



**HAL**  
open science

# Tagging Proteins with Fluorescent Reporters Using the CRISPR/Cas9 System and Double-Stranded DNA Donors

Sylvain Geny, Simon Pichard, Alice Brion, Jean-Baptiste Renaud, Sophie  
Jacquemin, Jean-Paul Concordet, A. Poterszman

## ► To cite this version:

Sylvain Geny, Simon Pichard, Alice Brion, Jean-Baptiste Renaud, Sophie Jacquemin, et al.. Tagging Proteins with Fluorescent Reporters Using the CRISPR/Cas9 System and Double-Stranded DNA Donors. Arnaud Poterszman. Multiprotein Complexes, 2247, Springer; Humana, pp.39-57, 2021, Methods in Molecular Biology, 978-1-0716-1125-8. 10.1007/978-1-0716-1126-5\_3 . hal-03092017

**HAL Id: hal-03092017**

**<https://hal.science/hal-03092017v1>**

Submitted on 19 Dec 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Chapter 3

Tagging proteins with fluorescent reporters using the CRISPR/Cas9 system and double stranded DNA donors–

Sylvain Geny, Simon Pichard, Alice Brion, Jean-Baptiste Renaud, Sophie Jacquemin, Jean-Paul Concordet\* and Arnaud Poterszman \*,

\*Correspondence

## Running head

Macromolecular complexes govern the majority of biological processes and are of great biomedical relevance as factors that perturb interaction networks underlie a number of diseases and inhibition of protein–protein interactions is a common strategy in drug discovery. Genome editing technologies enable precise modifications in protein coding genes in mammalian cells, offering the possibility to introduce affinity tags or fluorescent reporters for proteomic or imaging applications in the bona fide cellular context. Here we describe a streamlined procedure which uses the CRISPR/Cas9 system and a double-stranded donor plasmid for efficient generation of homozygous endogenously GFP-tagged human cell lines. Establishing cellular models that preserve native genomic regulation of the target protein is instrumental to investigate protein localization and dynamics using fluorescence imaging but also to affinity purify associated protein complexes using anti-GFP antibodies or nanobodies.

## Keywords

CRISPR/Cas9, genome editing, double-stranded DNA donors, fluorescent protein

## 1. 1 Introduction

Molecular complexes of interacting proteins govern virtually all biological processes such as metabolism, cell signaling, DNA repair and gene expression. Macromolecular assemblies are also of great biomedical relevance as their **dysfunctions** underlie a number of diseases and deliberate inhibition of protein–protein interactions is an increasingly common strategy in drug discovery [1-3]. To fully understand their biological roles, it is essential to study the structure and function of intact protein assemblies. Although advanced recombinant protein technologies are available to reconstitute multiprotein complexes composed of 10 or more subunits, many protein complexes are difficult to obtain using recombinant methods. An additional hurdle is that the subunit composition of complexes is not always known well enough to proceed to reconstitution of functional entities and therefore to structure determination.

The CRISPR/Cas9 system has revolutionized many fields of life sciences by making it possible to modify the genome sequence with unprecedented efficiency and precision [4, 5]. For example, it is now possible to insert fusion tags at endogenous loci of mammalian cells, providing an efficient way to undertake affinity purification of macromolecular complexes and/or visualize their distribution and dynamics in a cellular context [6, 7]. We have recently detailed a simplified protocol for CRISPR/Cas9-mediated gene tagging in human cell lines using chemically-modified single-stranded oligonucleotides encoding a small affinity tag as donor template and a co-selection strategy targeting the ATP1A1 gene [8] (To be published in MiB, Editor R Owens). Here we describe a procedure based on the use of the CRISPR/Cas9 system and double-stranded DNA donors for tagging the protein of interest with a fluorescent reporter and detail the generation of an U2-OS::XPB-GFP knocked-in cell line expressing a XPB-GFP fusion protein. XPB is a subunit of the TFIIH complex essential in initiation of DNA transcription by RNA polymerase II and DNA repair by Nucleotide Excision Repair [9, 10].

This XPB-GFP tagged U2-OS cell line was instrumental to establish a partnership between TFIID and the histone acetyl transferase GCN5 and the impact of TFIID on GCN5 activity with important consequences on gene expression and chromatin structure [11].

## 1. 2 Materials

Procedures described here need access to standard equipment for molecular biology (PCR amplification, agarose gel analysis, bacteria transformation, access to a DNA sequencing/synthesis service,...), cell culture (cryo-container and liquid nitrogen source, temperature and CO<sub>2</sub>-controlled incubator, laminar flow hood, centrifuge with adaptor for 15 and 50 ml tubes, cell counter), cell microscopy (fluorescence microscope with 63x or higher magnification objectives...) and protein analysis (refrigerated micro-centrifuge, a small scale ultrasonic homogenizer, Nano-UV spectrometer, protein gel electrophoresis and Western Blotting transfer system...). Basic knowledge in these fields is expected.

Experiments detailed below were performed in U2-OS cells (ATCC HTB-96). These adherent cells are grown at 37°C, 5% CO<sub>2</sub> in McCoy's 5a or DMEM-based culture medium (see **Note 1**). For experiments in **Sections 3.2, 3.3, 3.4** and **3.5**, in addition to specific materials, you will need

1. T75 cell culture flasks (75 cm<sup>2</sup>) (Corning cat. no. 431464U)
2. McCoy's 5a or DMEM medium supplemented with 10% fetal bovine serum (FBS), 4mM GlutaMAX™ (Gibco cat. no. 35050061) and 1% PenStrep (Gibco, cat. no. 15140-122).
3. Phosphate buffered saline (PBS)
4. Trypsin-EDTA solution (0.025% w/v Trypsin, 0.05 mM EDTA in 1x PBS) (Gibco cat. no. R001100)

### 2.1 Reagents for the preparations of sgRNA expression plasmids

1. Oligonucleotides for preparing (sgRNA-FW, sgRNA-RV) and sequencing (U6-seq) the single guide RNA expression plasmids (sgRNA) (**Table 1**).
2. Guide RNA expression plasmid: MLM3636 (Addgene plasmid #43860)
3. T4 DNA ligase with buffer for the preparation of the guide RNA expression plasmid (New England BioLabs cat. no. M0202S)
4. BsmBI restriction enzyme (New England BioLabs cat. no. R0580S)
5. LB and LB agar medium, ampicillin, competent DH5 $\alpha$  *E. coli* cells (New England BioLabs cat. no. C2987I)
6. Plasmid DNA Mini and Midiprep purification kits (Macherey-Nagel™ NucleoSpin Plasmid QuickPure™ Kit and Macherey-Nagel™ NucleoBond™ Xtra Midi)

## 2.2 Reagents for nucleofection and assessment of gene editing

### 2.2.1 Nucleofection with AMAXA nucleofector machine

1. Expression plasmid for SpCas9: JDS246 (Addgene plasmid #43861)
2. Expression plasmid for sgRNA (Addgene plasmid #43860) (from **Section 3.1**, step 4)
3. AMAXA nucleofector machine® (Lonza) and consumables including cuvettes (Lonza cat. no. AAB-1001), pipettes and solution V kit (Lonza cat. no. VVCA-1003)
4. Tissue culture plate, 6 wells (Eppendorf, cat. no. EP0030720121)

### 2.2.2 Pool and clone Analysis

1. QuickExtract DNA solution (Viagen cat. no. 302C)
2. Oligonucleotides for PCR amplification and sequencing (see **Table 1**)
3. Phusion High-Fidelity DNA polymerase with buffer and dNTPs (New England BioLabs cat. no. M0530S)
4. PCR isolation kit (Macherey-Nagel™ NucleoSpin™ Gel and PCR Clean-up Kit)

## 2.3 Reagents for gene editing and FACS sorting

### 2.3.1 Nucleofection and antibiotic selection

In addition to the reagents described in **Section 2.2.1**

1. Donor plasmid (from **Section 3.3**, step 2)
2. Puromycin Dihydrochloride (InvivoGen, cat. no. ant-pr-1)

### 2.3.2 FACS sorting

Sorting of GFP-fluorescent cells was carried out on a Melody FACS (BD) and analysis of GFP-positive cells was carried on an Accuri C6 analyzer (BD). You will need:

1. Cytometry Falcon tubes (Becton Dickinson)
2. Tissue culture plate, 96 wells (Eppendorf, cat. no. EP0030730119)

## 2.4 Reagents for validation of selected clones

### 2.4.1 Genotyping of single cell clones

In addition to reagents described in **Section 2.2.2** you will need:

1. Tissue culture plates, 6 and 96 wells (Eppendorf, cat. no. EP0030720121, EP0030730119)
2. Oligonucleotides from the inserted sequence and from the target chromosomal locus.

