

# LiPS-A3S, a human genomic site for robust expression of inserted transgenes

Andriana G Kotini<sup>1–4</sup>, Michel Sadelain<sup>5</sup> and Eirini P Papapetrou<sup>1–4</sup>

Transgenesis of human pluripotent stem cells (hPSCs) can enable and empower a variety of studies in stem cell research, including lineage tracing and functional genetics studies. While in recent years much progress has been made in the development of tools for gene targeting, little attention has been given to the identification of sites in the human genome where transgenes can be inserted and reliably expressed. In order to find human genomic sites capable of supporting long-term and high-level transgene expression in hPSCs, we performed a lentiviral screen in human induced pluripotent stem cells (iPSCs). We isolated 40 iPSC clones each harboring a single vector copy and characterized the level of transgene expression afforded by each unique integration site. We selected one clone, LiPS-A3 with an integration site in chromosome 15 maintaining robust expression without silencing and demonstrate that different transgenes can be inserted therein rapidly and efficiently through recombinase-mediated cassette exchange (RMCE). The LiPS-A3 line can greatly facilitate the insertion of reporter and other genes in hPSCs. Targeting transgenes in the LiPS-A3S genomic locus can find broad applications in stem cell research and possibly cell and gene therapy.

*Molecular Therapy—Nucleic Acids* (2016) 5, e394; doi:10.1038/mtna.2016.99; published online 29 November 2016

**Subject Category:** Gene addition, deletion and modification

## Introduction

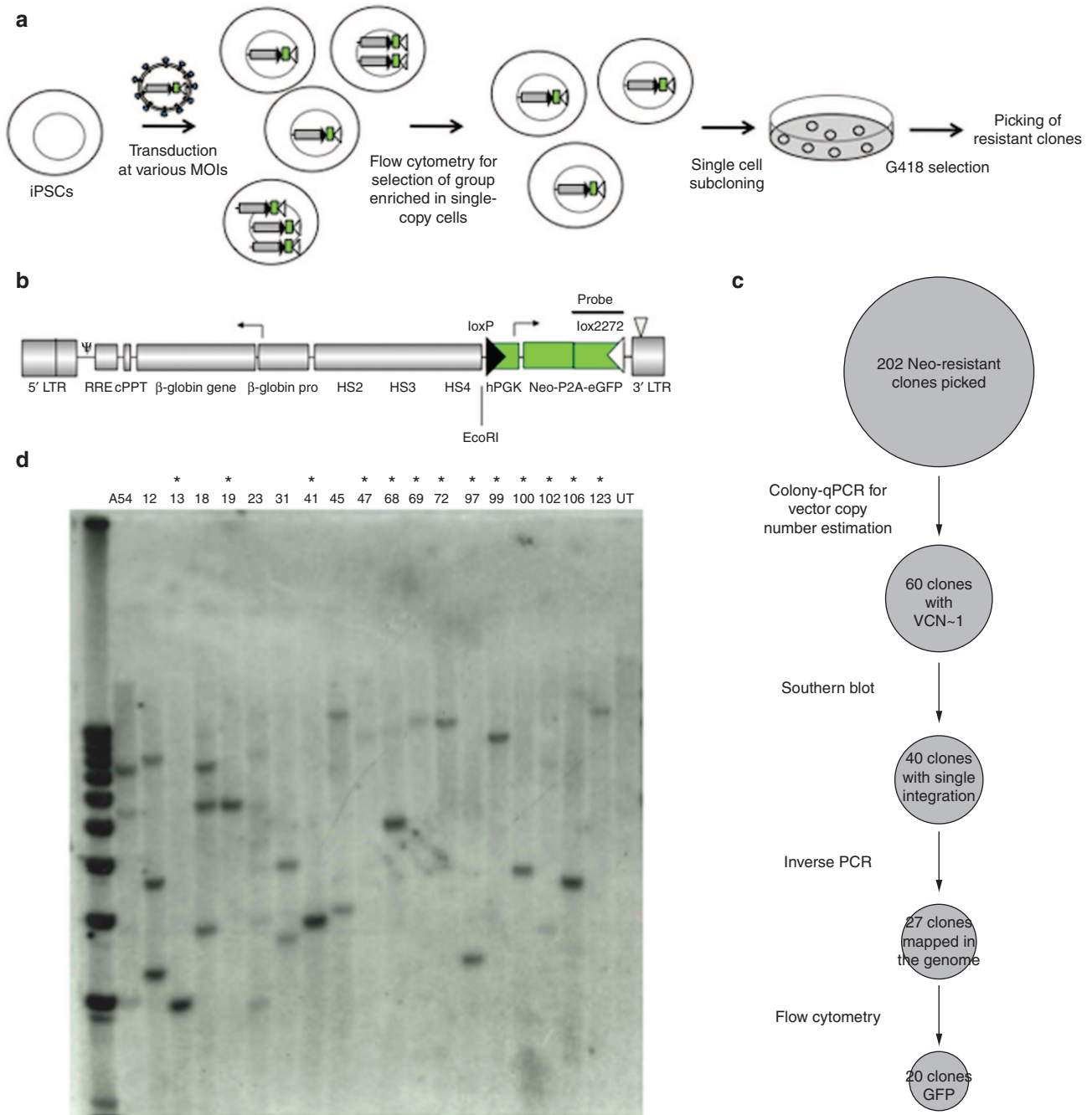
Many stem cell research studies require insertion of genes or genetic sequences in human pluripotent stem cells (hPSCs) in a way that ensures their expression is predictable, reliable and sustained over time. Random insertion by retroviral vectors or plasmid transfection usually results in unpredictable expression subject to variegation and silencing.<sup>1</sup> Bacterial artificial chromosome transgenesis can overcome these issues in some cases, but it is an inefficient and cumbersome process.<sup>2,3</sup> Furthermore, the majority of gene therapy applications currently in the clinic or in advanced preclinical development rely on gene addition and, while *in situ* gene correction may prove feasible and practical for some genetic diseases in the future, gene addition will always be the approach of choice for a number of gene therapy applications.<sup>4</sup>

