Stem Cell Reports
Report
OPEN ACCESS

FACS-Assisted CRISPR-Cas9 Genome Editing Facilitates Parkinson's Disease Modeling

Jonathan Arias-Fuenzalida,[1,2,7] Javier Jarazo,[1,7] Xiaobing Qing,[1] Jonas Walter,[1] Gemma Gomez-Giro,[1,4] Sarah Louise Nickels,[1,3] Holm Zaehres,[4,5] Hans Robert Schöler,[4,6] and Jens Christian Schwamborn[1,*]

[1]Luxembourg Centre for Systems Biomedicine (LCSB), Developmental and Cellular Biology, University of Luxembourg, 7 Avenue des Hauts-Fourneaux, Luxembourg City 4362, Luxembourg
[2]Graduate School of Biostudies, Kyoto University, Kyoto 606-8502, Japan
[3]Life Science Research Unit (LSRU), Systems Biology, University of Luxembourg, 6 Avenue du Swing, Luxembourg City 4367, Luxembourg
[4]Max Planck Institute for Molecular Biomedicine, Laboratory of Cell and Developmental Biology, Roentgenstrasse 20, Muenster, Germany
[5]Ruhr-University Bochum, Medical Faculty, Department of Anatomy and Molecular Embryology, 44801 Bochum, Germany
[6]Medical Faculty, Westphalian Wilhelms University Muenster, 48149 Muenster, Germany
[7]Co-first author
*Correspondence: jens.schwamborn@uni.lu
http://dx.doi.org/10.1016/j.stemcr.2017.08.026

SUMMARY

Genome editing and human induced pluripotent stem cells hold great promise for the development of isogenic disease models and the correction of disease-associated mutations for isogenic tissue therapy. CRISPR-Cas9 has emerged as a versatile and simple tool for engineering human cells for such purposes. However, the current protocols to derive genome-edited lines require the screening of a great number of clones to obtain one free of random integration or on-locus non-homologous end joining (NHEJ)-containing alleles. Here, we describe an efficient method to derive biallelic genome-edited populations by the use of fluorescent markers. We call this technique FACS-assisted CRISPR-Cas9 editing (FACE). FACE allows the derivation of correctly edited polyclones carrying a positive selection fluorescent module and the exclusion of non-edited, random integrations and on-target allele NHEJ-containing cells. We derived a set of isogenic lines containing Parkinson's-disease-associated mutations in α-synuclein and present their comparative phenotypes.

INTRODUCTION

Parkinson's disease (PD) is a multifactorial neurodegenerative disorder characterized by motor and non-motor symptoms (Caligiore et al., 2016). Some cases of PD cases result from autosomal dominant mutations in the *SNCA* gene, which encodes α-synuclein. Physiologically, α-synuclein is implicated in synaptic transmission and vesicle transport, while pathologically it is part of the protein aggregates known as Lewy bodies and Lewy neurites (Goedert et al., 2013). Patients carrying mutations in the *SNCA* gene suffer from early onset of PD. Mutations in *SNCA* include increase in gene dosage (Devine et al., 2011) and heterozygous missense mutations such as p.A30P and p.A53T (Bendor et al., 2013; Soldner et al., 2011). Mutations in *SNCA* can account for up to 15% of cases of familial early-onset PD (Bozi et al., 2014).

Importantly, genome editing tools can assist in parsing PD phenotypes. The reliability of CRISPR-Cas9 as an editing tool has been extensively validated by whole-genome sequencing (Veres et al., 2014). Furthermore, Cas9 specificity has been improved with high-fidelity variants (Kleinstiver et al., 2016). However, eliminating uncertainties in genotype outcomes of edited lines has remained challenging. Screening of correctly edited clones is a labor-consuming process that entails the selection of on-target knockin clones and the exclusion of random integrations, on-target indels, and second-allele indel events. To leverage the power of genome editing tools in the evaluation of polygenic diseases such as PD, it is necessary to overcome such labor- and time-consuming limitations. Hence, the fast generation of genome-edited populations carrying a known genotype outcome is highly necessary.

RESULTS

Deterministic Genotype Outcomes for the Generation of Isogenic Lines

The use of donors containing fluorescent protein (FP) reporters associated with defined SNP variants enables editing outcomes of known genotype (Figures 1A and 1E): heterozygous, homozygous healthy, and homozygous pathogenic (Figure 1E). One-step biallelic targeting occurs with a mean frequency of 37.5% using double-stranded DNA (dsDNA) templates (Table S1). Donor vectors for *SNCA* exon 2 and exon 3 were cloned with an internal positive selection module (PSM) coding EGFP or dTOMATO, and an external negative selection module (NSM) containing tagBFP (Figure 1A). *SNCA* mutations are dominant, and missense *SNCA* PD patients are heterozygote. Hence, donor pairs for the *SNCA* mutations rs104893878 (p.A30P) and rs104893877 (p.A53T) were designed to match heterozygous genotype outcomes (Figure 1E). In the case of *SNCA*e2,

This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

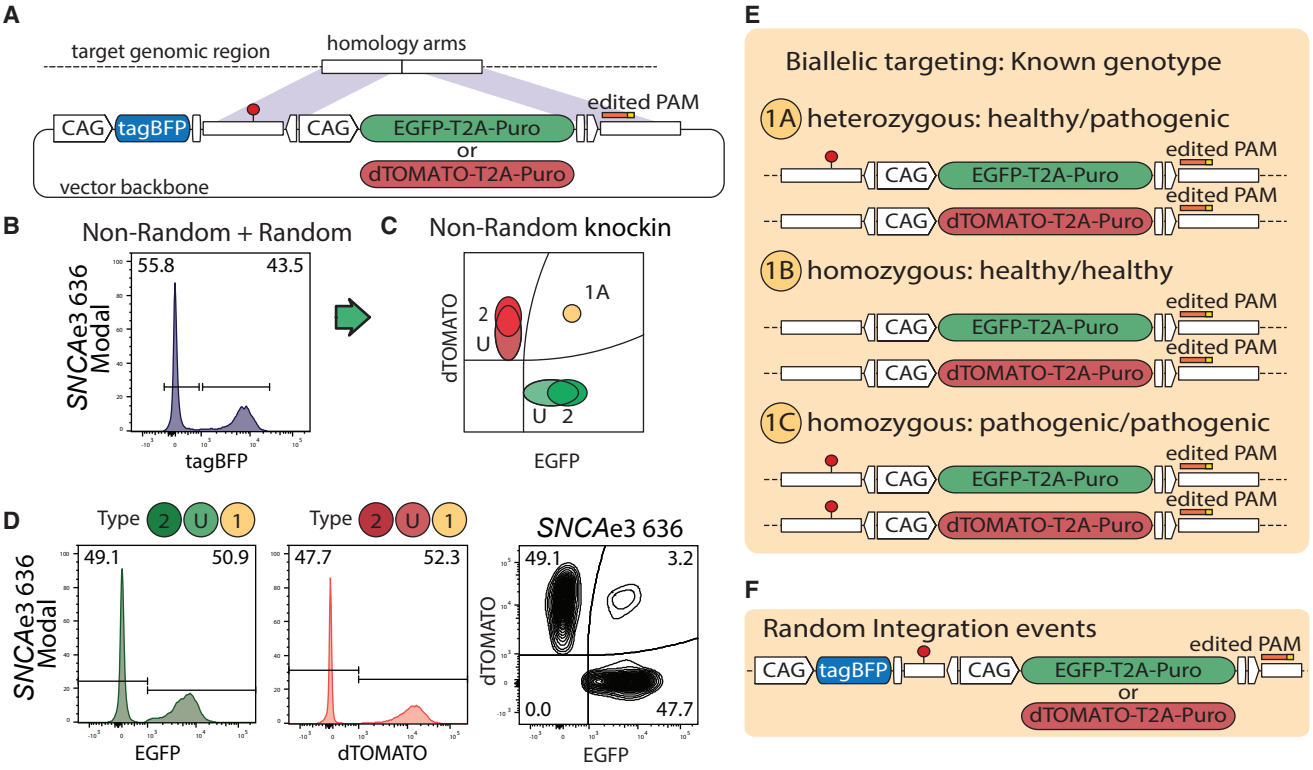


Figure 1. Biallelic Integration of FP-SNP Pairs Enable Deterministic Genotype Outcomes

(A) Donor vectors contain a PSM expressing EGFP or dTOMATO, and an NSM expressing tagBFP. PSMs contain puromycin resistance gene (Puro).

(B) Representative example of *SNCA*e3 polyclone 636. Random integration tagBFP^{pos} cells are excluded from the correctly edited on-target cells tagBFP^{neg}.

(C) Theoretical distribution of populations for non-random outcomes.

(D) Representative example of *SNCA*e3. On-target cells include homozygous populations, EGFP^{pos}/EGFP^{pos} or dTOMATO^{pos}/dTOMATO^{pos} (type 2), and heterozygous populations of undefined second-allele state EGFP^{pos}/WT-NHEJ or dTOMATO^{pos}/WT-NHEJ (type U). WT, wild-type.

(E) Outcomes of the derived population are defined according to the donor vector design.

(F) The tagBFP NSM allows removal of random integration events, assisting in the derivation of defined outcomes.

an EGFP donor carried the transversion c.88g>c. For *SNCA*e3, an EGFP donor carried the transition c.209g>a. For each locus, a corresponding dTOMATO donor carried the healthy variant, as shown for population type 1A (Figure 1E). A similar expression level of the FP reporters was observed from each allele in *SNCA* chromosome 4, as demonstrated by a symmetric fluorescence-activated cell sorting (FACS) analysis (Figure 1D). In order to test whether similar PSM expression levels are observed in other loci, the gene *PINK1* exon 5 of chromosome 1 was targeted. In contrast to *SNCA* mutant PD patients, *PINK1* PD patients are homozygote or compound heterozygote (Ishihara-Paul et al., 2008). Hence, for *PINK1*e5, both donors, EGFP and dTOMATO, carried the pathogenic transversion c.1197t>a, matching population type 1C (Figure 1E). FACS analysis showed that biallelic targeted populations separated clearly from other genotype outcomes for *PINK1* chromosome 1 (Figure S1C) and *SNCA* chromosome 4 (Figure 1D). These results validate the approach to target both alleles of a gene of interest, independent of the locus.

