Gene Therapy
https://doi.org/10.1038/s41434-020-0144-x

ARTICLE

Targeted gene therapy into a safe harbor site in human hematopoietic progenitor cells

Fatima Rodriguez-Fornes[1] · Oscar Quintana-Bustamante[1] · M. Luz Lozano[1] · Jose C. Segovia[1] · Juan A. Bueren[1] · Guillermo Guenechea[1]

Received: 9 July 2019 / Revised: 10 March 2020 / Accepted: 12 March 2020

Abstract

Directed gene therapy mediated by nucleases has become a new alternative to lead targeted integration of therapeutic genes in specific regions in the genome. In this work, we have compared the efficiency of two nuclease types, TALEN and meganucleases (MN), to introduce an EGFP reporter gene in a specific site in a safe harbor locus on chromosome 21 in an intergenic region, named here SH6. The efficiency of targeted integration mediated by SH6v5-MN and SH6-TALEN in HEK-293H cells was up to 16.3 and 15.0%. A stable expression was observed both in the pool of transfected cells and in established pseudoclones, with no detection of off-target integrations by Southern blot. In human hematopoietic stem and progenitor $CD34^+$ cells, the nucleofection process preserved the viability and clonogenic capacity of nucleofected cells, reaching up to 3.1% of specific integration of the transgene in colony forming cells when the SH6-TALEN was used, although no expression of the transgene could be found in these cells. Our results show the possibility to specifically integrate genes at the SH6 locus in $CD34^+$ progenitor cells, although further improvements in the efficacy of the procedure are required before this approach could be used for the gene editing of hematopoietic stem cells in patients with hematopoietic diseases.

Introduction

Although lentiviral vector (LV)-mediated gene therapy (GT) is offering clinical data demonstrating the safety and efficacy of this approach in monogenic diseases [1, 2], targeted GT approaches offer further improvements in the safety of GT [3]. The main approach used for targeted GT is based on homologous recombination (HR) [4, 5] promoted by the generation of double strand breaks (DSB) generated by engineered nucleases, such as Zinc-finger nucleases (ZFN), CRISPR/Cas system, meganucleases (MN) and transcription activator-like effector nucleases (TALEN) [6, 7]. Three different strategies of gene editing have been considered for the integration of a donor template in the target cells: gene correction [8, 9], targeted Knock-In [10–13], and gene targeting in a Safe Harbor (SH) [9, 14, 15]. The main advantage of the first two strategies is that the physiological expression of the corrected gene is preserved. However, gene targeting into SH loci could lead to the treatment of different diseases using the same engineered nucleases and just swapping the donor transgene.

Genomic SHs are defined as intragenic or extragenic chromosomal locations where therapeutic transgenes can be integrated to facilitate their function in a predictable manner, without perturbing the activity of the endogenous or surrounding genes [16]. The three main SH used in GT are located in intragenic regions (AAVS1 and CCR5 in human and Rosa26 in murine) and follow three main criteria, consisting in facilitating sufficient transgene expression, not predisposing targeted cells to a malignant transformation or to alterations of their cellular functions and offering a predictable outcome of targeted cells [16]. However, extragenic locations of the cell genome constitute alternative SH

Supplementary information The online version of this article (https://doi.org/10.1038/s41434-020-0144-x) contains supplementary material, which is available to authorized users.

✉ Guillermo Guenechea
g.guenetxea@ciemat.es

[1] Division of Hematopoietic Innovative Therapies, Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas (CIEMAT)/Centro de Investigación Biomédica en Red de Enfermedades Raras (CIBERER)/Instituto de Investigación Sanitaria Fundación Jiménez Díaz (IIS-FJD, UAM), 28040 Madrid, Spain

Published online: 26 March 2020

sites. Five criteria have been considered for the definition of an extragenic SH [17, 18]: (1) distance of more than 50 kb from the 5′ end of any gene, (2) distance of more than 300 kb from cancer-related genes, (3) distance of more than 300 kb from any microRNA, (4) outside a gene transcript unit and (5) outside of ultraconserved regions.

In this study, we have worked with several engineered nucleases, two MNs and one TALEN, all of them targeting a SH, called safe harbor 6 (SH6) locus, characterized by fitting the five criteria exposed above. SH6 is located in an intergenic region in chromosome 21, being the closest ORF (*C21orf37*) more than 350 kb away. SH6 was first studied by Eyquem et al., whose results suggested that it could be a useful locus for transgene expression, although this expression was lower than in other intergenic SHs due to the less permissive epigenetic status of its chromatin [15].

We have developed an approach for gene editing this SH6 locus in a human cell line and in human hematopoietic progenitors. SH6 locus was modified with MNs and TALEN specific nucleases in combination with a donor matrix that carries the *EGFP* reporter gene in HEK-293H cell line and in purified hematopoietic progenitors. In addition, we also used a donor matrix harboring the *FANCA* gene to find out the real therapeutic potential of this gene editing protocol.

Materials and methods

Cell cultures

HEK-293H cell line (Gibco, cat n° 11631017 and mycoplasma free) was maintained in culture medium consisted of DMEM (Dulbecco's Modified Eagle Medium) with GlutaMAX (Gibco/Life Technologies/Thermo Fisher Scientific, Waltham, MA, USA) +10% HyClone (GE Healthcare Life Sciences, Fairfield, USA) +1% penicillin/streptomycin (P/S, Gibco). Cells were grown in adherence at 37 °C, 21% O2, 5% CO_2 and 95% relative humidity (RH) in 75 cm^2 flask (Corning, NY, USA). One million and a half of these cells were seeded in p100 plates and on the next day, plates carrying cells with a confluence around 80% were transfected with Lipofectamine 2000 (Invitrogen) following vendor guidelines. Three days post transfection, cells of each condition were harvested, counted, and plated in two p96 plates (10 cells per well). The rest of the cells were maintained in culture in the same conditions and their EGFP expression was measured by flow cytometry at different time points up to 30 days. P96 plates were duplicated after 7 days in culture and cultured until the wells were confluent. At this point, the duplicated plates were frozen in viability to stablish pseudoclones from the wells carrying the edited cells. The other plates were maintained to extract the gDNA of the cells in order to perform specific targeted analysis.

Umbilical cord blood (UCB) samples from healthy donors were obtained from the Centro de Transfusiones de la Comunidad de Madrid, according to the protocol approved by the Ethical Committee and under informed consent. Mononuclear cells were purified by Ficoll-Paque PLUS (GE Healthcare) density by gradient centrifugation. $CD34^+$ cells were then selected using CD34 MicroBead Kit. Magnetic-labeled cells were isolated with an LS column in QuadroMACS^{TM} separator (MACS, Miltenyi Biotec, Bergisch Gladbach, Germany) following manufacturer's instructions. Purified $CD34^+$ were then evaluated for their purity by flow cytometry. Purities from 95 to 99% were routinely obtained. $CD34^+$ cells were cultured in prestimulation media: StemSpan (StemCell Technologies, Vancouver, Canada) + 1% P/S + 100 ng/mL stem cell factor (SCF, EuroBiosciences, Friesoythe, Germany) + 100 ng/mL FMS-like tyrosine kinase 3 ligand (Flt3L, EuroBiosciences) + 100 ng/mL thrombopoietin (TPO, EuroBiosciences) or in culture media: StemSpan + 1% P/S + 300 ng/mL SCF + 300 ng/mL Flt3L + 300 ng/mL TPO. Cells were cultured in suspension in tissue nontreated 96-well and 24-well plates (Gibco), at 10^6 cells/mL in normoxia: 37 °C, 21% O_2, 5% CO_2 and 95% RH. Around 10^6 cells were nucleofected using the AMAXA Nucleofector kit for human $CD34^+$ cells (Lonza, Basel, Switzerland).

Plasmids

The two SH6-MNs and the SH6-TALEN were transferred by Cellectis (Paris, France). Two plasmids encoding for the SH6v2 or the SH6v5 I-CreI derived MNs isosquizomers, targeting the TTAATACCCCGTACCTAAT ATTGC sequence in the SH6 locus, were used in this work. MNs are constituted by two subunits that are fused by a linker and transcribed by a unique plasmid. The SH6v5 MN was also delivered as mRNA, synthesized in vitro by amplifying the MNs sequence by PCR and cloned into the plasmid of the Zero Blunt PCR Cloning Kit (Thermo Fisher Scientific) carrying a T7 promoter. MN expression was driven by the cytomegalovirus promoter.

Left and right TALEN subunits, targeting SH6 locus, are present in two plasmids provided by Cellectis and targeting the TCTAAAGATTAATACCCCGTACCTAATATTGCA TTTCCTTCTACCAGTA sequence. The TALEN will only perform the cleavage after dimerization of the two FokI subunits from each half TALEN. TALEN expression was driven by human elongation factor 1-α (EF1α).

The two donor templates employed in this work, called Green Matrix (GM) and Therapeutic Matrix (TM), were a plasmid that was codelivered with the nucleases. The GM carries an *EGFP* gene which expression was controlled by

the EF1α with intron promoter, both of them flanked by two homology sequences, called homology arms (HAs), to the SH6 target. The TM carries the WT cDNA sequence of the *FANCA* gene with a 3xFlag signaling peptide in the 5′N-termini of the protein, a self-cleaving 2 A peptide (E2A) sequence [19] and a puromycin resistant gene as selectable marker. The long EF1α promoter controls the expression of both previous genes. The four elements are flanked by the same two HAs of the SH6 target as the GM.

Gene targeting experiments

One million and a half of HEK-293H cells were seeded in p100 plates and on the next day, plates carrying cells with a confluence around 80% were transfected with Lipofectamine 2000 (Invitrogen, Carlsbad, CA, USA) following vendor guidelines. Different amounts of the MNs or the TALEN were cotransfected with 4 µg of the GM.

Human $CD34^+$ cells obtained from UCB as described in "Cell Cultures" section were thawed and prestimulated in prestimulation media during 48 h at a density of 10^6 cells/mL. Then, around $1–2 \times 10^5$ (strip) to 10^6 (cuvette) cells were nucleofected using the AMAXA Nucleofector kit for human $CD34^+$ cells or the P3 Primary Cell 4D-Nucleofector L or S Kit (Lonza), depending on the experiment. Different amounts of the MN or the TALEN alone or with 4 µg of the GM donor were used in different experiments. EO-100 program was used with the AMAXA nucleofector. After nucleofection cells were maintained for 10–20 min in the incubator and then transferred to a culture plate with culture media at a density of 5×10^5 cells/mL. At different time points, an aliquot of the cells was processed and the rest was diluted with culture media to 5×10^5 cells/mL. All experimental groups were codified and were blinded until the results were processed.