With recent advances in technologies for targeted gene insertion (gene targeting), it is becoming increasingly feasible to introduce a transgene into a predetermined genomic site. The advent of CRISPR as a genome-editing tool, in particular, dramatically reduced the complexity of gene targeting making it more accessible to the wider scientific community. However, these advances have not been paralleled by advances in the identification and validation of appropriate genomic sites. Very few sites have so far been explored for their ability to accommodate newly introduced transgenes. The adeno-associated virus site 1 (AAVS1) in chromosome 19 is the site most commonly used today for gene insertion. It was originally recovered as a natural site of integration of wild-type AAV in cultured human cell lines<sup>5</sup> and gained popularity

due to, on one hand, its ability to afford robust expression of inserted reporter transgenes in various cell types, including hPSCs (embryonic stem cells and induced pluripotent stem cells (iPSCs)) and, on the other hand, the development of efficient commercially available tools for its targeting.<sup>6–9</sup> However, more systematic recent studies have revealed that the AAVS1 locus is prone to silencing in some cell lineages.<sup>10</sup> Furthermore, a number of applications require dual transgene or reporter expression or transgene insertion in hPSC lines where the AAVS1 locus is already occupied by other transgenic sequences, for example by an inducible genome editing enzyme.<sup>11–13</sup> In the latter example, recently generated hPSC lines harboring a doxycycline-inducible Cas9 (iCas9) inserted in one allele of the AAVS1 locus (and the reverse tetracycline transactivator, in the other allele) could be further exploited to mediate high-efficiency CRISPR-Cas9-facilitated insertion of a transgene in a genomic locus provided one was known and available. Two other candidate loci for gene insertion that have been investigated in past studies, the chemokine (C-C motif) receptor 5 (*CCR5*) gene—a coreceptor of wild-type HIV-1—and the human ortholog of the mouse Rosa26 locus, lack robust expression in most cell types including hPSCs and hence have not found use by the scientific community.<sup>14–16</sup> A more recently described intergenic locus, *H11*, located on chromosome 22, can support expression of multiple transgenes in hPSCs and appears to be a promising candidate worthy of further evaluation.<sup>17</sup> At this stage, the identification of additional candidate loci is therefore warranted if we were to deliver to the research community genomic sites that can support reliable expression and

<sup>1</sup>Department of Oncological Sciences, Icahn School of Medicine at Mount Sinai, New York, New York, USA; <sup>2</sup>The Tisch Cancer Institute, Icahn School of Medicine at Mount Sinai, New York, New York, USA; <sup>3</sup>The Black Family Stem Cell Institute, Icahn School of Medicine at Mount Sinai, New York, New York, USA; <sup>4</sup>Division of Hematology, Department of Medicine, University of Washington, Seattle, Washington, USA; <sup>5</sup>Center for Cell Engineering, Memorial Sloan-Kettering Cancer Center, New York, New York, USA. Correspondence: Eirini P Papapetrou, Department of Oncological Sciences, Icahn School of Medicine at Mount Sinai, One Gustave L. Levy Place, Box 1044A, New York, New York 10029, USA. E-mail: [eirini.papapetrou@mssm.edu](mailto:eirini.papapetrou@mssm.edu)

Received 23 April 2016; accepted 6 October 2016; published online 29 November 2016. doi:10.1038/mtna.2016.99

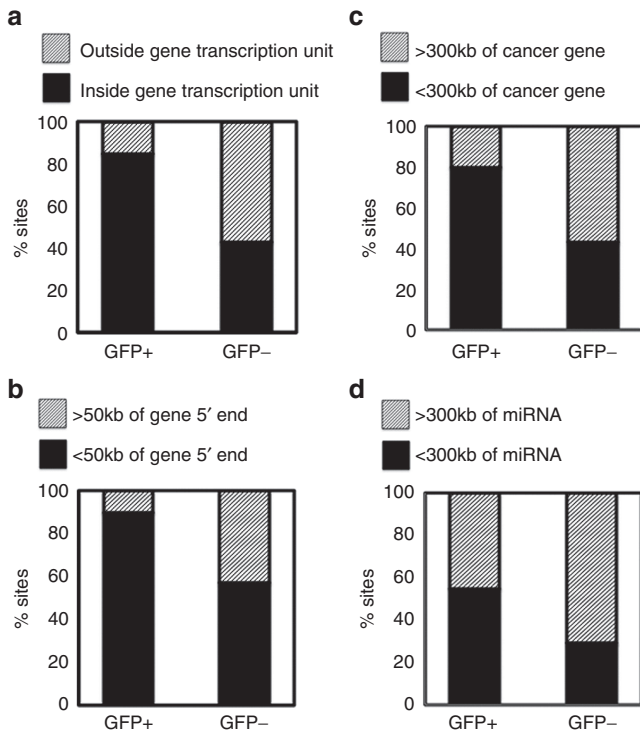


**Figure 1 Isolation of iPSC clones with unique lentiviral integrations.** (a) Experimental strategy for generation of a cohort of iPSC clones with lentiviral integrations enriched in single vector copy integrants. (b) The lentiviral vector TNS9.3/eNG used in this study. The vector carries a human phosphoglycerate kinase (hPGK)-driven cassette consisting of the Neomycin resistance gene linked by a P2A peptide to eGFP and flanked by a loxP and a lox2272 site in inverted orientation. This exchangeable cassette (eNeo-GFP, eNG) can be exchanged with a donor cassette flanked by two loxP and lox2272 sites identical and in the same orientation relative to each other as in the eNG cassette. (c) Scheme of sequential selection steps for the derivation of clones thoroughly verified to each contain a single integration in a known genomic site that supports transgene expression. (d) Representative Southern blot analysis to ascertain single integrations of the TNS9.3/eNG vector and clonality. Each band corresponds to a unique integration site. Genomic DNA was digested with EcoRI. The probe consists of the eGFP sequence and is shown in B. Asterisks depict clones with single integrations. MOI, multiplicity of infection; VCN, vector copy number; UT, untransduced.

can be used alternatively or in combination to the AAVS1 locus.

We previously used lentiviral integration to isolate genomic sites in human iPSCs and examine their location with respect to known annotated genomic elements with the goal to

predict their potential for insertional oncogenesis.<sup>18–20</sup> However, in this previous study, we did not assess transgene expression across many different sites or the likelihood that a given site will afford sufficient expression based on its position relative to endogenous genes. Here, we use a lentiviral



**Figure 2** Location and expression of single lentiviral integration sites. (a–d) Analysis of single integration sites in 20 eGFP<sup>+</sup> and 7 eGFP<sup>-</sup> iPSC clones with respect to the four “safe harbor” criteria.

screen to analyze the capacity for transgene expression in a large number of genomic sites in human iPSCs. We select a specific site on chromosome 15 affording robust expression and establish an iPSC line with a preexisting cassette inserted therein for efficient and versatile gene insertion through recombinase-mediated cassette exchange (RMCE).