Repetitive Elements Reduce On-Target Genome Editing Efficiency by Increasing Random Integration

Silent point mutations were introduced in protospacer adjacent motif (PAM) sequences of the donors (Table S2). The PAM edited template is resistant to Cas9-induced linearization, avoiding linear DNA-induced random integration. Thus, properly targeted alleles are shielded from Cas9-induced secondary incisions, eliminating the risk of on-target indels (Merkle et al., 2015) (Table S2). Two weeks after electroporation, each edited population was expanded up to 15 × 10^6 puromycin-resistant and FP-positive cells. The

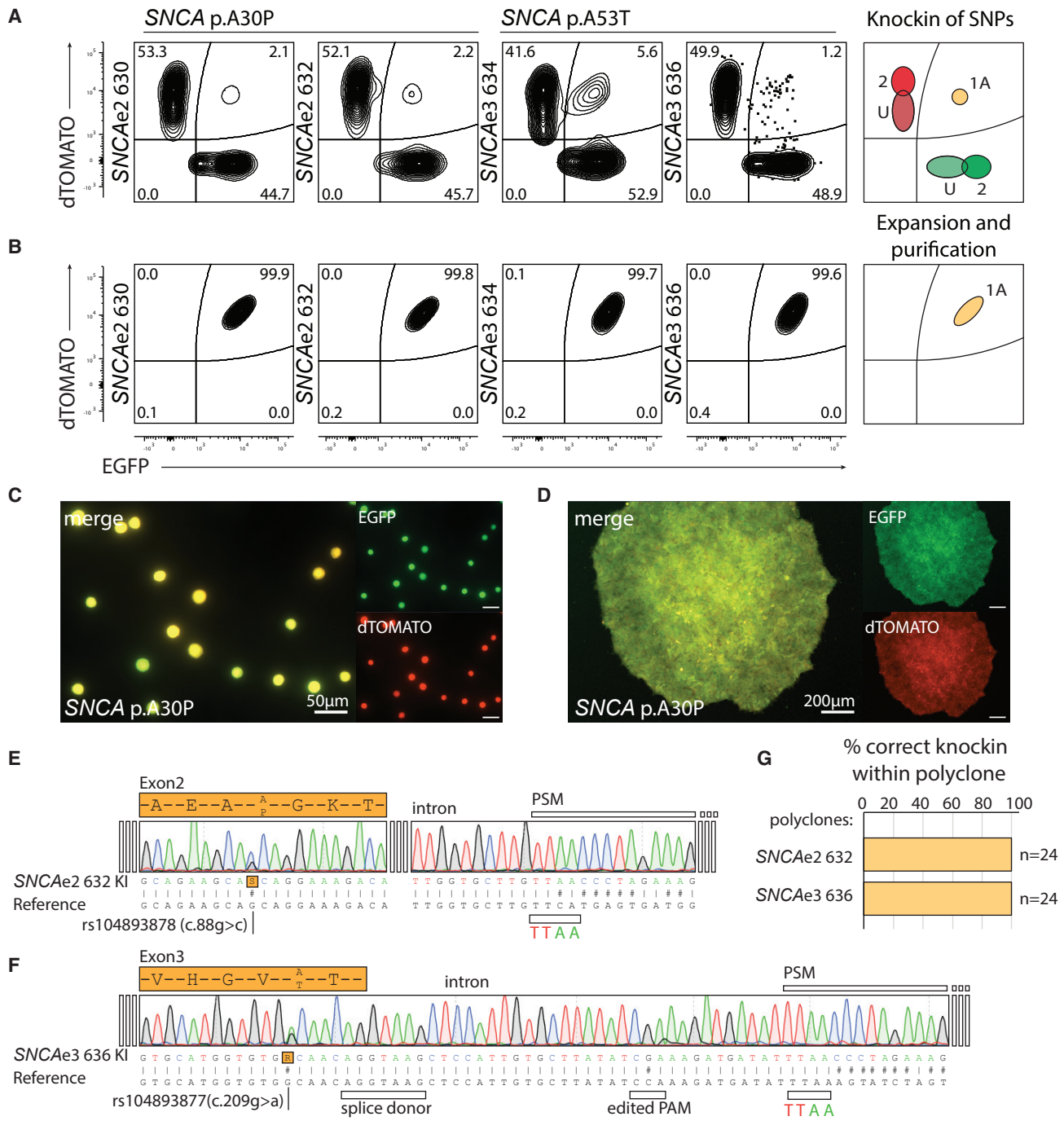


Figure 2. FACS Purification Increases the Speed and Yield of Isogenic Derivation

(A) Post-selection sorting of double-positive biallelic edited cells for *SNCA*e2 and *SNCA*e3 using independent sgRNAs. FACS plots are represented with 2% contour lines. For *SNCA*e3 sgRNA-636, a dotplot is included to show the distribution of 1.2%. Diagram of knockin population types is shown (right).

(B) Yield-purity and purity-purity sorting strategies permit the generation of a homogeneous biallelic knockin population. Diagram of purification population types is shown (right).

(C) Representative post-sorting of single cells for *SNCA* polyclone. Single-cell gating structures yields high-purity biallelic edited cells. Scale bar, 50 µm.

(legend continued on next page)

inclusion of tagBFP in the NSM allowed random integration events to be quantified, visualized, and excluded (Figures 1A, 1B, and 1F). The tagBFP NSM avoids bystander toxicity or incomplete negative selection from systems such as thymidine kinase (Ruby and Zheng, 2009). The percentage of tagBFP^{pos} random integration ranged from 5.8% to 14.6% for *SNCA*e2, from 42.6% to 64.2% for *SNCA*e3, and from 27.2% to 30.4% for *PINK1*e5 (Figures S2A–S2C). The extent of random integration correlated with the type and proportion of repetitive elements present in the homology arms of the donors. We assessed random integration using donors for six loci with known repetitive element composition and tested 12 single-guide RNAs (sgRNAs). The loci evaluated included chromosome 1 (*PINK1* exon 5), chromosome 4 (*SNCA* exon 2 and exon 3), and chromosome 16 (ceroid lupofuscinosis 3, *CLN3* exon 5–8, exon 10–13, and exon 14–15) (Figure S2G). For the analysis, we performed a linear optimization model of the form $Ax = b$ (Figure S2H). The resulting matrix A corresponds to the frequency of repetitive elements in the homology arms (Figure S2G). The vector x corresponds to the type of repetitive elements present in the analyzed dataset (Figure S2G) and a variable of all non-included repetitive elements (upsilon). The vector b corresponds to the experimentally measured random integration level, given by the percentage of tagBFP^{pos} cells (Figures S2A–S2F). Based on this, we derived a model to predict random integration frequency intrinsic to the composition of repetitive elements in the homology arms (Figures S2H and S2I). The solution allows assigning weight coefficients to each repetitive element. Their value indicates which repetitive element contributes the most to the random integration frequency observed. The solution space is constrained for a maximum of 100% random integration and sequence length boundaries of each repetitive element. The optimization solution indicates that the most relevant repetitive elements correspond to the Short Interspersed Nuclear Elements (SINE) family, specifically Alu and Mir (Figures S2H and S2I).

FACS Purification Increases the Speed and Yield of Isogenic Derivation

For the on-target tagBFP^{neg} cells, the ratio of EGFP to dTOMATO was ~50% in all cases analyzed, which is consistent with a comparable efficiency for both donors (Figures 2A and S1C). The initial percentage of double-positive EGFP^{pos}/dTOMATO^{pos} cells ranged from a mean 2.15% for *SNCA*e2, 3.4% for *SNCA*e3, to 3.75% for *PINK1*e5 (Figures 2A and S1C). Quantifications were conducted independently using different sgRNAs (Figures 2A and S1C). One sorting step yielded a population of up to 3×10^5 EGFP^{pos}/dTOMATO^{pos} cells (Figures 2C and 2D). The gating position of the double-positive population afforded nearly complete purity with either purity-purity or yield-purity sorting masks (Figures 2B and S1D). Although it is possible to isolate single-channel double-positive EGFP$^{pos/pos}$ or dTOMATO$^{pos/pos}$ populations (type 2) (Figure 2A) using the FSC-A dimension, there is an extensive overlap with the indel-bearing single-positive population (type U) (Figures S1G–S1J). A high frequency of non-homologous end joining (NHEJ) events was detected in the non-targeted allele of the single-positive population (type U) (Figures S1J and S1K). Hence, purification of the double-positive EGFP$^{pos/pos}$ or dTOMATO$^{pos/pos}$ populations presents the risk of co-purifying overlapping indel-bearing cells (Figures S1J and S1K). In this combination of events, only the biallelic EGFP^{pos}/dTOMATO^{pos} group offers a deterministic genotype outcome. Sanger sequencing of biallelic targeted *SNCA* mutations demonstrated the heterozygous integration of the pathogenic SNP rs104893878 (p.A30P) and rs104893877 (p.A53T) in each polyclone (Figures 2E and 2F), the homozygous integration of the edited PAM, and the transition from genome to PSM (Figures 2E and 2F). Sequencing-isolated single clones from the polyclonal populations permitted composition analysis (Figures 2G, S3F, and S3G and Table S3).

Transposase-Mediated Generation of Footprint-free Isogenic Lines

The PSMs in each double-positive polyclone were excised using a codon-optimized hyperactive and excision-only variant of the piggybac transposase (Li et al., 2013b; Yusa et al., 2011) (Figures 3A and 3B). Even though the excision-only variant presents an activity of 0.85 times that of wild-type (Figures S1M and S1N), it is preferred as it lacks the reintegration cycle of wild-type variants (Li et al., 2013a). The heterozygous *SNCA*e2 and *SNCA*e3 EGFP^{pos}/dTOMATO^{pos} polyclonal populations were transfected with *in vitro* transcribed mRNA encoding excision-only transposase. Subsequently, the excised EGFP^{neg}/dTOMATO^{neg} population was sorted (Figures 3A and 3B). Using the excision-only variant and two transfection steps, we observed average excision efficiencies of 3.65% for *SNCA*e2, 2.15% for *SNCA*e3, and 6.5% for *PINK1*e5 (Figures 3A and S1E). A second sorting step to purify cells that underwent selection module removal yielded up to

(D) Representative post-sorting culture for biallelic EGFP^{pos}/dTOMATO^{pos} *SNCA* polyclone. Scale bar, 200 μm.
(E) Sanger sequencing chromatogram of *SNCA*e2 p.A30P polyclone 632 knockin (KI).
(F) Sanger sequencing chromatogram of *SNCA*e3 p.A53T polyclone 636 knockin.
(G) Analysis of the polyclone composition as in Figure S3.

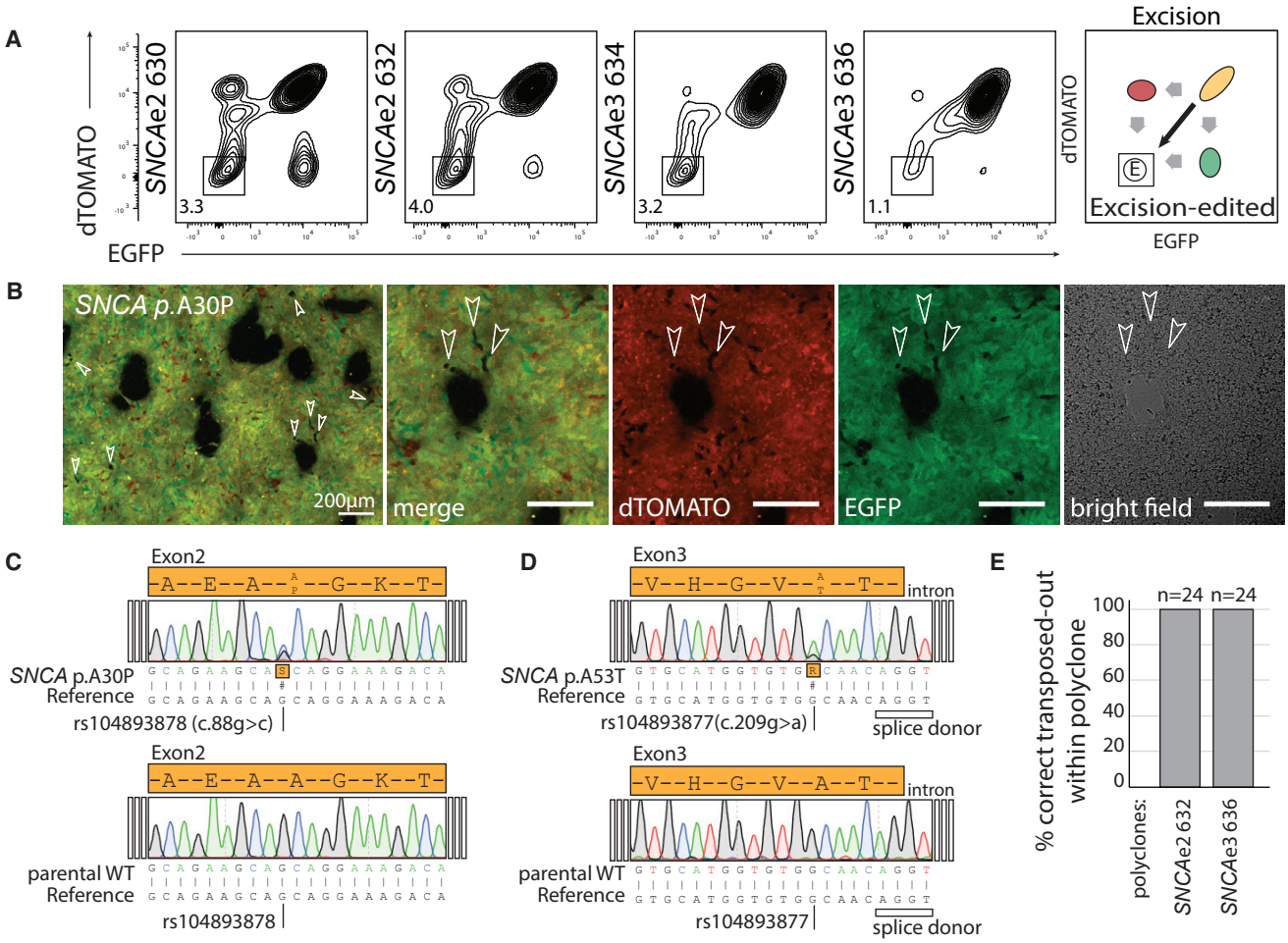


Figure 3. Transposase-Mediated Excision of PSMs

(A) FACS analysis for PSM removal. Two transfection steps of excision-only transposase result in removal of the PSMs for *SNCA*e2 and *SNCA*e3. Purification of EGFP^{neg}/dTOMATO^{neg} cells yields footprint-free edited lines. Diagram of removal population types is shown (right).