Molecular analyses

In total, 40 µL of a mix of lysis buffer + proteinase K (100 µg/mL final) were added per well to the HEK-293H cells in the p96 plates and incubated at 55 °C 2 h and at 95 °C 10 min. To prepare 500 mL of lysis buffer: 495.5 mL of 10 mM Tris HCl pH 8 + 2.25 mL of 0.45% (v/v) NP40 (Nonidet P-40 substitute, Roche Diagnostics, Basel, Switzerland) + 2.25 mL of 0.45% (v/v) of Tween 20 (Sigma-Aldrich, Saint Louis, MO, USA), storage at 4 °C. Then the plate was centrifuged and 3 µL of the supernatant was used to perform the PCR for the targeted integration analysis. Primers used for the PCR to detect specific integration in the SH6 *locus*, hybridize inside and outside of the cassette and are indicated in Fig. 1a and described in Table S1. PCR amplicons were resolved in 1% agarose gel stained with EtBr. The gDNA from samples of the established pseudoclones was obtained using the phenol-chloroform protocol [20]. In total, 50 µg of genomic DNA were isolated from each studied pseudoclone and digested with *BamHI* (New England Biolabs, MA, USA) in separate reactions for 8 h at 37 °C. The digested DNA was concentrated by precipitation with ethanol and sodium acetate and was measured with Nanodrop®. In total, 20 µg of each digestion was resolved on a 0.8% TBE agarose gel. The transfer to a NytranSuPerCharge nylon membrane was carried out overnight by turbo Blotter transfer system (Whatman, GE healthcare) following vendor guidelines and then incubated at 80 °C to fix the DNA to the membrane. Prehybridization and hybridization were performed by quick hybridization solution (Agilent technologies, Santa Clara, CA, USA) in rolling cylinders inside the hybridization oven at 65 °C. The probe used was the sequence of the *EGFP* gene, which was generated by digesting the GM DNA with *BamHI*/*SbfI* restriction enzymes and isolating the 892 bp band from an agarose gel. It was labeled with radioactive P^{32}dCTPs by Prime-It II Random Primer Labeling kit (Agilent) following vendor guidelines and the nonincorporated radioactivity was removed by cleaning the probe with IllustraNICK Columns (GE Healthcare). After 15 min of prehybridization, the probe was added to the quick hybridization solution and left overnight. After 12 h, the membrane was washed twice with 2× SSC, 0.1% SDS for 10 min at room temperature, then for 15 min with prewarmed 1× SSC, 0.1% SDS at 65 °C and finally for 50 min with 0.1× SSC, 0.1% SDS at 65 °C with two media changes. Then, the membrane was covered with plastic wrap and introduced in a Hyperscreen autoradiography cassette together with an Amersham Hyperfilm ECL (GE Healthcare) that was exposed overnight at −80 °C. After a minimum of 10 h, the film was revealed in an automated processor.

gDNA from the pool of nucleofected $CD34^+$ cells and from picked single $CD34^+$ derived colonies (that were distributed in p96 plates) were washed in 100–200 µl of PBS, pelleted and resuspended in 10 µl of PBS. Their gDNA was extracted using 20 µl of lysis buffer (0.3 mM tris HCl pH 7.5, 0.6 mM $CaCl_2$, 1,5% glycerol, 0.675% Tween-20 and 0.3 mg/ml proteinase K) and incubated at 65 °C for 30 min, 90 °C for 10 min and 4 °C. After the lysis, 30 µl of water was added and the mixes were centrifuged. Primers used for the Nested-PCR to detect specific integration in the SH6 *locus*, hybridize inside and outside of the cassette and are indicated in Fig. 1a and described in Table S1. PCR amplicons were resolved in 1% agarose gel stained with EtBr.

For the Surveyor assay, samples from HEK-293H or $CD34^+$ cells, in which nucleases were delivered, were harvested at 2 days and their gDNA extracted using the NucleoSpin® Tissue kit (Macherey-Nagel, Düren, Germany). The target site was amplified by PCR using a pair of

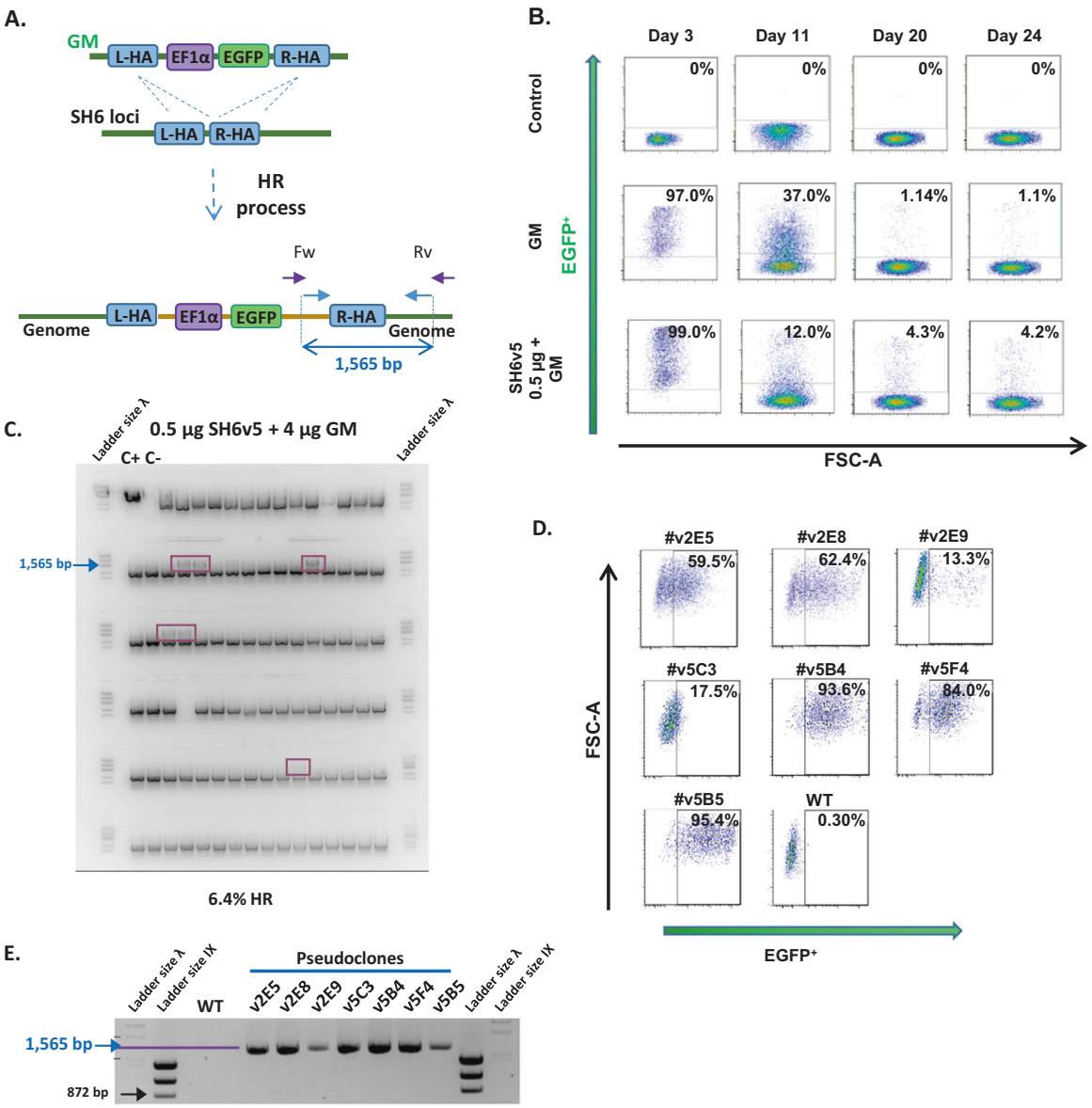

Fig. 1 Integration of a Green Matrix in the SH6 safe harbor site mediated by SH6-meganucleases in HEK-293H cells. a. Representative scheme of the gene targeting strategy in the SH6 using the GM circular plasmid donor. Discontinue blue lines represent the homology areas of the matrix that will recombine with the target site. Purple arrows represent the first pair of primers used in the Nested-PCR to check the specific integration. Blue arrows represent the pair of primers used in the PCR, and the second pair used in the Nested-PCR, to test the specific integration of the GM in the target site. The green arrow shows the size of the amplicon obtained as a positive result in the PCR analyses. **b**. Dot plots representation of the percentages of EGFP$^+$ HEK-293H cells transfected with SH6v5-MNs over time. GM Green matrix. **c**. Agarose electrophoresis gels showing positive bands, inside the purple boxes, corresponding to HR events in p96 plates by the SH6v5-MN. **d**. Representative dot plots of the EGFP expression corresponding to transfected HEK-293H cells pseudoclones. **e**. The agarose gel electrophoresis shows the positive band in tested pseudoclones after the specific PCR (scheme represented in **a**). Ladder size corresponds to λ DNA cut with BstII restriction enzyme and to DNA Molecular Weight Marker IX (0.072–1.35 kbp).

primers binding in the HAs sequence surrounding the target site in the SH6 locus, obtaining a 351 bp band (primers described in Table S1). After the denaturalization and rehybridation of the amplicons, were digested with the CelI enzyme (Surveyor® mutation detection kit, IDT, Coralville, IA, USA) and this product was resolved in 10% TBE Gels 1.0 mm (Invitrogen). A positive result is represented by the appearance of two additional bands of around 179 and 171 bp. The relative intensity of these bands, compared with the band of 351 bp in the top, was indicative of the percentage of activity of specific cleavage of the nuclease in the SH6 site.

For the off-target analyses, in-silico analyses of possible off-targets with five or six mismatches were performed (Table S2). Sequencing of the genomic DNA from the pool of cells of the different conditions samples were used to

perform the PCR for the off-target analysis. Primers used for the PCR to detect integrations in the different off-targets are indicated in Table S1. PCR amplicons were sent for sequencing and the resulting sequence was introduced in the ICE program from Synthego (https://ice.synthego.com/#/), entering 15–20 nt of the off-target sites containing NGG sequence as hypothetic gRNA.

mRNA synthesis

In order to be linearized, T7-MN plasmid was digested with BamHI and XhoI restriction enzymes. This digestion was purified using sodium acetate buffer solution, pH = 5.2 (Sigma-Aldrich) and ethanol. The amount of DNA was measured using a Nanodrop®. The in vitro synthesis of the mRNA was performed with the mMESSAGE mMACHINE® T7 Ultra Kit, Synthesis of Translation Enhanced Capped Transcripts (Ambion/Life Technologies/Thermo Fisher Scientific). First, 1 µg of the digested and purified product was used to synthesize mRNA following the vendors' guidelines and was purified by LiCl precipitation. To obtain a proper quantity of mRNA, three reactions were performed for each plasmid. For this reason, the quantities of each reactive during the protocol is three times higher. Finally, the amount of mRNA was measured with Nanodrop®, aliquoted and stored at −80 °C.