## Results

### A lentiviral screen of the human genome for sites that afford adequate transgene expression

In order to isolate iPSC clones with single vector integrations, which could subsequently be “re-used” through RMCE, we modified our previously described lentiviral vector TNS9.3/fNG<sup>18</sup>, encoding  $\beta$ -globin in an erythroid-specific manner, as well as a selection cassette consisting of a neomycin resistance gene and eGFP driven by the constitutive human phosphoglycerate kinase promoter. In the derivative TNS9.3/eNG vector we use in this study, the hPCK-eGFP cassette was flanked by two heterospecific loxP sites—a wild-type loxP and a lox2272 site<sup>21</sup>—in inverse orientation towards each other (Figure 1a,b). This cassette can be exchanged by RMCE with a cassette harboring the same configuration of two loxP sites, as heterospecific loxP sites cannot recombine with each other but can recombine with homologous sites. The placement of the two loxP sites in inverse orientation relative to each other further reduces the probability of recombination events between them (which would result in cassette excision). We transduced the TNS9.3/eNG vector into the iPSC line thal5.10-Cre8, derived from a patient with  $\beta$ -thalassemia and devoid of permanently integrated

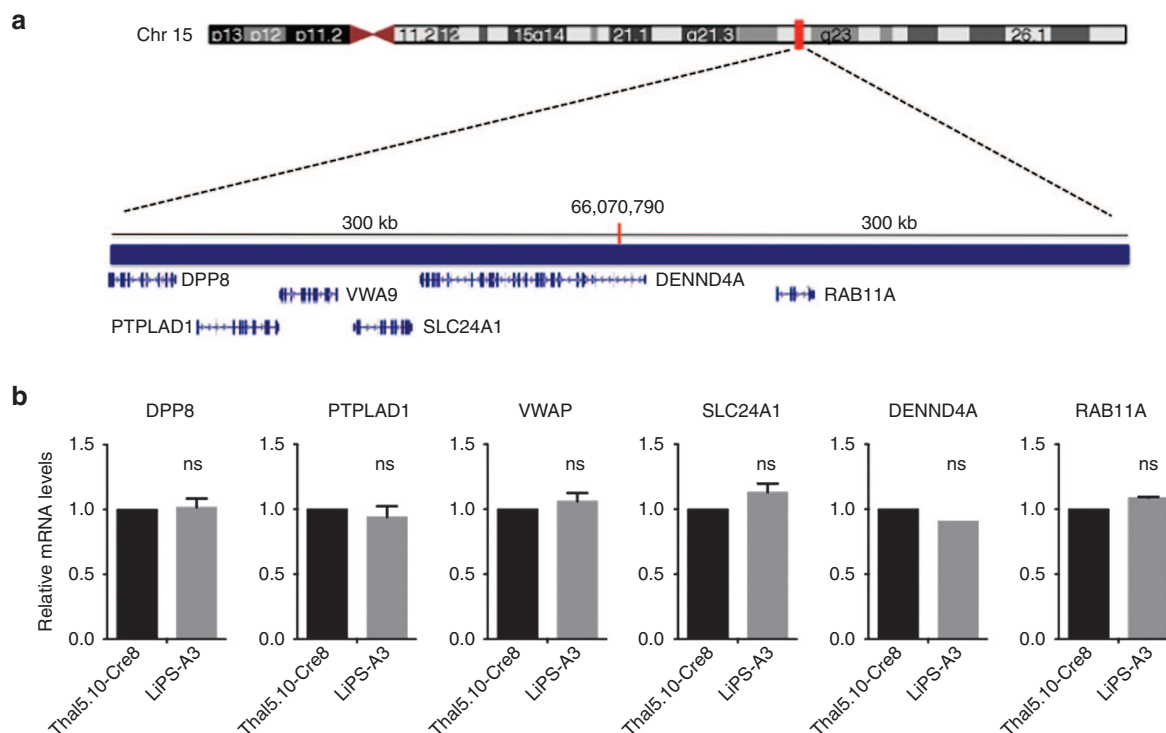
transgenes.<sup>18</sup> After transduction at a range of multiplicity of infection, the groups of cells more likely to be enriched in single vector copy clones (transduction efficiency 10–30%, as assessed by % eGFP<sup>+</sup> cells)<sup>19</sup> were single-cell subcloned. After selection, 202 neomycin-resistant clones were picked. Sixty of them were found to be single-copy by quantitative polymerase chain reaction (qPCR) (Figure 1c). Of these 40 were confirmed by Southern blotting to be single clones with single vector integrations (Figure 1d), 27 of which could be mapped to the genome (Supplementary Table S1).

### Dynamics of transgene expression in different genomic sites

The 40 single copy clones were grown for 3–4 weeks after the end of antibiotic selection to allow sufficient time for eGFP expression to reach steady state. Analysis of eGFP expression by flow cytometry showed that of the 40 single copy clones, 25 were predominantly eGFP<sup>+</sup> (more than 50% of cells) eGFP<sup>+</sup> and 15 were predominantly eGFP<sup>-</sup> (Supplementary Figure S1). We previously proposed criteria for selection of “safe harbor” sites.<sup>18,20</sup> We analyzed the location of the integration sites in the 27 mapped clones with respect to the four core criteria: (i) Inside a gene transcription unit; (ii) At a distance of less than 50 kb from a gene 5' end (transcriptional start site); (iii) At a distance of less than 300 kb from a gene associated with cancer in humans or model organisms and (iv) At a distance of less than 300 kb from a microRNA gene (Supplementary Table S1). Two of the 27 clones (clones 72 and 99) met the four core criteria in agreement with our previous estimates regarding the frequency of lentiviral integration sites genome-wide that are located in “safe harbor” sites by these criteria.<sup>18</sup> Neither of these expressed eGFP at levels detectable by flow cytometry. We thus next asked whether sites that can support expression are less likely to meet each of the four criteria. By comparing all 40 eGFP<sup>+</sup> and eGFP<sup>-</sup> clones against the criteria, we observed that sites supporting sustained expression are more likely to reside inside transcription units and in proximity to genes, including cancer genes (Figure 2). In contrast, sites wherein transgenes silence are randomly distributed with regards to these criteria. On the other hand, proximity to miRNA genes was not predictive of the probability of expression, unlike the other three criteria. These results show that transgene insertion sites that support expression—at least those accessible to lentiviral integration—are enriched in genomic sites that are located inside or in close proximity to endogenous genes.