#### 2.4.2 Western Blot analysis

1. 150mm Petri dishes (Falcon® 150 mm TC-treated dishes, Corning cat. no. 353025)
2. Cell scrapers (Dominique Dutscher cat. no. 353085)
3. PBS containing 30% w/v glycerol (Sigma Aldrich, cat. no. G5516)
4. RIPA buffer: 20 mM Tris-HCl or HEPES pH 7.5 120 mM KCl 1% NP-40 0,1% SDS, 1 mM EDTA, 0.5% Na-Deoxycholate, supplemented with protein inhibitor cocktail (Roche™) and 0.5 mM 1,4-dithreothiol (DTT) (Sigma Aldrich)
5. Lysis buffer: 20 mM Tris or HEPES, 250 mM KCl, NP-40 0,05% supplemented with protein inhibitor cocktail (Roche™) and 0.5 mM DDT
6. Protein assay dye reagent (Bio-Rad cat. no. 5000006)
7. Laemmli buffer 4X: 60mM Tris-HCl pH 6.8, 10% glycerol, 2% SDS, 0,0005% Bromophenol Blue, 355mM  $\beta$ -mercaptoethanol.
8. SDS PAGE gels, PVDF transfer membranes, 3MM Whatman paper
9. TBST: 20 mM Tris/HCl pH, 150mM NaCl, 0,1% Tween-20



10. Non-fat dry milk or BSA powder
11. Western blotting detection reagents (Amersham ECL prime cat no. 3030-931)
12. Primary antibodies against the target protein or against the GFP tag and corresponding secondary antibodies coupled to horseradish peroxidase (HRP). For detection of XPB protein, we used a mouse monoclonal anti-XPB antibody (1B3, [MABE1123](#), [Sigma-Aldrich](#)) and a donkey anti-mouse antibody coupled to HRP.

### 2.4.3 Fluorescence microscopy

1. Glass-bottom dishes (Glass Bottom Dish 35 mm, Clinisciences/Ibidi cat. no. 81218-200)
2. Epifluorescence inverted microscope with filters compatible with DAPI/Hoechst and the fluorescent reporter used (here GFP)
3. [Paraformaldehyde](#) (PFA) 16% ultrapure methanol free (Euromedex cat. no. 15710)
4. Hoechst (Sigma Aldrich cat. no. 32670)
5. Antibodies to confirm expression of the target protein or for co-localization experiments

## 3 Methods

Targeted gene knock-in is achieved through the following steps. First, the Cas9 endonuclease is directed by a guide RNA to a specific target site in the genome to generate a double-strand break (DSB). The DSB is then repaired with a donor template through a homology-directed repair (HDR) pathway [12]. Below, we first describe the design of the single-guide RNAs and donor plasmid template to edit the genome (**Section 3.1-3.3**) and the clonal isolation of CRISPR/Cas9-modified cells (**Section 3.4**). Selected clones are finally characterized at the genomic and protein levels (**Section 3.5**) (**Fig. 1**).

To facilitate the generation of homozygous knock-in cells, we favored a tagging strategy based on the use of a donor plasmid containing 300-1000 bp homology arms flanking the eGFP-2A-PURO coding sequence [13] (**Fig. 2**). This strategy allows to combine transient antibiotic selection with fluorescence-assisted cell sorting (FACS) to isolate CRISPR/Cas9-modified cells [14].

### 3.1 Design of the sgRNAs

The design of the sgRNA and HDR donor depends on the nature and position of the tag (see **Note 2**). As the efficiency of HDR insertion varies with the distance from the DSB, the DSB (generated 3 bp upstream of the PAM sequence corresponding to the sgRNA) should be as close as possible to the insertion site of tag. For example, to tag a protein at **its** C-terminal end, the DSB should be as close as possible to the stop codon of the gene of interest.

1. Retrieve the nucleotide sequence of the gene of interest (GOI) and focus on the site of tag insertion. The UCSC genome browser is a convenient source for this step ([genome.ucsc.edu](http://genome.ucsc.edu)).

2. Screen the genomic regions of interest using an online CRISPR design tool such as CRISPOR (<http://crispor.tefor.net/>) to identify and rank the guide sequences (**Table 1**). We recommend testing several sgRNAs with cut sites less than 20 bp away from the insertion site to find an optimal sgRNA. Guide RNAs for the human genome can also be visualized directly in the UCSC genome browser with pre-calculated data from CRISPOR (select “Full” display mode of the “CRISPR targets” in the “Genes and Gene Predictions” list of features).
3. Order the sense and antisense oligonucleotides for cloning the sgRNA sequences into a sgRNA expression plasmid such as MLM3636. Note that the oligonucleotides contain overhangs for ligation into the pair of BsmBI sites with the sense and antisense sequences matching the genomic target. The oligonucleotide sequences can be accessed by clicking on the “PCR/cloning primers” button in CRISPOR. If you have used the UCSC genome browser for sgRNA selection, you can directly transfer the sgRNA sequence to CRISPOR by a hyperlink feature or simply by copy-pasting to the CRISPOR input page.
4. Generate the oligo duplex, clone into the BsmBI pre-digested plasmid MLM3636, sequence the resulting plasmid. Transform and amplify the guide expression plasmid in DH5 competent cells using a Midi or Maxi Prep kit depending on the amount of plasmid required.

### 3.2 Delivery of CRISPR reagents and assessment of gene editing

As knock-in using double-stranded donor template has relatively low efficiency, successful HDR-mediated experiments require experimental optimization. In particular, it is recommended to control and optimize delivery of CRISPR reagents by evaluating levels of gene editing in absence of donor template. In some cases, GFP fluorescence can be used to directly characterize and optimize HDR events, but low to very low levels of fluorescent tagged proteins driven by

endogenous promoters can be challenging to detect. An attractive alternative consists in delivering the nuclease with a single strand donor designed to fuse a HiBit peptide to the target protein. The abundance of HiBiT-tagged proteins in transfected cells is proportional to the proportion of edited gene copies and is readily detected using the Nano-Glo HiBit luciferase based detection system (see [8]).

In this section we detail the delivery of CRISPR/Cas9 reagents by nucleofection (see **Note 3**) and PCR-based analysis of the cleavage efficiency of guide RNAs.

### 3.2.1 Nucleofection.

1. Thaw and maintain U2-OS cells using your favorite protocol (see **Note 1**). We usually split cells 1:5 every 3-4 days, verify that the viability is always higher than 95% and limit the number of passages to 20.
2. Two or three days before nucleofection, seed two 75 cm<sup>2</sup> flasks and incubate at 37°C, 5% CO<sub>2</sub>.
3. When 50-70% confluence is reached, detach the cells with trypsin, spin-down in a 15 ml tube at 90 g at room temperature for 10 minutes, remove the supernatant and re-suspend the cells in PBS to a concentration of 1-3 x10<sup>6</sup> cells/ml.
4. Aliquot 10<sup>6</sup> cells into 1.5 ml micro-centrifuge tubes (one tube per guide RNA to be tested plus an additional tube for a negative control), pellet the cells, discard supernatant completely and re-suspend cells in Nucleofector® solution (100 µl per sample).