(B) Cultures after transposase transfection for *SNCA*e2 present single and double PSM removal events (in arrowheads) as shown in (A). Scale bar, 200 µm.

(C) Sanger sequencing chromatogram of transposed *SNCA*e2 p.A30P polyclone 632 and parental control.

(D) Sanger sequencing chromatogram of transposed *SNCA*e3 p.A53T polyclone 636 and parental control.

(E) Analysis of the respective polyclone composition as in Figure S3.

2.5 × 10^6 EGFP^{neg}/dTOMATO^{neg} SNP knocked in cells. In the FACS analysis, it is possible to observe transition states for single-copy excision and complete removal of both selection modules (Figures 3A, 3B, and S1E). We observed a curved population shifting from the double-positive EGFP^{pos}/dTOMATO^{pos} quadrant to the double-negative EGFP^{neg}/dTOMATO^{neg} quadrant in all cases. Sanger sequencing of the *SNCA* targeted and transposed genomic region demonstrated the heterozygous integration of the pathogenic SNP rs104893878 (p.A30P) and rs104893877 (p.A53T) in each polyclone (Figures 3C and 3D). Isolation of single-cell-derived clones from the polyclones and sequencing permitted quantification of their composition (Figures 3E, S3H, and S3I and Table S3). The polyclone composition analysis demonstrated that PSMs were excised and the edited SNPs and edited PAM sites remained in the non-coding sequence (Figures S3H and S3I). Karyotype assessment was conducted for each polyclone and parental control (Figures S4A–S4C). Pluripotency of lines was assessed by immunostaining for OCT4, SOX2, TRA1-81, and SSEA4 (Figures S4D–S4F).

SNCA Mutants Present Early Mitochondrial Impairment

In order to validate the edited *SNCA* lines, a phenotypic characterization was conducted (Figure 4). Isogenic

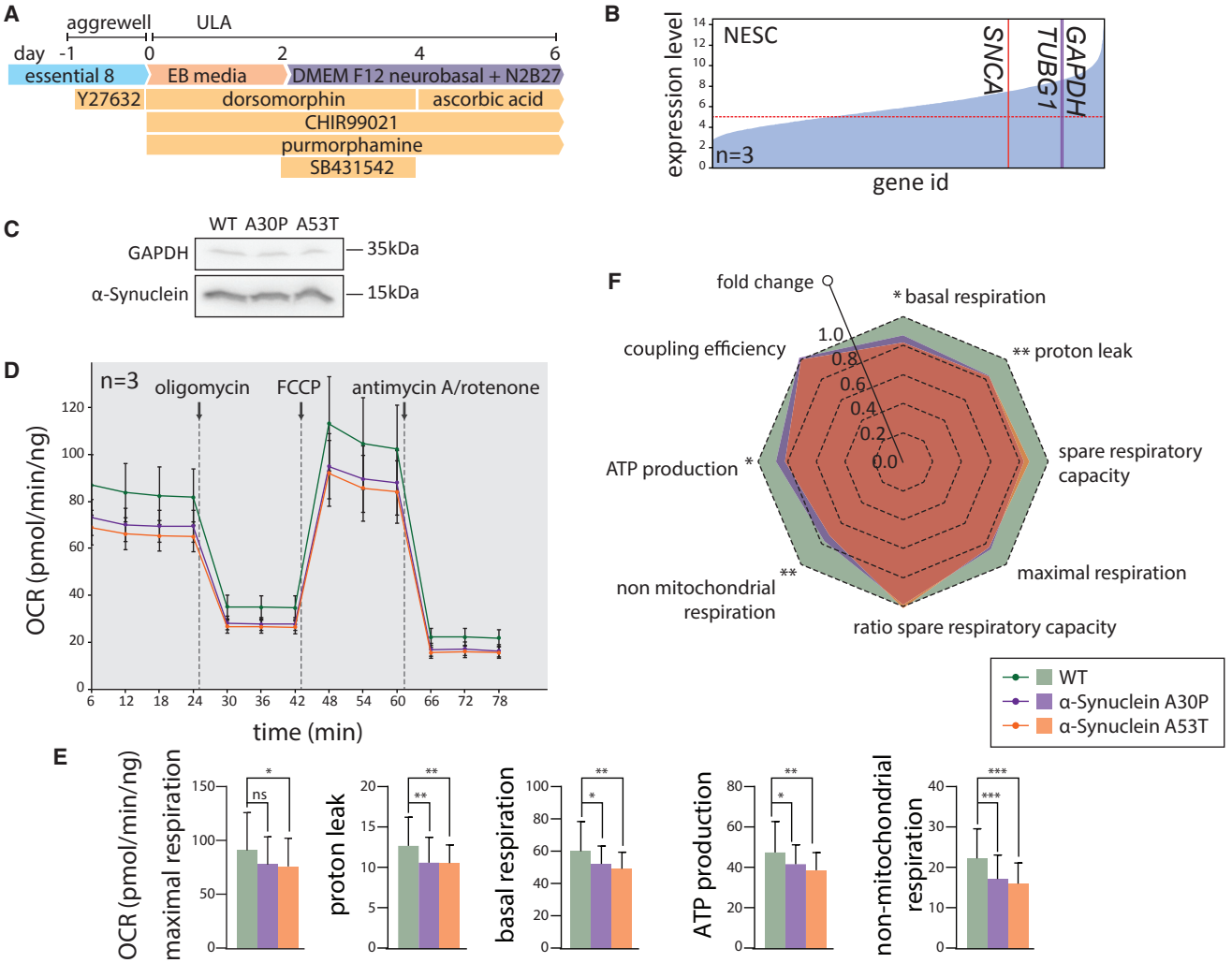


Figure 4. Edited *SNCA* Isogenic Lines Present PD-Associated Phenotypes

(A) NESC differentiation protocol.

(B) Microarray expression level for *SNCA*, *TUBG1*, and *GAPDH* in healthy control NESCs. Data represent three replicates.

(C) Western blot subsequent to denaturing SDS-PAGE for α-synuclein and GAPDH for NESCs.

(D) Wave plot of oxygen consumption rates for the α-synuclein isogenic set. Each wave corresponds to three biological replicates. SD of the sample is included.

(E) Maximal respiration, proton leak, basal respiration, ATP production, and non-mitochondrial respiration for the extracellular energy flux analysis in (D).

(F) Radar plot of fold changes for the parameters in (E). Significance levels correspond to the higher p value assigned to a mutant per category. Significance determined by unpaired Student's t-test.

Significance levels are *p < 0.05, **p < 0.01 and ***p < 0.001; n.s., not significant.

induced pluripotent stem cells (iPSCs) were differentiated into neuroepithelial stem cells (NESCs) (Reinhardt et al., 2013) (Figures 4A, S4G, and S4H). NESCs typically express the *SNCA* transcript at 0.86 and 0.7 times the level of *GAPDH* and *TUBG1*, respectively (Figures 4B and S4I). Western blot analysis indicated a similar protein level of monomeric α-synuclein for all genotypes (Figure 4C). Extracellular energy flux analyses were conducted for parental healthy NESCs, and mutant isogenic α-synuclein p.A30P and p.A53T NESCs (Figure 4D). Cells expressing the α-synuclein mutation p.A53T showed a significantly reduced maximal respiration capacity compared with the parental isogenic control (Figures 4D–4F). Moreover, both the p.A30P and p.A53T α-synuclein mutant NESCs showed comparatively reduced energy performance, manifested by a lower basal

respiration, ATP production, and non-mitochondrial respiration (Figures 4D–4F).

DISCUSSION

Overall, FACE constitutes a robust method to achieve deterministic genotype outcomes for the generation of isogenic cell lines. The selection of biallelic editing events ensures a defined genotype. It should be noted that, due to transient disruption of the coding sequence, this approach is restricted to genes with non-essential function in the target cell type. The use of fluorescent NSM excludes random integration events, enabling clearer sorting gates and isolated biallelic populations. This constitutes an advancement over similar approaches (Eggenschwiler et al., 2016). However, potential limitations are that PSMs could be subjected to position-effect variegation or promotor silencing. Nevertheless, usage of the FP markers expedites the selection, reducing the timescale in comparison with potential position-effect variegation (Norrman et al., 2010). It should be noted that editing approaches that use single-stranded DNA (ssDNA) or dsDNA could be subjected to cleavage within non-functional or functional sequences. Hence, donor break points within PSM cannot be fully excluded. The advantage of dsDNA approaches, in comparison with ssDNA, are their flexibility to carry larger cargos in order to deposit designer insertions, designer deletions, or PSMs. In addition, larger sequences of donors are easier to detect by conventional methods in comparison with short ssDNA. Similarly, potential imperfect integration of dsDNA donor templates can be readily detected by simple methods such as PCR, in comparison with ssDNA-based methods.

Conventional derivation of single nucleotide mutations, not associated with a direct selection phenotype or selection marker, can require screening an average of 911 ± 375 clones and using 8.8 ± 5.9 sgRNAs. Conversely, early elimination of undesirable outcomes obviates the need to perform extensive colony screening and results in a faster, more efficient derivation process. Thus, FACE constitutes an attractive alternative to conventional methods. The efficiency of homology-directed repair is influenced by the length of the homology arms used (Hasty et al., 1991). We and others have used homology arms of ~1 kbp, which provides a balance between efficiency and specificity (Soldner et al., 2011). The sequence conversion from an endogenous sequence to that carried in donor templates extends from ~400 bp in dsDNA (Elliott and Jasin, 2001) to ~30 bp in ssDNA donors (Paquet et al., 2016). Hence, it is of critical importance to include the edited bases close to the dsDNA break point and close to the PSM unit, independently of the length of the homology arms or the type of template used. Post-knockin and post-transposition clonal composition analysis confirmed that FACE enables the derivation of polyclones and significantly reduces the screening efforts if individual clones are needed. On the other hand, the derivation of edited polyclones presents the advantage of avoiding the risk inherent with clone-specific biases. Extensive expansion, required for clonal derivation, is reported to subject cells to culture aberrations (Martins-Taylor and Xu, 2012). It is widely accepted that single-cell passaging for any type of cell-culture application, including the process of FACS-based enrichment described here, imposes an unavoidable risk of genome instability (Chan et al., 2008). The derivation of polyclones reduces the culture time needed for each step, since sufficient material is available earlier. Karyotype analysis of the edited lines demonstrated that the process did not induce chromosomal abnormalities when compared with the parental line. Previous reports also support the possibility of achieving a low incidence of modification with genome editing tools (Tsai and Joung, 2014; Veres et al., 2014).