### Rapid and versatile gene exchange by RMCE

We selected clone A3 (hereafter referred to as Lentiviral-iPS-A3 or LiPS-A3) harboring an integration site in chromosome 15 (hereafter referred to as LiPS-A3 site, LiPS-A3S), on the basis of its sustained eGFP expression over more than 20 passages and its distance from cancer-related genes (Figure 3a, Supplementary Table S1). The LiPS-A3S site is within the 3' UTR region of the *DENND4A* gene. We did not detect any perturbation of expression of the *DENND4A* gene or of any of five other coding genes located within 300 kb on either side of the integration site (Figure 3b). In order to test whether LiPS-A3S could support expression of different transgenes



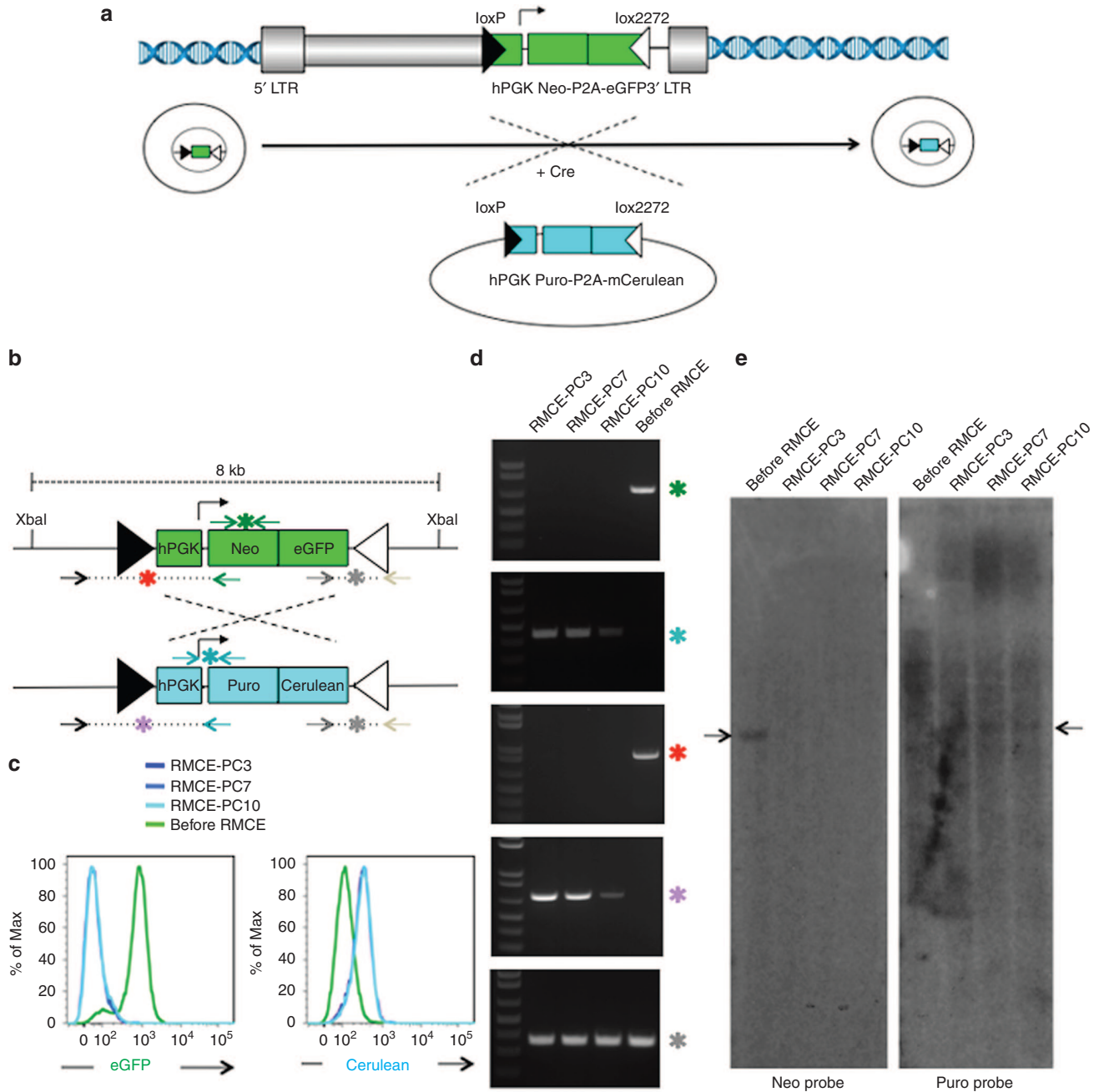
**Figure 3 The LiPS-A3S locus.** (a) Chromosome ideogram (upper panel) and graphic (lower panel) depicting 300 kb of human genome on both sides of the LiPS-A3S site on chromosome 15 (shown as a vertical red line), generated with the University of California, Santa Cruz (UCSC) Genome Graphs tool. All UCSC genes whose TSS is located within the 600 kb genomic region illustrated in the graphic are shown. Boxes depict exons. Arrows indicate the direction of transcription. (b) Expression levels by qRT-PCR of the 6 coding genes whose TSSs are located within the 600 kb genomic region shown in a in the LiPS-A3 clone relative to those in the parental Thal5.10-Cre8 iPSC line. Mean and SEM of three experiments are shown.

we used RMCE to exchange the eGFP reporter in the LiPS-A3 line. As a donor for cassette exchange, we constructed an integrase-deficient lentiviral vector (IDLV) carrying a cassette consisting of the puromycin resistance and the mCerulean genes driven by human phosphoglycerate kinase flanked by loxP and lox2272 sites (Figure 4a). This was codelivered with a second integrase-deficient lentiviral vector expressing Cre recombinase together with mCherry (Supplementary Figure S2). After transduction with both vectors at high efficiency (over 85%), and selection with puromycin, 33 resistant clones were picked. Seven clones were found to have lost the neomycin resistance gene by PCR. Three of these (PC3, PC7, PC10) were further tested. They were shown to express mCerulean and to have lost eGFP expression (Figure 4c). Cassette exchange was further tested and confirmed by multiple PCR primer sets (Figure 4b,d) and by Southern blotting (Figure 4e). mCerulean expression was maintained homogeneously in all cells (Figure 4c). Cre recombinase expression or DNA could not be detected by flow cytometry or qPCR, respectively (Supplementary Figure S2d, e). The cells maintained all features of pluripotency, expressed pluripotent cell markers and could form teratomas comprising tissues of all three embryonic germ layers (Figure 5a,b) and were karyotypically normal (Figure 5c).

These results show that the LiPS-A3 line can be used to efficiently and rapidly introduce transgenes at a known genomic site that can maintain robust and long-term expression.

