Title:

In vitro and *In vivo* Genetic Disease Modelling via NHEJ-precise deletions using CRISPR/Cas9

Authors:

Sergio López-Manzaneda[1,2,#], Isabel Ojeda-Pérez[1,2], Nerea Zabaleta[3], Aída Garcia-Torralba[1,2], Omaira Alberquilla[1,2], Raúl Torres[4], Rebeca Sanchez-Dominguez[1,2], Laura Torella[3], Emmanuel Olivier[5], Joanne Mountford[5], Juan C. Ramírez[6], Juan A. Bueren[7,2], Gloria González-Aseguinolaza[3], Jose-Carlos Segovia[1,2,*]

Affiliations:

[1]Cell Differentiation and Cytometry Unit. Hematopoietic Innovative Therapies Division, Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas (CIEMAT) and Centro de Investigación Biomédica en Red de Enfermedades Raras (CIBERER), Madrid, Spain

[2] Unidad Mixta de Terapias Avanzadas. Instituto de Investigación Sanitaria Fundación Jiménez Díaz (IIS-FJD, UAM), Madrid, Spain

[3] CIMA, Pamplona, Spain

[4]Molecular Cytogenetics Group, Human Cancer Genetics Program, Spanish National Cancer Research Centre – CNIO, Madrid, Spain.

[5] Scottish National Blood Transfusion Service and ICAMS, University of Glasgow, Glasgow, UK

[6] VIVEbioTECH, San Sebastian, Spain

[7]Hematopoietic Innovative Therapies Division, Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas (CIEMAT) and Centro de Investigación Biomédica en Red de Enfermedades Raras (CIBERER), Madrid, Spain

[#]Present address:

Thérapie cellulaire et génique des maladies neurologiques de l'enfant et de l'adulte – NeuroGenCell

Equipe CARTIER

ICM - Institut du Cerveau et de la Moelle épinière

Hôpital Pitié-Salpêtrière

47 boulevard de l'Hôpital

75013 Paris, France

Tfn: +33 0157274413

E-mail: sergio.lopez-manzaneda@inserm.fr

*Corresponding author:

Jose-Carlos Segovia

Differentiation and Cytometry Unit

CIEMAT (Building 70)

Av. Complutense, 40

28040 Madrid

Tfn: +34913466268

Fax: +34913466484

E-mail: jc.segovia@ciemat.es

Short title: Disease Model via NHEJ-Precise Deletions.

ABSTRACT

Development of advanced gene and cell therapies for the treatment of genetic diseases requires confident animal and cellular models to test their efficacy and is crucial in the cases where no primary samples from patients are available. CRISPR/Cas9 technology, has become one of the most spread endonuclease tools for editing the genome at will. Moreover, it is known that the use of two guides tends to produce the precise deletion between the guides via NHEJ. Different distances between guides were tested (from 8 to 500 base pairs). We found that distances between the two cutting sites and orientation of Cas9 protein-DNA interaction are important for the efficiency within the optimal range (30-60 bp), we could obtain new genetically reproducible models for two rare disease, a Pyruvate Kinase Deficiency model, using human primary cells, and a (for *in vivo* primary hyperoxaluria therapeutic mouse model. We demonstrate that the use of a 2-guideRNAs at the optimal distance and orientation is a powerful strategy for disease modelling in those diseases where the availability of primary cells is limited.

INTRODUCTION

Basic biology and the study of the function of the genome has been based on the availability of function deficient models, either cells or organisms, and the association of these loss-of-function with gene mutations. The availability of these models has allowed the research of human genetic diseases and, even, the development of gene therapy strategies for their treatment (Keeler, ElMallah, and Flotte 2017) . The availability of endonucleases that can be directed to interact with very precise position in the cell genome has allowed the generation of knock-out models of any desired gene or position (Sakuma and Woltjen 2014; Silva et al. 2011; Deng et al. 2012).

Among all, CRISPR technology has become one of the most spread gene editing tools in the last years thanks to its easy design and its low cost. The action of these endonucleases produces the cleavage in both DNA strands in a precise manner according with the guide RNA (gRNA) position (Mojica and Montoliu 2016; Shibata et al. 2017; Nishimasu et al. 2014). The DNA cell machinery repairs these breaks either by non-homologous end joining (NHEJ), an error prone process, or by homology directed repair, that precisely corrects the damage. Both mechanisms have been extensively used to either eliminate the expression of a specific gene or to introduce new genetic material in a precise position of the cell genome, respectively. Repair by the non-homologous end joining machinery results in high variety of insertions, deletions (InDels) and sporadically inversions. This capacity to alter the original sequence has made these nucleases an excellent tool for the generation of knock out models, from bacteria to the human genome (Wassef et al. 2017; Islam 2018; Gilles, Schinko, and Averof 2015). Moreover, the deletion of specific regions that results in a recovery of function, either by eliminating a premature stop codon (Mettananda et al. 2017) or by deletion of specific gene regulators or silencers (Mojica and Montoliu 2016; Canver et al. 2014; Geisinger et al. 2016) has been suggested as an alternative therapy for specific diseases (Mettananda et al. 2017; Bonafont et al. 2019; Ousterout et al. 2015). In fact, a gene therapy

clinical trial for the re-expression of the fetal globin in adult b-thalassemia patients by means of knocking out the BCL11A protein regulator is ongoing("A Study to Assess the Safety, Tolerability, and Efficacy of ST-400 for Treatment of Transfusion-Dependent Beta-Thalassemia (TDT) - Full Text View - ClinicalTrials.Gov" n.d.).

NHEJ repairs double strand breaks in a non-predictive way, introducing InDels that can be very variable. Different reports have shown that the generation of two double strand breaks can facilitate the outcome of the NHEJ repair action(Canver et al. 2014; Zheng et al. 2017; Geisinger et al. 2016). Moreover, it has been proposed as a potential therapeutic option by eliminating mutated exons and recovering an almost normal although functional protein(Ousterout et al. 2015; Bonafont et al. 2019). Guo et al studied the efficacy of the NHEJ-precise deletion (NHEJ-PD) and how this process acts when guides are separated 23-148 apart, being the precise deletion of the DNA material between the two guide cuts the most common event(Guo et al. 2018). Thus, the use of two guides that could delete a defined DNA fragment and alter the open reading frame of the affected locus in an efficient and pseudo-controlled way could be used to generate KO models for the study of the biology of the cell or for the generation of cellular or animal models of rare diseases.

Here we reported the use of NHEJ-PD to generate cellular and animal models of genetic diseases. Moreover, we studied the optimal distances and the impact of the CRISPR-Cas9 machinery orientation in the NHEJ-PD.

MATERIAL AND METHODS

Cell lines

HEK293T cell line (human embryonic kidney cell line competent to replicate vectors carrying the SV40 T antigen; ATCC: CRL-3219) was cultured in Iscove's modified Dulbecco's medium (IMDM; Gibco), HyClone (10%; GE Healthcare) and penicillin/streptomycin (1%; Gibco). Cells were maintained at $5x10^5$ cell/mL concentration.

K562 cell line (chronic myelogenous leukemia; ATCC: CCL-243) was cultured in Iscove's modified Dulbecco's medium (IMDM; Gibco), HyClone (10%; GE Healthcare) and penicillin/streptomycin (1%; Gibco). Cells were maintained at $1x10^5$-$1x10^6$ cell/mL.

Incubation conditions were the same for all cell lines used: 37ºC, 5% CO_2 and 95% relative humidity.

Animals

All experimental procedures were approved by the Ethics Committee of the University of Navarra and the Institute of Public Health of Navarra according to European Council Guidelines. All the experiments using animal models complied with all relevant ethical regulations. Agxt1-/- mice (B6.129SvAgxttm1Ull) were bred and maintained in a pathogen-free facility with free access to standard chow and water. Agxt1-/- mice were genotyped as described(Salido et al. 2006). Age-matched C57BL/6J mice (Harlan laboratories) were used as control animals. A final dose of $1x10^{13}$ vg/kg ($5x10^{12}$ vg/kg/vector) was administered to 12-14-week-old Agxt1-/- male animals by intravenous injection. Animals were sacrificed and livers harvested 21 days after the administration of the vectors.

Design and generation of CRISPR/Cas9 tools

For the experiments performed *in vitro* in cell lines, two different plasmids were used to clone specific guides. The plasmids used contained guide RNA and RNA scaffold expression driven under U6 promoter and also the Cas9-ZsGreen sequences linked by a P2A (porcine teschovirus-1´s motif that produce co-translational "cleavage" of both proteins) under the CMV promoter. The different guides were designed based on Zhang Lab´s program (MIT) against PKLR gene and chosen by score and its location in the gene.