In order to protect the dsDNA donor template from Cas9-induced linearization and to avoid post-integration cleavage of targeted sequences, we introduced silent mutations in the PAM sequences. This requires special attention to the design in order to introduce the edited PAM in a non-coding sequence or as a synonymous mutation. We and others have successfully used this mechanistic insight to protect post-integration targeted sequences from secondary cleavage events (Inui et al., 2014; Paquet et al., 2016). Similarly, design considerations are needed to identify adjacent transposase excision sequences, or to generate a *de novo* TTAA sequence in non-coding regions or by silent editions. Protocol optimization for the use of an excision-only transposase variant (Li et al., 2013b) allowed the derivation of footprint-free isogenic sets for disease modeling. We were able to observe transition states that represent the removal of one or both PSMs. The transition populations presented a curve pattern that accounts for the dissimilar stability of the FPs (Snapp, 2009) and transcripts after the CDS module was removed.

The influence of repetitive elements on the efficiency of genome editing has been reported previously (Ishii et al., 2014). Recognizable repetitive elements constitute up to 45% of the human genome (Lander et al., 2001). Repetitive elements in humans can be classified in four families: SINE, LINE, LTR retrotransposons, and DNA transposons. Each category presents multiple sub-families. Using linear optimization modeling, we determined that, in our dataset, the repetitive elements of the SINE family, Alu and Mir, contribute the most to random integration events. These repetitive elements have 1.5 million copies and constitute ~13% of the human genome (Lander et al., 2001). Although this discrete dataset does not include all existing human repetitive

elements, it demonstrates their direct contribution to random integration. Other aspects, such as the composition of repetitive elements and distance to the dsDNA break point, might modulate the frequency of random integration. Our data confirm previous reports that repetitive elements act as templates for off-target homologous recombination (Ishii et al., 2014). These sequences should be avoided when designing homology arms in order to enhance on-target recombination and edition.

In summary, we generated an isogenic set of human *SNCA* mutants for PD-specific cellular modeling. The set carries disease-associated mutations p.A30P or p.A53T in the *SNCA* gene. We observed energy metabolism phenotypes in human NESCs, an early neurodevelopment disease model. Such traits have been previously described in *SNCA* p.A30P mutant differentiated neurons (Ryan et al., 2013). This validates the applicability of the approach described here for the generation of disease-relevant models. We envision that FACE could be efficiently implemented for automated high-throughput genome editing, enabling fast phenotype assessment in the future.

EXPERIMENTAL PROCEDURES

Stem cells A13777 were cultured in Essential 8 medium on Geltrex. Cells were passed with accutase and plated with Y27632 (10 μM) for 24 hr after dissociation. Cells were electroporated using 4D-Nucleofector. Selection was conducted with puromycin (0.5 μg/mL). FACS was conducted on an ARIA III sorter. Cells were purified with single-cell exclusive gating. Post-knockin cells were transfected with *in vitro* transcribed mRNA coding transposase. Human iPSCs were characterized for OCT4, SOX2, TRA1-81, and TRA1-61. Microarray karyotype was conducted using Illumina iScan technology. NESCs were differentiated as represented in Figure 4A. NESCs were characterized for NESTIN and SOX2. Transcription levels for NESCs were evaluated using Affymetrix human gene arrays (GEO: GSE101534). Extracellular energy flux analysis was conducted on NESCs using a Seahorse XFe96 assay as indicated in Figure 4D. Comprehensive information on the experimental procedures is described in the Supplemental Information.

ACCESSION NUMBERS

The accession number for the gene expression data (Figure 4B) reported in this paper is GEO: GSE101534.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, four figures, and four tables and can be found with this article online at http://dx.doi.org/10.1016/j.stemcr.2017.08.026.

AUTHOR CONTRIBUTIONS

J.A.-F., J.J., X.Q., G.G.-G., J.W., and S.L.N. designed the study and conducted the experiments. J.C.S., H.Z., and H.R.S. supervised. J.A.-F., J.J., and J.C.S. wrote the manuscript and organized the display items. All the authors read and agreed to the final version of the manuscript.

ACKNOWLEDGMENTS

We would like to thank Prof. F. Zhang from the McGovern Institute for Brain Research for providing the Cas9 vector. We acknowledge G. Preciat for valuable feedback on optimization. We acknowledge Prof. J. Hejna, E. Berger, and S. Bolognin for their valuable comments on the manuscript. This project was supported by the LCSB Pluripotent Stem Cell Core Facility. J.J., J.W., and X.Q. were supported by fellowships from the FNR (AFR, Aides à la Formation-Recherche). G.G.-G. was supported by NCL-Stiftung. J.J. is supported by a Pelican award from the Fondation du Pelican de Mie et Pierre Hippert-Faber. This is an EU Joint Programme-Neurodegenerative Disease Research (JPND) project (INTER/JPND/14/02; INTER/JPND/15/11092422). Further support comes from the SysMedPD project, which has received funding from the European Union's Horizon 2020 Research and Innovation Program under grant agreement no. 668738. Jens Schwamborn is founder, CSO, and shareholder of Braingineering Technologies Sarl. J.C.S., J.A.F., J.J., and X.Q. are inventors on the patent (LU92964).

Received: February 3, 2017
Revised: August 31, 2017
Accepted: August 31, 2017
Published: October 5, 2017

REFERENCES

Bendor, J.T., Logan, T.P., and Edwards, R.H. (2013). The function of alpha-synuclein. Neuron *79*, 1044–1066.

Bozi, M., Papadimitriou, D., Antonellou, R., Moraitou, M., Maniati, M., Vassilatis, D.K., Papageorgiou, S.G., Leonardos, A., Tagaris, G., Malamis, G., et al. (2014). Genetic assessment of familial and early-onset Parkinson's disease in a Greek population. Eur. J. Neurol. *21*, 963–968.

Caligiore, D., Helmich, R.C., Hallett, M., Moustafa, A.A., Timmermann, L., Toni, I., and Baldassarre, G. (2016). Parkinson's disease as a system-level disorder. NPJ Parkinsons Dis. *2*, 16025.

Chan, E.M., Yates, F., Boyer, L.F., Schlaeger, T.M., and Daley, G.Q. (2008). Enhanced plating efficiency of trypsin-adapted human embryonic stem cells is reversible and independent of trisomy 12/17. Cloning Stem Cells *10*, 107–118.

Devine, M.J., Ryten, M., Vodicka, P., Thomson, A.J., Burdon, T., Houlden, H., Cavaleri, F., Nagano, M., Drummond, N.J., Taanman, J.W., et al. (2011). Parkinson's disease induced pluripotent stem cells with triplication of the alpha-synuclein locus. Nat. Commun. *2*, 440.

Eggenschwiler, R., Moslem, M., Fraguas, M.S., Galla, M., Papp, O., Naujock, M., Fonfara, I., Gensch, I., Wahner, A., Beh-Pajooh, A., et al. (2016). Improved bi-allelic modification of a transcriptionally silent locus in patient-derived iPSC by Cas9 nickase. Sci. Rep. *6*, 38198.

Elliott, B., and Jasin, M. (2001). Repair of double-strand breaks by homologous recombination in mismatch repair-defective mammalian cells. Mol. Cell. Biol. *21*, 2671–2682.

Goedert, M., Spillantini, M.G., Del Tredici, K., and Braak, H. (2013). 100 years of Lewy pathology. Nat. Rev. Neurol. *9*, 13–24.

Hasty, P., Rivera-Perez, J., and Bradley, A. (1991). The length of homology required for gene targeting in embryonic stem cells. Mol. Cell. Biol. *11*, 5586–5591.

Inui, M., Miyado, M., Igarashi, M., Tamano, M., Kubo, A., Yamashita, S., Asahara, H., Fukami, M., and Takada, S. (2014). Rapid generation of mouse models with defined point mutations by the CRISPR/Cas9 system. Sci. Rep. *4*, 5396.

Ishihara-Paul, L., Hulihan, M.M., Kachergus, J., Upmanyu, R., Warren, L., Amouri, R., Elango, R., Prinjha, R.K., Soto, A., Kefi, M., et al. (2008). PINK1 mutations and parkinsonism. Neurology *71*, 896–902.

Ishii, A., Kurosawa, A., Saito, S., and Adachi, N. (2014). Analysis of the role of homology arms in gene-targeting vectors in human cells. PLoS One *9*, e108236.

Kleinstiver, B.P., Pattanayak, V., Prew, M.S., Tsai, S.Q., Nguyen, N.T., Zheng, Z., and Joung, J.K. (2016). High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. Nature *529*, 490–495.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., et al. (2001). Initial sequencing and analysis of the human genome. Nature *409*, 860–921.

Li, M.A., Pettitt, S.J., Eckert, S., Ning, Z., Rice, S., Cadinanos, J., Yusa, K., Conte, N., and Bradley, A. (2013a). The piggyBac transposon displays local and distant reintegration preferences and can cause mutations at noncanonical integration sites. Mol. Cell. Biol. *33*, 1317–1330.

Li, X., Burnight, E.R., Cooney, A.L., Malani, N., Brady, T., Sander, J.D., Staber, J., Wheelan, S.J., Joung, J.K., McCray, P.B., Jr., et al. (2013b). piggyBac transposase tools for genome engineering. Proc. Natl. Acad. Sci. USA *110*, E2279–E2287.

Martins-Taylor, K., and Xu, R.H. (2012). Concise review: genomic stability of human induced pluripotent stem cells. Stem Cells *30*, 22–27.

Merkle, F.T., Neuhausser, W.M., Santos, D., Valen, E., Gagnon, J.A., Maas, K., Sandoe, J., Schier, A.F., and Eggan, K. (2015). Efficient CRISPR-Cas9-mediated generation of knockin human pluripotent stem cells lacking undesired mutations at the targeted locus. Cell Rep. *11*, 875–883.

Norrman, K., Fischer, Y., Bonnamy, B., Wolfhagen Sand, F., Ravassard, P., and Semb, H. (2010). Quantitative comparison of constitutive promoters in human ES cells. PLoS One *5*, e12413.

Paquet, D., Kwart, D., Chen, A., Sproul, A., Jacob, S., Teo, S., Olsen, K.M., Gregg, A., Noggle, S., and Tessier-Lavigne, M. (2016). Efficient introduction of specific homozygous and heterozygous mutations using CRISPR/Cas9. Nature *533*, 125–129.

Reinhardt, P., Glatza, M., Hemmer, K., Tsytsyura, Y., Thiel, C.S., Hoing, S., Moritz, S., Parga, J.A., Wagner, L., Bruder, J.M., et al. (2013). Derivation and expansion using only small molecules of human neural progenitors for neurodegenerative disease modeling. PLoS One *8*, e59252.

Ruby, K.M., and Zheng, B. (2009). Gene targeting in a HUES line of human embryonic stem cells via electroporation. Stem Cells *27*, 1496–1506.

Ryan, S.D., Dolatabadi, N., Chan, S.F., Zhang, X., Akhtar, M.W., Parker, J., Soldner, F., Sunico, C.R., Nagar, S., Talantova, M., et al. (2013). Isogenic human iPSC Parkinson's model shows nitrosative stress-induced dysfunction in MEF2-PGC1alpha transcription. Cell *155*, 1351–1364.

Snapp, E.L. (2009). Fluorescent proteins: a cell biologist's user guide. Trends Cell Biol. *19*, 649–655.

Soldner, F., Laganiere, J., Cheng, A.W., Hockemeyer, D., Gao, Q., Alagappan, R., Khurana, V., Golbe, L.I., Myers, R.H., Lindquist, S., et al. (2011). Generation of isogenic pluripotent stem cells differing exclusively at two early onset Parkinson point mutations. Cell *146*, 318–331.

Tsai, S.Q., and Joung, J.K. (2014). What's changed with genome editing? Cell Stem Cell *15*, 3–4.