5. Prepare the DNA mixes containing 2  $\mu\text{g}$  of guide sgRNA plasmid (sgXPB) and 2  $\mu\text{g}$  of plasmid expressing the SpCas9 protein (JDS246) in a total volume less than 10  $\mu\text{L}$ .
6. Mix the DNA solution with 100  $\mu\text{L}$  of cell suspension and transfer to a cuvette. Process the samples quickly to avoid storing the cells longer than 10 min in Nucleofector® Solution V.
7. Insert the cuvette into the Nucleofector®, select the cell-type specific program X-001 and press the start button. Using the provided pipette, immediately remove the sample from the cuvette and transfer into the 6-well plate containing 2 mL of culture medium.
8. Incubate the 6-well plates at 37°C, 5% CO<sub>2</sub> and proceed to analysis or collect the cells and store the cell pellet at -20 °C.

### 3.2.2 PCR based analysis of the cleavage efficiency for the sgRNA

CRISPR/Cas9 cleavage is usually assessed by amplifying the target region by PCR and analyzing the mutation rate in the resulting product. This can be performed using various assays such as the polymerase chain reaction (PCR)/restriction enzyme (RE) assay, T7 endonuclease I (T7EI) assay, Surveyor nuclease assay or with TIDE [15-18].

1. Three days after transfection, wash cells with PBS, extract genomic DNA (typically from 1. 10<sup>5</sup> cells) using QuickExtract DNA extraction kit or similar according to the manufacturer's recommendations. Dilute genomic DNA at a final concentration around 50 ng/ $\mu\text{L}$ .

2. Design a pair of PCR primers (XPB-FW and XPB-RV) which hybridize on either side of the target site to amplify a ~500 bp stretch of DNA. The sgRNA cut site should ideally be located ~200 bp downstream from one the primers for convenient analysis by DNA sequencing.
3. Amplify the locus of interest by PCR (see **Table 2**), run 2  $\mu$ L of PCR products on 1% agarose gel TBE 0.5x at 100 Volts for 90 minutes and visualize DNA using a UV trans-illuminator. A sharp single band of the expected size should be visible in control samples. Additional bands may be detected in test samples corresponding to insertion and/or deletion.
4. Purify the PCR products using an appropriate PCR purification kit and measure its concentration.

A wide panel of assays can be used to analyze PCR products and verify the efficacy of programmable nucleases. The enzymatic Surveyor and T7 endonuclease I cleavage assays [19, 20] are commonly used but these approaches are semi-quantitative and their sensitivity is limited [15]. We favor the TIDE assay that accurately quantifies editing efficacy and simultaneously identifies the predominant types of insertions and deletions (indels) in the pool of treated cells [21].

5. Send the PCR products (control and experimental samples) for Sanger sequencing and retrieve the sequence trace files in .ab1 or .scf format.
6. Upload the trace files and the guide RNA sequence (20 nt) into the TIDE web tool (available at <http://tide.nki.nl> and <https://deskgen.com>) and perform the analysis with default parameters.

Use the TIDE assay to compare the efficiency of the different guides and select the best experimental conditions for further experiments. An example of TIDE analysis is provided on

the TIDE website.

### 3.3 Design of plasmid donor

Several types of DNA repair template can be utilized in HDR-mediated CRISPR/Cas9 gene editing. Donor plasmids, with homology arms of a 300-1000 bps allow for large insertions in the 1-2 kbp range. Other types of DNA repair templates such as linear DNAs with short homology arms obtained by PCR or gene synthesis can also be used for inserting fluorescent labels or other protein tags such as the Halo or Snap tags [22, 23]. Linear DNA, however, may concatemerize before insertion [24], which may perturb gene expression or make the modification unstable during cell replication. Different types of donor DNA will give similar or different KI efficiencies depending on the gene targeted and cell line used and may be tested if needed.

This section details the design and preparation of the donor plasmid which requires the introduction of the coding sequence for the insertion flanked by two homology arms of a few hundred bases (typically between 0.3 and 1.0 kbp).

1. Define the DNA fragment that you wish to introduce, which in this example case consists in the eGFP-P2A-PURO coding sequence where eGFP, the 2A self-cleaving peptide (P2A) and the puromycin N-acetyl transferase coding sequence are separated by flexible glycine/serine-rich linkers (Fig. 4). References [25] and [26] provide useful information on the use of 2A peptides in a polycistronic vector and the design of protein linkers, respectively. Note that when using 2A self-cleaving peptide, antibiotic resistance is expressed from the target gene, which minimizes selection of clones with non-targeted integration. However, in some cases, expression levels of the target gene may be insufficient to provide antibiotic resistance and it may be necessary

to provide antibiotic resistance from a constitutive promoter; higher numbers of clones with random integration of the donor plasmid into the cell genome may be isolated [27].

2. Retrieve the sequence of a fragment encompassing upstream and downstream homology arms of approximately 700 bp and replace the STOP codon of the GOI by the eGFP-P2A-PURO coding sequence. Carefully verify that the eGFP-P2A-PURO coding sequence is in frame with the last exon of the target protein and that the guide RNA or PAM sequence have been modified to prevent cleavage by Cas9. If that is not the case, introduce point mutations to avoid unwanted mutations being introduced after integration.

Having precisely defined the donor sequence, different options can be considered to prepare the donor plasmid.

3. Purchasing the designed sequence from your preferred supplier is the easiest approach. Alternatively, you can amplify the eGFP-P2A-PURO coding sequence from an existing plasmid, the homology arms from genomic DNA and assemble the three fragments into your preferred cloning vector using restriction enzyme-free technologies (see **Note 4**).

### 3.4 Gene targeting and Isolation of gene edited cells by FACS sorting

When the CRISPR/Cas9 reagents are ready, you can proceed to genome editing with CRISPR/Cas9 to generate cell lines that express an endogenously tagged protein. The strategy detailed below relies on antibiotic selection combined with fluorescence-activated cell sorting



(FACS) for enrichment in HDR during genome editing. FACS is an efficient technology which allows to simultaneously measure the knock-in efficiency by counting fluorescently positive cells and sort isolated cells into 96-well plates. As cell sorting relies on the expression levels of the fluorescent tagged proteins driven by endogenous promoters, isolation of positive clones can be challenging for proteins expressed at low levels.

#### 3.4.1 Nucleofection and antibiotic selection

1. Proceed as detailed in **Section 3.2.1** with a master mix containing 2  $\mu\text{g}$  of the selected guide sgRNA plasmid (sgXPB), 2  $\mu\text{g}$  of plasmid expressing the SpCas9 protein (JDS246) and 6  $\mu\text{g}$  of donor plasmid in a total volume lower than 10  $\mu\text{L}$ .
2. Three days after transfection, replace the culture medium and add puromycin to a concentration of 5  $\mu\text{g}/\text{ml}$  and further incubate for an additional week. The use of puromycin selection can be avoided if the signal of the fluorescently tagged protein is sufficiently strong and the proportion of fluorescent cells is sufficient for FACS sorting.

#### 3.4.2 Isolation of gene edited cells by FACS sorting

Proceed to FACS sorting seven days after the beginning of the puromycin treatment (or three days after nucleofection in absence of antibiotic selection). Use non-transfected cells as negative control for gating.

1. Prepare 96-well plates containing 100  $\mu\text{l}$  of growth medium in each well.
2. Wash cells with PBS, add 1 ml of trypsin and transfer the re-suspended cells in a 1.5 ml micro-centrifuge tube containing 0.5 ml of medium.
3. Centrifuge the cell suspension for 5 minutes at 90g at room temperature,

remove the supernatant and re-suspend the cells in 1 ml PBS containing 2% FBS and filter the cells through a 50- $\mu$ m cell strainer into a sterile flow cytometry tube.

4. Turn on the cytometer, run the cleaning/calibration procedures and load settings to detect GFP. Refer to the user manual of your instrument for detailed procedures.
5. Run the non-transfected cells to check the cell sorter parameters and establish gates to select single live cells (**Fig. 5**). For single cells, plot forward scatter linear (FCS) on the y axis and the fluorescent channel (GFP-FL) on the x axis. Ideally, set a background threshold where less than 0.1% of the non-transfected control cells are counted as GFP positive.
6. Repeat with the cell population to be sorted and verify that the CRISPR/Cas9-modified cells can be differentiated from parental cells. Run the transfected cells to quantify the amount of fluorescently labeled cells and define the collection gate. The fluorescent signal depends on expression levels of the target gene and can be relatively dim (**Fig. 6B**).
7. Sort the positive cells into the 96 wells plates and incubate at 37°C, 5% CO<sub>2</sub>.
8. After ~5 days, identify 24 single cell clones for validation and keep them growing. Maintain the cells in 96-well plates until you can proceed to validation. Split cells by trypsinization when required.