In order to test the quality of the synthesized mRNA before and after the polyadenylation reaction, an aliquot was run in an agarose gel in denaturing conditions (with formaldehyde).

Flow cytometry analyses

Analyses were performed in the LSRFortessa cell analyzer (BD/Becton, Dickinson and Company, NJ, USA). Off-line analysis was conducted with the FlowJo Software v7.6.5 (Tree star, Ashland, USA).

The measurement of the percentage of the population of cells expressing the *EGFP* gene was performed by using an aliquot of the pool of cells transfected/nucleofected with the nucleases or cells from the pseudoclones or from the saturated methylcellulose plates. Cells were washed and resuspended in the flow cytometry buffer (PBS containing 0.5% BSA and 0.05% sodium azide) containing 3.3 µg/mL of 4′,6-diamidino-2-phenylindole (DAPI) as a viability marker.

For the primitive population analysis, 2×10^5 $CD34^+$ cells that had been nucleofected with the nuclease and donor combination were harvested, resuspended in 100 µL of flow cytometry buffer and stained during 30 min at 4 °C with CD34-PE-Cy5 (Immunotech, Vaudreuil-Dorion, Quebec, Canada), CD38-PE (BD), CD45RA-APC eFluor 780 (eBioscience, San Diego, CA, USA) and CD90-APC (BD). Then, cells were washed and resuspended again in the flow cytometry buffer containing 3.3 µg/mL DAPI as a viability marker. The percentage of $EGFP^+$ cells was measured in the total population of cells and also in the different subpopulations marked with different antibodies. $CD34^+/CD38^-$ cells were selected at first and, inside this population, $CD45RA^-$ cells were gated. The most primitive compartment, the HSCs (hematopoietic stem cells with long term engraftment) correspond to cells positive to CD90 ($CD34^+/CD38^-/CD45RA^-/CD90^+$). The $CD90^-$ cells correspond to multipotent progenitors (MPP, short term engraftment) ($CD34^+/CD38^-/CD45RA^-/CD90^-$) [21].

Clonogenic assays

In total, 300 to 1500 $CD34^+$ cells were resuspended in 3 mL of methylcellulose (Miltenyi Biotec, StemMACS HSC-CFU). Each mL of the triplicate was seeded in a M35 plate and cells were incubated in normoxia: 37 °C, 21% O_2, 5% CO_2 and 95% RH. Fourteen days later, the number of colonies was counted using an inverted microscope (Olympus IX70 WH10x/22, objective 4×) and the CFU-GMs (granulocyte-macrophage colony forming unit) and BFU-Es (Burst Forming Unit-Erythroid) were identified. Single colonies were picked to analyze by specific PCR the integration of the GM donor in the SH6 locus.

NSG mice

NOD.Cg-Prkdc^{scid} Il2rg^{tm1Wjl}/SzJ (NSG) mice were obtained from the Jackson Laboratory (Bar Harbor, ME) and maintained at the CIEMAT animal facility (registration number 28079–21 A). All experimental procedures were carried out according to European Directive 2010/63/UE on the use and protection of mammals used for experimentation and other scientific purposes, approved by the ethical committee and the General Direction of Environment (PROEX 070/15). NSG female mice of 8–10 weeks old were irradiated with 1.5 Gy 24 h prior to transplantation. Among 10^5 and 4.5×10^5 nucleofected $CD34^+$ cells were transplanted intravenously into the retrobulbar venous plexus at day 6 of the experiment. Depending on the number of cells obtained we could transplant between 1 and 2 mice per experiment. Thirty and 60 days after transplant, bone marrow aspirations were performed and 90 days post transplant, mice were sacrificed and whole bone marrow was collected. Mice were codified and the analyses were blinded until results were processed. All bone marrow samples were processed to analyze human reconstitution, percentage of gene targeted cells and also multilineage engraftment by flow cytometry with huCD45 (#304014, BioLegend, San Diego, CA, USA), huCD19 (#25–0198, eBioscience), huCD33 (#A07775, Beckman Coulter), and huCD34 (#555824, BD) monoclonal antibodies.

Results

Integration of a Green Matrix in the SH6 safe harbor site mediated by SH6-meganucleases

A donor plasmid harboring HAs to the SH6 site was generated (Fig. 1a). This matrix, referred throughout as the GM, carries an *EGFP* reporter gene regulated by the EF1α promoter. In order to evaluate the transfection efficiency and functionality of the GM, $1–2 \times 10^6$ HEK-293H cells were transfected with 4 µg of the GM, or 4 µg of a control plasmid carrying both, the *GFP* and the *Cherry* reporter genes (GFP/Cherry plasmid). The percentage of cells expressing the *EGFP* gene was measured by flow cytometry at 3, 11, 20, and 24 days post transfection. Up to 97% of HEK-293H cells were EGFP-positive at 3 days post transfection, while the expression of the transgene almost disappeared at 24 days post transfection (Fig. S1).

In the following experiments, the efficiency and specificity of the SH6v5 MN (SH6v5-MN) to induce HR events of the GM in the SH6 site was investigated. With this aim, 0.5 µg of plasmid DNA encoding SH6v5-MN was transfected together with 4 µg of the GM in HEK-293H cells. This culture was analyzed by flow cytometry up to 37 days post transfection to determine the proportion of EGFP-positive cells. In parallel, another cell sample was counted at 3 days post transfection and 10 cells per well were plated and cultured in 96-well plates for 1–2 weeks. At this time, plates were duplicated and after 1–2 additional weeks, half of the plates were frozen to preserve the cell viability and the other half were used to quantify HR events by PCR, using specific primers binding sequences inside and outside of the cassette (in/out PCRs). Frozen samples were used to establish pseudoclone cell lines from wells with SH6 targeted cells to perform confirmatory in/out PCRs, flow cytometry and Southern blot analyses (See schematic protocol in Fig. S2).

Analyses of EGFP expression in HEK-293H cells transfected with the GM, either with or without the MN, showed a progressive reduction in the percentage of $EGFP^+$ cells (Fig. 1b). Nevertheless, when cells were cotransfected with the GM and the SH6v5-MN, a higher proportion of $EGFP^+$ cells (4.0%) was observed at 24 days post nucleofection, compared with the other conditions (See Fig. 1b).

In/out PCR analyses performed in 96-well plates at 2–4 weeks post transfection (Fig. 1a, c) showed that the SH6v5 MN mediated percentages of HR of 6.4% (Fig. 1c), consistent with the 4% of $EGPF^+$ cells determined in the flow cytometry studies. When EGFP expression analyses were performed in those plates by UV microscope we could detect 12 positive wells. On the other hand, six wells were positive for HR. Therefore, six wells corresponded to unspecific integration (6.4%). This result is in accordance with other works that described 10–40% of unspecific integration [10, 22].

As described in Figs. S2, 7 pseudoclones could be established from wells carrying targeted cells as deduced from initial in/out PCR analyses. The presence of cells expressing the *EGFP* gene in these pseudoclones, visualized by flow cytometry analyses, was indicative of the activity of the long EF1α promoter to drive stable EGFP expression (Fig. 1d). Percentages of $EGFP^+$ cells in those pseudoclones were variable depending on the number of seeded cells contributing to the pseudoclones and the MFI was also variable because of different multiple copy integrations in the resulting pseudoclones. gDNA was obtained from these pseudoclones and the specific PCR performed. The specific bands obtained in the seven pseudoclones were indicative of a specific integration of the GM mediated by the SH6v5-MNs, in the SH6 site (Fig. 1e).

In order to determine if SH6v5-MNs induced off-target integrations of the GM, Southern blot analyses were performed using a probe for the *EGFP* transgene (Figure S3A). As shown in Fig. S3B, the expected band of 3946 bp was observed in v5-pseudoclones, indicating the specific integration of the GM in the SH6 site. An additional band of around 7242 bp was observed in pseudoclones v5B4, v5F4, and v5B5 (Fig. S3B), which could correspond to additional integrations of the GM or multiple copy integration at SH6 locus (Fig. S3C).

Similar experiments were performed using another MN, SH6v2-MN, which results were alike to those obtained previously with SH6v5-MN (Figs. 1d, e, S3B and S4).

Thus, the cotransfection of a GM together with specific MNs targeting SH6 locus facilitates the stable HR at this locus of an EGFP expression cassette in a human cell line.

Comparative analysis of TALE nucleases versus MNs to mediate homologous recombination in the SH6 safe harbor site

In a next set of experiments we investigated the efficiency of a designed SH6-TALE nuclease (SH6 TALEN) to induce targeted integration of the GM in the SH6 site of HEK-293H cells, as compared with the SH6v5-MN

The comparative activity of both nucleases to target the *SH6* locus was first tested by performing Surveyor assays. With this aim, four doses of each SH6-TALEN subunit were used to compare its activity with respect to 0.5 µg of the SH6v5-MN. The intensity of the 179 and 171 bp bands, compared with the parental band of 351 bp, represents the activity of the different doses and type of nucleases used in these experiments (Fig. S5). The quantification of these bands showed a similar nuclease activity (between 40.0 and 60.0%) of all tested conditions to target the SH6 site (Fig. 2a).