Veres, A., Gosis, B.S., Ding, Q., Collins, R., Ragavendran, A., Brand, H., Erdin, S., Cowan, C.A., Talkowski, M.E., and Musunuru, K. (2014). Low incidence of off-target mutations in individual CRISPR-Cas9 and TALEN targeted human stem cell clones detected by whole-genome sequencing. Cell Stem Cell *15*, 27–30.

Yusa, K., Zhou, L., Li, M.A., Bradley, A., and Craig, N.L. (2011). A hyperactive piggyBac transposase for mammalian applications. Proc. Natl. Acad. Sci. USA *108*, 1531–1536.

Stem Cell Reports, Volume *9*

Supplemental Information

FACS-Assisted CRISPR-Cas9 Genome Editing Facilitates Parkinson's Disease Modeling

Jonathan Arias-Fuenzalida, Javier Jarazo, Xiaobing Qing, Jonas Walter, Gemma Gomez-Giro, Sarah Louise Nickels, Holm Zaehres, Hans Robert Schöler, and Jens Christian Schwamborn

FACS assisted CRISPR-Cas9 genome editing facilitates Parkinson's disease modeling

- Supplemental Information -

Jonathan Arias-Fuenzalida [1,2,7], Javier Jarazo [1,7], Xiaobing Qing [1], Jonas Walter [1] Gemma Gomez-Giro [1,4], Sarah Louise Nickels [1,3], Holm Zaehres [4,5], Hans Robert Schöler [4,6], Jens Christian Schwamborn [1]

[1] Luxembourg Centre for Systems Biomedicine (LCSB), Developmental and Cellular Biology, University of Luxembourg, L-4362, 7 avenue des Hauts-Fourneaux, Luxembourg

[2] Graduate School of Biostudies, Kyoto University, Kyoto 606-8502, Japan

[3] Life Science Research Unit (LSRU), Systems Biology, University of Luxembourg, L-4367, 6 avenue du swing, Luxembourg

[4] Max Planck Institute for Molecular Biomedicine, Laboratory of Cell and Developmental Biology, Roentgenstrasse 20, Muenster, Germany

[5] Ruhr-University Bochum, Medical Faculty, Department of Anatomy and Molecular Embryology, 44801 Bochum, Germany

[6] Westphalian Wilhelms University Muenster, Medical Faculty, 48149 Muenster, Germany

[7] These authors equally contributed to the article

Correspondence should be addressed to J.S. (jens.schwamborn@uni.lu)

Supplemental Experimental procedures

Stem cell culture and electroporation. The following human induced pluripotent stem (iPS) cells reprogrammed with non-integrative episomal methods were used: A13777 (Gibco cat no. A13777) from female cord blood-derived $CD34^{pos}$ cells. Cell lines were cultured in Essential 8 medium (Thermo Fisher cat no. A1517001) on Geltrex (Thermo Fisher cat no. A1413301) or matrigel. Cells were normally dissociated with accutase (Thermo Fisher cat no. A1110501) and plated in media containing ROCK inhibitor Y27632 (Sigma cat no. Y0503) at 10µM for 24h after dissociation. Cells were subjected to positive selection with puromycin (Sigma cat no. P9620) at a concentration of 0.5µg/mL. Cells were electroporated using 4D-Nucleofector System (Lonza) and a 4D kit for human dermal fibroblast (Lonza cat no. V4XP). Parental pre-electroporation line presents micro-duplication 20q11.21.

Construction of sgRNA vectors and donor plasmids. Cas9 target sequences with predicted high catalytic activity were selected (Doench et al., 2014) (Table S2) and cloned into pX330 vector (Addgene 42230) as previously described (Ran et al., 2013). Primers used are indicated in Table S4. The donor vectors were pDONOR-SNCAe2-WT (Addgene 85845), pDONOR-SNCAe2-A30P (Addgene 85846), pDONOR-SNCAe3-WT (Addgene 85847), pDONOR-SNCAe3-A53T (Addgene 85848) and pDONOR-PINK1e5-I368N (Addgene 86154) in EGFP and dTOMATO containing versions. Homology arms were assembled by conventional methods (Gibson, 2011) on donor scaffolds pDONOR-tagBFP-PSM-EGFP (Addgene 100603) and pDONOR-tagBFP-PSM-dTOMATO (Addgene 100604).

In vitro RNA transcription and mRNA transfection. The coding sequence of codon-optimized hyperactive transposase Piggybac from Trichoplusia ni (Yusa et al., 2011) and the excision-only mutant (R372A/K375A) (Li et al., 2013) were amplified to incorporate the T7 promoter. Primers used are indicated in Table S4. The PCR product was used as template for in vitro transcription with an mMESSAGE mMACHINE T7 kit (Thermo Fisher cat no. AM1344) according to the manufacturer's protocol. The transcript was poly-adenylated with a Poly(A) tailing kit (Thermo Fisher cat no. AM1350) and purified with a MEGAclear transcription clean-up kit (Thermo Fisher cat no. AM1908). The transcript quality was evaluated with a Bionalayzer RNA 6000 nano (Aglient cat no. 5067-1511). Transfection was performed with Stemfect RNA transfection kit (Stemgent cat no. 00-0069) according to the manufacturer's protocol.

Fluorescent Activated Cell Sorting. FACS was conducted using sterile line sorting on a baseline and CST calibrated BD FACS ARIA III. Drop delay calibrations were ensured prior to each sample. For all human iPS cells an 85µm nozzle, a yield or purity-sorting mask and neutral density filter 2.0 were used. Cells were pre-separated with 35µm and 20µm strainers (Corning cat no. 352235 and Miltenyi cat no. 130-101-812). Sorting was conducted with single cell exclusive gating hierarchies on FSC and SSC wide and high (Figure S1A). Use of strainers and single cell gating is highly recommended (Figure S1B). For efficiency analysis, live cells were quantified by SYTOX Blue Dead Cell Stain (Thermo Fisher cat no. S34857).

Characterization of polyclones. Composition of polyclones was assessed by sub-cloning. Single cell clones were expanded and genomic DNA extracted using QuickExtract solution (Epicentre cat no. QE09050). Clones were genotyped for the left homology arm junction, right homology arm junction, and wild type junction as indicated in Figure S3, using primers in Table S4. PCR products of the left homology arm were used for Sanger sequencing of subclones of *SNCA*e2(p.A30P) polyclone 632 and *SNCA*e3(p.A53T) polyclone 636 as shown in Figure S3. The wild type junction was used for Sanger sequencing of subclones of transposed *SNCA*e2(p.A30P) polyclone 632 and transposed *SNCA*e3(p.A53T) polyclone 636 as shown in Figure S3.

Microarray Karyotype. Genomic DNA from the pre-electroporation parental, and isogenic polyclones was purified using GenElute Blood genomic DNA Kit (Sigma cat no. NA2020). Samples were processed at Bonn Univesity Life&Brain genomics facility using Illumina iScan technology (Illumina).

Immunostaining. Cells were fixed on PFA and permeabilized on PBS triton-X 0.2%. For characterizing human iPS cells, primary antibodies used were OCT4 (Santa cruz cat no. sc-5279) dilution 1:100, TRA1-81 (Millipore cat no. MAB4381) dilution 1:50, SOX2 (Abcam cat no. ab97959) dilution 1:100 and SSEA4 (Millipore cat no. MAB4304) dilution 1:50. Secondary antibodies used were donkey anti-mouse alexa fluor 488 (Thermo Fisher cat no. A-21202) and donkey anti-rabbit alexa fluor 488 (Thermo Fisher cat no. A-21206), both at dilution 1:1000. For characterizing NESCs, primary antibodies used were NESTIN (BD cat no. 611659) dilution 1:600 and SOX2 (Abcam cat no. ab97959) dilution 1:200. Secondary antibodies used were donkey anti-mouse 488 (Thermo Fisher cat no. A-21202) and donkey anti-rabbit 647 (Thermo Fisher cat no. A-31573), both at dilution 1:1000. For nuclear staining, Hoechst-33342 (Thermo Fisher cat no. 62249) was used at dilution 1:1000. Images were acquired in an inverted microscope (Zeiss Axio ObserverZ1).

NESCs differentiation and culture. Human iPS cells were clustered on aggrewell plates (Stem cell technologies cat no. 27845) for 12 hours. Embryoid bodies were transferred to ultra-low attachment plates and differentiated with the program in Figure 4A. Briefly, cells were cultured on KO-DMEM (Gibco cat no. 10829018) supplemented with 20% knock-out serum replacement (Gibco cat no. A3181501), 2mM glutamax (Gibco cat no. 35050061), 1x non-essential amino acids (Gibco cat no.11140035), 1µM dorsomorphine (Sigma cat no. P5499), 3µM CHIR99021 (Sigma cat no. SML1046) and 0.5µM purmorphamine (Sigma cat no. SML0868). From day three onwards, cells were cultured on DMEM-F12:neurobasal media (1:1) supplemented with N2 (Gibco cat no. 17502048), B27 without vitamin A (Gibco cat no. 12587001) and 2mM glutamax. For day three and four, media was supplemented with 10µM SB431542 (Sigma cat no. S4317). From day five onwards, the culture was maintained with 150µM ascorbic acid (Sigma cat no. A5960), 3µM CHIR99021 and 0.5µM purmorphamine. At day six, embryoid bodies were dissociated with accutase and plated on matrigel coated plates.

Extracellular energy flux analysis. NESCs were plated on Seahorse XFe96 assay plates (Aglient) at a density of 65k cells per well and the oxygen consumption rate was quantified in a Seahorse XFe96 Analyzer. Four baseline measurements were performed before any treatment injection. Three measurements were performed after each injection as shown in Figure 4. Final concentrations of compounds were 1µM for oligomycin (Sigma cat no. 75351), FCCP (Sigma cat no. C2920), antimycin A (Sigma cat no. A8674) and rotenone (Sigma cat no. R8875). DNA was quantified using CyQUANT kit (Thermo Fisher cat no. C7026) and normalization based on DNA content as previously described (Silva et al., 2013).

Western Blotting. For western blot analysis of NESCs total protein, an antibody against α-Synuclein (C-20)-R (Santa cruz cat no. sc-7011) was used at a dilution of 1:100, and an antibody against GAPDH (abcam cat no. ab9485) was used at a dilution of 1:1000 overnight. Blots were developed using anti-rabbit IgG HRP-linked secondary antibody (GE Healthcare Life Sciences cat no. NA934V) and west-pico chemiluminescent substrate (Thermo Fisher cat no. 34080). Membranes were imaged in a Raytest Stells system with exposure of 30s for both alpha-Synuclein and GAPDH.

Microarray. RNA was extracted from healthy control NESCs using quiazol (Qiagen cat no. 79306) and miRNeasy (Qiagen cat no. 217004). Samples were processed at the EMBL Genomics Core Facility using

Affymetrix human Gene 2.0 arrays. Results were processed using GC-RMA analysis. Gene expression omnibus accession code GSE101534.

Supplemental Figure legends

Figure S1. **Gating structures, second locus, subpopulation types and transposition optimization. Related to Figure 1, Figure 2 and Figure 3**. (A) Hierarchical single cell gating structure of SSC and FSC wide and high for single cell preparations. Representative example of *PINK1*e5(p.I368N) polyclone 517. (B) Preparation of cells with a cell strainer and single cell gating structure is essential to ensure high quality sorting. Scale bar 25µm. (C) Post-selection sorting for knock-in *PINK1*e5, using independent sgRNAs: sgRNA 517 and sgRNA 526. FACS plots are represented with 2% contour lines. (D) Purity-purity sorting allows the generation of a homogenous biallelic knock-in population. (E) Post-transposition sorting. Excision-only transposase expression removes the positive selection module for *PINK1*e5. (F) FACS analysis of parental wild type (WT) control. (G) FACS analysis for homozygous and heterozygous dTOMATO^{pos} clones for *SNCA*e2. Diagram of knock-in overlapping population types is shown (right). (H) Schematic representation of overlapping populations type U and type 2. (I) Type U (one copy integration) and type 2 (two copy integration) single cell clones present high overlap in the FSC-A dimension. Clone type U population overlaps 26.1% with the gate established by the type 2 clone. (J) Indel and wild type frequency from non-targeted allele. Sanger sequencing of the non-targeted allele from population type U presents a high frequency of indels (n=20) for *SNCA*e2 and *SNCA*e3. (K) Representative chromatogram for an indel bearing non-targeted allele of population type U. Cas9 cleavage site indicated in arrowhead. (L) Schematic representation of the transition states for PSM excision. The transition is recapitulated in E and Figure 3A-B. Puromycin resistance gene (Puro). (M) Optimization of conditions for transposase mediated excision. Representative histograms for excision of the PSM using wild type (WT) and excision-only (EO) transposase variants, and one to three transfection steps. (N) Quantification of the excision efficiency shown in M. Each condition represents three replicates. For transposase optimization assays, the EGFP^{pos} populations type 2 + type U were used. Significance determined by a one-way ANOVA. Significance level * $p<0.05$.