### 3.5 Validation of selected clones

This section details: (i) genotyping of the knock-in cell clones by PCR to verify that the fluorescent marker was correctly inserted at the target locus and to test for homozygosity, (ii)

Western Blot analysis to verify that the GFP fusion protein is effectively expressed and (iii) fluorescence microscopy to validate expression and localization of the tagged protein.

### 3.5.1 Validation at the genomic level

1. Split cells from each positive well into two 96-wells plates. The day after, spin down and wash with PBS the cells in one of the plates.
2. Extract genomic DNA directly in the wells using 25  $\mu$ l of commercially available DNA lysis buffer.
3. Proceed to PCR amplification from genomic DNA of each clone with the primer pairs XPB-FW/XPB-RV or XPB-FW/GFP-RV following the conditions used for PCR analysis of the pool of transfected cells. Run 2  $\mu$ L of the PCR on 1% agarose TBE 1x gel at 100 volts for 1 hour. Visualize the PCR products using a UV transilluminator (see **Note 5**).
4. Sequence the PCR fragments of clones to ensure correct tag sequence insertion.

### 3.5.2 Western blot analysis

1. Seed 5 150 mm Petri dishes containing 25 ml of growth medium and incubate at 37°C, 5% CO<sub>2</sub> until a confluency of 80% is reached.
2. Wash the plate with PBS containing 30% w/v glycerol, detach the cells with a scraper and after centrifugation at 1000 g for 10 minutes at 4°C, snap freeze the pellet in liquid nitrogen and store at -80°C. We usually prepare batches of 15  $10^6$  and 25  $10^6$  cells.

3. Resuspend  $15 \times 10^6$  cells in 150  $\mu$ l of RIPA buffer and incubate for 10 minutes by pipetting up and down.
4. Sonicate two times 30s on ice. We use a Bioruptor<sup>TM</sup> sonication system (amplitude 30 and 0.5 seconds pulse on ice). (Optional)
5. Centrifuge 15 minutes at 14 000 g at 4°C, collect the supernatant and estimate the total protein concentration using a Bradford assay. A concentration of 3-4 mg/mL is expected.
6. Heat 20  $\mu$ g of total proteins from the soluble extract mixed with 2xLaemmli buffer at 95°C for 5 minutes and centrifuge samples at 10 000 g for 30 seconds to bring down the condensate and remove insoluble debris.
7. Load the centrifuged sample on a SDS polyacrylamide gel, electrophorese and transfer proteins from the gel matrix to a nitrocellulose or PVDF membrane using your favorite device.
8. Block the membrane for 1 hour at room temperature or overnight at 4°C using 3 % w/v dry skimmed milk or BSA solution in PBS, incubate the membrane with an appropriate dilution of a primary antibody directed against the affinity-tag or the tagged subunit in the same buffer for 1h at room temperature or overnight at 4°C.
9. Wash the membrane three times in TBST 5 minutes each, incubate with the recommended dilution of conjugated secondary antibody in TBST at room temperature for 1 hour and develop the western blot.

The western blot analysis of protein lysates from the parental and from a modified cell line where XPB and XPB-GFP are detected with an anti-XPB antibody is shown in **Fig. 6A**. As expected in the case of a homozygous modification, the 90 kDa wild-type XPB protein (lane 1)

is replaced by the 130 kDa XPB-GFP fusion (lane 2). A heterozygous modification would result in detection of both a 90 kDa and a 130 kDa protein.

### 3.5.3 Fluorescence microscopy

We typically select 4 positive clones and include the non-modified cell line as control. The target protein being fused to GFP, expression of a fluorescent fusion protein can be checked using an epifluorescence inverted microscope. Live or fixed cell imaging can be performed.

1. Seed glass-bottom dishes/plates with clones that have been tested positive until 70-80% confluency is reached.
2. For live cell imaging, change the medium to a phenol red-free medium and observe cells using an inverted epifluorescence microscope with adapted filters (i.e. an FITC filter for GFP detection). Maintain cells at 37°C in the incubation chamber of the microscope. Record images with a 63x/1.4 NA immersion objective. Cells can be counter-stained with Hoechst (0.1 µg/ml) to observe subcellular localization of the fusion protein.
3. For imaging of fixed cells, wash the cells with PBS, fix them with 4% PFA in PBS and again twice with PBS. Record images with a 63x/1.4 NA immersion objective. Although fixed cells can be stored for a few weeks at 4°C, try to observe them 1-2 days after fixation. Cells can be counter-stained with antibodies to confirm expression of the target protein or for co-localization experiments.

A live cell image of U2-OS <sup>::XPB-GFP</sup> cells in which the XPB gene is fused to eGFP is shown in **Fig. 6B**.

## 4 Notes

1. Detailed information is available on the ATCC web site (<https://www.lgcstandards-atcc.org/>). A guide containing general technical information for working with animal cells in culture, including media, sub-culturing, cryopreservation and thawing is also accessible.
2. Affinity tags can be introduced at the N- or C- end of the target protein or inserted into internal loops. It is wise to perform extensive literature search for functionality studies and expression data in human cells as well as in other species such as mouse, *C. elegans* or yeast to avoid trying to tag an extremity that would compromise protein function. In absence of information on the accessibility of the extremities of the target protein, a C-terminal tag is generally preferred [28].
3. Nucleofection relies on permeabilization of the cellular membranes by electroporation. Although, Amaxa nucleofection is often superior to other widely used transfection approaches, lipid-based transfection with reagents such as Lipofectamine 3000™ or jet-CRISPR™ can also be used.
4. Homology-based cloning techniques including Sequence and Ligation Independent cloning (SLIC) and commercially available technologies such as In-Fusion, Gibson assembly are perfectly adapted to assemble 3-4 DNA fragments in a single step reaction. These technologies are now routinely used in many laboratories (see [29] in this issue) and can be used to rapidly generate the donor plasmid. The homology arms can be amplified from genomic DNA and the insert from plasmids which can be obtained from non-profit organizations (for example, see Addgene plasmids #112848, #52379). Note that a growing number of companies also offer donor

plasmid design and preparation as a service.

5. With the XPB-FW2/XPB-RV2 primers which hybridize on either side of the insertion site, both knock-in and non-modified alleles should be amplified: A 2800 bp PCR product is expected if the eGFP-P2A-PURO was successfully inserted while a 1300 bp fragment should be observed for a non-modified allele, allowing to discriminate between homozygous, heterozygous and non-modified clones. With the XPB-FW/GFP-RV and Puro-FW/XPB-RV2 pairs which amplify 540 bp and 1000 bp fragments only modified alleles will be detected as the GFP-RV and Puro-RV2 oligonucleotides are complementary to the eGFP-P2A-PURO cDNA from the inserted sequence.

## Acknowledgements

We thank IGBMC cell culture and imaging facilities for assistance and fruitful discussions. This work was supported by the Centre National pour la Recherche Scientifique (CNRS), the Institut National de la Santé et de la Recherche Médicale (INSERM), Université de Strasbourg (UdS), Association pour la Recherche sur le Cancer (ARC), the Ligue nationale contre le cancer, Institut National du Cancer (INCa; INCA 9378), Agence National pour la Recherche (ANR-12-BSV8-0015-01 and ANR-10-LABX-0030-INRT under the program Investissements d'Avenir ANR-10-IDEX-0002-02), Instruct (R&D Project Funding) and by Instruct-ULTRA as part of the European Union's Horizon 2020 (Grant ID 731005) by Instruct-ERIC.