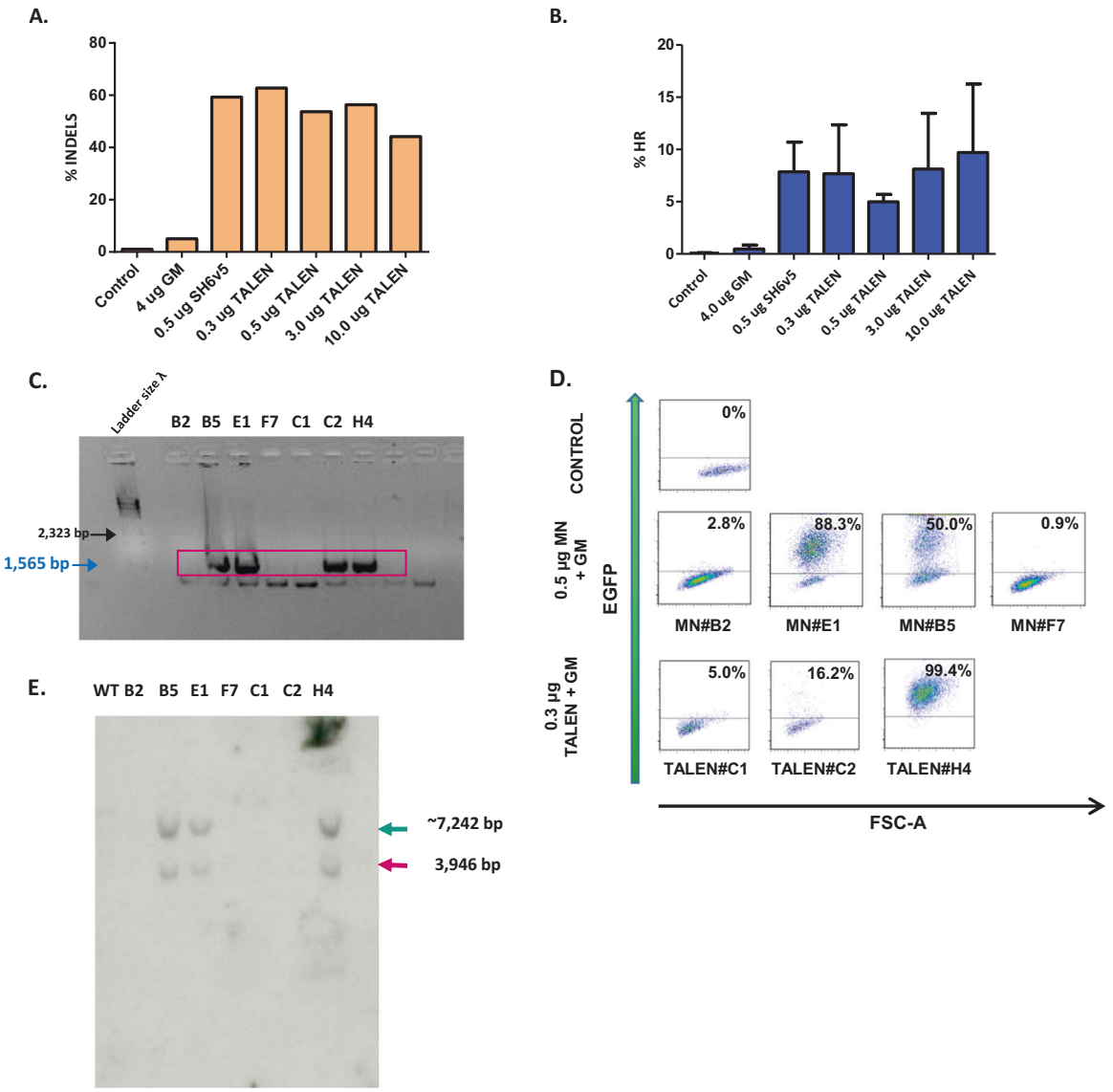

Fig. 2 Comparative analysis of TALE nucleases versus SH6v5-MNs to mediate homologous recombination in the SH6 safe harbor site in HEK-293H cells. a. Percentage of specific cleavage performed by the SH6v5-MN and the SH6-TALEN, in the pool of cells, analyzed by the Surveyor analysis, indicative of the activity of the nucleases. **b**. Percentages of HR obtained after cotransfection of the GM either with the SH6-MN and different doses of the SH6-TALEN in HEK-293H cells. Data are the mean ± SEM ($n = 4$–9 experiments). Not statistical differences were found among groups with a one-way ANOVA followed by a post hoc Tukey test. **c**. The agarose gel electrophoresis represents which of the pseudoclones carry a specific integration of the GM, represented by a band of 1,565 bp. **d**. The panel shows the presence of EGFP^{+} cells in some of stablished pseudoclones. **e**. Southern blot film, following the strategy explained in Supplementary Fig. 2A, corresponding to digested gDNA of some pseudoclones. The red and green arrows mark, respectively, the expected size bands and the additional bigger band explained in Supplementary Fig. 2C. Ladder size corresponds to λ DNA cut with BstII restriction enzyme.

Gene targeting experiments were then performed under the same conditions as described in the previous experiments. In/out PCRs were performed in cells cultured in 96-well plates. The efficiencies of HR obtained with the SH6v5-MN in four different experiments was around 7.8%, while the efficiencies achieved in nine gene-editing experiments using SH6-TALEN (doses ranged between 0.3 and 10.0 µg) ranged between 7.7 and 9.7% (Fig. 2b).

Pseudoclones were established with transfected cells that showed HR. As measure by in/out PCR, HR occurred in four out of the seven pseudoclones and the FACs analysis corroborate that only these four clones expressed eGFP (Fig. 2c, d), indicating the stable expression of the transgene in targeted HEK-293H cells.

Southern blot analysis performed in pseudoclones showed the expected band of 3946 bp in three out of seven

analyzed pseudoclones (B5, E1, and H4), the ones with higher percentage of EGFP$^+$ cells, indicating the specific integration in the SH6 site (Fig. 2e). As previously observed, an additional band of 7242 bp was seen in these clones (Figs. 2e and S3C). Surprisingly, no band was obtained in pseudoclone C2, in which a specific integration by PCR was identified. In this case, however, the level of EGFP expression was very low. No band was detected in pseudoclones B2, F7, and C1, consistent with the absence of the specific band in the previous PCR analyses, which showed a low expression of EGFP (Fig. 2c–e).

Taken together these experiments demonstrate that both, SH6v5-MN and SH6-TALEN mediate the specific targeting of a reporter GM matrix in the SH6 safe harbor locus in HEK-293 H cells, facilitating the stable expression of the targeted matrix in these cells.

Analysis of SH6v5-meganuclease-mediated homologous recombination in CD34$^+$ hematopoietic progenitor cells

In order to investigate the feasibility of performing targeted GT in the *SH6* locus of human hematopoietic stem progenitor cells (HSPC), we designed the experimental protocol shown in Fig. S6A. UCB CD34$^+$ cells were prestimulated for 48 h and then nucleofected with the MN or the TALEN plus the GM. Seven days after nucleofection, the percentage of EGFP$^+$ cells was analyzed by flow cytometry to determine the nucleofection efficiency. In order to enrich nucleofected cells, EGFP$^+$ cells were sorted and then methylcellulose assays were performed with sorted cells. After 14 days, green colonies were picked and DNA from these colonies was extracted and PCR analyses were performed to detect specific integrations.

To this aim, CD34$^+$ cells were nucleofected with 3 or 10 µg of nuclease as DNA, or with 10 or 20 µg of the MN as mRNA that is reported to be less toxic and more efficient [23, 24]. The surveyor assay showed that the delivery of the MN as mRNA was more active for the cleavage of the *SH6* site, where up to 18.0% of INDELs were observed (Figs. S6B and 3a).

Once this protocol was set up, CD34$^+$ cells were nucleofected with 4.0 µg of the GM alone or with the SH6v5-MN (either as DNA or mRNA), or with the SH6-TALEN used as DNA. Most of the incubated cells remained positive for CD34$^+$ expression at 2 and 4 days of culture, while the percentage of EGFP expression decreased in all tested conditions (Fig. 3b). At 7 days post nucleofection the proportion of EGFP$^+$ cells ranged between 1.5 and 8.9% (Fig. 3b). EGFP$^+$/CD34$^+$ cells obtained in each condition were sorted out and seeded in methylcellulose cultures before and after the sorting (Fig. 3c). As shown in Fig. 3c, when the number of colonies generated by EGFP-sorted cells was scored, we observed that the number of EGFP$^+$ colonies was very low, around 1/10,000, with respect to the total number of colonies obtained in the EGFP-sorted population (Fig. 3c, d), explaining the low percentages showed in Fig. 3d. That's why the efficiency of EGFP$^+$ colonies from different experiments were picked and their DNA used to perform Nested-PCR (Fig. 1a) to check if EGFP$^+$ colonies carried the specific integration of the GM in the *SH6* locus. Strikingly, none of the EGFP$^+$ colonies obtained in these experiments showed the presence of a positive band (Fig. S7).

Because sorting of EGFP$^+$ cells may have selected cells in which nontargeted integration of the GM occurred at loci supporting particularly high expression of the transgene, subsequent experiments were conducted in the absence of the cell sorting process.

Thus, CD34$^+$ cells were nucleofected with 4 µg of the GM alone or either with 2.5 µg of SH6v5-MN, used as DNA or mRNA, or with 2.5 or 5 µg of the SH6-TALEN used as DNA. Nucleofected cells were maintained in culture for 4 days. At this time, the percentage of EGFP$^+$ cells ranged between 20 and 40% (Fig. 4a). These percentages were similar to the ones obtained in previous experiments, indicating the consistency of our experimental protocol to nucleofect CD34$^+$ cells without a marked toxicity and a significant efficacy to mediate the delivery and expression of the transgene. Different hematopoietic subpopulations were studied by flow cytometry, using the protocol proposed by Doulatov et al. [21] (Fig. S8), to analyze if the different manipulations of the cells were inducing the differentiation of CD34$^+$ cells. The most primitive HSPC (hematopoietic stem and progenitor cells) percentages (HSPC: CD34$^+$/38$^-$/45RA$^-$/90$^+$) in nucleofected cells ranged from 5 to 10%, slightly lower to values observed in the nonelectroporated control sample (14%, Fig. 4b). EGFP$^+$ HSPC percentages were among 2–4% of the total population and, depending on the condition, 16–46% of the HSPC cells were EGFP$^+$, indicating that a significant proportion of the HSPCs carried the GM (Fig. 4b). The percentage of (MPP: CD34$^+$/38$^-$/45RA$^-$/90$^-$) was lower, although the proportion of EGFP$^+$ cells within this population was similar to the proportion of EGFP$^+$ cells found in the HSPC fraction (Fig. 4c). These results suggested that our experimental conditions were efficient to preserve the HSPC phenotype of the nucleofected CD34$^+$ cells carrying the GM, although we could observe that our gene editing approach implied certain degree of toxicity to HSPCs.