The vectors and gRNAs used for the generation of a KO model in vivo were previously designed and individually characterized(Zabaleta et al. 2018)

Human hematopoietic progenitor medium

Human hematopoietic progenitors were cultured in StemSpan SFEM I (Stem Cell Technologies) supplemented with human stem cell factor (hSCF, 100 ng/mL; EuroBioSciences), recombinant human thrombopoeitin (hTPO, 100 ng/mL; EuroBioSciences), human FMS-like tyrosine kinase 3 ligand (hFlt3-ligand, 100 ng/mL; EuroBioSciences), Glutamax (1%; Gibco), penicillin/streptomycin (1%; Gibco). Cells were maintained at $1x10^6$ cell/mL concentration. Incubation conditions were the following: 37ºC, 5% CO_2 and 95% relative humidity.

Erythroid differentiation protocol

Human hematopoietic progenitors were differentiated toward erythroid lineage using 4 different media. The first medium, days 1 to 7, was based on StemSpan SFEM I (Stem Cell Technologies) supplemented with hSCF (50 ng/mL; EuroBioSciences), hFlt3-ligand (16.7 ng/mL; EuroBioSciences), bone morphogenetic protein 4 (BMP-4, 6.7 ng/mL; Peprotech), human interleukin 3 (hIL-3, 6.7

158 ng/mL; EuroBioSciences), human interleukin 11 (hIL-11, 6.7 ng/mL; EuroBioSciences) and human
159 erythropoietin (h-EPO, 1.3 U/mL; Amgen). Cells were then transferred to the second medium in
160 which they were culture on days 7 to 14. Second medium was based on Iscove's modified
161 Dulbecco's medium Glutamine supplemented (IMDM GlutaMAX supplemented; Gibco) enriched
162 with BSA (1%; Sigma), insulin (0.01 mg/mL; Sigma), human transferrin (0.2 mg/mL; Sigma), β-
163 mercaptoethanol (91 µM; Gibco), penicillin/streptomycin (1%; Gibco), lipid mixture 1 (1x; Sigma),
164 ethanolamine (0.004%, Sigma), hSCF (5 ng/mL; EuroBioSciences), hIL-3 (6.7 ng/mL;
165 EuroBioSciences), hIL-11 (6.7 ng/mL; EuroBioSciences), hEPO (1.3 U/mL; Amgen), insulin-like
166 growth factor-1 (IGF-1, 20ng/mL; PeproTech) and hydrocortisone (1 µM; Sigma). After day 14 and
167 for 2 days, cells were transferred to the third medium that was also based on IMDM GlutaMAX
168 medium (Gibco) supplemented with BSA (1%; Sigma), insulin (0.01 mg/mL; Sigma), human
169 transferrin (0.2 mg/mL; Sigma), β-mercaptoethanol (91 µM; Gibco), penicillin/streptomycin (1%;
170 Gibco), lipid mixture 1 (1x; Sigma), ethanolamine (0.004%; Sigma) and hEPO (10 U/mL; Amgen).
171 Final enucleation was carried out for five days on the forth medium that was also based on IMDM
172 GlutaMAX medium (Gibco) supplemented with BSA (1%; Sigma), insulin (0.01 mg/mL; Sigma),
173 human transferrin (0.2 mg/mL; Sigma), β-mercaptoethanol (91 µM; Gibco), penicillin/streptomycin
174 (1%; Gibco), lipid mixture 1 (1x; Sigma) and ethanolamine (0.004%; Sigma).
175
176 All media were filtered by 0.22 µm before use. Sampling and analysis were done within this third
177 week. During erythroid differentiation, cells were maintained between $4x10^5$ and $4x10^6$ cell/mL
178 concentration. Incubation conditions were also 37ºC, 5% CO_2 and 95% relative humidity.
179
180 **Erythroid differentiation immunophenotype**

Immunophenotype of hematopoietic progenitors differentiated toward erythroid lineage was analyzed by flow cytometry using the following antibody combination: CD36-PE, clone CB38 (NL07), BD Pharmingen, 1µL ab/100µL; CD45-APCCy7, clone HI30, BioLegend, 1µL ab/100µL; CD71-PECy5, clone M-A712, BD Pharmingen, 1µL ab/100µL; CD235a-PECy7, clone HI264, BioLegend, 1µL ab/100µL). Cells were resuspended in a final volume of 100 µL, labeled for 30 minutes at 4ºC and washed with PBA (PBS with 0.2% sodium azide [Merck]; 1% BSA [Sigma]).DAPI (final concentration of 1 µg/ml) were added and events were recorded with a LSR Fortessa (BD Biosciences). Data were analyzed by FlowJo software (BD Biosciences).

Electroporation of K562 cell line

DNA delivery was carried out on 20 µL volume format using Amaxa 4D-Nucleofector System and Cell line nucleofector solution SF (Lonza). Total amount of DNA electroporated was 3 µg (1.5 µg each plasmid in the case of using two plasmids at the same time). $2x10^5$ cells were electroporated following manufacturer´s protocol under FF-120 program.

Electroporation of primary cells

Hematopoietic progenitor cells were electroporated with CRISPR/Cas9 plasmids as previously described, using Amaxa Nucleofector I (Lonza) and following manufacturer´s protocols. Cells were electroporated in cuvettes (electroporation volume of 100 µL) using Human $CD34^+$ Cells kit (Lonza) and U08 electroporation program. Total amount of plasmid used was 11.26 µg, hence, when two plasmids were electroporated at the same time, 5.63 µg of each were added to the mix. After electroporation, cells were cultured on 24-well plate (FALCON) at 2 mL of final volume. In the case of the experiments with selection based on ZsGreen reporter expression, sorting was performed

24 hours after electroporation. Aggregates were eliminated before sorting using 40 µm Cell Strainer (FALCON).

Deep sequencing analysis

Next generation sequencing (NGS) for a single cell suspension was carried out by STABVIDA and GeneWiz Companies. PCR products were treated and purified according to manufactures´s protocols. Samples were used for a library construction and sequenced with v3 chemistry in the lllumina MiSeq platform. Only variants represented in more than 1% were considered.

For deep sequencing of liver cells, total genomic DNA was extracted from the livers using NucleoSpin Tissue DNA extraction kit (Macherey Nagel) according to the manufacturer's instructions. A first PCR round was performed using primers that amplified 390bp of mouse Hao1 gene (containing the targeted region in the middle of the amplicon) carrying the universal adaptor M13 sequence (underlined) in their 5' end (forward 5' GTTGTAAAACGACGGCCAGTAGACCAATGTTTGTCAGAGG 3' and reverse 5' CACAGGAAACAGCTATGACCTAAAAGCATCCTAGGAAGGG 3'). Platinum Taq Polymerase (Invitrogen) was used. Subsequently, a second PCR was performed using primers targeting the universal M13 sequence and adding unique MID (multiplex identifier) barcodes to each sample. Sample pools were NGS-sequenced on the MiSeq Platfrom (Illumina, San Diego, United States) following manufacturer's protocol.

RESULTS

The use of guide pairs allows precise deletion of the DNA sequences between the two induced double strand breaks

As a first attempt we studied if the use of two guides simultaneously could reducere the heterogeneity of the NHEJ activity to inactivate a certain gene. We choose the *PKLR* gene, responsible for the expression of the liver and the erythroid pyruvate kinase enzymes (LPK and RPK, respectively). Mutations in this gene are responsible of the rare disease Pyruvate Kinase Deficiency. Aiming to mimic mutations occurring in PKD patients, we selected exon 9 of the *PKLR* gene, the second exon with a highest number of mutations reported (33 different along the exon). Two guide RNAs were designed to cut in two hotspots of exon 9 where several mutations where grouped (Fig 1A; Fig EV1A).

Human HEK293T cells were electroporated with all-in-one DNA plasmids containing sequences for one guide RNA, spCas9 nuclease and the green fluorescent protein (ZsGreen). Plasmids were electroporated individually or together. Seventy two hours after the electroporation cells were sorted based on ZsGreen expression (Fig 1B) and expanded up to confluence in 6-well plates. Surveyor assay was conducted in order to determine the activity of the designed guides and of the combination of both of them together. When guides were electroporated individually, the expected pattern of surveyor DNA fragments was obtained with high cutting efficiencies (40% for guide I and 59% for guide II; Fig 1C, 1D). When both guides were electroporated together, surveyor DNA fragments obtained were the combination of the cut of the two guides with the loss of the DNA fragment between the cut sites of the two guides (31 bp fragment). No evidence of fragments resulting from the cut of only one guide was observed (Fig 1C, 1D). These results confirmed previous evidences, demonstrating that the deletion of the DNA sequences between

two guides, when used simultaneously, utilizing the CRISPR/Cas9 system, was highly frequent and the most common event of the NHEJ repair mechanism.

In order to define the joint between the two generated DSB we amplified the region of the joint and cloned it in PCR plasmids for their analysis by Sanger sequencing. All colonies analyzed that contained non-wt sequences presented the precise joint between the two DNA sides (NHEJ-PD; Fig EV1B).

