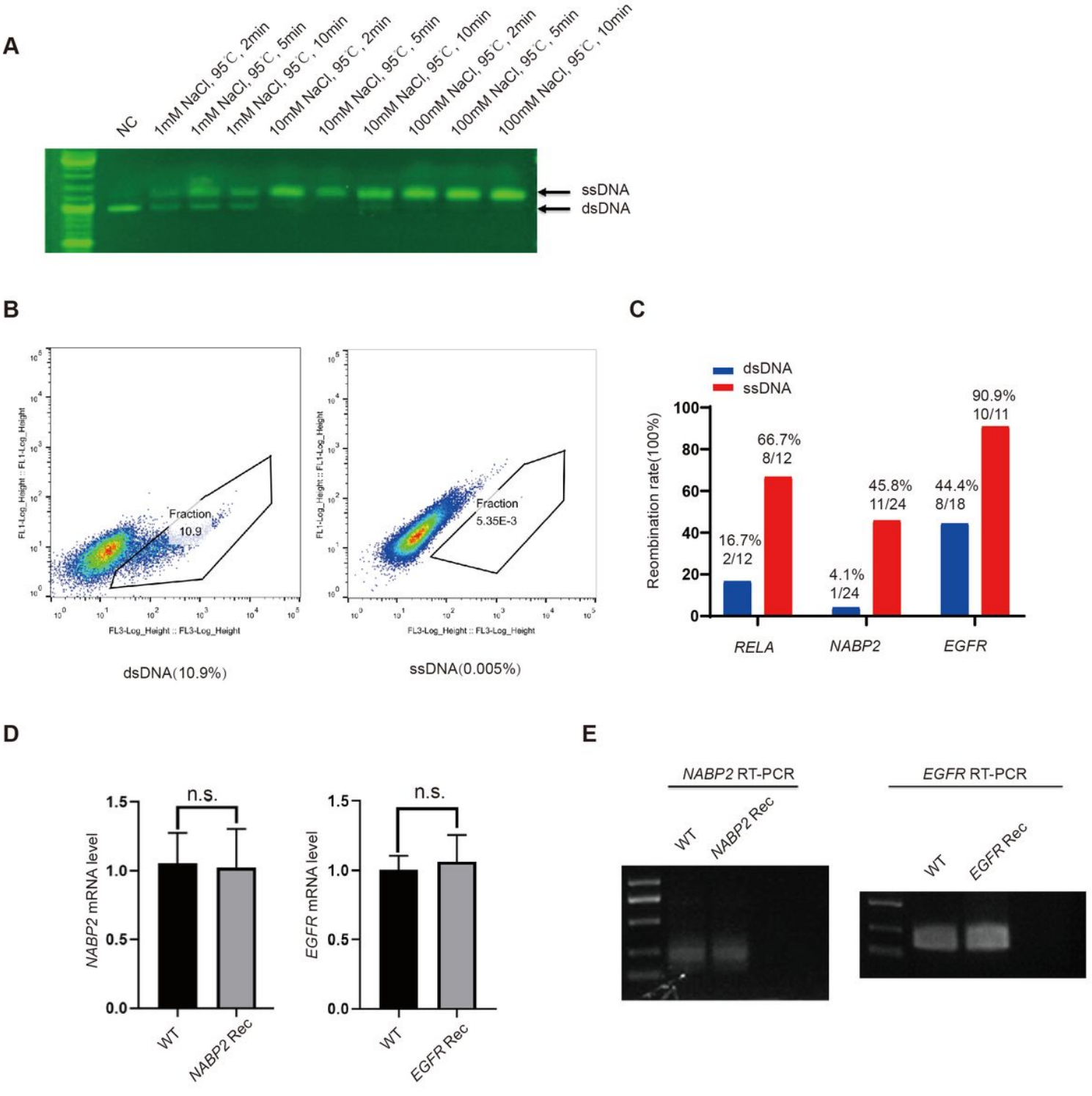

Figure 2

A. Agarose gel electrophoresis of denaturing dsDNA under the indicated conditions. B. Percentage of DsRed-positive cells from dsDNA- or ssDNA-transfected HEK293T cells determined by FACS. C. HDR

efficiency of RELA, NABP2 and EGFR using GEIS using ssDNA (denatured) or dsDNA (nondenatured) as donor DNA. D. qPCR of NABP2 and EGFR in WT and GEIS-recombined (Rec) cells. Data are mean ± s. d. of n=3 biological independent experiments. No significant (n. s.) variation was found by Student's t-test for either NABP2 or EGFR mutants. E. Agarose gel electrophoresis of RT-PCR products of NABP2 and EGFR in WT- and GEIS-modified cells. No alternative variant was found.