Figure S2. Repetitive elements decrease on-target efficiency and increase random integration events. Related to Figure 1. Flow cytometry histogram for tagBFP: (A) *SNCA* exon 2 sgRNA 630 and 632, (B) *SNCA* exon 3 sgRNA 634 and 636, (C) *PINK1* exon 5 sgRNA 517 and 526, (D) *CLN3* exon 14-15 sgRNA 788, 789 and 909, (E) *CLN3* exon 5-8 sgRNA 781 and 783, and (F) *CLN3* exon 10-13 sgRNA 561 and 563. (G) Distribution

and type of repetitive elements in the homology arms of the dsDNA donors for *SNCA*e2, *SNCA*e3, *PINK1*e5, *CLN3*e5-6, *CLN3*e10-13 and *CLN3*e14-15. (H) Predictive model for random integration. The predictive model *PR* uses the matrix of repetitive element frequency in the homology arms *A*, the repetitive elements vector *x*, and the observed incidence of tagBFP^{pos} random integration *b*. The mathematical model generates coefficients for each repetitive element and the constant of the system for random integration prediction. (I) The space of non-zero coefficients derived from H: SINE Alu and SINE Mir, allows inferring expected random integration frequencies.

Figure S3. Quantification of polyclones composition post-knock-in and post-transposition. Related to Figure 2 and 3. (A) Schematic representation of the genomic structure after knock-in and genomic structure after transposition. PSM, left homology arm (LHA) and right homology arm (RHA). The binding sites of the genotyping primers are represented (Table S4), as well as the left homology arm junction, right homology arm junction, and WT junction. (B) Genotyping PCR products of 24 clones derived from the polyclone *SNCA*e2(p.A30P) 632, and WT control. (C) Genotyping PCR products of 24 clones derived from the polyclone *SNCA*e3(A53T) 636, and WT control. (D) Genotyping PCR products of 24 clones derived from the transposed polyclone *SNCA*e2(p.A30P) 632, pre-removal polyclone, and WT control. (E) Genotyping PCR products of 24 clones derived from the transposed polyclone *SNCA*e3(p.A53T) 636, pre-removal polyclone, and WT control. (F) Representation of the left homology arm junction of *SNCA*e2 including the SNP rs104893878 and PSM interface. Sanger sequencing chromatograms of 24 clones (not shown for space limitations) derived from the polyclone *SNCA*e2(p.A30P) 632 as in B. Chromatograms show the transversion *SNCA* c.88g>c and the TTAA interface to the PSM. Knock-in (KI). (G) Representation of the left homology arm junction of *SNCA*e3 including the SNP rs104893877 and PSM interface. Sanger sequencing chromatograms of 24 clones (not shown for space limitations) derived from the polyclone *SNCA*e3(p.A53T) 636 as in C. Chromatograms show the transition *SNCA* c.209g>a and the TTAA interface to the PSM. (H) Representation of the WT junction of *SNCA*e2 including the SNP rs104893878, and the TTAA interface to the genomic region. Sanger sequencing chromatograms of 24 clones (not shown for space limitations) derived from the transposed polyclone *SNCA*e2(p.A30P) 632. Chromatograms show the transversion *SNCA* c.88g>c and the TTAA interface to the genomic region. (I) Representation of the WT junction of *SNCA*e3 including the SNP rs104893877, and the TTAA interface to the genomic region. Sanger sequencing chromatograms of 24 clones (not shown for space limitations) derived from the transposed polyclone

*SNCA*e3(p.A53T) 636. Chromatograms show the transition *SNCA* c.209g>a and the TTAA interface to the genomic region.

Figure S4. Characterization of human iPS cells and NESCs: Microarray karyotype, pluripotency, differentiation and expression levels. Related to Figure 3 and Figure 4. (A) Microarray karyotype analysis of the parental line before electroporation, (B) polyclone 6321421 *SNCA* p.A30P and (C) polyclone 6361868 *SNCA* p.A53T. (D) Immunostaining for the pluripotency markers OCT4, TRA1-81, SOX2 and SSEA4 for parental control, (E) polyclone 6321421 *SNCA* p.A30P and (F) polyclone 6361868 *SNCA* p.A53T. Scale bar 200µm. (G) Differentiation of human iPS cells to NESCs in 3D culture as shown in Figure 4A. Scale bar 500µm. (H) Immunostaining of NESCs for the neuroepithelial stem cell markers NESTIN and SOX2. Scale bar 50µm. (C) Relative expression of *SNCA* mRNA with respect to *TUBG1* and *GAPDH* transcripts in microarray expression analysis in Figure 4B. Independent samples of NESCs A13777 were used (n=3).

Supplemental Table 1. Biallelic targeting frequency

Polyclone sample	Frequency composed biallelic[a]	Frequency single channel biallelic[b]	Frequency total biallelic[c]
*SNCA*e3 636	0.032	0.179	0.390
*SNCA*e2 630	0.021	0.145	0.311
*SNCA*e2 632	0.022	0.148	0.319
*SNCA*e3 634	0.056	0.237	0.529
*SNCA*e3 636	0.012	0.110	0.231
*PINK1*e5 517	0.033	0.182	0.396
*PINK1*e5 526	0.042	0.205	0.452
Mean global			**0.375**

[a]Frequency composed biallelic is defined as the experimentally measured $EGFP^{pos}dTOMATO^{pos}$ population.
[b]Frequency of single channel biallelic represents separately the $EGFP^{pos}EGFP^{pos}$ and $dTOMATO^{pos}dTOMATO^{pos}$ population, calculated as $\sqrt{\text{frequency composed biallelic}}$.
[c]Frequency total biallelic correspond to frequency composed biallelic $+ 2 *$ frequency of single channel biallelic.

Supplemental Table 2. Genome Cas9 binding sites tested and design for PAM edited donor sequences.

sgRNA	Genomic sequence	Sequence in donor
628	gtaaaggaattcattagcca**tgg**	gtaaaggaattcattagcca**tgg**
629	ggactttcaaaggccaagga**ggg**	ggactttcaaaggccaagga**ggg**
630	gctgctgagaaaaccaaaca**ggg**	gctgctgagaaaaccaaaca**ggg**
631	agggtgttctctatgtaggt**agg**	agggtgttctctatgtaggt**agg**
632[a]	ggtgcttgttcatgagtgat**ggg**	ggtgcttgTTAAtgagtgat**g**C**g**
633	**gga**agaagatcaaaatcctatat	**gga**agaagatcaaaatcctatat
634	tgtaggctccaaaaccaagg**agg**	tgtaggctccaaaaccaagg**agg**
635	**ggt**aacacgaatataggtttcta	**ggt**aacacgaatataggtttcta
636[a]	**ggt**ttctactataaatttcatag	**g**C**t**ttctactataAATTtcatag
637	atacttgccaagaataatga**ggg**	atacttgccaagaataatga**ggg**

[a]Edited PAM sequence.

Supplemental Table 3. *SNCA* polyclones summary.

Polyclone	PAM shielded	sgRNA	FACS % non-random	FACS % composed biallelic knock-in	% correct genotype post-knock-in (n correct/total)	FACS % transposition	% correct genotype post-transposition (n correct/total)
*SNCA*e2(p.A30P) 632	YES	632	94.1	2.2	100 (24/24)	4.0	100 (24/24)
*SNCA*e3(p.A53T) 636	YES	636	56.8	1.2	100 (24/24)	1.1	100 (24/24)
*SNCA*e2(p.A30P) 630	NO	630	85.1	2.1	Not determined	3.3	Not determined
*SNCA*e3(p.A53T) 634	NO	634	34.2	5.6	Not determined	3.2	Not determined

Supplemental Table 4. Oligonucleotides used in this study.

Primer	Sequence (5' to 3')	Region (Purpose)
SNCAe2_F1 (no1615)	gaggagtcggagttgtggagaag	SNCAe2 (Genotyping)
SNCAe2_R1 (no1616)	ttcccccactgatctatgttgaagag	SNCAe2 (Genotyping)
SNCAe3_F1 (no1617)	actgaaaaatccaacattagagagg	SNCAe3 (Genotyping)
SNCAe3_R1 (no1036)	ccagaacttgccacatgctt	SNCAe3 (Genotyping)
ITR_R1 (no861)	agatgtcctaaatgcacagcg	ITR (Genotyping)
ITR_F1 (no1310)	cgtcaattttacgcatgattatctttaac	ITR (Genotyping)
SNCAe2 (no1065)	tccgtggttaggtggctaga	SNCAe2 (Sequencing)
SNCAe3 (no1034)	gggccccggtgttatctcat	SNCAe3 (Sequencing)
T7-transposase_F (no1673)	gaaattaatacgactcactataggg ccgccaccatgggcagcagcctggac	transposase CDS (T7 fusion IVT)
Transposase_R (no1693)	ggcaaacaacagatggctgg	transposase CDS (IVT)
SNCAe2_628F	caccggtaaaggaattcattagcca	synthetic (sgRNA cloning)
SNCAe2_629F	caccgggactttcaaaggccaagga	synthetic (sgRNA cloning)
SNCAe2_630F	caccggctgctgagaaaaccaaaca	synthetic (sgRNA cloning)
SNCAe2_631F	caccgagggtgttctctatgtaggt	synthetic (sgRNA cloning)
SNCAe2_632F	caccgggtgcttgttcatgagtgat	synthetic (sgRNA cloning)
SNCAe3_633F	caccgtatatcctaaaactagaaga	synthetic (sgRNA cloning)
SNCAe3_634F	caccgtgtaggctccaaaaccaagg	synthetic (sgRNA cloning)
SNCAe3_635F	caccgatctttggatataagcacaa	synthetic (sgRNA cloning)
SNCAe3_636F	caccggatactttaaatatcatctt	synthetic (sgRNA cloning)
SNCAe3_637F	caccgatacttgccaagaataatga	synthetic (sgRNA cloning)
SNCAe2_628R	aaactggctaatgaattcctttacc	synthetic (sgRNA cloning)
SNCAe2_629R	aaactccttggcctttgaaagtccc	synthetic (sgRNA cloning)
SNCAe2_630R	aaactgtttggttttctcagcagcc	synthetic (sgRNA cloning)
SNCAe2_631R	aaacacctacatagagaacaccctc	synthetic (sgRNA cloning)
SNCAe2_632R	aaacatcactcatgaacaagcaccc	synthetic (sgRNA cloning)
SNCAe3_633R	aaactcttctagttttaggatatac	synthetic (sgRNA cloning)
SNCAe3_634R	aaacccttggttttggagcctacac	synthetic (sgRNA cloning)
SNCAe3_635R	aaacttgtgcttatatccaaagatc	synthetic (sgRNA cloning)
SNCAe3_636R	aaacaagatgatatttaaagtatcc	synthetic (sgRNA cloning)
SNCAe3_637R	aaactcattattcttggcaagtatc	synthetic (sgRNA cloning)
U6_F	gagggcctatttcccatgattcc	U6 (sequencing)

Supplemental references

Doench, J.G., Hartenian, E., Graham, D.B., Tothova, Z., Hegde, M., Smith, I., Sullender, M., Ebert, B.L., Xavier, R.J., and Root, D.E. (2014). Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation. Nat Biotechnol 32, 1262-1267.