Distance between the two guides (30-500 bp apart) do not affect the generation of precise deletions

To test if the process was cell dependent, we performed a similar study using K562 (chronic myelogenous leukaemia cell line). Furthermore, in order to determine if the generation of the NHEJ-PD, being the most common event, was dependent of the distance between the guides used, five additional pairs were designed separated 30 bp to 500 bp, sharing always the upstream guide. Guide RNA pairs were designed to delete regions at exon 9, intron 9 and exon 10 of the *PKLR* gene (Fig 2A).

DNA plasmid constructs containing gRNAs and Cas9 were electroporated into K562 cells. Twenty-four hours post-transduction, cells were sorted based on the ZsGreen expression and expanded as described above. Editing efficacy was tested by PCR amplification of the affected region. Guide pairs showed a good capacity to delete gDNA, decreasing the size of the PCR product at the expected sizes. Indeed, wild type size products were not found (Fig 2B).

To deeply elucidate the outcomes that happened in the gene editing with the different guide pairs, PCR amplicons of the joint sequences were analyzed by next generation sequencing. Guide pair

number 1, which precise deletion was predicted on 31 bp, presented only two events, being the most represented event the mentioned precise deletion with an 84.60% frequency. The other recorded event was also a deletion of 32 bp that was the same 31 bp deletion without an additional A at 5' of the deleted sequence (precise deletion+1). Similar results were obtained with the other guide pair combinations (guide pair 2-63bp deletion: 69.70%; guide pair 3-119bp deletion: 74.05%; guide pair 4-239bp deletion: 69.18%; guide pair 5-490bp deletion: 95.84%) (Fig 2C). In all the other combinations, the remaining events, different from the precise deletion, were slight modifications surrounding it (Fig EV2A). When considering precise deletion plus precise deletion+1 together, the efficiency was always higher than 75% of the sequences analyzed in all the cases analyzed (Fig EV2C)

An additional guide pair with cutting sites separated 8 bp was also tested. Due to the proximity of the two cutting sites, the gRNAs were overlapping one each other. In this case, InDels detected where around one or the other cut site. No precise deletions or deletions containing the fragment between the cut sites were observed (Fig EV2B).

Finally, to corroborate the capacity of these deletions to knock-out the gene, a western blot assay of the edited populations was performed. No RPK protein was detected in any of the edited samples, indicating that any combination of guides was capable of eliminating completely the production of the protein (Fig 2D).

PAM orientation of the guides affects the efficacy of precise deletions

One of the potential reasons to explain why a short distance between the two guides is less efficient in generating precise deletions could be the difficulties to access the two cutting sites at

296 the same time due to allosteric or space impediments. To study these potential difficulties, we
297 designed and cloned gRNAs spaced approximately 30 bp that facilitated the Cas9-DNA interactions
298 in all 4 possible orientations, PAM-in/PAM-in (or-1), PAM-in/PAM-out (or-2), PAM-out/PAM-in (or-
299 3) and PAM-out/PAM-out (or-4) (Fig 3A). Human K562 cells were electroporated with pairs of all-
300 in-one DNA plasmids containing sequences for one gRNA, spCas9 nuclease and ZsGreen. Twenty
301 four hours after electroporation cells were sorted based on ZsGreen expression and expanded as
302 above. Total DNA was extracted from the 4 different combinations. Three different experiments
303 were conducted with each combination. NGS was performed and bioinformatics analysis was
304 performed to define the incidence NHEJ-PD and additional InDels in the genomic region. The
305 combination or-1 (PAM-in/PAM-in) showed the best NHEJ-PD efficiency and the or-4 combination
306 (PAM-out/PAM-out) was the worst one, including a high proportion of non-edited sequences.
307 Combinations or-2 (PAM-in/PAM-out) and or-3 (PAM-out/PAM-in) were also efficient but the
308 percentage of InDels different from the precise deletion were higher (Fig 3B and Fig EV3A).
309 Interestingly, if we analyze the cases where only one guide cut, the guide oriented PAM-in was
310 always more efficient (data not shown). As observed above, when we considered together the
311 NHEJ-PD plus the NHEJ-PD+1 the percentage of edited sequences were close to 100% in the or-1
312 combination (PAM-in/PAM-in) (Fig EV3B). Altogether, these experiments showed that steric
313 difficulties could be taking place in the efficacy of the cutting activity of the guides, suggesting that
314 orientation of the PAM and positioning of the Cas9 nuclease on the DNA are important in the
315 generation of short deletions by NHEJ.
316
317 **NHEJ-PD generates a cellular model of PKD-like CD34 primary cells**
318 Next, we wanted to test if the NHEJ-PD would also occur in primary cells and if this mechanism
319 could be used to generate cellular models of genetic diseases in primary cells. For that, we applied

320 the guide pair 1 to knock-out the PKLR gene in human cord blood CD34+ (CB-CD34+) cells. CD34+
321 cells are responsible for the generation of all the hematopoietic lineages, including the erythroid
322 lineage where RPK expression is required for the proper production of mature erythroid cells.
323 Reduction or lack of RPK function produces Pyruvate Kinase Deficiency (PKD). Healthy CB-CD34+
324 where electroporated, sorted and differentiated towards erythroid lineage. Afterwards, the
325 modified region was analyzed by deep sequencing. Similarly to the results obtained in
326 immortalized cell lines, NHEJ-PD was consistently achieved in edited CB-CD34+ cells, easily
327 traceable following the decrease of the size of the PCR product (Fig 4A). Furthermore, the
328 proportion of precise deletions was followed by the analysis of the cleavage produced by BstXI
329 which target sequence is generated after precise deletion and is not present in the wild type
330 neither in other deletions such as the 32 bp one (Fig 4B). Interestingly, the most represented
331 events recorded in this primary cells were the same found in the cell line studies (31 bp deletion:
332 82±2 %; deletion 32 bp: 16±4 %). In addition, in one of the samples two other variants were found,
333 representing around 1.5% of the events. Both were deletions, one of 40 and another of 55 bp, also
334 centered in the precise deletion region (Fig 4C). Thus, the use of this two selected guides
335 demonstrated the capacity of consistently limit the editing events to just this two that, afterwards,
336 generate stop codons (Fig 4D).
337
338 In order to check if the edited cells reproduced the *in vitro* PKD phenotype, CD34-PKLR$^{-/-}$ cells were
339 cultured in specific media to allow erythroid differentiation. After a two-week differentiation
340 protocol, cells were analyzed in terms of cell morphology, immunophenotype and pyruvate kinase
341 activity. Non-edited CB-CD34+ cells and bone marrow CD34 cells obtained from a PKD patient
342 were also differentiated following the same protocol. No differences were observed in the *in vitro*
343 differentiation of the different samples demonstrating that PKD progenitors differentiate normally

344 *in vitro* under the conditions used (Fig 5A). Similarly, analysis of cytospins performed at the end of
345 the culture corroborated also the same degree of differentiation in all of the groups,
346 independently of its deficiency in PKD, having the orthochromatic erythroblast population as the
347 most represented cell type (Fig 5B).
348
349 Finally, edited, control and PKD patient cells were tested for pyruvate kinase activity. PKD activity
350 was tested together with the hexokinase (HK) activity, to normalize the increased PK activity of the
351 most immature erythroblasts. Ratio PK/HK of the edited cells showed a decreased enzymatic
352 activity comparable with the values obtained in the PKD patient sample differentiated in parallel
353 and, approximately, 0.6 times the values of the healthy control cells (Fig 5C). Overall, NHEJ-PD
354 after Cas9 DNA breaks by means of the use of guide pair combination generates an *in vitro* cellular
355 model, resembling the characteristics of patient deficient cells, which would be very valuable for
356 further testing alternative therapeutic strategies.
357
358 **NHEJ-PD generates *in vivo* knock-out animal model**
359 In order to study if the use of two closed guides to minimize the variability of NHEJ could also work
360 *in vivo*, we addressed the knock-out the mouse *Hao1* gene, which encodes for the glycolate
361 oxidase (GO), in the liver of mice injected with AAV vectors expressing 2 different gRNAs and the
362 *Staphiloccocus Aureus* Cas9 protein. Both gRNAs targeted opposite strands of exon 2 of mouse
363 *Hao1* gene, were oriented in PAM-out orientation and DSBs were 64bp apart (Fig 6A). NGS analysis
364 of the livers of animals treated with both vectors revealed that inside the edited cells, the majority
365 of the sequences presented the deletion of the 64bp harbored between the two DSBs (Fig 6B).
366 Moreover, these genetic modifications led to a decrease in *Hao1* mRNA levels (Fig 6C) and to the
367 complete elimination of GO protein analyzed by WB (Fig 6D).