In the total nucleofected population with 4 µg of the GM and 5 µg of the SH6-TALEN, a clear band of 1565 bp corresponding to the specific integration was observed by Nested-PCR (Fig. S9). When we performed methylcellulose cultures, the number of total colonies generated in each condition was similar, with the exception of the condition in

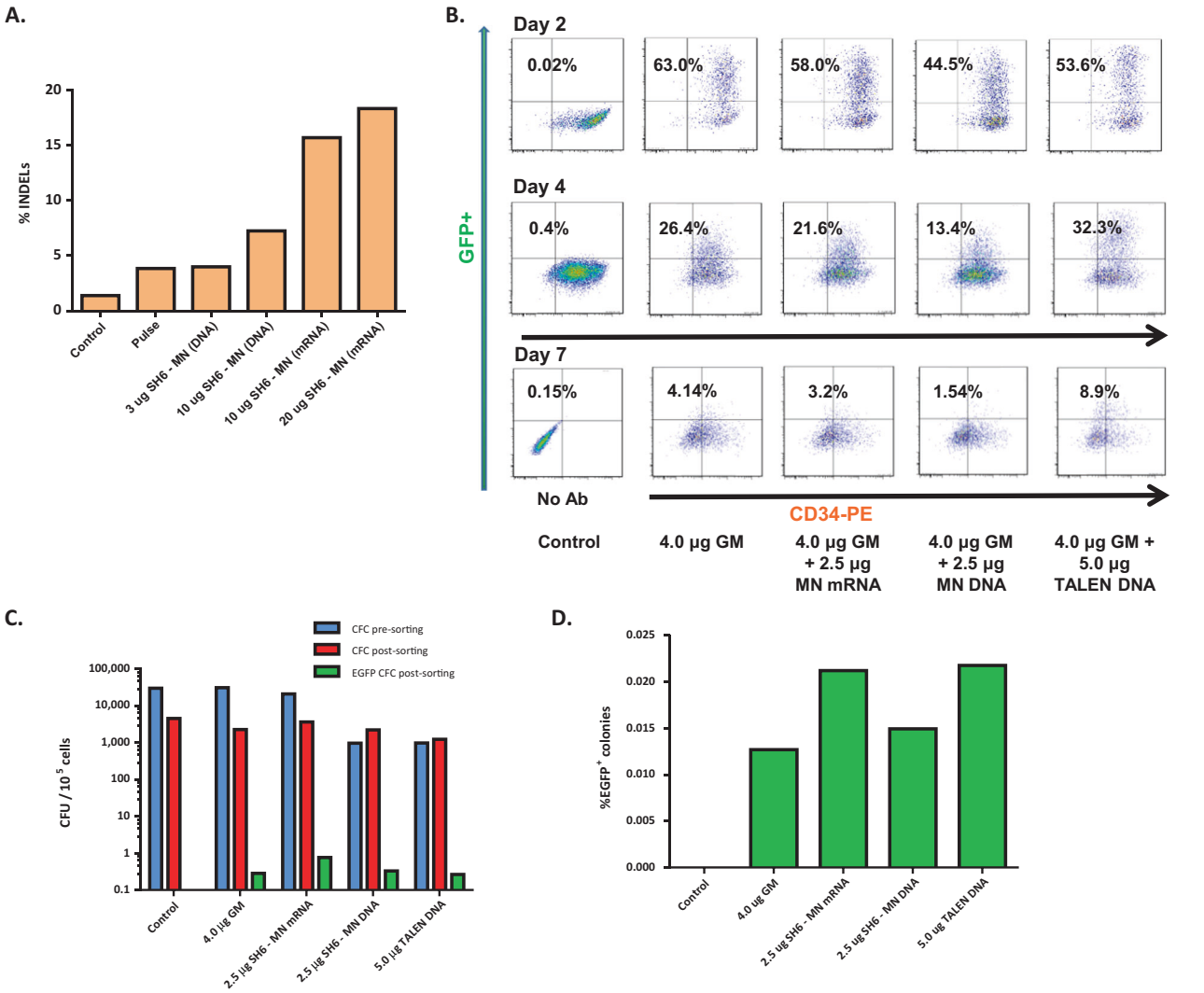

Fig. 3 SH6v5-meganuclease and TALE nuclease mediated homologous recombination in EGFP$^+$ sorted CD34$^+$ hematopoietic progenitor cells. a. Percentage of specific cleavage performed by the SH6v5-MN as DNA or mRNA, in the pool of CD34$^+$ cells, analyzed by the surveyor analysis, indicative of the activity of the nucleases. **b**. Flow cytometry analysis of CD34$^+$ and EGFP$^+$ cells after 2, 4, and 7 days post nucleofection with 4.0 µg of GM alone or in combination with 2.5 µg of SH6v5-MN as DNA or mRNA and 5.0 µg of TALEN as DNA. **c**. Number of CFU/10^5 cells obtained after 14 days in methylcellulose cultures. **d**. Percentage of EGFP$^+$ colonies among the total number of colonies obtained. **e** and **f** correspond to a representative experiment. In all figures control condition are thawed, cultured and not nucleofected CD34$^+$ cells.

which 5 µg of TALEN was used, in which a lower number of colonies was obtained (Fig. 4d). Interestingly, this was the only condition in which a colony presented a specific integration in the *SH6* site (Fig. 4e), corresponding to 2.1% of targeted integration in the hematopoietic progenitor cells. PCR analyses for internal EGFP revealed no positive colonies, indicating no unspecific integration in those conditions (data not shown).

Five experiments using these doses of the TALEN as DNA were performed. At 24 h post nucleofection, the viability of the cells decreased with the dose of DNA used, and the percentage of EGFP$^+$ cells was around 60% (Fig. 5a, b). At 4 days post nucleofection primary populations were measured with no significant differences between the percentages of HSPC and MPP populations (5–8% and 8–10.5%, respectively) for all the nucleofected conditions (Fig. 5c, d). EGFP$^+$ percentages were around 2.0–2.8% in the HSPC population, and 1.7–3.8% for the EGFP$^+$ MPP, in the conditions nucleofected with the GM alone or plus the nucleases. However, no significant differences were found among these nucleofected conditions (Fig. 5c, d). In all the experiments performed, a band representing a specific integration of the GM in the *SH6* locus was found in the pool of cells nucleofected with GM and TALEN, indicating that a fraction of these cells was edited. After 14 days, colonies obtained in methylcellulose assays were

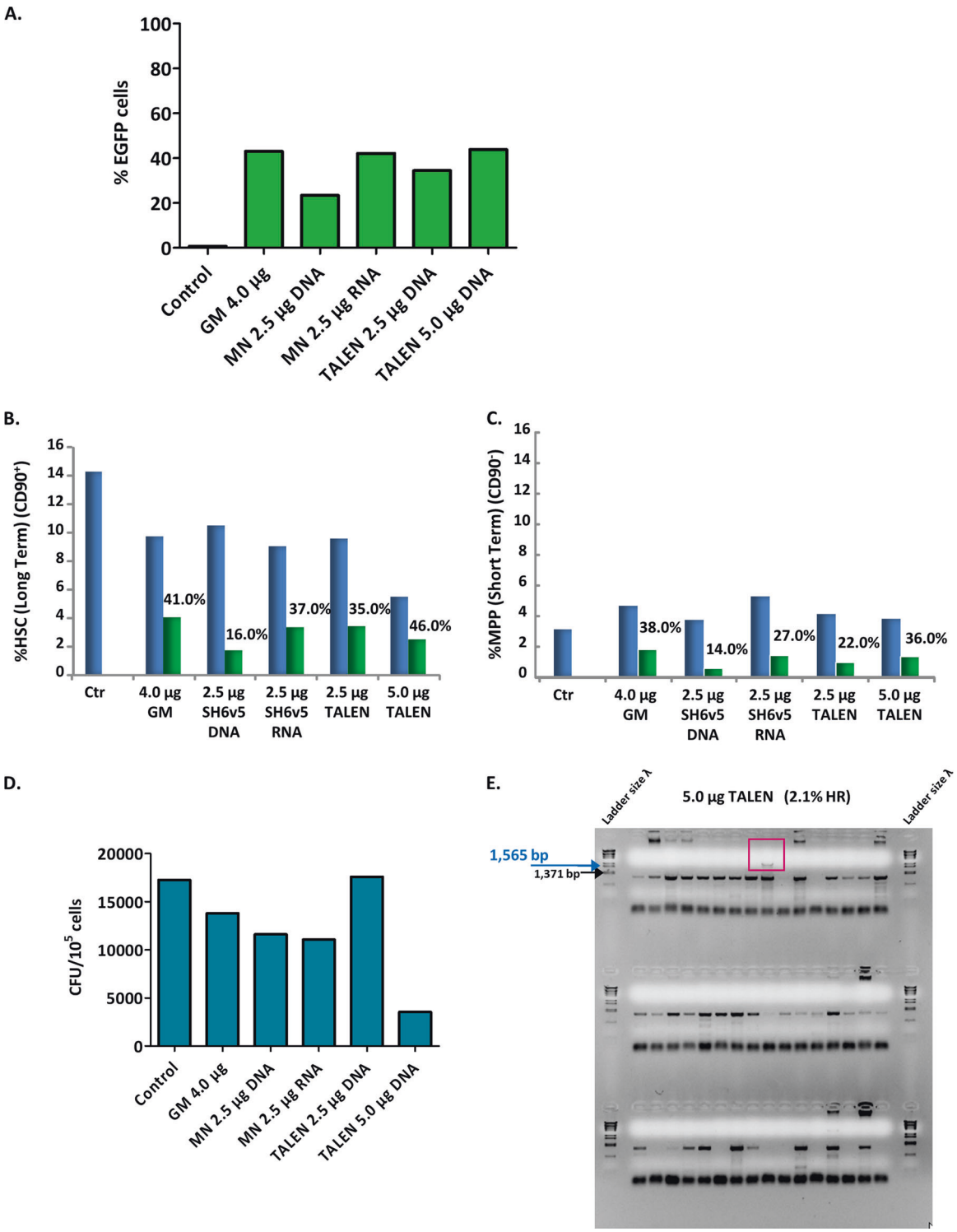

Fig. 4 SH6v5-meganuclease and TALE nuclease mediated homologous recombination in nonsorted CD34$^+$ hematopoietic progenitor cells. a. Percentages of EGFP$^+$ cells determined at 4 days postnucleofection. **b**. Percentages of HSPC (blue bars) and of EGFP$^+$ HSPC (green bars) in cells analyzed at 4 days post nucleofection. Numbers above the green bars mean the percentage of HSPC that were EGFP$^+$. **c**. Percentages of MPP (blue bars) and EGFP$^+$ MPP (green bars), and the percentages of MPP that were EGFP$^+$. **d**. Number of colonies/10^5 cells generated after nucleofection. **e**. The agarose gel electrophoresis displays the result of the Nested-PCR which shows a positive band inside the purple square. Ladder size corresponds to λ DNA cut with BstII restriction enzyme.