## Bibliography

1. Doudna, J.A. and E. Charpentier, *Genome editing. The new frontier of genome engineering with CRISPR-Cas9*. Science, 2014. **346**(6213): p. 1258096.
2. Jinek, M., K. Chylinski, I. Fonfara, M. Hauer, J.A. Doudna, and E. Charpentier, A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science, 2012. **337**(6096): p. 816-21.
3. Dalvai, M., J. Loehr, K. Jacquet, C.C. Huard, C. Roques, P. Herst, J. Cote, and Y. Doyon, A Scalable Genome-Editing-Based Approach for Mapping Multiprotein Complexes in Human Cells. Cell Rep, 2015. **13**(3): p. 621-633.
4. Dewari, P.S., B. Southgate, K. McCarten, G. Monogarov, E. O'Duibhir, N. Quinn, A. Tyrer, M.C. Leitner, C. Plumb, M. Kalantzaki, C. Blin, R. Finch, R.B. Bressan, G. Morrison, A.M. Jacobi, M.A. Behlke, A. von Kriegsheim, S. Tomlinson, J. Krijgsveld, and S.M. Pollard, An efficient and scalable pipeline for epitope tagging in mammalian stem cells using Cas9 ribonucleoprotein. Elife, 2018. **7**.
5. Geny, S., S. Pichard, A. Poterszman, and J. Concordet, *Gene tagging with the CRISPR-Cas9 system to facilitate macromolecular complex purification*. Methods Mol Biol, In press. **XXXX**: p. XXXX.
6. Compe, E. and J.M. Egly, *Nucleotide Excision Repair and Transcriptional Regulation: TFIIH and Beyond*. Annu Rev Biochem, 2016. **85**: p. 265-90.
7. Kolesnikova, O., L. Radu, and A. Poterszman, *TFIIH: A multi-subunit complex at the cross-roads of transcription and DNA repair*. Adv Protein Chem Struct Biol, 2019. **115**: p. 21-67.
8. Sandoz, J., Z. Nagy, P. Catez, G. Caliskan, S. Geny, J.B. Renaud, J.P. Concordet, A. Poterszman, L. Tora, J.M. Egly, N. Le May, and F. Coin, *Functional interplay between TFIIH and KAT2A regulates higher-order chromatin structure and class II gene expression*. Nat Commun, 2019. **10**(1): p. 1288.
9. Yeh, C.D., C.D. Richardson, and J.E. Corn, *Advances in genome editing through control of DNA repair pathways*. Nat Cell Biol, 2019. **21**(12): p. 1468-1478.
10. Chu, V.T., T. Weber, B. Wefers, W. Wurst, S. Sander, K. Rajewsky, and R. Kuhn, *Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells*. Nat Biotechnol, 2015. **33**(5): p. 543-8.
11. Sluch, V.M., X. Chamling, C. Wenger, Y. Duan, D.S. Rice, and D.J. Zack, *Highly efficient scarless knock-in of reporter genes into human and mouse pluripotent stem cells via transient antibiotic selection*. PLoS One, 2018. **13**(11): p. e0201683.
12. Sentmanat, M.F., S.T. Peters, C.P. Florian, J.P. Connelly, and S.M. Pruetz-Miller, *A Survey of Validation Strategies for CRISPR-Cas9 Editing*. Sci Rep, 2018. **8**(1): p. 888.
13. Vouillot, L., A. Thelie, and N. Pollet, *Comparison of T7E1 and surveyor mismatch cleavage assays to detect mutations triggered by engineered nucleases*. G3 (Bethesda), 2015. **5**(3): p. 407-15.
14. Renaud, J.B., C. Boix, M. Charpentier, A. De Cian, J. Cochenne, E. Duvernois-Berthet, L. Perrouault, L. Tesson, J. Edouard, R. Thinard, Y. Cherifi, S. Menoret, S. Fontaniere, N. de Croze, A. Fraichard, F. Sohm, I. Anegon, J.P. Concordet, and C. Giovannangeli, *Improved Genome Editing Efficiency and Flexibility Using Modified Oligonucleotides with TALEN and CRISPR-Cas9 Nucleases*. Cell Rep, 2016. **14**(9): p. 2263-2272.

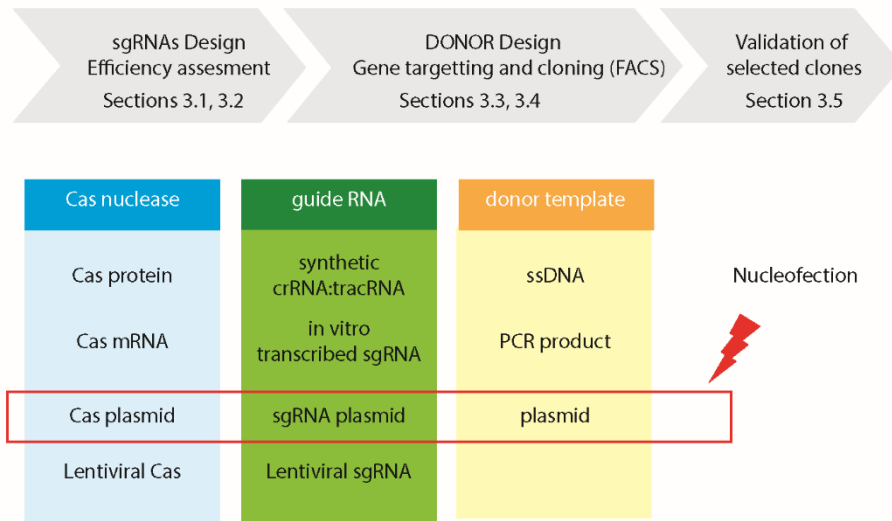


15. Shan, Q., Y. Wang, J. Li, and C. Gao, *Genome editing in rice and wheat using the CRISPR/Cas system*. Nat Protoc, 2014. **9**(10): p. 2395-410.
16. Kim, H.J., H.J. Lee, H. Kim, S.W. Cho, and J.S. Kim, *Targeted genome editing in human cells with zinc finger nucleases constructed via modular assembly*. Genome Res, 2009. **19**(7): p. 1279-88.
17. Guschin, D.Y., A.J. Waite, G.E. Katibah, J.C. Miller, M.C. Holmes, and E.J. Rebar, *A rapid and general assay for monitoring endogenous gene modification*. Methods Mol Biol, 2010. **649**: p. 247-56.
18. Brinkman, E.K., T. Chen, M. Amendola, and B. van Steensel, *Easy quantitative assessment of genome editing by sequence trace decomposition*. Nucleic Acids Res, 2014. **42**(22): p. e168.
19. Paix, A., D. Rasoloson, A. Folkmann, and G. Seydoux, *Rapid Tagging of Human Proteins with Fluorescent Reporters by Genome Engineering using Double-Stranded DNA Donors*. Curr Protoc Mol Biol, 2019. **129**(1): p. e102.
20. Paix, A., H. Schmidt, and G. Seydoux, *Cas9-assisted recombineering in C. elegans: genome editing using in vivo assembly of linear DNAs*. Nucleic Acids Res, 2016. **44**(15): p. e128.
21. Auer, T.O. and F. Del Bene, *CRISPR/Cas9 and TALEN-mediated knock-in approaches in zebrafish*. Methods, 2014. **69**(2): p. 142-50.
22. Liu, Z., O. Chen, J.B.J. Wall, M. Zheng, Y. Zhou, L. Wang, H. Ruth Vaseghi, L. Qian, and J. Liu, *Systematic comparison of 2A peptides for cloning multi-genes in a polycistronic vector*. Sci Rep, 2017. **7**(1): p. 2193.
23. Chen, X., J.L. Zaro, and W.C. Shen, *Fusion protein linkers: property, design and functionality*. Adv Drug Deliv Rev, 2013. **65**(10): p. 1357-69.
24. Koch, B., B. Nijmeijer, M. Kueblbeck, Y. Cai, N. Walther, and J. Ellenberg, *Generation and validation of homozygous fluorescent knock-in cells using CRISPR-Cas9 genome editing*. Nat Protoc, 2018. **13**(6): p. 1465-1487.
25. Kimple, M.E., A.L. Brill, and R.L. Pasker, *Overview of affinity tags for protein purification*. Curr Protoc Protein Sci, 2013. **73**: p. 9 9 1-9 9 23.
26. Rossolillo, P., O. Kolesnikova, K. Essabri, G. Ramon Zamorano, and A. Poterszman, *Production of multiprotein complexes using the Baculovirus expression system: Homology based and Restriction Free cloning strategies for construct design*. Methods Mol Biol, In press. **XXXX**: p. XXXX.
1. Villicana, C., G. Cruz, and M. Zurita, *The basal transcription machinery as a target for cancer therapy*. Cancer Cell Int, 2014. **14**(1): p. 18.
2. Scott, D.E., A.R. Bayly, C. Abell, and J. Skidmore, *Small molecules, big targets: drug discovery faces the protein-protein interaction challenge*. Nat Rev Drug Discov, 2016. **15**(8): p. 533-50.
3. Bushweller, J.H., *Targeting transcription factors in cancer - from undruggable to reality*. Nat Rev Cancer, 2019. **19**(11): p. 611-624.
4. Doudna, J.A. and E. Charpentier, *Genome editing. The new frontier of genome engineering with CRISPR-Cas9*. Science, 2014. **346**(6213): p. 1258096.
5. Jinek, M., K. Chylinski, I. Fonfara, M. Hauer, J.A. Doudna, and E. Charpentier, *A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity*. Science, 2012. **337**(6096): p. 816-21.
6. Dalvai, M., J. Loehr, K. Jacquet, C.C. Huard, C. Roques, P. Herst, J. Cote, and Y. Doyon, *A Scalable Genome-Editing-Based Approach for Mapping Multiprotein Complexes in Human Cells*. Cell Rep, 2015. **13**(3): p. 621-633.
7. Dewari, P.S., B. Southgate, K. McCarten, G. Monogarov, E. O'Duibhir, N. Quinn, A. Tyrer, M.C. Leitner, C. Plumb, M. Kalantzaki, C. Blin, R. Finch, R.B. Bressan, G. Morrison, A.M. Jacobi, M.A. Behlke, A. von Kriegsheim, S. Tomlinson, J. Krijgsveld, and S.M. Pollard, *An efficient and scalable pipeline for epitope tagging in mammalian stem cells using Cas9 ribonucleoprotein*. Elife, 2018. **7**.
8. Geny, S., S. Pichard, A. Poterszman, and J. Concordet, *Gene tagging with the CRISPR-Cas9 system to facilitate macromolecular complex purification*. Methods Mol Biol, In press. **XXXX**: p. XXXX.