DISCUSSION

Herein, we have shown that the use of two guide RNAs to drive Cas9 cutting in adult cells facilitates the precise deletion of the sequences between them and reduces, or even eliminates, the heterogeneity of the error-prone NHEJ, allowing the knock down or knock out of the desired gene and the generation of cellular or animal models, with an almost homogeneous genotype. Previous works have indicated the preference for the NHEJ repair by the precise deletion between the two DSB generated when two guides are used simultaneously(Guo et al. 2018). Here, we demonstrate that this characteristic facilitates the generation of cellular and animal models with a more homogeneous and controlled genotype modification.

We have observed that the NHEJ-PD process generated by the two cuts occurs in different cell types, in primary and in immortalized cell lines, in different species and using different Cas9 proteins (either spCas9 or saCas9). Thus, the process seems to be mainly dependent on the NHEJ mechanism.

The combination of NHEJ-PD and the design of gRNAs to drive the activity of the Cas9 protein to very specific genome sites allow the prediction and generation of genotypes that could resemble very precisely mutations already described in humans. Here, we selected a specific region in exon 9 that generated mutations in the same location that a cluster of genetic mutations founds in PKD patients. This procedure is 10-times more efficient that the introduction of specific mutations by means of Homology Directed Repair and has allowed us to apply it to primary human hematopoietic progenitors. This procedure allows then to generate primary cells carrying almost similar human mutations in normal cells from healthy donors for those samples that are difficult or not ethically justified, and that could be used as cellular models for testing new available

392 therapeutic strategies. Also, NHEJ-PD allow the selection of concrete sites which alterations could
393 generate stop codons nearby, facilitating the generation of complete knock-out cellular models in
394 a more efficient way. Moreover, we demonstrated that the same strategy applied to animal
395 models is also very efficient.
396
397 Previous evidences have suggested that a minimum distance between the two cutting sites is
398 required for the optimal NHEJ-PD. We have observed a similar requirement with an almost
399 avoidance of precise deletion when cutting site guides were designed 8 bp apart. In addition, we
400 demonstrate that also the orientation of the guides, and consequently the orientation of the Cas9
401 enzyme, is also important to allow optimal NHEJ-PD. A higher efficiency was obtained when both
402 guides had the PAM sequence located in the sequence to be deleted, what we called PAM-in. The
403 worst cutting efficiency was observed in the opposite orientation, PAM-out/PAM-out. Taking into
404 account that the last 3 bp where common for both guides, steric interferences could be the
405 responsible of this worse efficacy. Moreover, in those cases where PAM-in and PAM-out guides
406 were used simultaneously, the PAM-in guide was more efficient than the PAM-out. Thus, factors
407 like steric interactions but also orientation towards the sequence to be deleted are important for
408 the efficiency of the CRISPR/Cas9 system to generate NHEJ-PD.
409
410 The use of guide pairs and the generation of NHEJ-PD has been also proposed as a therapeutic tool
411 to precisely eliminate specific sequences that allowed the restoration of an almost completely
412 functional protein. Mutations at dystrophin gene have been precisely eliminated by removing the
413 sequence between exons 45-55 with the use of two guides restoring the function, and the
414 Duchenne muscular dystrophy phenotype(Ousterout et al. 2015). The same strategy has been
415 followed in epidermolysis bullosa, a recessive disorder caused by frameshift mutations at COL7A1

gene. The excision of the mutated sequences at exon 80 restores the collagen function and recovers a normal phenotype in keratinocytes(Bonafont et al. 2019). Our results clearly support this strategy and points out the importance of the design of the strategy in terms of CAs9 positioning.,

A potential drawback of the use of two guides is that the amount of off target mutations is multiplied by two. The optimization of the on line tools for the design of guide RNAs and the use of better systems to characterize them will help to improve their use.

In summary, we demonstrate here that the use of guide pairs in combination with the Cas9 nuclease generate NHEJ-PD in adult cells that facilitate the generation of cellular and animal models of specific mutations found human diseases. Moreover, these combinations could be used as a therapeutic tool for the treatment of genetic diseases where the excision of a defined sequence restores the function of the modified protein.

436 **REFERENCES**

437 "A Study to Assess the Safety, Tolerability, and Efficacy of ST-400 for Treatment of Transfusion-
438 Dependent Beta-Thalassemia (TDT) - Full Text View - ClinicalTrials.Gov." n.d. Accessed
439 February 17, 2020. https://clinicaltrials.gov/ct2/show/NCT03432364?term=BCL11A&cond=B-
440 Thalassemia&draw=2&rank=1.

441 Bonafont, Jose, Ángeles Mencía, Marta García, Raúl Torres, Sandra Rodríguez, Marta Carretero,
442 Esteban Chacón-Solano, et al. 2019. "Clinically Relevant Correction of Recessive Dystrophic
443 Epidermolysis Bullosa by Dual SgRNA CRISPR/Cas9-Mediated Gene Editing." *Molecular*
444 *Therapy* 27 (5): 986–98. https://doi.org/10.1016/j.ymthe.2019.03.007.

445 Canver, Matthew C., Daniel E. Bauer, Abhishek Dass, Yvette Y. Yien, Jacky Chung, Takeshi Masuda,
446 Takahiro Maeda, Barry H. Paw, and Stuart H. Orkin. 2014. "Characterization of Genomic
447 Deletion Efficiency Mediated by Clustered Regularly Interspaced Palindromic Repeats
448 (CRISPR)/Cas9 Nuclease System in Mammalian Cells." *Journal of Biological Chemistry* 289
449 (31): 21312–24. https://doi.org/10.1074/jbc.M114.564625.

450 Deng, Dong, Chuangye Yan, Xiaojing Pan, Magdy Mahfouz, Jiawei Wang, Jian Kang Zhu, Yigong Shi,
451 and Nieng Yan. 2012. "Structural Basis for Sequence-Specific Recognition of DNA by TAL
452 Effectors." *Science* 335 (6069): 720–23. https://doi.org/10.1126/science.1215670.

453 Geisinger, Jonathan M., Sören Turan, Sophia Hernandez, Laura P. Spector, and Michele P. Calos.
454 2016. "In Vivo Blunt-End Cloning through CRISPR/Cas9-Facilitated Non-Homologous End-
455 Joining." *Nucleic Acids Research* 44 (8): 1–15. https://doi.org/10.1093/nar/gkv1542.

456 Gilles, Anna F., Johannes B. Schinko, and Michalis Averof. 2015. "Efficient CRISPR-Mediated Gene
457 Targeting and Transgene Replacement in the Beetle Tribolium Castaneum." *Development*
458 *(Cambridge)* 142 (16): 2832–39. https://doi.org/10.1242/dev.125054.

459 Guo, Tao, Yi Li Feng, Jing Jing Xiao, Qian Liu, Xiu Na Sun, Ji Feng Xiang, Na Kong, et al. 2018.

460 "Harnessing Accurate Non-Homologous End Joining for Efficient Precise Deletion in
461 CRISPR/Cas9-Mediated Genome Editing." *Genome Biology* 19 (1): 1–20.
462 https://doi.org/10.1186/s13059-018-1518-x.

463 Islam, Waqar. 2018. "CRISPR-Cas9; an Efficient Tool for Precise Plant Genome Editing." *Molecular*
464 *and Cellular Probes* 39: 47–52. https://doi.org/10.1016/j.mcp.2018.03.006.

465 Keeler, A. M., M. K. ElMallah, and T. R. Flotte. 2017. "Gene Therapy 2017: Progress and Future
466 Directions." *Clinical and Translational Science* 10 (4): 242–48.
467 https://doi.org/10.1111/cts.12466.

468 Mettananda, Sachith, Chris A. Fisher, Deborah Hay, Mohsin Badat, Lynn Quek, Kevin Clark, Philip
469 Hublitz, et al. 2017. "Editing an α-Globin Enhancer in Primary Human Hematopoietic Stem
470 Cells as a Treatment for β-Thalassemia." *Nature Communications* 8 (1).
471 https://doi.org/10.1038/s41467-017-00479-7.

472 Mojica, Francisco J.M., and Lluis Montoliu. 2016. "On the Origin of CRISPR-Cas Technology: From
473 Prokaryotes to Mammals." *Trends in Microbiology* 24 (10): 811–20.
474 https://doi.org/10.1016/j.tim.2016.06.005.

475 Nishimasu, Hiroshi, F. Ann Ran, Patrick D. Hsu, Silvana Konermann, Soraya I. Shehata, Naoshi
476 Dohmae, Ryuichiro Ishitani, Feng Zhang, and Osamu Nureki. 2014. "Crystal Structure of Cas9
477 in Complex with Guide RNA and Target DNA." *Cell* 156 (5): 935–49.
478 https://doi.org/10.1016/j.cell.2014.02.001.