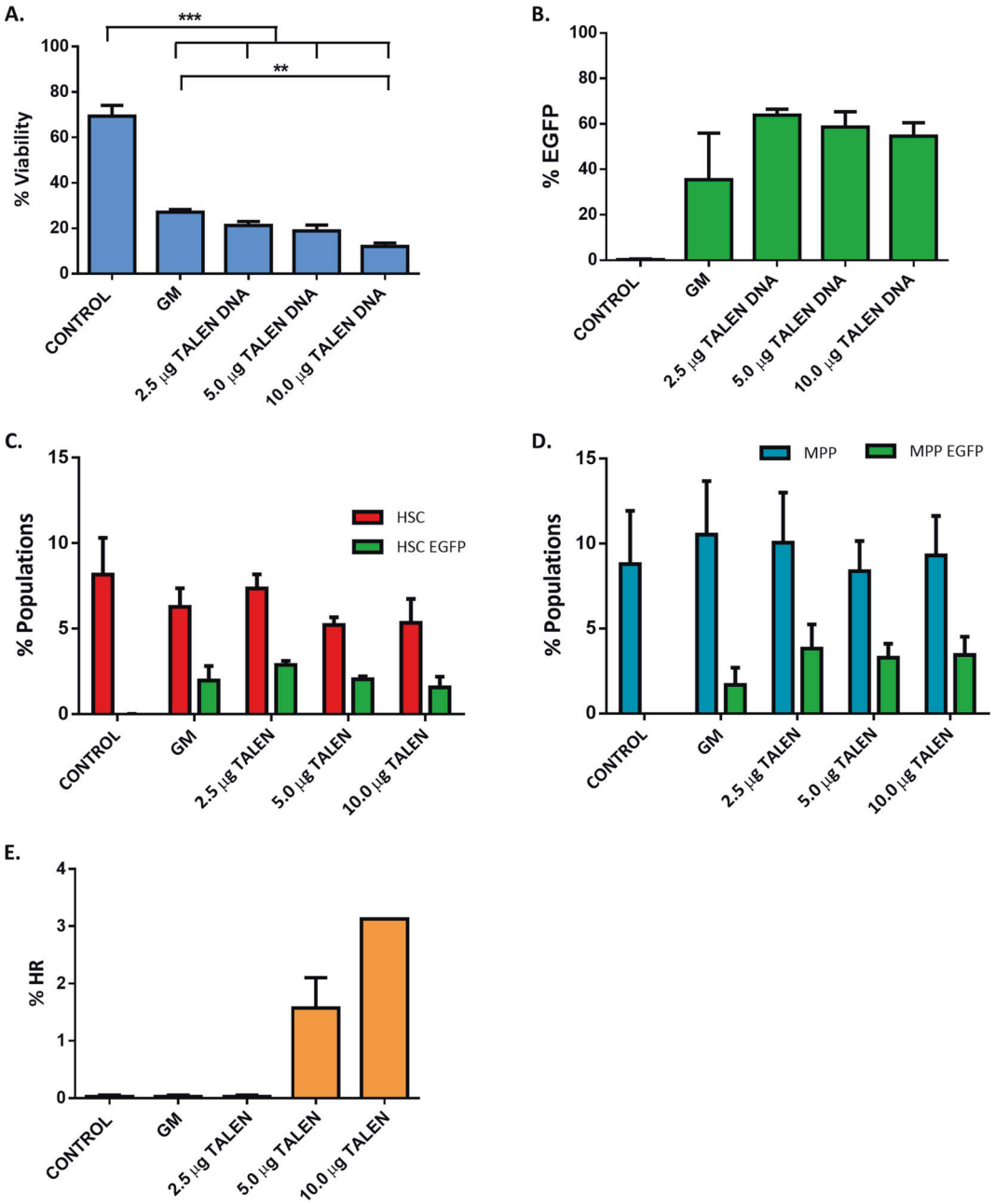

Fig. 5 SH6 TALE nuclease mediated homologous recombination in CD34$^+$ hematopoietic progenitor cells. a. Viability of nucleofected CD34$^+$ cells with the TALEN as DNA 24 h post nucleofection. **b**. Percentage of EGFP$^+$ cells at 24 h post nucleofection. **c**. Percentages of HSPC (red bars) and EGFP$^+$ HSPC (green bars) populations in cells analyzed at 4 days post nucleofection. **d**. Percentages of MPP (blue bars) and EGFP$^+$ MPP (green bars) populations. Data are the mean ± SEM ($n = 4$–5 experiments). (*) P value < 0.05, (**) P value < 0.01, (***) P value < 0.001 indicate significant differences with a one-way ANOVA followed by a post hoc Tukey test. **e**. HR percentages induced by the SH6-TALEN using the GM alone or with the TALEN®. Data are mean ± SEM ($n = 2$).

quantified and there were no significant differences in the number of CFU per 10^5 cells (Fig. S10A). The number of erythroid colonies was higher than the number of myeloid ones, although no significant differences were found (Fig. S10B). We were unable to detect EGFP-positive colonies by fluorescence microscope. These colonies were analyzed for specific integrations of the GM in the *SH6* site and this integration was mainly found in colonies from progenitors nucleofected with 5 or 10 µg of the TALEN as DNA, showing a mean of 1.57 and 3.13% of HR events, respectively (Fig. 5e).

In addition, we checked possible off-targets with 5–6 mismatches (Table S1) and we decided to analyze the ones with five mismatches as relevant off-target integrations [25].

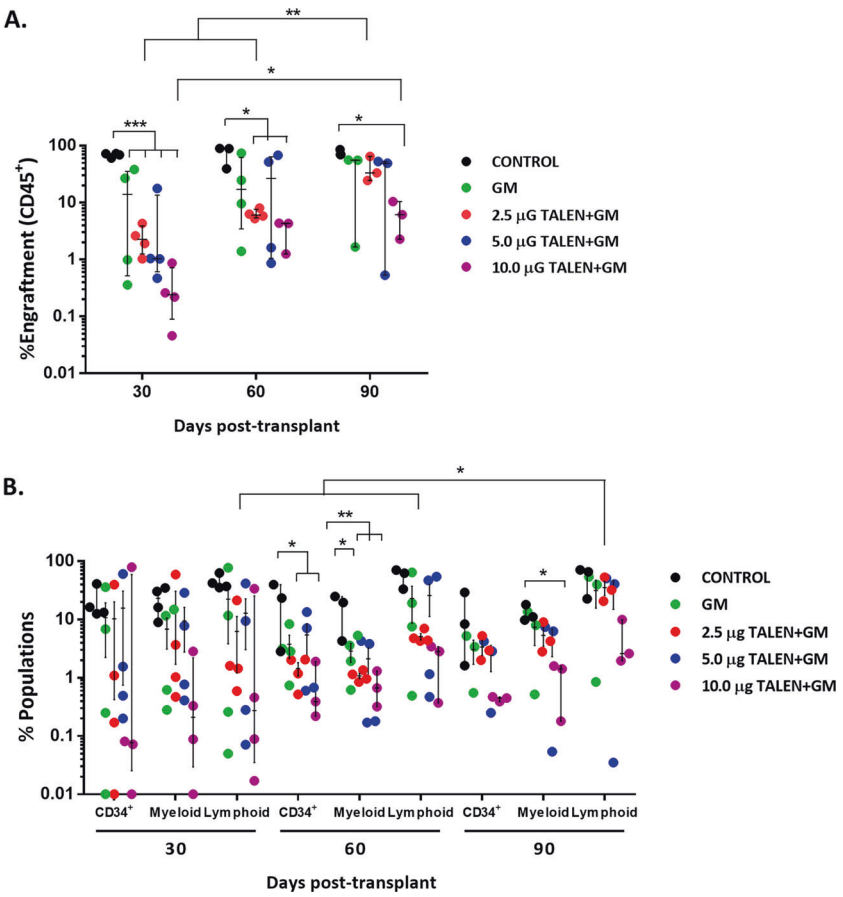

Fig. 6 Reconstitution potential of CD34$^+$ cells subjected to SH6 TALE nuclease mediated homologous recombination protocol. a. The graph shows the percentage of cells, obtained from the bone marrow (BM) of engrafted NSG mice at 30, 60, and 90 days post transplant. Data are the mean ± interquartile range. ($n = 3$–4 experiments). **b**. The graft shows the percentage of CD34$^+$, myeloid and lymphoid cells, obtained from the bone marrow (BM), observed in NSG mice after the differentiation of the engrafted of the nucleofected cells, at 30, 60, and 90 days post transplant. Data are the mean ± interquartile range. ($n = 3$–4 experiments). (*) P value < 0.05, (**) P value < 0.01 indicate significant differences with a one-way ANOVA followed by a post hoc Tukey test.

After sequencing nucleofected CD34$^+$ samples with different conditions (SH6 TALEN + GM donor) we could not detect any INDELS (Figs. S11, S12).

Four days post nucleofection, nucleofected cells were transplanted into NSG mice in order to check the engraftment and differentiation capacity of the cells at 30, 60, and 90 days post transplant. Cells were nucleofected only with GM or with GM and TALEN as DNA at different doses (2.5, 5.0, and 10.0 µg) (Fig. 6). At day 30, the engraftment capacity measured as percentage of CD45$^+$ cells decreased in mice transplanted with nucleofected cells in comparison with the engraftment shown in mice transplanted with control cells (not nucleofected) (Fig. 6a). This decrease was even more evident as the dose of DNA used also increased (Fig. 6a). However, at day 90, the level of engraftment increased for all the conditions, and was similar to that of the control, except for the highest dose of TALEN, indicating that nucleofected cells retained their long term repopulating properties (Fig. 6a). This human engraftment was multilineage as we detected progenitor cells (CD34$^+$), myeloid (CD33$^+$) and lymphoid lineages (CD19$^+$) at 30, 60, and 90 days post transplant (Fig. 6b). However, none of the CD45$^+$ cells were EGFP$^+$ (data not shown). Finally, sorted CD45$^+$ and CD45$^+$/CD34$^+$ cells from the engrafted mice were cultured in methylcellulose. After 14 days, CFU colonies were picked and analyzed for HR events by the Nested PCR. The band indicating the specific integration of the GM in the SH6 site was not observed neither in the pool of cells nor in the colonies obtained (data not shown).

Taken together, these final results indicate that under our conditions, consisting on the nucleofection of the GM with a relatively high dose of TALEN-encoding plasmid, a targeted integration of 3.13% can be achieved in HSPCs able to give rise to hematopoietic colonies in vitro, although no editing in very primitive HSPCs could be confirmed in these experiments.