## Discussion

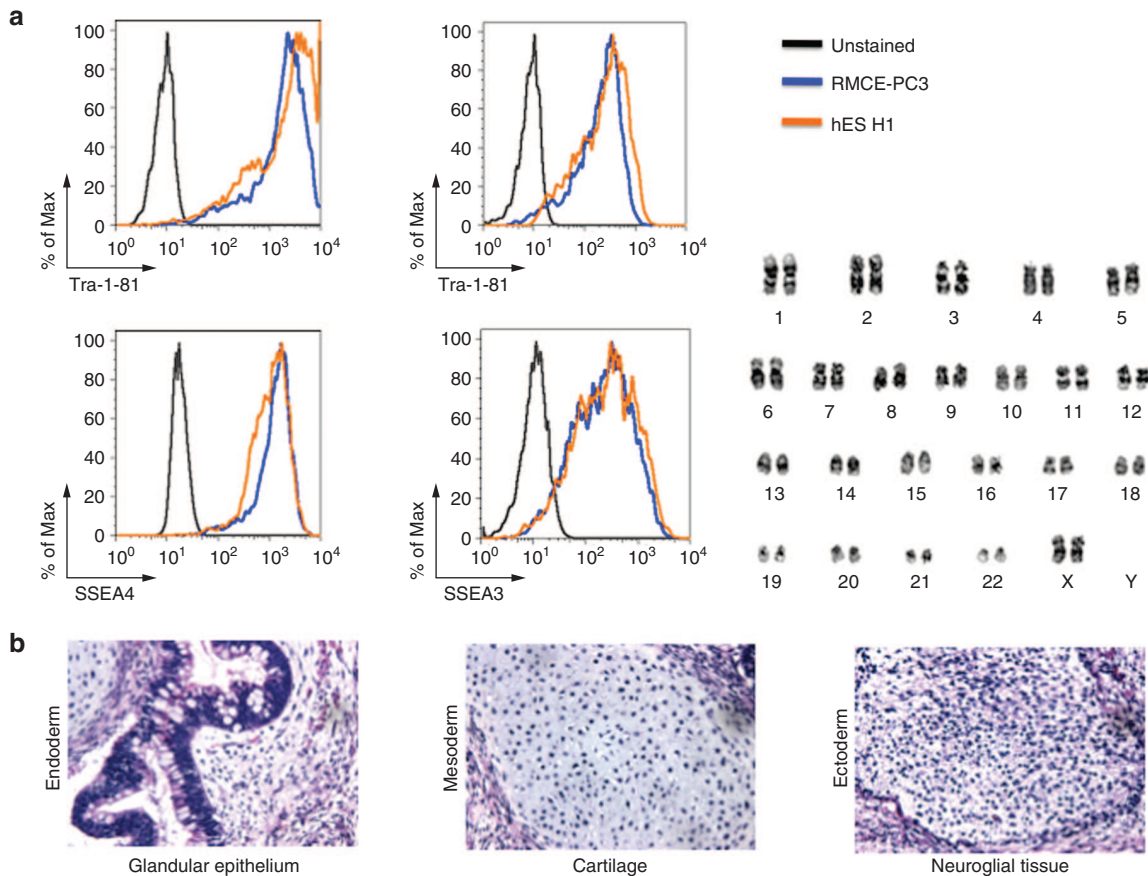
The mouse Rosa26 locus, a locus widely used for transgenesis in the mouse, was discovered through a retroviral gene trap screen.<sup>22</sup> Because lentiviral vectors preferentially integrate in transcriptionally active genomic sites,<sup>23</sup> here we used a lentiviral screen to capture genomic sites enriched in actively transcribed sites. As expected from this, over half of the sites we retrieved (25 out of 40) expressed eGFP (Supplementary Table S1 and Supplementary Figure S1). The locus control region (LCR) elements (DNase-I-hypersensitive sites) from the beta-globin locus contained in the TNS9.3/eNG vector that we used for this screen may have further contributed to preserving an open chromatin configuration and active transcription in the neighboring human phosphoglycerate kinase-driven reporter gene.<sup>24,25</sup> Of the 15 sites that did not support eGFP expression, approximately half were silenced early on, whereas the other half contained mostly eGFP<sup>+</sup> cells shortly after transduction, but showed gradual silencing over several days. These kinetics are consistent with two distinct silencing events, also described by others,<sup>26</sup> one early and rapid and one late that extinguishes transgene expression over a prolonged period of time. In some of the clones that showed late gradual silencing, but not in others, eGFP<sup>+</sup> cells could be sorted or picked under a fluorescent microscope and maintained stably expressing eGFP thereafter. While lentiviral integrations are enriched in transcriptionally active genes, they are at the same time more



**Figure 4 Recombinase-mediated cassette exchange.** (a) Scheme of the RMCE approach. The cassette flanked by two heterospecific loxP sites in the vector already inserted in the LiPS-A3S locus (hPGK-Neo-eGFP) can be exchanged with a compatible donor cassette (hPGK-Puro-mCerulean) containing the same configuration of loxP sites in the presence of Cre recombinase. In this study, both the Cre recombinase and the donor cassette were delivered by integrase-deficient lentiviral vectors (see also **Supplementary Figure S2**). (b) Scheme of the cassette inserted in the LiPS-A3S locus before and after RMCE. (c) Confirmation of cassette exchange by flow cytometric evaluation of eGFP and mCerulean. (d) PCR for confirmation of successful exchange by RMCE in 3 clones of the LiPS-A3 line (PC3, PC7, PC10) with primers shown in b. (e) Southern blot for further verification of cassette exchange in the 3 clones and the maternal line after XbaI digestion with a Neo and a Puro probe as indicated. The scheme of the locus with the position of the XbaI restriction sites is shown in b.

likely to reside in the vicinity of endogenous genes and hence perturb their expression. We previously found that less than 20% of lentiviral integrations genome-wide meet our core “safe harbor” criteria.<sup>18</sup> In agreement with this, the present set of sites contained 2 out of 27 sites that met the criteria, none of which expressed eGFP. Furthermore, we found an

overall inverse correlation between probability of transgene expression and location outside of or far from coding (but not noncoding) genes (Figure 2). In other words, genomic sites inside or in proximity to coding genes were enriched in eGFP<sup>+</sup> compared to eGFP<sup>-</sup> clones. In agreement with our previous estimates, our experimental scheme enabled us



**Figure 5 The LiPS-A3 line is pluripotent and forms trilineage teratomas.** (a) Flow cytometry of pluripotency markers Tra-1-81, Tra-1-60, SSEA3 and SSEA4 in the LiPS-A3 RMCE-PC3 line. (b) Hematoxylin-eosin staining of a representative teratoma derived from the LiPS-A3 RMCE-PC3 line showing differentiation into tissues of all embryonic germ layers. (c) G-band karyotyping of the LiPS-A3 RMCE-PC3 line.