479 Ousterout, David G., Ami M. Kabadi, Pratiksha I. Thakore, William H. Majoros, Timothy E. Reddy,
480 and Charles A. Gersbach. 2015. "Multiplex CRISPR/Cas9-Based Genome Editing for Correction
481 of Dystrophin Mutations That Cause Duchenne Muscular Dystrophy." *Nature*
482 *Communications* 6: 1–13. https://doi.org/10.1038/ncomms7244.

483 Sakuma, Tetsushi, and Knut Woltjen. 2014. "Nuclease-Mediated Genome Editing: At the Front-Line

484 of Functional Genomics Technology." *Development Growth and Differentiation* 56 (1): 2–13.
485 https://doi.org/10.1111/dgd.12111.
486 Salido, Eduardo C., Xiao M. Li, Yang Lu, Xia Wang, Alfredo Santana, Namita Roy-Chowdhury,
487 Armando Torres, Larry J. Shapiro, and Jayanta Roy-Chowdhury. 2006. "Alanine-Glyoxylate
488 Aminotransferase-Deficient Mice, a Model for Primary Hyperoxaluria That Responds to
489 Adenoviral Gene Transfer." *Proceedings of the National Academy of Sciences of the United*
490 *States of America* 103 (48): 18249–54. https://doi.org/10.1073/pnas.0607218103.
491 Shibata, Mikihiro, Hiroshi Nishimasu, Noriyuki Kodera, Seiichi Hirano, Toshio Ando, Takayuki
492 Uchihashi, and Osamu Nureki. 2017. "Real-Space and Real-Time Dynamics of CRISPR-Cas9
493 Visualized by High-Speed Atomic Force Microscopy." *Nature Communications* 8 (1): 1–9.
494 https://doi.org/10.1038/s41467-017-01466-8.
495 Silva, George, Laurent Poirot, Roman Galetto, Julianne Smith, Guillermo Montoya, Philippe
496 Duchateau, and Frederic Paques. 2011. "Meganucleases and Other Tools for Targeted
497 Genome Engineering: Perspectives and Challenges for Gene Therapy." *Current Gene Therapy*
498 11 (1): 11–27. https://doi.org/10.2174/156652311794520111.
499 Wassef, M., A. Luscan, A. Battistella, S. Le Corre, H. Li, M. R. Wallace, M. Vidaud, and R.
500 Margueron. 2017. "Versatile and Precise Gene-Targeting Strategies for Functional Studies in
501 Mammalian Cell Lines." *Methods* 121–122: 45–54.
502 https://doi.org/10.1016/j.ymeth.2017.05.003.
503 Zabaleta, Nerea, Miren Barberia, Cristina Martin-Higueras, Natalia Zapata-Linares, Isabel Betancor,
504 Saray Rodriguez, Rebeca Martinez-Turrillas, et al. 2018. "CRISPR/Cas9-Mediated Glycolate
505 Oxidase Disruption Is an Efficacious and Safe Treatment for Primary Hyperoxaluria Type I."
506 *Nature Communications* 9 (1): 5454. https://doi.org/10.1038/s41467-018-07827-1.
507 Zheng, Qiupeng, Xiaohong Cai, Meng How Tan, Steven Schaffert, Christopher P Arnold, Xue Gong,

508 Chang Zheng Chen, and Shenglin Huang. 2017. “Precise Gene Deletion and Replacement
509 Using the CRISPR/Cas9 System in Human Cells.” *BioTechniques* 62 (1): xxii.
510 https://doi.org/10.2144/000114508.

511

512 **ACKNOWLEDGEMENTS**

513 The authors would like to thank Miguel A. Martin for the careful maintenance of NSG mice, and
514 Mrs Aurora de la Cal, María del Carmen Sánchez, Soledad Moreno, Nadia Abu-sabha, Montserrat
515 Aldea and Mr Sergio Losada for their dedicated administrative help. This work was supported by
516 grants from “Ministerio de Ciencia, Innovación y Universidades y Fondo Europeo de Desarrollo
517 Regional (FEDER)” (SAF2017-84248-P), “Fondo de Investigaciones Sanitarias, Instituto de Salud
518 Carlos III” (Red TERCEL; RD16/0011/0011) and Comunidad de Madrid (AvanCell, B2017/BMD-
519 3692). The authors also thank Fundación Botín for promoting translational research at the
520 Hematopoietic Innovative Therapies Division of the CIEMAT. CIBERER is an initiative of the
521 “Instituto de Salud Carlos III” and “Fondo Europeo de Desarrollo Regional (FEDER)”.

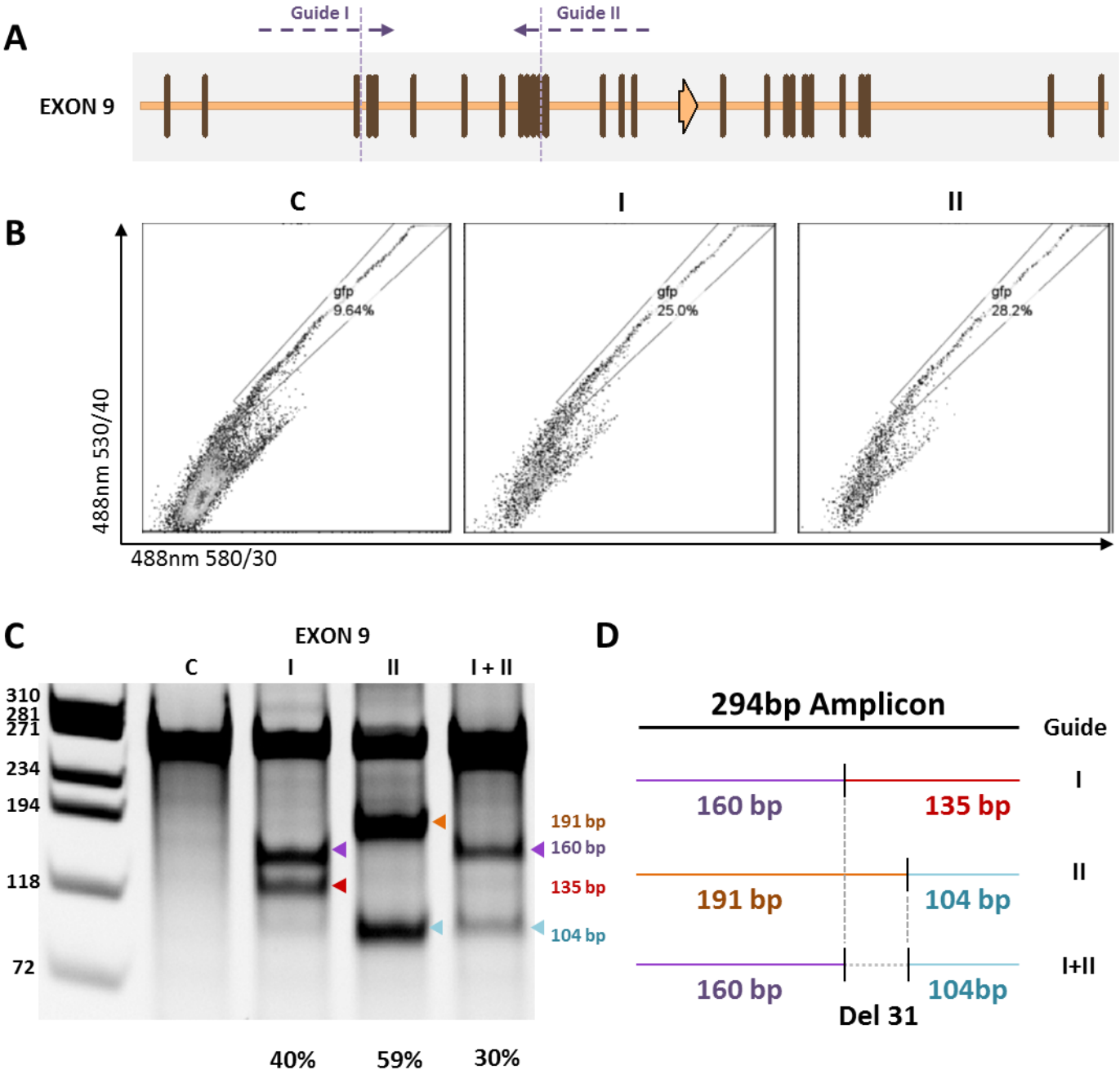
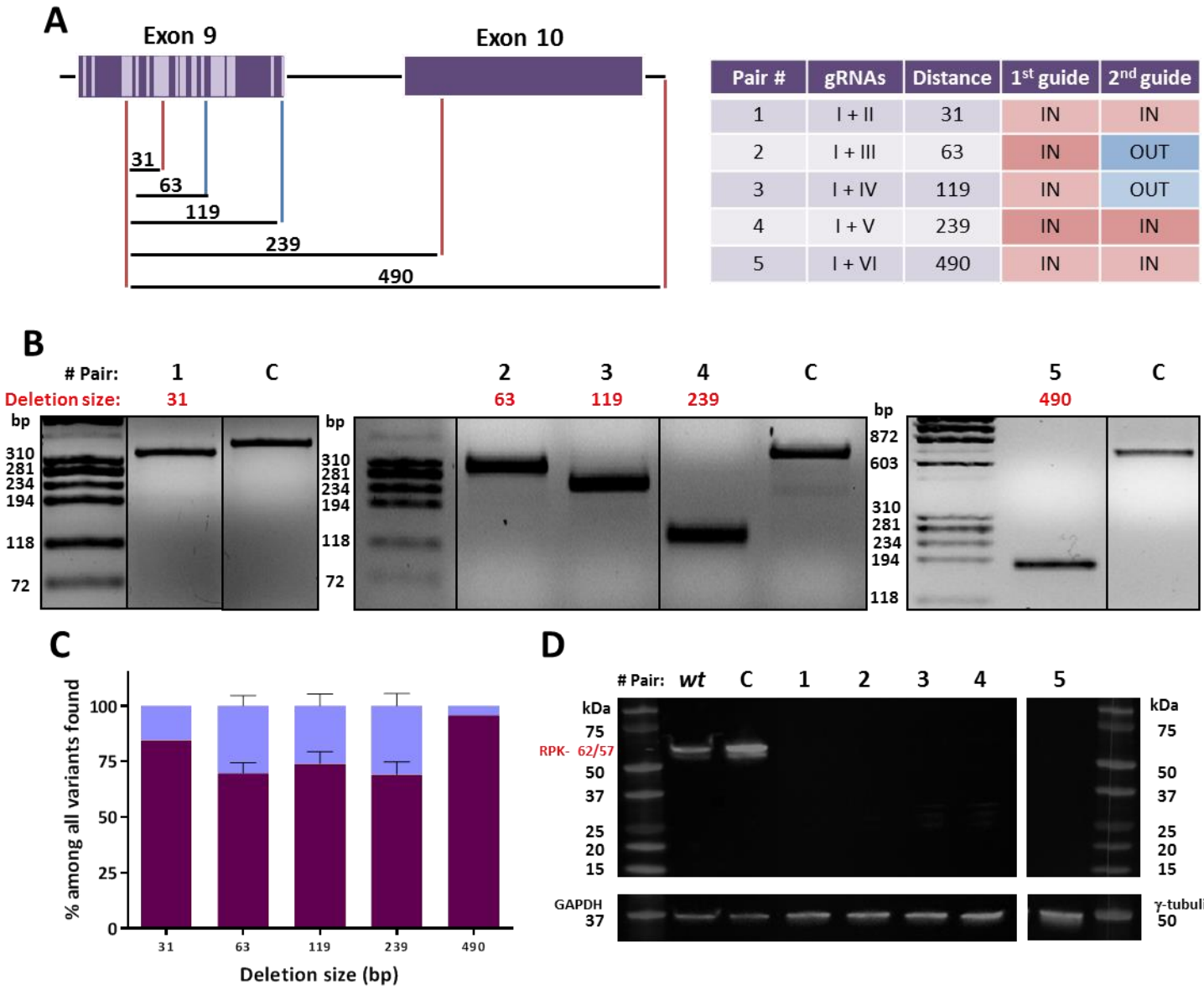