Integration of a therapeutic matrix in the SH6 safe harbor site mediated by SH6-meganucleases and TALEN

Once targeting integration of a reporter GM in the SH6 site was performed, we aimed at developing a similar approach, performing gene targeting under the same conditions as described in the previous experiments, with a donor harboring the *FANCA* gene that we called the TM. This matrix carried the WT *FANCA* gene, followed by an E2A sequence and a puromycin resistant gene, both of them

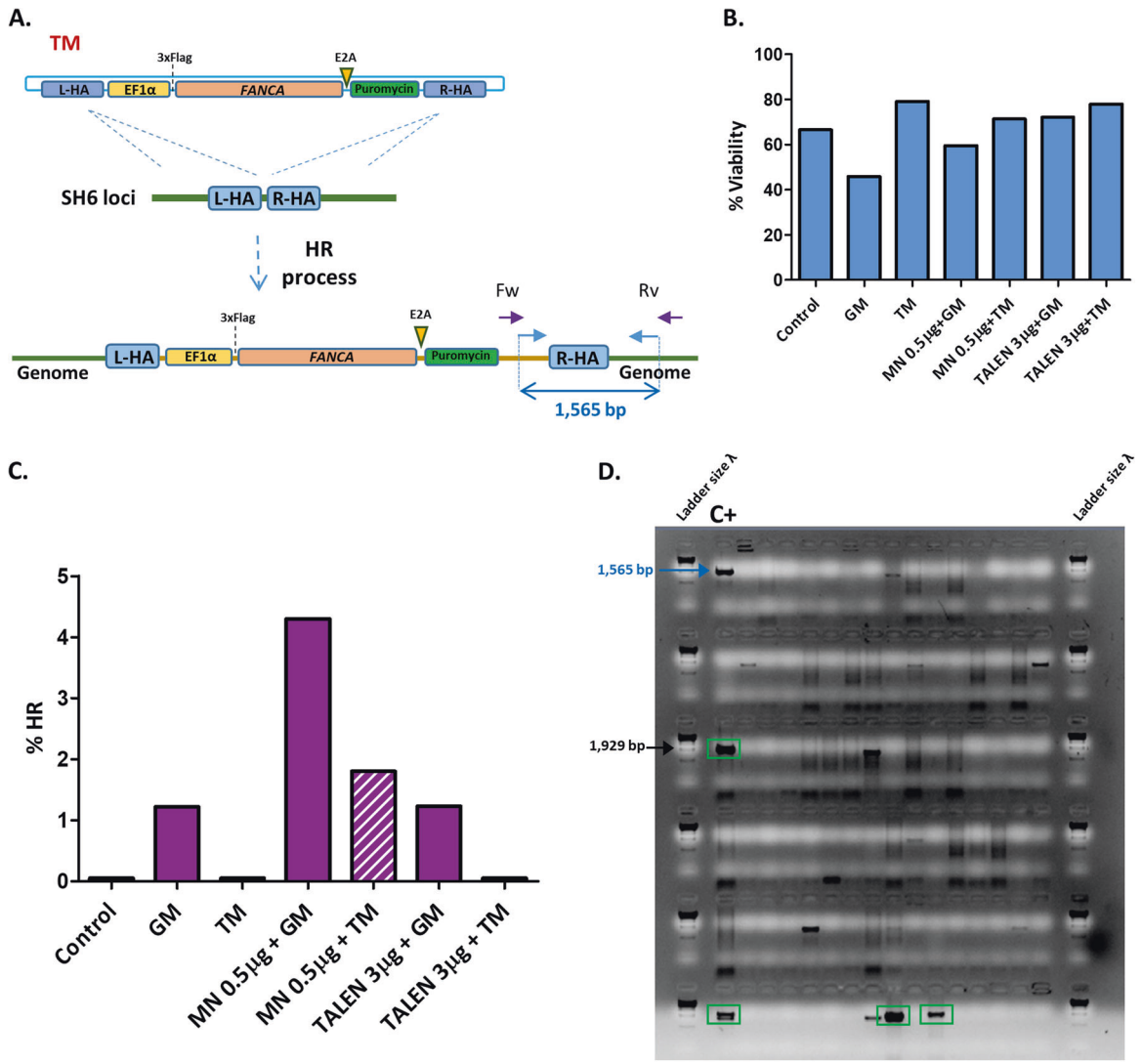

Fig. 7 Integration of a therapeutic matrix in the SH6 safe harbor site mediated by SH6-meganucleases and TALEN. a. Representative scheme of the gene targeting strategy in the SH6 using the TM circular plasmid donor. Discontinue blue lines represent the homology areas of the matrix that will recombine with the target site. Purple arrows represent the first pair of primers used in the Nested-PCR to check the specific integration. Blue arrows represent the pair of primers used in the PCR, and the second pair used in the Nested-PCR, to test the specific integration of the TM in the target site. The green arrow shows the size of the amplicon obtained as a positive result in the PCR analyses. **b**. Viability corresponding to cells transfected with the two matrixes and the nucleases at three days post transfection. In all instances, the viability is represented by the proportion of DAPI-negative cells. **c**. HR percentages induced by the MN and the TALEN using the GM or the TM donors. The lined bar represents the condition transfected with TM with positive HR. **d**. Agarose electrophoresis gels showing positive bands, inside the green boxes, corresponding to HR events in the MN + GM condition. Ladder size corresponds to λ DNA cut with BstII restriction enzyme.

regulated by the EF1α promoter. As in the GM this construct is flanked by the two HAs for the SH6 target site generated (Fig. 7a).

A gene targeting experiment was performed in HEK-293H cells. Cells were transfected with 4 µg of the GM or the TM alone or with 0.5 µg of the SH6v5-MN or 3 µg of the SH6-TALEN. Three days after transfection, the cell viability was compared among all conditions. The transfection with the TM (size 11,848 bp) indicated that this matrix was not more cytotoxic for the transfected cells, as compared with the GM (size 7,440 bp) (Fig. 7b). In/out PCRs were performed in cells cultured in 96-well plates (Fig. 7c). HR events were obtained with both nucleases when the GM was used, being the HR percentage of 4% for the SH6v5-MN and 1% for the SH6-TALEN. When the TM donor was used, HR percentages decreased to 1.8% for the MN and to <0.3% for the TALEN (Fig. 7c), indicating that the integration of the TM was possible although being a larger donor (Fig. 7c, d). PCR analyses for internal puromycin sequence showed unspecific integration in all

conditions: 11.45% with TM, 2.08% MN + TM and 8.3% TALEN + TM (data not shown). The highest percentage obtained without nuclease would suggest that the presence of MN o TALEN decreases the possibility of unspecific integrations.

Discussion

Gene editing mediated by nucleases using a specific matrix remains the new challenge in GT. In this work, two MNs and one TALEN targeting a new safe harbor (SH6) have been investigated together with a functional matrix carrying an *EGFP* gene which expression was controlled by the EF1α promoter.

With respect to the HR frequencies obtained with the MNs, Eyquem et al. [15] were the first authors using MNs targeting SH6 site and a similar EGFP matrix. The doses they used were similar or lower to the ones we employed, and the HR percentages were comparable to ours when we used the SH6v5-MNs.

This is the first work in which the SH6-TALEN has been used. HEK-293H cells transfected with increasing doses of the nuclease showed that the use of higher doses of the SH6-TALEN seems to reduce the efficiency of this nuclease. This result could be explained by an increased toxicity induced by the TALEN or the DNA transfection itself in HEK-293H cells (Fig. 2a, b). When pseudoclones were established, the ones carrying the specific integration of the donor maintained the EGFP expression along time, indicating that the EF1α promoter was able to confer long term expression in the *SH6* site. These results are similar to those obtained in Eyquem et al. [15], work in which the expression intensity of the transgene is dependent on the number of copies integrated, the target site (SH6) and the promoter (EF1α with intron). A unique copy of the transgene controlled by the same promoter showed low intensity of the EGFP expression in the SH6 site (in Eyquem et al. compared with other targeted sites), which is a hypermethylated and a not transcriptionally active area described by ENCODE database [15]. This low expression driven in the *SH6* site could not be problematic for the GT in Fanconi anemia patients with mutations in *FANCA* because low levels of expression of this gene have been observed to be sufficient to reverse the FA-A phenotype [26]. Maybe this expression is not sufficient for other therapeutic applications in which low expression levels could fail to produce the desire therapeutic effect. To improve the low expression described for the SH6 locus we propose the use of chromatin insulators in the donor matrix which protect chromosomal domains from heterochromatinization [27, 28], although their efficacy in the SH6 locus should be addressed.

Southern blot analyses showed the expected band in most of the pseudoclones that were positive for the targeted integration in the *SH6* site by the specific PCR. Negative pseudoclones by Southern blot corresponded with those with lower percentage of $EGFP^+$ cells, suggesting that the frequency of cells with the integration was lower than the sensitivity of the Southern blot analysis (Fig. 2e). Among the pseudoclones presenting the specific integration in the Southern, some of them showed a larger band of 7,242 bp, in line with those presenting a higher percentage and intensity of $EGFP^+$ cells (and maybe containing more of one integrated copy). It could be a consequence of the concatemers formation by our GM and mediated by the HR process. This issue has been shown in other gene editing works using other nucleases and donors such as ZFNs and IDLVs [14] and in Eyquem et al. [15], with a higher level of expression of the transgene in these cases.

Several experiments were performed to check the maintenance of the pluripotent capacity of $CD34^+$ cells after several manipulations as nucleofection, transfection method, sorting, nuclease delivery and the in vitro culture. Although all these processes affected the viability of the cells and their pluripotent capacity, especially the nucleofection and quantity of DNA, the results obtained were good enough to perform the experiments and with similar results to those that have been observed in other studies with these cells [8, 10, 29–32]. Even when the primitive populations were measured by flow cytometry following the protocol proposed by Doulatov et al. [21], the percentage of total and $EGFP^+$ primitive HSPCs obtained was very good, indicating that the cocktail of cytokines used in our culture media is able to favor the pluripotent capacity of treated cells (Fig. 4c, d). Other works used different cytokine cocktails, as Hoban et al. [8]. They used ZFNs for the treatment of sickle cell disease (SCD), and did not measure primitive populations although they also obtained a majority of erythroid colonies after methylcellulose assays. Genovese et al. observed an increase of the self-renewal capacity of $CD34^+$ cells adding to the culture SR1 (1 μM) and dmPGE2 (10 μM) measured as hematopoietic reconstitution of NSG mice [10]. Wang et al. used the traditional cocktail to culture edited $CD34^+$ cells with a GFP cassette integrated using ZFN mRNA at CCR5 and AAVS1 loci and observed similar $CD34^+CD133^+CD90^+$ percentages to those obtained in our experiments [29]. Dever et al. were able to get a very high engraftment of gene edited human hematopoietic cells in NSG mice, when $CD34^+$ were cultured in presence of SCF, FLT3L, TPO and UM171 [12]. Their gene editing approach is based on the electroporation of CRISPR/Cas9 system as ribonucleoproetin, followed by transduction with AAV6 viral vectors carried the HR donor. Our results indicate that, although our culture conditions are able to maintain primitive HSPCs, the gene editing level is

either very low or the gene editing with our conditions is toxic to be detected in xenograft analysis in NSG mice.

When HR frequencies were studied, we realized that after the sorting of the $EGFP^+$ cells no HR events were obtained in the $EGFP^+$ colonies. As in Eyquem et al. are reported, SH6 is a hypermethylated region that impairs the expression of the transgene. Then, a higher expression of the *EGFP* gene is indicative of tandem or random integrations. As the PCR for specific integrations did not show the expected bands, maybe we were facilitating the enrichment of cells with random integration [15]. Checking the specific integrations of the GM in the SH6 site in the pool of cells before methylcellulose assays and analyzing all the colonies obtained without the enrichment with the sorting, HR events were obtained for all the conditions when higher doses of SH6-TALEN were delivered (Fig. 5e). Although no gene targeting was observed using the SH6v5-MN, up to 2.1% of HR events was obtained with 5 µg of SH6-TALEN, and up to 3.1% with 10 µg of the nuclease. This observation was different than the results obtained with HEK-293H cells, showing that in primary $CD34^+$ cells increasing the dose of SH6-TALEN is not detrimental as observed in HEK-293H cells (Fig. 2a, b). This contradiction could be due to the different cell types and delivery systems used.