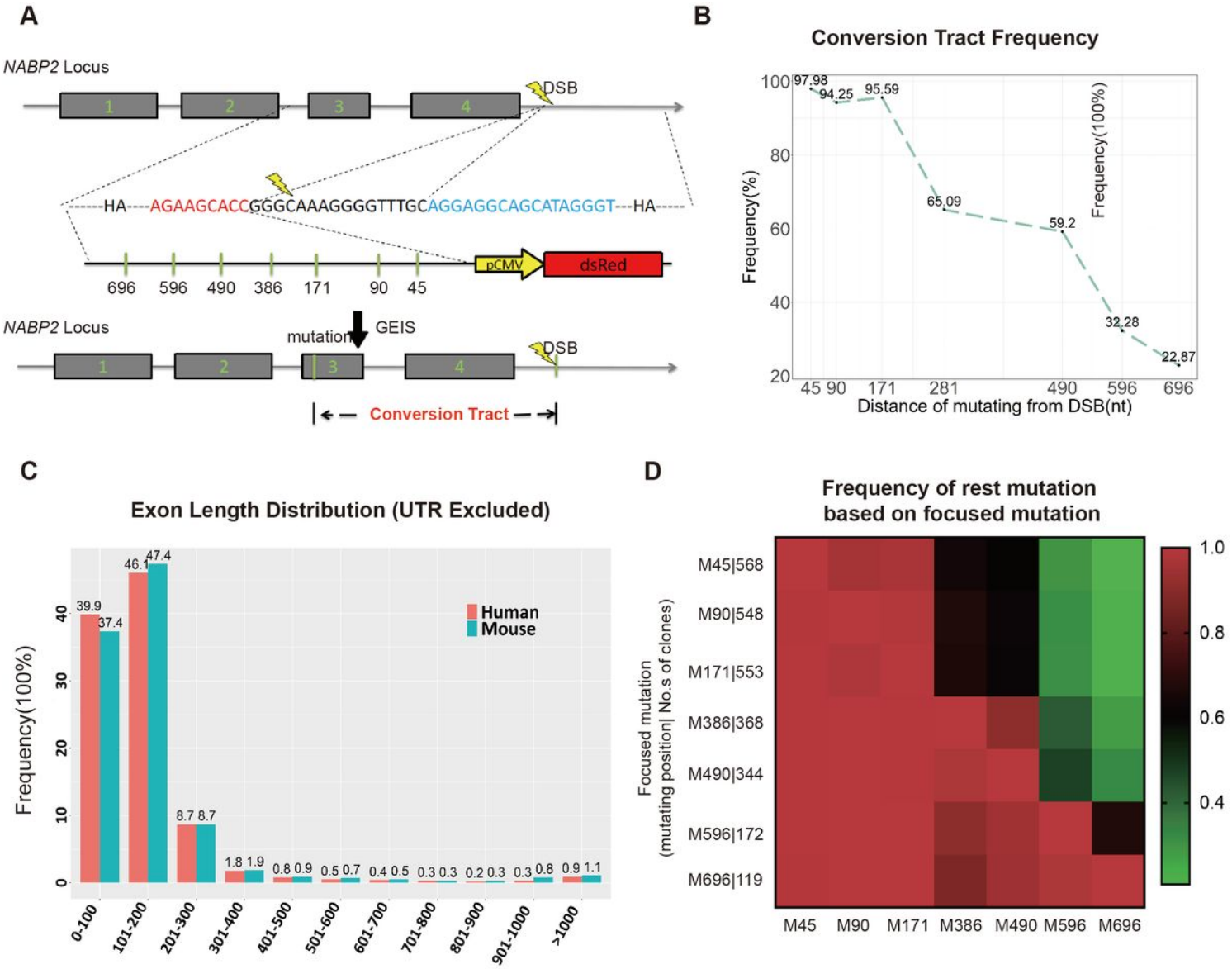

Figure 3

A. Schematic overview of the genome editing conversion tract modification experiments: Seven different DNA variations were included in the left HA of donor DNA in GEIS to test how many of these variations could be introduced in the genome locus of NABP2. B. Percentage of DNA variations from Fig. 3A that were introduced in the genome. C. Calculation of exon lengths of all human and mouse exons (UTR excluded). D. Calculation of editing efficiency of each variation indicated in Fig. 3B from Sanger sequencing of 624 clones that were PCR amplified from the GEIS-edited cell mix. When alteration A (vertical axis, M stands for mutated position, numbers stand for numbers of mutated clones) occurs, the

percentage of the rest alterations (horizontal axis, M stands for mutated position) is presented on a heat map.

Supplementary Files

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

- NARSupplementary.docx