Figure 1. Design and use of CRISPR gRNAs in HEK293T cell line

(A) Exon 9 of the human *PKLR* gene is represented. Vertical lines and arrows are reported mutations, corresponding to one or more than one nucleotide affected, respectively. Horizontal dashed lines represent the two guide RNAs used. Vertical dashed lines represent predicted cut

sites of guide RNAs selected. **(B)** Sorting of green fluorescence based on the ZsGreen present in the electroporated plasmids. 530/40 vs 580/30 nm dot-plot was used to distinguish ZsGreen fluorescence from auto fluorescence. The figure represent the ZsGreen fluorescence of the cell electroporated with the plasmid DNA containing only the Cas9 and the ZsGreen cDNAs (C), or containing the Cas9, the ZsGreen cDNAs and the sequences of the guide I(I) or the guide II (II). **(C)** Surveyor assay of the DNA obtained from the cells electroporated with the pasmids containing no guide, guide I, guide II or both guides together (I+II). The gene editing generate a characteristic band pattern for guide I (160 bp + 135 bp), for guide II (191 bp + 104 bp) and a mixed pattern when both are used at the same time (160 bp + 104 bp). **(D)** Scheme of the size bands that appear in (C) showing the 31bp deletion resulting after the simultaneous electroporation of guide I and II.



Pair #	gRNAs	Distance	1st guide	2nd guide
1	I + II	31	IN	IN
2	I + III	63	IN	OUT
3	I + IV	119	IN	OUT
4	I + V	239	IN	IN
5	I + VI	490	IN	IN

Figure 2. DNA and protein analysis of K562 cells electroporated with 5 guide pairs.

(A) Left: Scheme of the 5 pairs tested. In dark purple: *PKLR* gene, exons are represented as boxes and introns as horizontal lines. Light purple represents mutations reported in exon 9. Vertical lines: position of the guides used, the different colors represent PAM position respect to the deletion, red lines: PAM "in" (inside the deletion sequence); blue lines: PAM "out" (outside the

deletion sequence). Numbers above black horizontal lines are the size of the deletion generated in each combination. Right: list of the guides used, size of the deletion generated and its PAM disposition **(B)** PCR amplicons of the region of all combinations. Up: Pair number. Red: size of the precise deletion. C: control sample, K562 cells electroporated with Cas9-P2A-ZsGreen without guides and sorted. Pairs 1 - 4 were evaluated performing a 320 bp PCR meanwhile pair 5 was checked by a 700 bp PCR. **(C)** Next generation sequencing results from variants of the pairs tested, represented in more than 1%. Dark purple: precise deletion. Light blue: variants with modifications around expected deletion. Data are presented as mean ± SEM and Tukey's Multiple Comparison statistical test was used to evaluate differences between groups. **(D)** Western blot of RPK pairs tested. Wt: non-electroporated K562 cells. C: control sample, K562 cells electroporated with Cas9-P2A-ZsGreen without guides and sorted. Black numbers: different pairs tested Red: sizes of RPK protein, complete (62 kDa) or after physiological N-terminal excision (57 kDa). GAPDH and γ-tubulin were used as loading control.