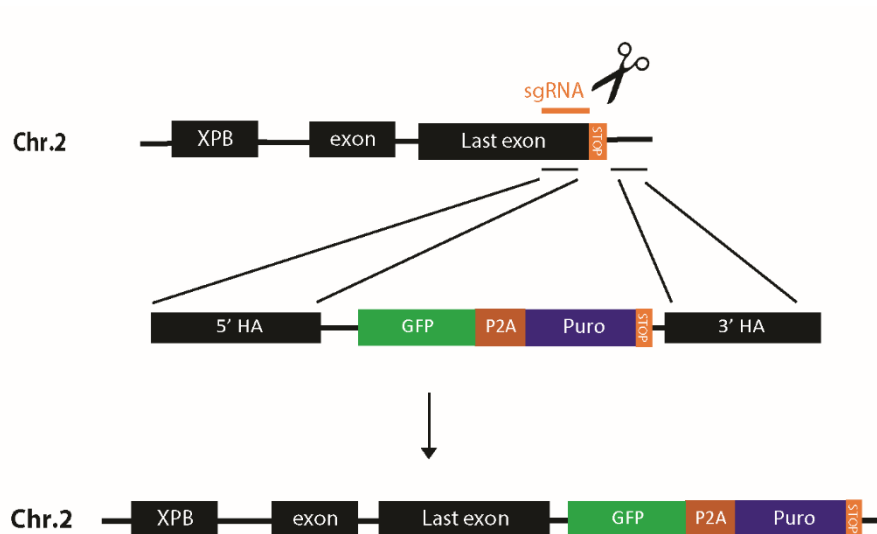
9. Compe, E. and J.M. Egly, *Nucleotide Excision Repair and Transcriptional Regulation: TFIIH and Beyond*. *Annu Rev Biochem*, 2016. **85**: p. 265-90.
10. Kolesnikova, O., L. Radu, and A. Poterszman, *TFIIH: A multi-subunit complex at the cross-roads of transcription and DNA repair*. *Adv Protein Chem Struct Biol*, 2019. **115**: p. 21-67.
11. Sandoz, J., Z. Nagy, P. Catez, G. Caliskan, S. Geny, J.B. Renaud, J.P. Concordet, A. Poterszman, L. Tora, J.M. Egly, N. Le May, and F. Coin, *Functional interplay between TFIIH and KAT2A regulates higher-order chromatin structure and class II gene expression*. *Nat Commun*, 2019. **10**(1): p. 1288.
12. Yeh, C.D., C.D. Richardson, and J.E. Corn, *Advances in genome editing through control of DNA repair pathways*. *Nat Cell Biol*, 2019. **21**(12): p. 1468-1478.
13. Chu, V.T., T. Weber, B. Wefers, W. Wurst, S. Sander, K. Rajewsky, and R. Kuhn, *Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells*. *Nat Biotechnol*, 2015. **33**(5): p. 543-8.
14. Sluch, V.M., X. Chamling, C. Wenger, Y. Duan, D.S. Rice, and D.J. Zack, *Highly efficient scarless knock-in of reporter genes into human and mouse pluripotent stem cells via transient antibiotic selection*. *PLoS One*, 2018. **13**(11): p. e0201683.
15. Sentmanat, M.F., S.T. Peters, C.P. Florian, J.P. Connelly, and S.M. Pruett-Miller, *A Survey of Validation Strategies for CRISPR-Cas9 Editing*. *Sci Rep*, 2018. **8**(1): p. 888.
16. Vouillot, L., A. Thelie, and N. Pollet, *Comparison of T7E1 and surveyor mismatch cleavage assays to detect mutations triggered by engineered nucleases*. *G3 (Bethesda)*, 2015. **5**(3): p. 407-15.
17. Renaud, J.B., C. Boix, M. Charpentier, A. De Cian, J. Cochenne, E. Duvernois-Berthet, L. Perrouault, L. Tesson, J. Edouard, R. Thinard, Y. Cherifi, S. Menoret, S. Fontaniere, N. de Croze, A. Fraichard, F. Sohm, I. Anegon, J.P. Concordet, and C. Giovannangeli, *Improved Genome Editing Efficiency and Flexibility Using Modified Oligonucleotides with TALEN and CRISPR-Cas9 Nucleases*. *Cell Rep*, 2016. **14**(9): p. 2263-2272.
18. Shan, Q., Y. Wang, J. Li, and C. Gao, *Genome editing in rice and wheat using the CRISPR/Cas system*. *Nat Protoc*, 2014. **9**(10): p. 2395-410.
19. Kim, H.J., H.J. Lee, H. Kim, S.W. Cho, and J.S. Kim, *Targeted genome editing in human cells with zinc finger nucleases constructed via modular assembly*. *Genome Res*, 2009. **19**(7): p. 1279-88.
20. Guschin, D.Y., A.J. Waite, G.E. Katibah, J.C. Miller, M.C. Holmes, and E.J. Rebar, *A rapid and general assay for monitoring endogenous gene modification*. *Methods Mol Biol*, 2010. **649**: p. 247-56.
21. Brinkman, E.K., T. Chen, M. Amendola, and B. van Steensel, *Easy quantitative assessment of genome editing by sequence trace decomposition*. *Nucleic Acids Res*, 2014. **42**(22): p. e168.
22. Paix, A., D. Rasoloson, A. Folkmann, and G. Seydoux, *Rapid Tagging of Human Proteins with Fluorescent Reporters by Genome Engineering using Double-Stranded DNA Donors*. *Curr Protoc Mol Biol*, 2019. **129**(1): p. e102.
23. Paix, A., H. Schmidt, and G. Seydoux, *Cas9-assisted recombineering in C. elegans: genome editing using in vivo assembly of linear DNAs*. *Nucleic Acids Res*, 2016. **44**(15): p. e128.
24. Remy, S., L. Tesson, S. Menoret, C. Usal, A. De Cian, V. Thepenier, R. Thinard, D. Baron, M. Charpentier, J.B. Renaud, R. Buelow, G.J. Cost, C. Giovannangeli, A. Fraichard, J.P. Concordet, and I. Anegon, *Efficient gene targeting by homology-directed repair in rat zygotes using TALE nucleases*. *Genome Res*, 2014. **24**(8): p. 1371-83.
25. Liu, Z., O. Chen, J.B.J. Wall, M. Zheng, Y. Zhou, L. Wang, H. Ruth Vaseghi, L. Qian, and J. Liu, *Systematic comparison of 2A peptides for cloning multi-genes in a polycistronic vector*. *Sci Rep*, 2017. **7**(1): p. 2193.
26. Chen, X., J.L. Zaro, and W.C. Shen, *Fusion protein linkers: property, design and functionality*. *Adv Drug Deliv Rev*, 2013. **65**(10): p. 1357-69.
27. Koch, B., B. Nijmeijer, M. Kueblbeck, Y. Cai, N. Walther, and J. Ellenberg, *Generation and validation of homozygous fluorescent knock-in cells using CRISPR-Cas9 genome editing*. *Nat Protoc*, 2018. **13**(6): p. 1465-1487.