to retrieve approximately 1 transgene-expressing clone for which clonality, single integration and the location of the latter in the genome could be thoroughly established for every 10 clones initially picked (20 out of 202 in the current study, **Figure 1c**).<sup>18,19</sup>

The process of RMCE provides the possibility of deriving a variety of iPSC lines expressing transgenes of choice from a single well-characterized paternal line.<sup>27,28</sup>

Although transgene expression in differentiated cells of various lineages needs to be further tested, the LiPS-A3 line described here can potentially be of use in a broad range of applications. These include constitutive or inducible knock-down or overexpression of genes for functional genetics studies, ectopic expression of genes for reprogramming or lineage conversion, insertion of reporters for lineage tracing studies and introduction of therapeutic transgenes for pre-clinical gene therapy studies.<sup>6,29,30</sup> RMCE multiplexing using different recombination systems (Cre/loxP and Flp/FRT) can further enable more complex serial RMCE-based modifications at the LiPS-A3S locus.<sup>12,31</sup> RMCE is at least as easy and user-accessible as endonuclease-based gene targeting, as it only requires expression of Cre recombinase and construction of a compatible donor, similarly to the requirement for expression of the nuclease and delivery of a donor DNA in a nuclease-based approach. A potential advantage of RMCE

over gene targeting mediated by site-specific nucleases is that off-target effects can be avoided and the need to test for them is eliminated.

The broader impact of this work lies in the identification of a genomic site—the LiPS-A3S locus on chromosome 15—for reliable transgene expression. With gene targeting tools becoming increasingly accessible to the wider scientific community the need for predetermined sites that can support reliable transgene expression is becoming an urgent need.<sup>32</sup> To the best of our knowledge, this is the only genomic site other than AAVS1 that has been shown to support robust and long-term expression of multiple transgenes in hPSCs. The future development of genome editing tools specific for this site could expand the applications and facilitate the generation of hPSC lines with multiple reporters. These may include any type of site-specific endonuclease: zinc-finger nucleases,<sup>33</sup> meganucleases,<sup>34</sup> transcription activator-like effector nucleases<sup>35</sup> or the CRISPR-Cas9 system.<sup>36,37</sup> Transgene expression afforded by the LiPS-A3S locus in multiple lineages, cell types, and developmental stages can easily be assessed using the existing LiPS-A3 line with *in vitro* differentiation and/or teratoma formation assays.

Finally, while this site does not meet our previously proposed “safe harbor” criteria, it is far from known cancer genes and its safety features, including the dysregulation of

endogenous genes in the neighborhood of the site by integrated transgenes, can be tested in the future. Its evaluation as a candidate “safe harbor” site may open possibilities for its use in human cell and gene therapy relying on gene addition. While we did not find evidence for perturbation of expression of nearby endogenous genes by our vector integrated in the LIPSC-A3S site, this would also need to be tested in diverse differentiated cell types. Also, integration of diverse expression cassettes with different constitutive or tissue-specific promoters and regulatory elements can impact the expression of transgenes from this site, as well as the interaction of the integrated cassette with neighboring genetic elements and will thus need to be tested on a case-by-case basis.

In conclusion, we report here a human iPSC line with a preinserted cassette that can easily be exchanged with any transgene of choice by RMCE, obviating the need to test for off-target effects, as well as a new locus of the human genome that can support reliable transgene expression. These new tools can facilitate the transgenesis of hPSCs and their differentiated progeny.

## Materials and Methods

**Cell culture.** The thal-iPSC line thal-5-10-Cre8 has previously been described.<sup>18</sup> Human iPSCs were cultured on a feeder layer of mitomycin C-treated mouse embryonic fibroblasts (GlobalStem, Rockville, MD) or in feeder-free conditions and passaged with dispase or accutase as previously described.<sup>18,38,39</sup>

**Lentiviral vector construction and production.** The TNS9.3/eNG vector was derived from TNS9.3/fNG<sup>18</sup> by substituting the 3' loxP site with a lox2272 site in inverse orientation relative to the loxP site. Lentiviral vector packaging was performed by triple co-transfection of the plasmid DNA encoding the vector, pUCMD.G and pCMVΔR8.91 into 293T cells, as described.<sup>18,19</sup> For packaging of the integrase-deficient lentiviral vector encoding the Cre recombinase, pCMVΔR8.91 was replaced by pCMVΔR8.91N/N.

**Flow cytometry.** Undifferentiated iPSCs were dissociated with accutase, and stained with Alexa Fluor 647-conjugated anti-Tra-1–81 or anti-Tra-1–60 or anti-SSEA3 or anti-SSEA4 (BD Biosciences, San Jose, CA). Data were acquired in a LSRII cytometer (BD Biosciences) and analyzed with the FlowJo software (Tree Star, Ashland, OR).

**Teratoma formation assay.** For teratoma formation assays, iPSCs were suspended in hES medium containing 10 μmol/l of the Rho-associated kinase (Rock) inhibitor Y-27632 (Tocris, Bristol, UK). Approximately 2 million cells were injected intramuscularly into NOD-SCID *IL2Rγ*-null (NSG) mice (Jackson Laboratory, Bar Harbor, ME). Five to six weeks later, the tumors were surgically dissected and fixed in 4% formaldehyde. Cryosectioned samples were stained with hematoxylin and eosin for histological analysis. All animal experiments were conducted in accordance with protocols approved by the University of Washington Institutional Animal Care and Use Committee and following National Institutes of Health guidelines for animal welfare.

**Transduction of iPSCs and screening of colonies.** Transduction of iPSCs and screening of colonies was performed as described previously.<sup>18,19</sup> Forty-eight hours after transduction, the cells were harvested with accutase and vigorously triturated into single cells. An aliquot of the cells was used for analysis of eGFP expression by flow cytometry. Transduced cells with gene transfer <30% (as estimated by the percentage of eGFP<sup>+</sup> cells) were replated at a density of 1,500 cells/cm<sup>2</sup> on a layer of Neo-resistant mitomycin C-treated mouse embryonic fibroblasts (GlobalStem). G418 (Invitrogen, Carlsbad, CA) was added at a concentration of 12.5 μg/ml between days 5 and 9 after transduction. Approximately 20 days post-transduction, Neo-resistant colonies were manually picked and replated into six-well dishes on mitomycin C-treated mouse embryonic fibroblasts. One week later, ~100–200 cells from each clone were manually picked and lysed as described.<sup>19</sup> Measurement of vector copy number was performed with multiplex qPCR using a standard curve, as described.<sup>18</sup> For Southern blot analysis, 5–10 μg of genomic DNA was digested with EcoRI and a radio-labeled fragment spanning the eGFP coding region was used as probe, as described.<sup>19</sup>

**Integration site analysis.** TNS9.3/eNG vector integration sites in single-copy clones were mapped by inverse PCR (iPCR), as described.<sup>19</sup> Integration sites were judged to be authentic if the sequences were adjacent to vector LTR ends and had a unique hit when aligned to the draft human genome (University of California Santa Cruz, UCSC hg19). Genomic annotations were also obtained from UCSC hg19 Genome Browser and mapped against the integration sites. The integration sites of clones A3, A29, 72 and 99 were confirmed by integration-specific PCR using an LTR universal forward primer and reverse primers corresponding to the specific genomic sequence adjacent to the vector integration.