Gibson, D.G. (2011). Enzymatic assembly of overlapping DNA fragments. Methods Enzymol 498, 349-361.

Li, X., Burnight, E.R., Cooney, A.L., Malani, N., Brady, T., Sander, J.D., Staber, J., Wheelan, S.J., Joung, J.K., McCray, P.B., Jr., et al. (2013). piggyBac transposase tools for genome engineering. Proc Natl Acad Sci U S A 110, E2279-2287.

Ran, F.A., Hsu, P.D., Wright, J., Agarwala, V., Scott, D.A., and Zhang, F. (2013). Genome engineering using the CRISPR-Cas9 system. Nature protocols 8, 2281-2308.

Silva, L.P., Lorenzi, P.L., Purwaha, P., Yong, V., Hawke, D.H., and Weinstein, J.N. (2013). Measurement of DNA concentration as a normalization strategy for metabolomic data from adherent cell lines. Anal Chem 85, 9536-9542.

Yusa, K., Zhou, L., Li, M.A., Bradley, A., and Craig, N.L. (2011). A hyperactive piggyBac transposase for mammalian applications. Proc Natl Acad Sci U S A 108, 1531-1536.

Figure S1

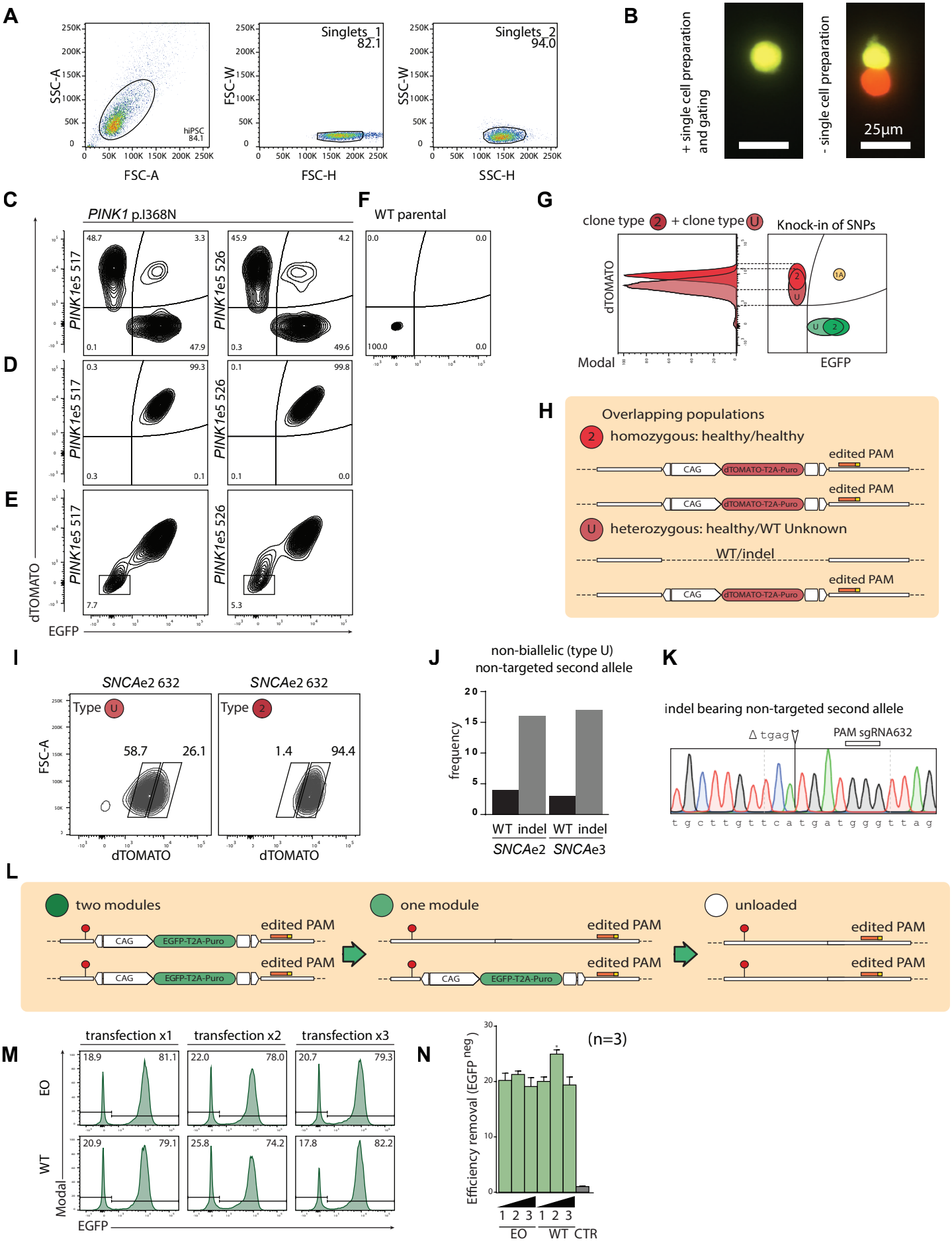
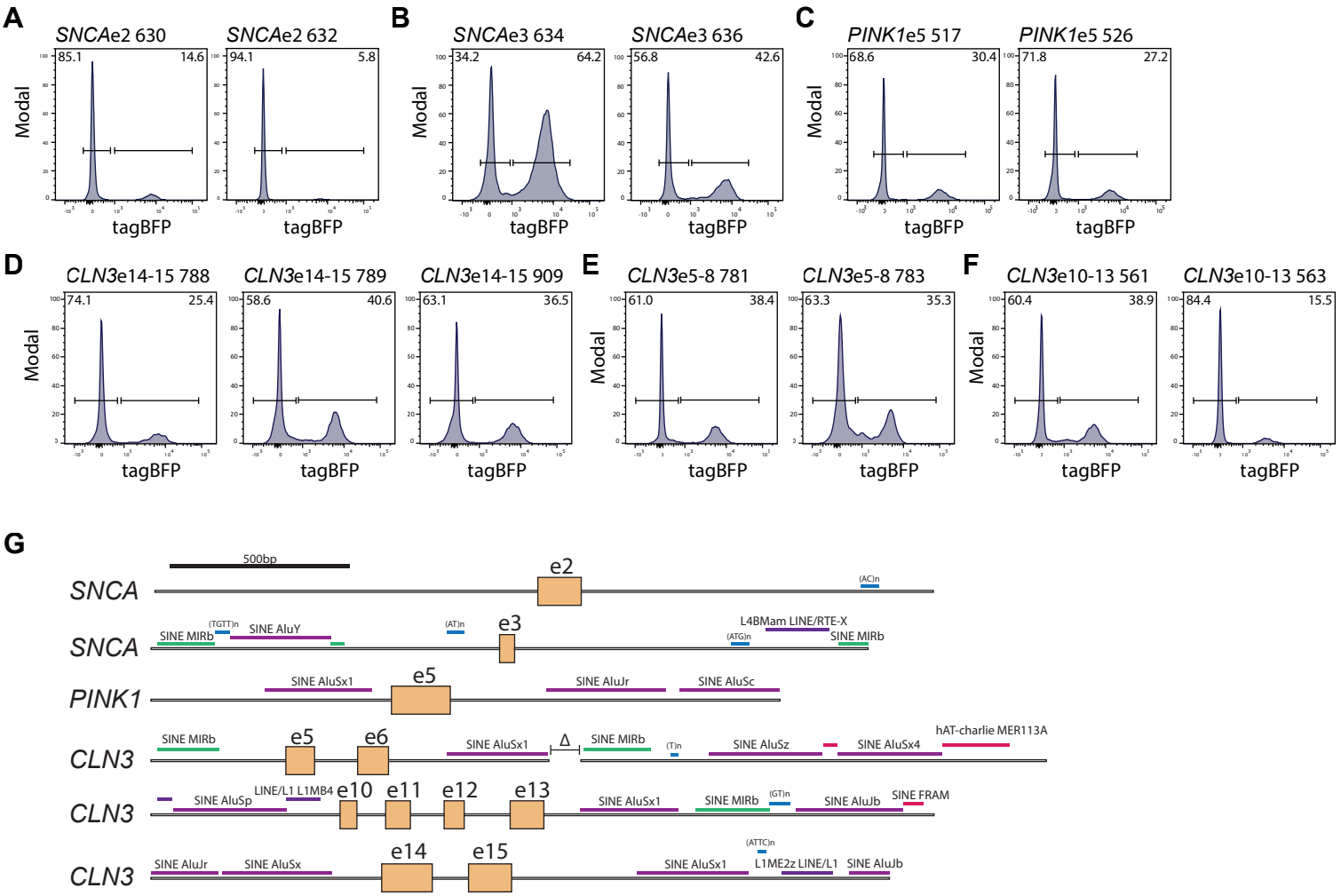

A
SSC-A
FSC-A
hiPSC
84.1
Singlets_1
82.1
FSC-W
FSC-H
Singlets_2
94.0
SSC-W
SSC-H
B
+ single cell preparation and gating
- single cell preparation
25µm
C
PINK1 p.I368N
PINK1e5 517
PINK1e5 526
48.7
3.3
0.1
47.9
45.9
4.2
0.3
49.6
D
0.3
99.3
0.3
0.1
0.1
99.8
0.1
0.0
E
7.7
5.3
dTOMATO
EGFP
F
WT parental
0.0
0.0
100.0
0.0
G
clone type 2 + clone type U
Knock-in of SNPs
dTOMATO
Modal
EGFP
1A
2
U
H
Overlapping populations
2 homozygous: healthy/healthy
edited PAM
CAG
dTOMATO-T2A-Puro
U heterozygous: healthy/WT Unknown
WT/indel
I
SNCAe2 632
Type U
Type 2
58.7
26.1
1.4
94.4
FSC-A
dTOMATO
J
non-biallelic (type U)
non-targeted second allele
frequency
WT indel WT indel
SNCAe2 SNCAe3
K
indel bearing non-targeted second allele
Δ tgag
PAM sgRNA632
t g c t t g t t c a t g a t g g g t t a g
L
two modules
one module
unloaded
edited PAM
CAG
EGFP-T2A-Puro
M
transfection x1
transfection x2
transfection x3
EO
WT
Modal
EGFP
18.9
81.1
22.0
78.0
20.7
79.3
20.9
79.1
25.8
74.2
17.8
82.2
N
(n=3)
Efficiency removal (EGFP neg)
1 2 3 1 2 3
EO WT CTR

Figure S2



H

Ax=b

frequency matrix

Loci	SINE Alu	LINE L4	SINE Mir	short repeats	Line L1	hAT charlie	υ	= measured random integration (tagBFP pos)
SNCAe3A	1	1	2	2	0	0	0	64.2
SNCAe3B	1	1	2	2	0	0	0	42.6
SNCAe2A	0	0	0	1	0	0	0	14.6
SNCAe2B	0	0	0	1	0	0	0	5.7
PINK1e5A	3	0	0	0	0	0	0	30.4
PINK1e5B	3	0	0	0	0	0	0	27.2
CLN3e14-15A	3	0	0	0	1	0	0	36.5
CLN3e14-15B	3	0	0	0	1	0	0	25.4
CLN3e14-15C	3	0	0	0	1	0	0	40.6
CLN3e10-13A	3	0	1	1	1	0	0	38.9
CLN3e10-13B	3	0	1	1	1	0	0	15.5
CLN3e5-8delA	3	0	2	0	0	1	0	38.4
CLN3e5-8delB	3	0	2	0	0	1	0	35.3
maximal-v	8.5	14.4	14.5	25	20.3	11.4	25	100.0