28. Kimple, M.E., A.L. Brill, and R.L. Pasker, *Overview of affinity tags for protein purification*. *Curr Protoc Protein Sci*, 2013. **73**: p. 9 9 1-9 9 23.
29. Rossolillo, P., O. Kolesnikova, K. Essabri, G. Ramon Zamorano, and A. Poterszman, *Production of multiprotein complexes using the Baculovirus expression system: Homology based and Restriction Free cloning strategies for construct design*. *Methods Mol Biol*, In press. **XXXX**: p. XXXX.
30. Anders, C., O. Niewoehner, A. Duerst, and M. Jinek, *Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease*. *Nature*, 2014. **513**(7519): p. 569-73.

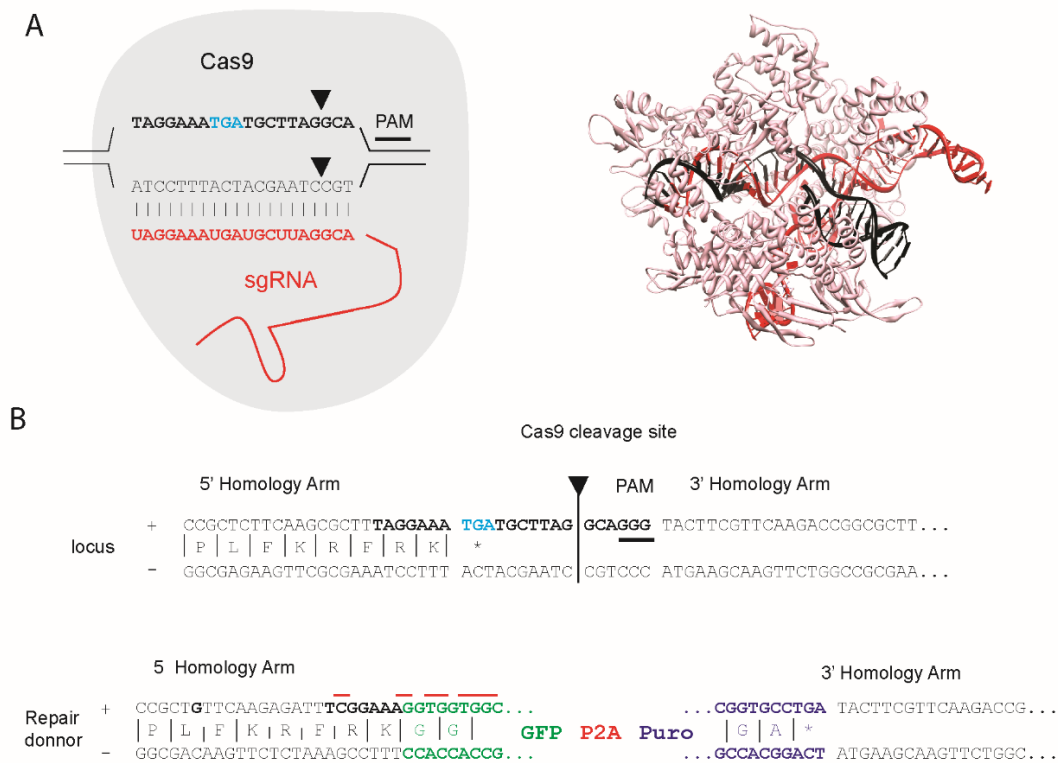
## Figures and Tables



**Figure 1. Experimental workflow for CRISPR/Cas9-mediated genome editing.** Several sources of Cas nuclease, guide RNA and donor template can be used. We detailed a protocol that utilizes electroporation (nucleofection) for delivery of plasmidic reagents (donor, Cas9, sgRNA).



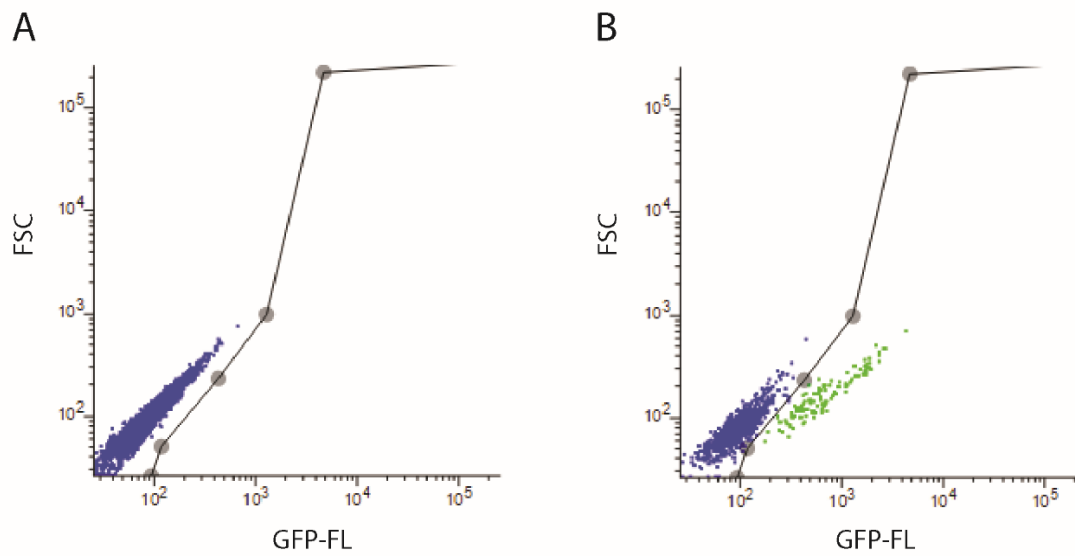
**Figure 2. Schematic of the tagging strategy for the XPB locus.** To introduce a GFP tag at the C-terminal end of the XPB gene, we select a sgRNA to generate a CRISPR/Cas9-induced DSB in vicinity of the stop codon of the XPB gene. The donor plasmid contains the eGFP-2A-Puro coding sequence flanked by respectively 300- and 800 bp homology arms (5' and 3' HA). The eGFP-2A-Puro coding sequence is inserted at the position of the XPB STOP codon and in frame with the last exon of the target protein.



**Figure 3. Design of sgRNAs and HDR donor sequences.** (A) Schematic representation of the Cas9 protein (grey) bound to sgRNA (red) and targeted to genomic DNA (black). The PAM sequence (GGG with overbar), the Cas9 cleavage sites (black triangles) and the stop codon of the XPB gene (TGA, in cyan) are indicated. A ribbon representation of the Cas9 nuclease from *Streptococcus pyogenes* bound to PAM-containing DNA target is shown in the right panel (PDB code, 4UN3;[30]). (B) Repair donor designed to insert the eGFP-2A-Puro coding sequence (all in frame with the last exon of the target protein) upstream of the XPB STOP codon (TGA). Nucleotides targeted by the sgRNA are in bold. After integration of the tag, the sequence is significantly modified (red line above the + strand), preventing cleavage of the repaired locus by Cas9.