Different HR frequencies could be obtained depending on the target site. The SH6 is an hypermethylated site [15] that maybe is blocking the access of the TALEN to the target. The delivery of the nuclease as DNA would maintain its expression into the cell for longer, increasing the probability of the TALEN to break the DNA and of the HR process during the synthesis phase, when the DSB repair mechanism is more active [33]. On the other hand, a smaller nuclease as the SH6v5-MN is expected to enter easily in this region, but no gene targeting has been observed in CFC colonies. Therefore, our results showed that MNs can induce HR events in SH6 when the specific PCR is performed in the pool of nucleofected $CD34^+$ cells, but no positive colonies were observed, which could suggest that this specific integration is performed in more differentiated cells. It is reported that the donor employed, the DNA repair machinery and the chromosomic context affect MN activity [34], and even CpG methylation dinucleotides in the MN target could impair their affinity and efficiency [35]. All these circumstances would be the cause of the inefficacy of the SH6-MNs in $CD34^+$ cells. Although genetic edition percentages obtained in other works are higher to those obtained in this work, they are not comparable due to the different nucleases employed in some cases, the different target sites and the use of other types of donors [8, 10, 29–32].

The results of this work reveal the possibility to perform genome editing in $CD34^+$ progenitor cells in a safe harbor called SH6 using TALEN gene editing technology. Although this SH is a genomic desert, it allowed expression of an *EGFP* transgene controlled by the EF1α promoter. Results obtained in this work integrating a TM in SH6 site could be enough for a possible future treatment of diseases in which a low expression of the therapeutic gene corrects the phenotype. In those cases in which a higher expression of the transgene is needed, the use of insulators could solve this challenge although this possibility has not been addressed in this work. This strategy would be useful to treat inherited hematologic monogenic diseases after improving transgene expression and targeted integration into true HSCs.

Acknowledgements The authors would like to thank Agnes Gouble, Roman Galetto and Laurent Poirot from CELLECTIS S.A (Paris) for their help providing us with reagents, tools, and their contribution with ideas, Miguel A. Martin for the careful maintenance of NSG mice, Rebeca Sánchez and Omaira Alberquilla for their technical assistance in flow cytometry and Centro de Transfusiones de la Comunidad de Madrid for the cord blood availability. The authors also thank Fundación Botín for promoting translational research at the Hematopoietic Innovative Therapies Division of the CIEMAT. CIBERER is an initiative of the Instituto de Salud Carlos III and Fondo Europeo de Desarrollo Regional (FEDER).

Funding This work was supported by grants from the 7th Framework Program European Commission (HEALTH-F5-2012-305421; EUROFANCOLEN), Ministerio de Ciencia, Innovación y Universidades y Fondo Europeo de Desarrollo Regional (FEDER) (SAF2015-68073-R, SAF2017-84248-P and RTI2018-097125-B-I00) and Fondo de Investigaciones Sanitarias, Instituto de Salud Carlos III (RD16/0011/0011).

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

1. Kaufmann KB, Buning H, Galy A, Schambach A, Grez M. Gene therapy on the move. EMBO Mol Med. 2013;5:1642–61.
2. Ghosh S, Thrasher AJ, Gaspar HB. Gene therapy for monogenic disorders of the bone marrow. Br J Haematol. 2015;171:155–70.
3. Shen H, Zhong X, Zhao F, Wang Y, Yan B, Li Q, et al. Overexpression of receptor-like kinase ERECTA improves thermotolerance in rice and tomato. Nat Biotechnol. 2015;33:996–1003.
4. Ciccia A, Elledge SJ. The DNA damage response: making it safe to play with knives. Mol Cell. 2010;40:179–204.
5. Jensen NM, Dalsgaard T, Jakobsen M, Nielsen RR, Sorensen CB, Bolund L, et al. An update on targeted gene repair in mammalian cells: methods and mechanisms. J Biomed Sci. 2011;18:10.
6. LaFountaine JS, Fathe K, Smyth HD. Delivery and therapeutic applications of gene editing technologies ZFNs, TALENs, and CRISPR/Cas9. Int J Pharm. 2015;494:180–94.
7. Naldini L. Ex vivo gene transfer and correction for cell-based therapies. Nat Rev Genet. 2011;12:301–15.
8. Hoban MD, Cost GJ, Mendel MC, Romero Z, Kaufman ML, Joglekar AV, et al. Correction of the sickle cell disease mutation

in human hematopoietic stem/progenitor cells. Blood. 2015;125:2597–604.
9. Wang J, Exline CM, DeClercq JJ, Llewellyn GN, Hayward SB, Li PW, et al. Homology-driven genome editing in hematopoietic stem and progenitor cells using ZFN mRNA and AAV6 donors. Nat Biotechnol. 2015;33:1256–63.
10. Genovese P, Schiroli G, Escobar G, Tomaso TD, Firrito C, Calabria A, et al. Targeted genome editing in human repopulating haematopoietic stem cells. Nature. 2014;510:235–40.
11. Garate Z, Quintana-Bustamante O, Crane AM, Olivier E, Poirot L, Galetto R, et al. Generation of a high number of healthy erythroid cells from gene-edited pyruvate kinase deficiency patient-specific induced pluripotent stem cells. Stem Cell Rep. 2015;5:1053–66.
12. Dever DP, Bak RO, Reinisch A, Camarena J, Washington G, Nicolas CE, et al. CRISPR/Cas9 beta-globin gene targeting in human haematopoietic stem cells. Nature. 2016;539:384–9.
13. Bak RO, Dever DP, Porteus MH. CRISPR/Cas9 genome editing in human hematopoietic stem cells. Nat Protoc. 2018;13:358–76.
14. Lombardo A, Cesana D, Genovese P, Di Stefano B, Provasi E, Colombo DF, et al. Site-specific integration and tailoring of cassette design for sustainable gene transfer. Nat Methods. 2011;8:861–9.
15. Eyquem J, Poirot L, Galetto R. Characterization of three loci for homologous gene targeting and transgene expression. Biotechnol Bioeng. 2013;110:11.
16. Sadelain M, Papapetrou EP, Bushman D. Safe harbors for the integration of new DNA in the human genome. Nature. 2012;12:51–8.
17. Bejerano G, Pheasant M, Makunin I, Stephen S, James KW, Mattick JS, et al. Ultraconserved elements in the human genome. Science. 2004;304:5.
18. Papapetrou EP, Lee G, Malani N, Setty M, Riviere I, Tirunagari LM, et al. Genomic safe harbors permit high beta-globin transgene expression in thalassemia induced pluripotent stem cells. Nat Biotechnol. 2010;29:73–8.
19. Szymczak AL, Workman CJ, Wang Y, Vignali KM, Dilioglou S, Vanin EF, et al. Correction of multi-gene deficiency in vivo using a single 'self-cleaving' 2A peptide-based retroviral vector. Nat Biotechnol. 2004;22:589–94.
20. Sambrook J, Fritsch EF, Maniatis T (eds). Molecular cloning. A laboratory manual. 2nd edn. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, 1989).
21. Doulatov S, Notta F, Laurenti E, Dick JE. Hematopoiesis: a human perspective. Cell Stem Cell. 2012;10:120–36.
22. Gutierrez-Guerrero A, Sanchez-Hernandez S, Galvani G, Pinedo-Gomez J, Martin-Guerra R, Sanchez-Gilabert A, et al. Comparison of zinc finger nucleases versus CRISPR-specific nucleases for genome editing of the Wiskott-Aldrich syndrome locus. Hum Gene Ther. 2018;29:366–80.
23. Maggio I, Goncalves MA. Genome editing at the crossroads of delivery, specificity, and fidelity. Trends Biotechnol. 2015;33:280–91.
24. Skipper KA, Mikkelsen JG. Delivering the goods for genome engineering and editing. Hum Gene Ther. 2015;26:486–97.
25. Boch J, Scholze H, Schornack S, Landgraf A, Hahn S, Kay S, et al. Breaking the code of DNA binding specificity of TAL-type III effectors. Science. 2009;326:1509–12.
26. Gonzalez-Murillo A, Lozano ML, Alvarez L, Jacome A, Almarza E, Navarro S, et al. Development of lentiviral vectors with optimized transcriptional activity for the gene therapy of patients with Fanconi anemia. Hum Gene Ther. 2010;21:623–30.
27. Chung JH, Bell AC, Felsenfeld G. Characterization of the chicken beta-globin insulator. Proc Natl Acad Sci USA. 1997;94:575–80.
28. Gaszner M, Felsenfeld G. Insulators: exploiting transcriptional and epigenetic mechanisms. Nat Rev Genet. 2006;7:703–13.
29. Schneidawind D, Federmann B, Buechele C, Helwig A, Schmohl J, Vogel W, et al. Reduced-intensity conditioning with fludarabine and busulfan for allogeneic hematopoietic cell transplantation in elderly or infirm patients with advanced myeloid malignancies. Ann Hematol. 2016;95:115–24.
30. Mao ZN, Liang CE. [Observation on therapeutic effect of acupuncture on hyperemesis gravidarum]. Zhongguo Zhen Jiu. 2009;29:973–6.
31. De Ravin SS, Reik A, Liu PQ, Li L, Wu X, Su L, et al. Targeted gene addition in human CD34(+) hematopoietic cells for correction of X-linked chronic granulomatous disease. Nat Biotechnol. 2016;34:424–9.
32. Diez B, Genovese P, Roman-Rodriguez FJ, Alvarez L, Schiroli G, Ugalde L, et al. Therapeutic gene editing in $CD34^{+}$ hematopoietic progenitors from Fanconi anemia patients. EMBO Mol Med. 2017;9:1574–88.
33. Mao Z, Bozzella M, Seluanov A, Gorbunova V. DNA repair by nonhomologous end joining and homologous recombination during cell cycle in human cells. Cell Cycle. 2008;7:2902–6.
34. Daboussi F, Zaslavskiy M, Poirot L, Loperfido M, Gouble A, Guyot V, et al. Chromosomal context and epigenetic mechanisms control the efficacy of genome editing by rare-cutting designer endonucleases. Nucleic acids Res. 2012;40:6367–79.
35. Valton J, Dupuy A, Daboussi F, Thomas S, Marechal A, Macmaster R, et al. Overcoming transcription activator-like effector (TALE) DNA binding domain sensitivity to cytosine methylation. J Biol Chem. 2012;287:38427–32.