Predicted random

$PR = \alpha(RE1) + \beta(RE2) + \gamma(RE3) + \bullet\bullet\bullet + \delta(REn) + C$

$PR = 10.0978(\text{SINE Alu}) + (0)(\text{LINE L4}) + (1.746)(\text{SINE Mir}) + (0)(\text{short repeats}) + (0)(\text{Line L1}) + (0)(\text{hAT charlie}) + (0)(\upsilon) + C$

$PR = 10.0978(\text{SINE Alu}) + (1.746)(\text{SINE Mir}) + C \quad \text{with} \quad C=10.15$

I

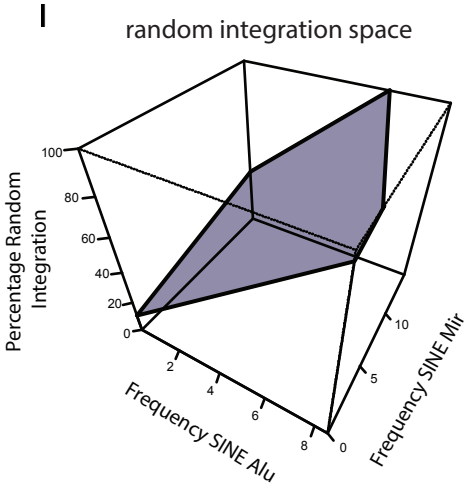


Figure S3

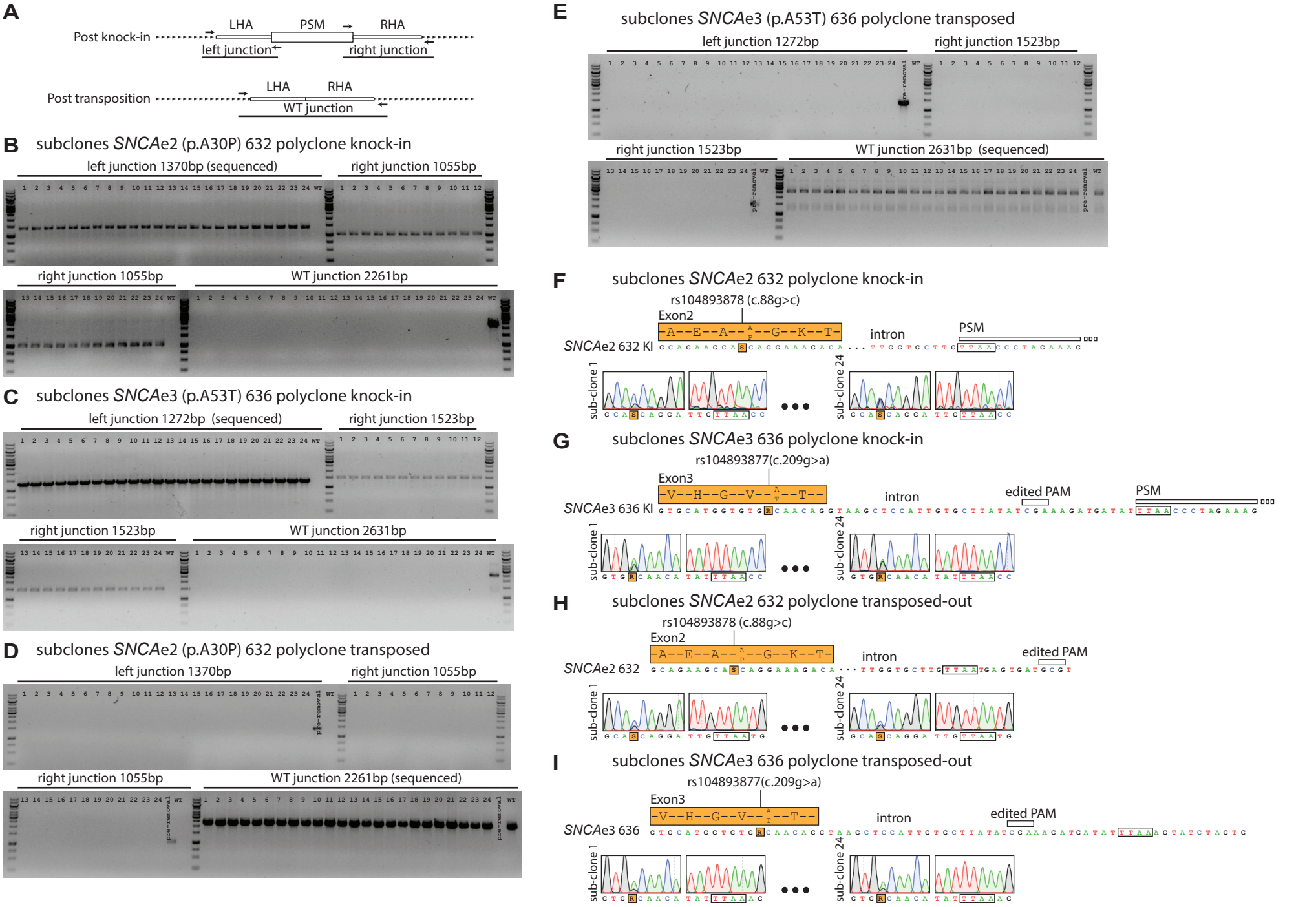
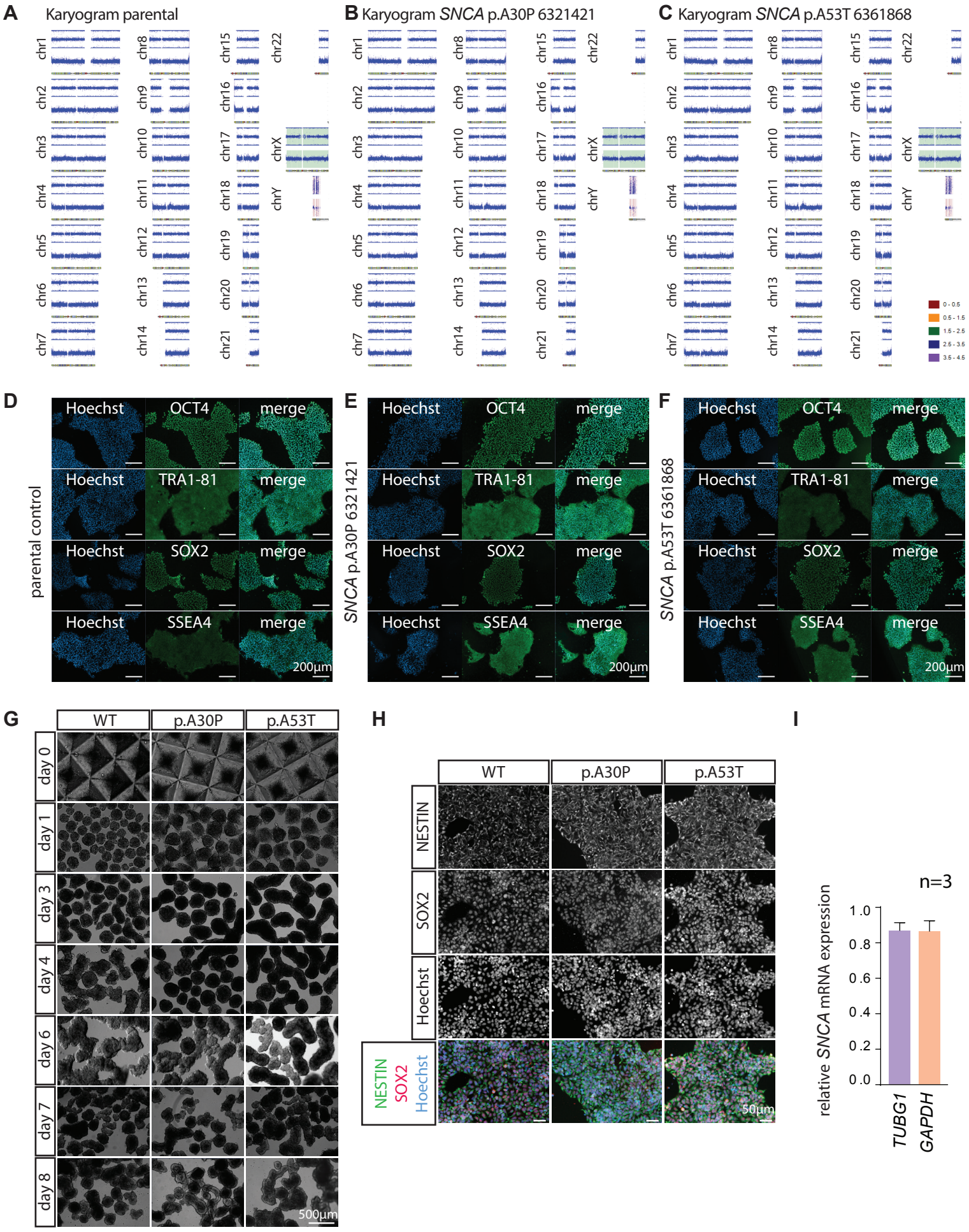

A
Post knock-in
LHA
PSM
RHA
left junction
right junction
Post transposition
LHA
RHA
WT junction
B
subclones SNCAe2 (p.A30P) 632 polyclone knock-in
left junction 1370bp (sequenced)
right junction 1055bp
right junction 1055bp
WT junction 2261bp
C
subclones SNCAe3 (p.A53T) 636 polyclone knock-in
left junction 1272bp (sequenced)
right junction 1523bp
right junction 1523bp
WT junction 2631bp
D
subclones SNCAe2 (p.A30P) 632 polyclone transposed
left junction 1370bp
right junction 1055bp
pre-removal
right junction 1055bp
WT junction 2261bp (sequenced)
pre-removal
pre-removal
E
subclones SNCAe3 (p.A53T) 636 polyclone transposed
left junction 1272bp
right junction 1523bp
pre-removal
right junction 1523bp
WT junction 2631bp (sequenced)
pre-removal
pre-removal
F
subclones SNCAe2 632 polyclone knock-in
rs104893878 (c.88g>c)
Exon2
-A--E--A--A/P--G--K--T-
intron
PSM
SNCAe2 632 KI
sub-clone 1
sub-clone 24
G
subclones SNCAe3 636 polyclone knock-in
rs104893877(c.209g>a)
Exon3
-V--H--G--V--A/T--T--
intron
edited PAM
PSM
SNCAe3 636 KI
sub-clone 1
sub-clone 24
H
subclones SNCAe2 632 polyclone transposed-out
rs104893878 (c.88g>c)
Exon2
-A--E--A--A/P--G--K--T-
intron
edited PAM
SNCAe2 632
sub-clone 1
sub-clone 24
I
subclones SNCAe3 636 polyclone transposed-out
rs104893877(c.209g>a)
Exon3
-V--H--G--V--A/T--T--
intron
edited PAM
SNCAe3 636
sub-clone 1
sub-clone 24

Figure S4


A
Karyogram parental
B Karyogram SNCA p.A30P 6321421
C Karyogram SNCA p.A53T 6361868
chr1
chr2
chr3
chr4
chr5
chr6
chr7
chr8
chr9
chr10
chr11
chr12
chr13
chr14
chr15
chr16
chr17
chr18
chr19
chr20
chr21
chr22
chrX
chrY
0 - 0.5
0.5 - 1.5
1.5 - 2.5
2.5 - 3.5
3.5 - 4.5
D
E
F
parental control
SNCA p.A30P 6321421
SNCA p.A53T 6361868
Hoechst
OCT4
merge
TRA1-81
SOX2
SSEA4
200µm
G
WT
p.A30P
p.A53T
day 0
day 1
day 3
day 4
day 6
day 7
day 8
500µm
H
NESTIN
SOX2
Hoechst
50µm
I
n=3
relative SNCA mRNA expression
1.0
0.8
0.6
0.4
0.2
0.0
TUBG1
GAPDH