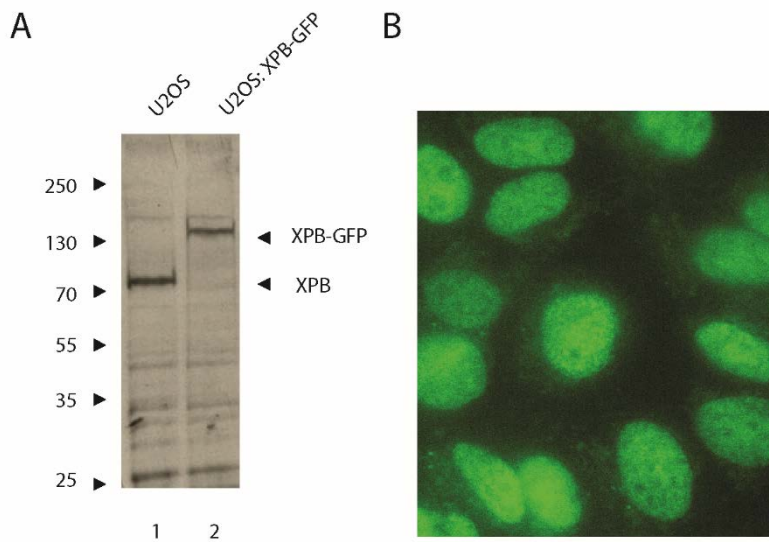
ggtggtggcgggttcaggcggaggtggctctggcgggtggcggatcggccATGGTGAGCAAGGGCGAGGAGCTGTTACCCGG  
 GGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTA AACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATG  
 CCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCACCCTCGTGACCACC  
 CTGACCTACGGCGTGCAGTGTCTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGA  
 AGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCG  
 ACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTAC  
 AACTACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAA  
 CATCGAGGACGGCAGCGTGCAGCTCGCCGACCCTACCAGCAGAACACCCCATCGGCGACGGCCCGGTGCTGCTGCCCG  
 ACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGGATCACATGGTCTGCTGGAGTTC  
 GTGACCGCCCGCGGATCACTCTCGGCATGGACGAGCTGTACAAGtacagcggtcgcgactctagagtcgacggttctGG  
 TAAGCAAAC TTTGAATTTTGACCTTCTTAAGCTTTCGGGAGACGTCGAGTCCAACCCCGGGCCCGcatgcaagcttcagc  
 tgaagcttaccATGACCGAGTACAAGCCACGGTGCGCCTCGCCACCCGCGACGACGTCCCAGGGCCGTACGCACCTC  
 GCCCGCGGTTTCGCCGACTACCCCGCCACGCGCCACACCGTTCGATCCGGACCGCCACATCGAGCGGGTCACCGAGCTGCA  
 AGA ACTCTTCTCAGCGCGTTCGGGCTCGACATCGGCAAGGTGTGGGTTCGGGACGACGGCGCCGCGGTGGCGGTCTGGA  
 CCACGCCGAGAGCGTCGAAGCGGGGCGGTGTTTCGCCGAGATCGGCCCGCGCATGGCCGAGTTGAGCGGTTCCCGGCTG  
 GCCGCGCAGCAACAGATGGAAGGCCTCCTGGCGCCGACCGGCCCAAGGAGCCCGGTGGTTCTTCTGGCCACCGTTCGGCGT  
 CTCGCCCGACCACCAGGGCAAGGGTCTGGGCAGCGCCGTCTGCTCCCGGAGTGGAGGGCGCGAGCGCGCGGGGTGC  
 CCGCCTTCTGGAGACCTCCGCGCCCGCAACCTCCCCTTCTACGAGCGGCTCGGCTTACCGTCAACCGCGACGTCGAG  
 GTGCCCGAAGGACCGCGCACCTGGTGCATGACCCGCAAGCCCGGTGCCTGA

**Figure 4. eGFP-2A-Puro nucleotide sequence.** cDNAs encoding gGFP and the puromycin resistance genes are in green and blue, respectively, the P2A peptide in red and linker regions in black.



**Figure 5. FACS analysis of wild-type and modified U2-OS cells.** Non-transfected U2-OS cells (A) and CRISPR/Cas9 modified cells in which a GFP tag is fused to the C-terminus of XPB (B) were analyzed for expression of a green fluorescent protein (B). Both plots show the green fluorescence intensity (GFP-FL) versus the forward scatter (FSC). U2-OS wild-type cells served as a negative control (blue dots) and define background signal; gates for the GFP positive cells (green dots) were set accordingly. The GFP positive cells were sorted into 96-well plates to obtain single clones expressing XPB-eGFP.





**Figure 6. Validation of engineered cell lines.** (A) Western Blot analysis of engineered U2-OS cells in which wild-type XPB was replaced by C-terminal GFP-tagged version (U2-OS<sup>XPB::GFP</sup>, lane 2). Un-modified U2-OS cells were used as control (U2-OS, lane 1) (Adapted from Sandoz et al. 2019). (B) GFP fluorescence of living U2OS<sup>XPB::GFP</sup> cells observed using an inverted microscope equipped with a 63x immersion objective.

	<b>U-2 OS</b> <sup>XPB::GFP</sup>
<b>Chromosome/Gene</b>	Chr. 2/XPB
<b>Homology arms</b>	5' Chr.2 : 127257599-127257872 (hg 38) 3' Chr.2 : 127257596-127256813 (hg 38)
<b>sgRNA</b>	5' <b>TAGGAAATGATGCTTAGGCA</b> <u>ggg</u> 3'
<b>Selection marker</b>	Puromycin
<b>sgRNA-FW</b>	5' ACACCG <u>TAGGAAATGATGCTTAGGC</u> G 3'
<b>sgRNA-RV</b>	5' AAAAC TGCCTAAGCATCATTTCCTA <u>CG</u>
<b>U6-seq</b>	5'-CAGGGTTATTGTCTCATGAGCGG-3
<b>XPB-FW</b>	5' AGACAGTAAGCGATCTGTAAACA 3'
<b>XPB-RV</b>	5' ACCCCACTCCCCAAAAGTT 3'
<b>XPB-FW2</b>	5' TCCTCTTCTTTCAGGTGTGGA 3'
<b>GFP-RV</b>	5' GAACTTCAGGGTCAGCTTGC 3'
<b>Puro-FW</b>	5' GCAACCTCC CTTCTACGAG 3'
<b>XPB-RV2</b>	5' GCGAATATGCCTTATGTGTG 3'

**Table 1. XPB locus.** Target sequence for the sgRNA used for editing the C-terminus of the XPB gene. Online web-applications such as CRISPOR (<http://crispor.tefor.net/>) aimed at optimizing CRISPR knock-in tag experimentation identify and rank potential guide RNAs in an input sequence. Oligonucleotides for sense and antisense strands designed with CRISPOR software for cloning into MLM3636 (sgRNA-FW and sgRNA-RV) as well as for genomic screening (XPB-FW, RV) are detailed. Efficient transcription from the U6 promoter requires a 5' G (Underlined).

<b>Genomic DNA(50ng/μL)</b>	1 μL
<b>dNTPs (25mM)</b>	0.4 μL
<b>DNA polymerase (2000U/mL)</b>	0.5 μL
<b>Buffer PCR 5X</b>	10 μL
<b>Forward Primer (10 μM)</b>	2.5 μL
<b>Reverse Primer (10 μM)</b>	2.5 μL
<b>H<sub>2</sub>O</b>	33.1 μL

**Table 2. Composition of the PCR mix used to amplify the genomic site XPB.** Use the following PCR program: 98 °C / 30s ; 98°C / 10 s 65 °C / 20 s 72 °C / 20 s (30 cycles); 72 °C / 5 min 4°C / Hold.