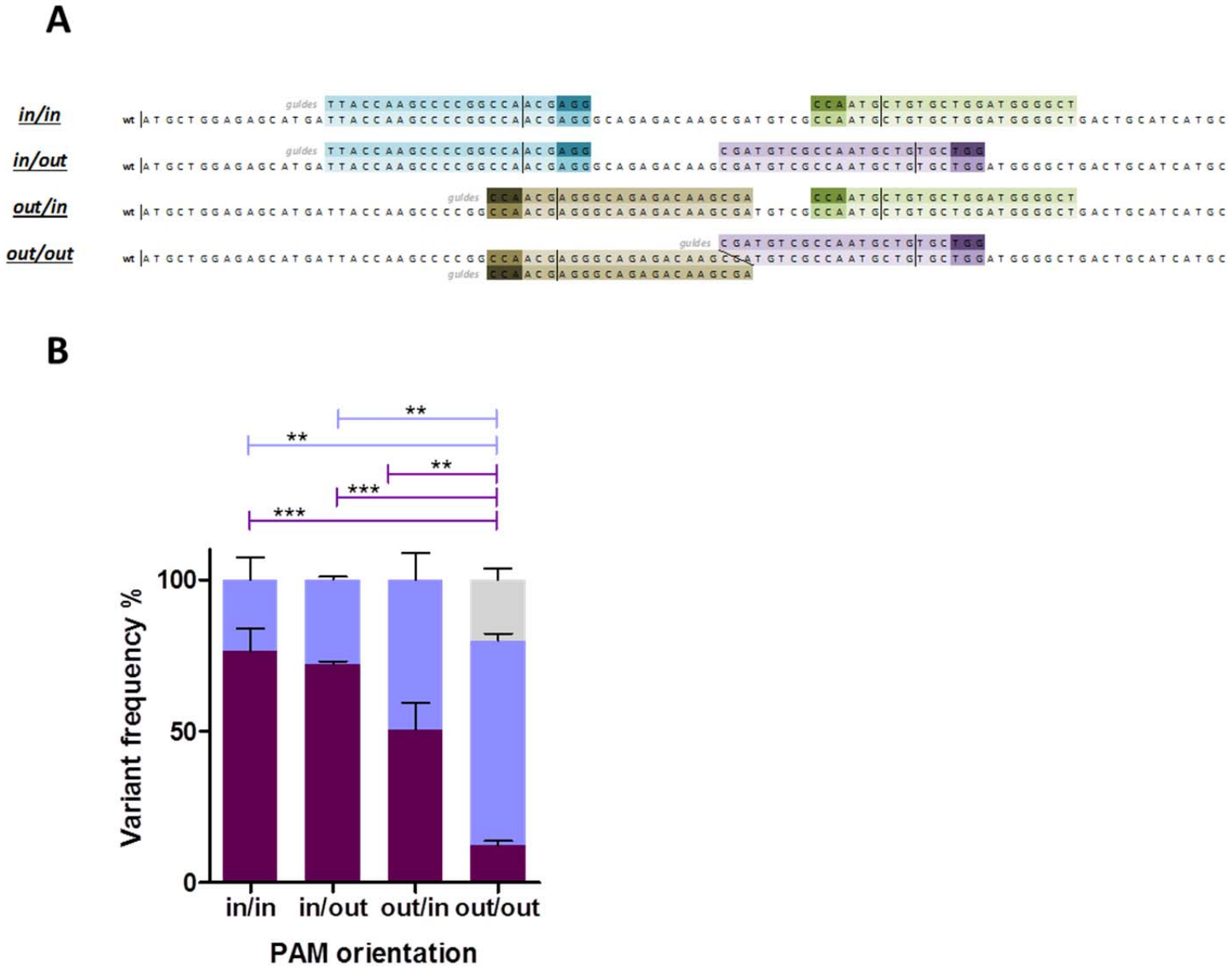


Figure 3. DNA analysis of K562 cells electroporated with 4 guide pairs separated 28-34pb apart providing all the PAM-In and PAM-out combinations.

(A) Scheme of the 4 different combinations tested. In light colors: guides. In dark colors: PAMs sequences. The orientations of the PAMs are represented on the left, being "in" when the PAM is inside of the precise deletion and "out" when it is outside. **(B)** Next generation sequencing results

from variants of the different combinations, represented in more than 1%. Dark purple: precise deletion. Light purple: variants with modifications around expected deletion. Grey: wild type. Data are presented as mean ± SEM and Tukey's Multiple Comparison statistical test was used to evaluate differences between groups.

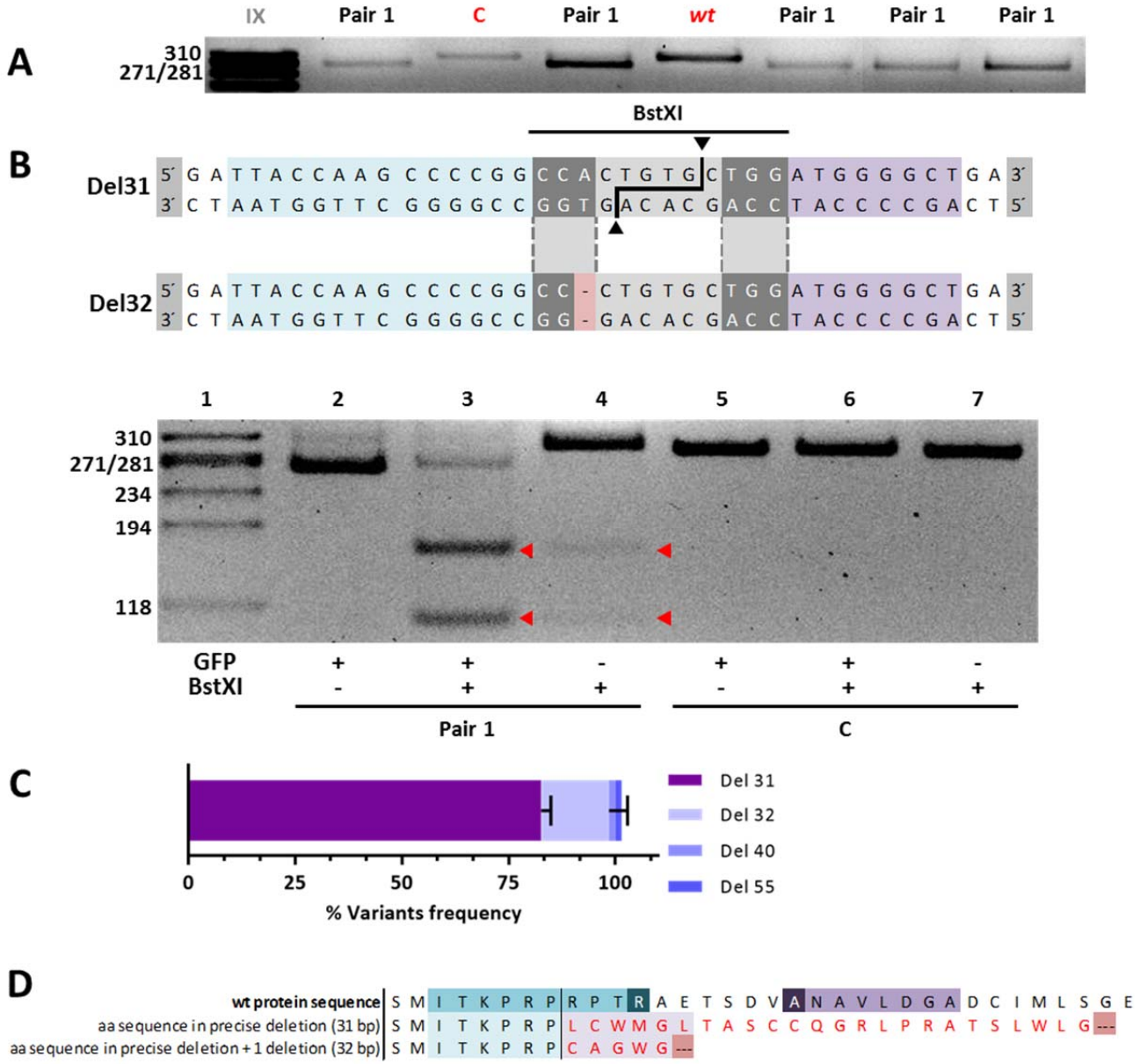


Figure 4. Primary CD34+ cells modified by Pair 1 gRNAs and differentiated towards erythroid lineage.

(**A**) Exon 9 PCR amplicons from different experiments of hematopoietic progenitor cells differentiated *in vitro* toward erythroid lineage and edited with guides I + II (Pair 1). C: control cells electroporated with Cas9- P2A-ZsGreen without guides. I + II and C are the ZsGreen^{+} fraction selected by sorting. Wt: non-electroporated cells. Pair 1: edited cells from different experiments.

(**B**) Up: Sequences resultant from the two main deletions produced after gene edit. Blue: region covered by guide I. Purple: region covered by guide II. Sequence in grey: BstXI recognition site. Light red: differential nucleotide, not present in 32 bp deletion, that is required for BstXI recognition. Down: Lane 1, molecular weight reference. 2-4 lanes: hematopoietic progenitor cells differentiated *in vitro* toward erythroid lineage and edited with guides I + II. 5-7 lanes: control cells electroporated with Cas9-P2A-ZsGreen without guides. ZsGreen, + or – fractions separated by sorting. + or – BstXI, treated or not with the enzyme. Red triangles, DNA cut after enzyme digestion. (**C**) Next generation sequencing results from two edited DNA samples, colors represent each of the deletions found named Del 31, Del 32, Del 40 and Del 55. (**D**) Amino acid sequence resultant after 31 and 32 bp deletions. Red amino acids are the alerted ones with respect to the wt. Both deletions generate premature stop codons represented by dots on red background.