**Gene expression analysis by qRT-PCR.** RNA was isolated with Trizol (Life Technologies, Carlsbad, CA). Reverse transcription was performed with Superscript III (Life Technologies) and qPCR was performed with the SsoFast EvaGreen Supermix (Bio-Rad, Hercules, CA) using primers DPP8-F: GCT TGG TCC ATC CTA CTA GAT CG, DPP8-R: TAG TGG CGT CAC AGA ATC AGG, PTPLAD1-F: CTT AGA CCT TGT GAA ACC AGA G, PTPLAD1-R: CTG CTT TGT GAG TCT CTC CCA, VWA9-F: CTG CTT TGT GAG TCT CTC CCA, VWA9-R: GTT GAT TTT GAG TGG CTA GGG, SLSC24A1-F: TCA AGG GAG ATC AGA AGG AGA AT, SLSC24A1-R: GCG TGG ATT TCA CCC TCA TCT T, DENND4A-F: CCA CAC ACG TTC TGC AAA GTC, DENND4A-R: AGG AAC AGA TTC CGG TAG TGA, RAB11A-F: CAA CAA GAA GCA TCC AGG TTG A, RAB11A-R: GCA CCT ACA GCT CCA CGA TAA T, ACTIN-F: TGA AGT GTG ACG TGG ACA TC, ACTIN-R: GGA GGA GCA ATG ATC TTG AT. Reactions were carried out in triplicate in a 7500 Fast Real-Time PCR System (Applied Biosystems, Foster city, CA).

**RMCE.** The donor pLM-ePC vector was derived from vector pLM-mCerulean-P2A-cMYC<sup>38</sup> after replacing the cMYC cDNA with the Puromycin resistance gene and inserting loxP sites. The vector was packaged as an integrase-deficient lentiviral vector, as described above. For Southern blot analyses,

5–10 µg of genomic DNA was digested with XbaI. Two radiolabeled fragments corresponding to the Neomycin resistance and the Puromycin resistance gene sequences were used as probes.

**Cre recombinase detection.** For mCherry detection, cells were assayed on a BD Fortessa and data were analyzed with FlowJo software (Tree Star). For qPCR, gDNA was isolated with the DNeasy Blood and Tissue kit (Qiagen, Hilden, Germany). qPCR was performed with the SsoFast EvaGreen Supermix (Bio-Rad) using primers Cre-F: TTA CGG CGC TAA GGA TGA CT and Cre-R: TGC ATG ATC TCC GGT ATT GA. Reactions were carried out in triplicate in a 7500 Fast Real-Time PCR System (Applied Biosystems).