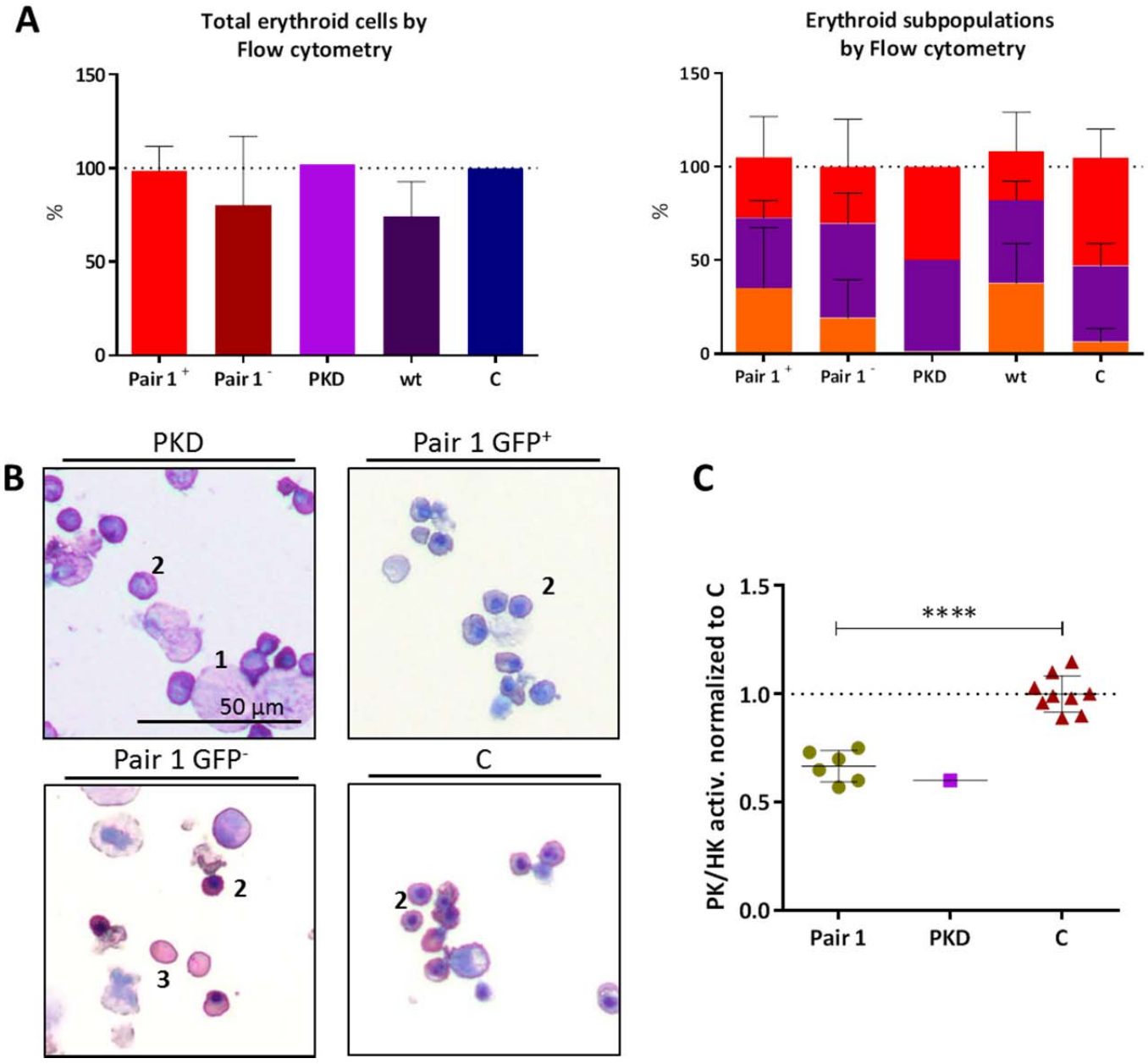


Figure 5. Immunophenotype and activity of hematopoietic cells edited and differentiated towards erythroid lineage.

(A) Left: Percentage of cells considered erythroid based on their CD36/CD45 staining (it includes cells in transition from $CD45^{+}/CD36^{+}$ to $CD45^{-}/CD45^{-}$; values normalized with respect to control samples). Right: erythroid maturation followed by CD71/CD235a staining. Immature:

$CD71^+/CD235a^+$ (orange); mature: $CD71^-/CD235a^+$ (red); transition from immature to mature: (purple). (n=6). PKD: cells differentiated from patient bone marrow sample. Pair 1: cells electroporated with guides I + II. C: control cells electroporated with Cas9-P2A-ZsGreen without guides. + or - : ZsGreen positive or negative fraction separated by sorting. Wt: non-electroporated/sorted cells. **(B)** Cytospin after 14 days of erythroid differentiation. 1: polychromatophilic erythroblast; 2: orthochromatic erythroblast; 3: reticulocyte. **(C)** PK/HK activity of edited cells after erythroid differentiation. Green circles: $ZsGreen^+$ fraction cells electroporated with the two guides from different experiments. Purple square: PKD cells differentiated from patient bone marrow sample. Red triangles: control cells. Measures are normalized with respect of PK/HK of control samples, which were consider =1 (**** P<0.0001; n=6)

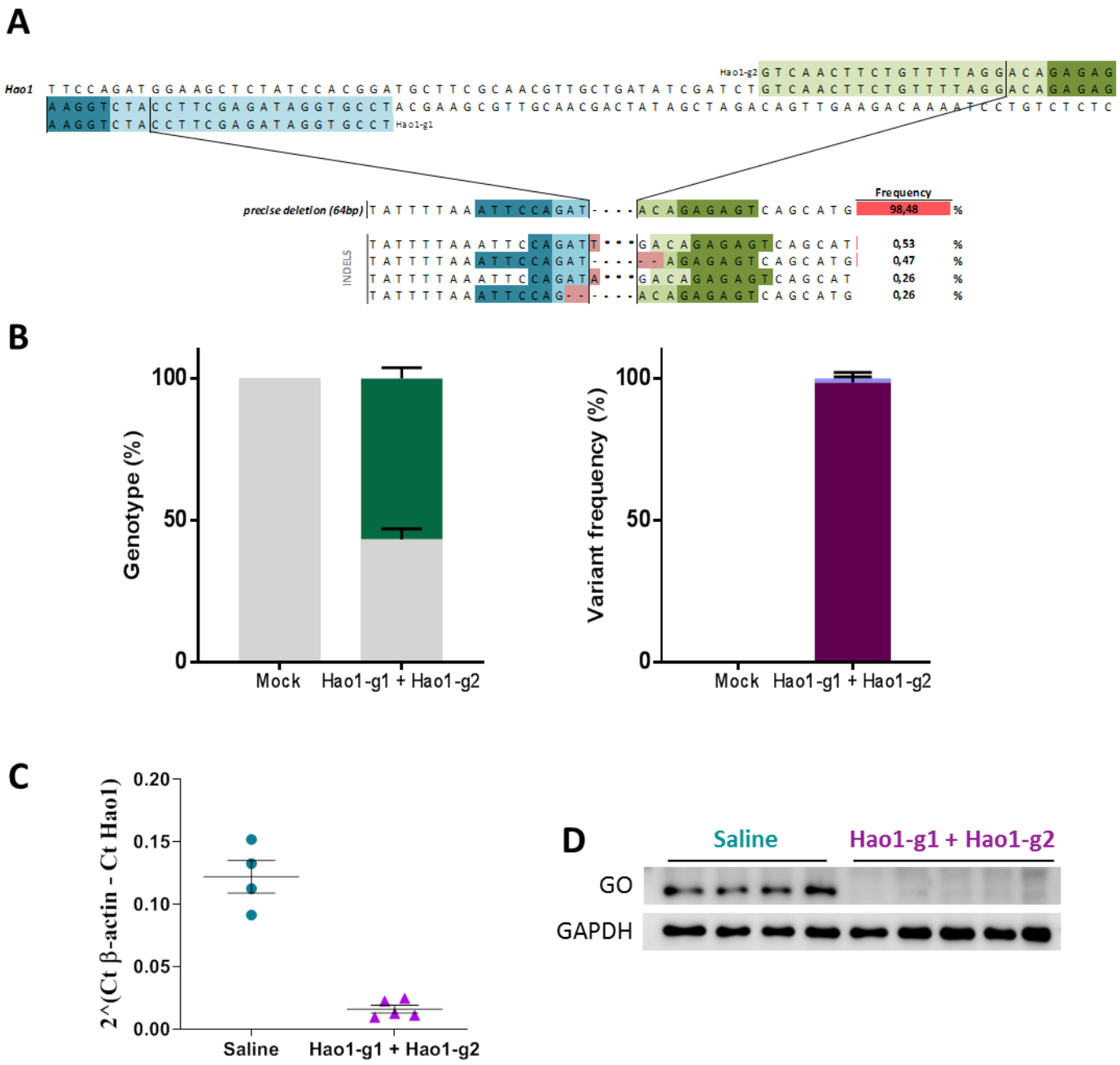


Figure 6. *In vivo* generation of NHEJ-PD generates a Hao1 gene knock-out in mice

(A) Scheme of the two guides tested in Hao1 gene. In light colors: guides (Hao1-g1 and Hao1-g2). In dark colors: PAMs sequences. Within the pair scheme there are three differentiated parts. Up: wt sequence and the guides used. Middle: precise deletion sequence and its frequency. Bottom:

InDels near precise deletion and its frequency. NGS was performed for two different groups of mice, ones treated with a saline solution, and the others treated with the two guides and the SaCas9. Stars represent the wild type sequence and scripts are deletions respective to the wild type sequence (**B)** Next generation sequencing results from variants of the different combinations. Left: genotype of the different groups. In grey: WT mice; in green: edited mice. Right: percentage of the different edition variant; Dark purple: precise deletion; Light purple: variants with modifications around expected deletion. **(C)** Quantification of Hao1 mRNA expression levels by RT-qPCR in animals treated with saline and the combination of Hao1-g1 and Hao1-g2. **(D)** Western blot analysis of GO protein levels in representative animals treated with saline and the combination of Hao1-g1 and Hao1-g2. GAPDH was used as loading control.