### Supplementary material

**Figure S1.** Transgene expression in single vector copy clones.

**Figure S2.** IDLV delivery of the Cre recombinase and the donor DNA.

**Table S1.** Integration sites mapped in single vector-copy clones.

**Acknowledgments.** This work was supported by the National Institutes of Health (NIH) grant R00 DK087923 (E.P.P.).

- Ellis, J (2005). Silencing and variegation of gammaretrovirus and lentivirus vectors. *Hum Gene Ther* **16**: 1241–1246.
- Placantonakis, DG, Tomishima, MJ, Lafaille, F, Desbordes, SC, Jia, F, Socci, ND et al. (2009). BAC transgenesis in human embryonic stem cells as a novel tool to define the human neural lineage. *Stem Cells* **27**: 521–532.
- Tomishima, MJ, Hadjantonakis, AK, Gong, S and Studer, L (2007). Production of green fluorescent protein transgenic embryonic stem cells using the GENSAT bacterial artificial chromosome library. *Stem Cells* **25**: 39–45.
- Papapetrou, EP and Schambach, A (2016). Gene Insertion Into Genomic Safe Harbors for Human Gene Therapy. *Mol Ther* **24**: 678–684.
- Kotin, RM, Linden, RM and Berns, KI (1992). Characterization of a preferred site on human chromosome 19q for integration of adeno-associated virus DNA by non-homologous recombination. *EMBO J* **11**: 5071–5078.
- Yang, L, Soonpaa, MH, Adler, ED, Roepke, TK, Kattman, SJ, Kennedy, M et al. (2008). Human cardiovascular progenitor cells develop from a KDR+ embryonic-stem-cell-derived population. *Nature* **453**: 524–528.
- Garçon, L, Ge, J, Manjunath, SH, Mills, JA, Apicella, M, Parikh, S et al. (2013). Ribosomal and hematopoietic defects in induced pluripotent stem cells derived from Diamond Blackfan anemia patients. *Blood* **122**: 912–921.
- Merling, RK, Sweeney, CL, Chu, J, Bodansky, A, Choi, U, Priel, DL et al. (2015). An AAVS1-targeted minigenome platform for correction of iPSCs from all five types of chronic granulomatous disease. *Mol Ther* **23**: 147–157.
- Sim, X, Cardenas-Diaz, FL, French, DL and Gadue, P (2016). A doxycycline-inducible system for genetic correction of iPSC disease models. *Methods Mol Biol* **1353**: 13–23.
- Ordovás, L, Boon, R, Pistoni, M, Chen, Y, Wolfs, E, Guo, W et al. (2015). Efficient recombinase-mediated cassette exchange in hPSCs to study the hepatocyte lineage reveals AAVS1 locus-mediated transgene inhibition. *Stem Cell Reports* **5**: 918–931.
- González, F, Zhu, Z, Shi, ZD, Lelli, K, Verma, N, Li, QV et al. (2014). An iCRISPR platform for rapid, multiplexable, and inducible genome editing in human pluripotent stem cells. *Cell Stem Cell* **15**: 215–226.
- Chen, Y, Cao, J, Xiong, M, Petersen, AJ, Dong, Y, Tao, Y et al. (2015). Engineering human stem cell lines with inducible gene knockout using CRISPR/Cas9. *Cell Stem Cell* **17**: 233–244.
- Mandegar, MA, Huebsch, N, Frolow, EB, Shin, E, Truong, A, Olvera, MP et al. (2016). CRISPR interference efficiently induces specific and reversible gene silencing in human iPSCs. *Cell Stem Cell* **18**: 541–553.
- Irion, S, Luche, H, Gadue, P, Fehling, HJ, Kennedy, M and Keller, G (2007). Identification and targeting of the ROSA26 locus in human embryonic stem cells. *Nat Biotechnol* **25**: 1477–1482.
- Lombardo, A, Cesana, D, Genovese, P, Di Stefano, B, Provasi, E, Colombo, DF et al. (2011). Site-specific integration and tailoring of cassette design for sustainable gene transfer. *Nat Methods* **8**: 861–869.
- Perez, EE, Wang, J, Miller, JC, Jouvenot, Y, Kim, KA, Liu, O et al. (2008). Establishment of HIV-1 resistance in CD4+ T cells by genome editing using zinc-finger nucleases. *Nat Biotechnol* **26**: 808–816.
- Zhu, F, Gamboa, M, Farruggio, AP, Hippenmeyer, S, Tasic, B, Schüle, B et al. (2014). DICE, an efficient system for iterative genomic editing in human pluripotent stem cells. *Nucleic Acids Res* **42**: e34.
- Papapetrou, EP, Lee, G, Malani, N, Setty, M, Riviere, I, Tirunagari, LM et al. (2011). Genomic safe harbors permit high β-globin transgene expression in thalassemia induced pluripotent stem cells. *Nat Biotechnol* **29**: 73–78.
- Papapetrou, EP and Sadelain, M (2011). Derivation of genetically modified human pluripotent stem cells with integrated transgenes at unique mapped genomic sites. *Nat Protoc* **6**: 1274–1289.
- Sadelain, M, Papapetrou, EP and Bushman, FD (2011). Safe harbours for the integration of new DNA in the human genome. *Nat Rev Cancer* **12**: 51–58.
- Lee, G and Saito, I (1998). Role of nucleotide sequences of loxP spacer region in Cre-mediated recombination. *Gene* **216**: 55–65.
- Zambrowicz, BP, Imamoto, A, Fiering, S, Herzenberg, LA, Kerr, WG and Soriano, P (1997). Disruption of overlapping transcripts in the ROSA beta geo 26 gene trap strain leads to widespread expression of beta-galactosidase in mouse embryos and hematopoietic cells. *Proc Natl Acad Sci USA* **94**: 3789–3794.
- Schröder, AR, Shinn, P, Chen, H, Berry, C, Ecker, JR and Bushman, F (2002). HIV-1 integration in the human genome favors active genes and local hotspots. *Cell* **110**: 521–529.
- Grosveld, F, van Assendelft, GB, Greaves, DR and Kollias, G (1987). Position-independent, high-level expression of the human beta-globin gene in transgenic mice. *Cell* **51**: 975–985.
- May, C, Rivella, S, Callegari, J, Heller, G, Gaensler, KM, Luzzatto, L et al. (2000). Therapeutic haemoglobin synthesis in beta-thalassaemic mice expressing lentivirus-encoded human beta-globin. *Nature* **406**: 82–86.
- Rival-Gervier, S, Lo, MY, Khattak, S, Pasceri, P, Lorincz, MC and Ellis, J (2013). Kinetics and epigenetics of retroviral silencing in mouse embryonic stem cells defined by deletion of the D4Z4 element. *Mol Ther* **21**: 1536–1550.
- Bouhassira, EE, Westerman, K and Leeboulch, P (1997). Transcriptional behavior of LCR enhancer elements integrated at the same chromosomal locus by recombinase-mediated cassette exchange. *Blood* **90**: 3332–3344.
- Ramachandra, CJ, Shahbazi, M, Kwang, TW, Choudhury, Y, Bak, XY, Yang, J et al. (2011). Efficient recombinase-mediated cassette exchange at the AAVS1 locus in human embryonic stem cells using baculoviral vectors. *Nucleic Acids Res* **39**: e107.
- Kuehle, J, Turan, S, Cantz, T, Hoffmann, D, Suerth, JD, Maetzig, T et al. (2014). Modified lentiviral LTRs allow Flp recombinase-mediated cassette exchange and *in vivo* tracing of “factor-free” induced pluripotent stem cells. *Mol Ther* **22**: 919–928.
- Haenebalkke, L, Goossens, S, Dierickx, P, Bartunkova, S, D’Hont, J, Haigh, K et al. (2013). The ROSA26-iPSC mouse: a conditional, inducible, and exchangeable resource for studying cellular (De)differentiation. *Cell Rep* **3**: 335–341.
- Turan, S, Qiao, J, Madden, S, Benham, C, Kotz, M, Schambach, A et al. (2014). Expanding Flp-RMCE options: the potential of Recombinase Mediated Twin-Site Targeting (RMTT). *Gene* **546**: 135–144.
- Hockemeyer, D, Wang, H, Kiani, S, Lai, CS, Gao, Q, Cassady, JP et al. (2011). Genetic engineering of human pluripotent cells using TALE nucleases. *Nat Biotechnol* **29**: 731–734.
- Porteus, MH and Carroll, D (2005). Gene targeting using zinc finger nucleases. *Nat Biotechnol* **23**: 967–973.
- Pâques, F and Duchateau, P (2007). Meganucleases and DNA double-strand break-induced recombination: perspectives for gene therapy. *Curr Gene Ther* **7**: 49–66.
- Boch, J (2011). TALEs of genome targeting. *Nat Biotechnol* **29**: 135–136.
- Mali, P, Yang, L, Esvelt, KM, Aach, J, Guell, M, DiCarlo, JE et al. (2013). RNA-guided human genome engineering via Cas9. *Science* **339**: 823–826.
- Doudna, JA and Charpentier, E (2014). Genome editing. The new frontier of genome engineering with CRISPR-Cas9. *Science* **346**: 1258096.
- Papapetrou, EP, Tomishima MJ, Chambers SM, Mica Y, Reed E, Menon J et al. (2009). Stoichiometric and temporal requirements of Oct4, Sox2, Klf4, and c-Myc expression for efficient human iPSC induction and differentiation. *Proc Natl Acad Sci USA* **106**: 12759–12764.
- Papapetrou, EP and Sadelain, M (2011). Generation of transgene-free human induced pluripotent stem cells with an excisable single polycistronic vector. *Nat Protoc* **6**: 1251–1273.



This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>

© The Author(s) (2016)

Supplementary Information accompanies this paper on the Molecular Therapy–Nucleic Acids website (<http://www.nature.com/mtna>)