Università degli Studi di Torino

Dipartimento di Biotecnologie Molecolari e Scienze per la Salute

Dottorato di ricerca in Medicina Molecolare
Ciclo XXXV

Application of gene editing methods in the biopharmaceutical field: CRISPR/Cas system assessment for large Knock-Out and large Knock-In.
PhD Thesis

Candidata:
Dott.ssa Ammoni Alessandra

Tutors:
Marena Luisa, PhD
Prof. Bertero Alessandro

Coordinatore del dottorato:
Prof. Novelli Francesco

Anni Accademici 2019/2023
Settore scientifico-disciplinare di afferenza: BIO/11

Abstract

Chinese Hamster Ovary (CHO) cells are largely employed as host cell lines in the biopharmaceutical field for therapeutic protein production. Although many efforts were made to improve the process of clone development, the workflow to identify highly-producing clones remains a time-consuming procedure due to the great number of clones to be screened. Indeed, the process itself introduces variability between clones which involves the random integration of the transgene and the genome rearrangements occurring during the selection steps. The growing need to exert more control over cell line development and improve the performances of producing cell lines has prompted the implementation of targeted cell line engineering strategies. This has been made feasible by the significant advancements in CRISPR/Cas technology, which enables a rapid and efficient execution of a wide array of gene modifications.
In this frame, our study wants to assess the feasibility of the CRISPR/Cas system in driving large Knock-Out (KO) and in performing targeted insertions (Knock-In, KI) of a large portion of DNA in our CHO cells.

To assess the CRISPR/Cas large KO, we characterized the insertion site in one of our highly-producing clones with the Nanopore long-reads to delete the inserted transgenes and uncovered genomic rearrangements, within a region spanning 38.5 kb. The transfected pools were analyzed to identify the most efficient CRISPR/Cas12a assay. Monoclonal populations were subsequently generated, and a small-scale screening using droplet digital PCR detected 4 KO clones, demonstrating an overall KO efficiency of 3%. To further validate our results, Nanopore technology was employed, confirming the successful deletion of 38,5 kb.
For the large KI assessment, we investigated two approaches that can be strategically applied to our cell line development workflow: the direct CRISPR/Cas12a integration of a large plasmid and a two-step integration strategy with the so-called landing pads.
For the direct integration, we obtained 2 KI clones integrating 8’758 bp donor DNA at the target site achieving an on-target integration efficiency of 0,89%, consistent with literature data on large CRISPR/Cas-mediated integration. For the landing-pads approach we generated 2 clones integrating the landing pad at the target site with the CRISPR/Cas12a system, calculating an overall targeted integration efficiency of 12,5%. Subsequently, one of the landing pad clones successfully served as a platform cell line for the recombinase integration of a large donor. Both CRISPR/Cas12a integrations were confirmed with Nanopore technology, validating the on-target integration while excluding off-target integrations along the genome.

Our findings have demonstrated the applicability of CRISPR/Cas technology to our processes, yielding efficiencies consistent with those reported in the literature for large deletions and integrations. The acquired knowledge in these types of genetic manipulations has enabled the development of robust protocols that can foster a more consistent application of the CRISPR/Cas technology for the customization of our CHO-producing cell lines. Notably, both large KO and KI can be leveraged to modulate productivity-related genes, thus resulting in a higher and more efficient production of recombinant protein.

Table of Contents

1. Introduction

1.1 Current manufacturing process for therapeutic protein production

Over the last three decades, recombinant proteins gained increased importance in the therapeutic field for the treatment of complex human diseases (e.g., inflammatory disorders, cancer, infectious diseases), and Monoclonal Antibodies (mAbs) among others are the most employed class of biologics, thus continuously approved for clinical trials or launched into the market[1,2].

Biotherapeutic protein production relies on recombinant DNA technology and needs biological systems for their production, such as bacteria, yeast, and mammalian cell lines due to the higher dimension and higher complexity of those molecules if compared to non-protein drugs. From the biochemical point of view, therapeutic proteins should have a human-specific glycosylation pattern and post-translational modification to be effective and avoid possible adverse effects such as immunogenic reactions. For this reason, most biopharmaceuticals are currently produced in mammalian cell lines due to their inherent ability to synthetize proteins with appropriate human-like attributes [1,3,4].

Among mammalian cell lines, the Chinese Hamster Ovary (CHO) cell line represents the expression system of choice in the manufacturing process of biopharmaceuticals because of their desirable features in terms of safety and industrial purpose. Notably, CHO cells ensure a human-like post-translational modification, a proper folding of proteins, and are less susceptible to most human infective viruses. Moreover, CHO cell lines are able to grow in suspension with serum-free media and can be maintained in culture at high densities always ensuring a high titer of therapeutic proteins (1-10g/L)[5–8]. For this reason, in the past years, scientists tried to optimize as much as possible the CHO expression system and after the first immortalized CHO cell line was generated, monoclonal populations were isolated and selected for advantageous characteristics. Some of them were further modified by introducing mutations or via adaptation to different culture conditions to improve their performances. Actually, these derived cell lineages bared specific properties promoting more effective protein production or more efficient cell line establishment as compared to the parental one, thus better accomplishing the needs of the industrial process[9].

1.1.1 Cell Line Development

The manufacturing process for therapeutic protein production in mammalian cell lines follows a well-established process that relies on the random integration of the DNA sequence of the biopharmaceutical into the host cell line genome. The entire workflow follows several steps, from the vector design to the purification of the therapeutical protein as depicted in Figure 1. More in detail, the steps are the following:

- *Vector design.* In the first phase of cell line development, the gene coding for the therapeutic protein is cloned into the expression vector. The plasmid vector is designed to guarantee high and stable expression of the construct and to facilitate the selection process of stable insertions. Strong promoters and enhancers are usually employed to ensure a high transcription of the gene of interest. Moreover, since the splicing process is known to enhance mRNA nuclear export and stability, introns are introduced in the construct sequence to mimic the structure of an endogenous gene. Another aspect that has to be taken into account is that transgene expression is also influenced by the context of the integration site: insertions in heterochromatic regions will not result in protein expression, but even if the transgene is inserted in a euchromatic region, its transcription could be negatively affected by gene silencing due to epigenetic modification (DNA methylation and chromatin condensation). For this reason, UTR (Untranslated Regions that flank the coding sequence) or UCOE (Ubiquitous Chromatin Opening Elements) sequences could be inserted in the expression vector in order to avoid gene silencing events[1,3,7]. Ultimately, important genes for the selection phases are inserted in the plasmid backbone.

- *Transfection and selection.* Once the vector is properly designed, the host cells are transfected by electroporation with the plasmid and then subjected to selective conditions to recover those that have stably integrated the transgene into their genome. The most used selection markers are antibiotic-resistance genes, but other effective systems are becoming increasingly established. These systems include Dihydrofolate reductase (DHFR) or Glutamine synthase (GS) genes, which are inserted into the plasmid sequence, thus linked to the gene of interest. The DHFR enzyme catalyses the conversion of folate to tetrahydrofolate. This precursor is necessary for de-novo synthesis of purines and pyrimidines allowing cells to replicate. MTX, a drug similar to folate, binds to DHFR, thereby inhibiting the production of tetrahydrofolate: only cells with sufficient levels of DHFR, the ones bearing the construct, will survive and grow. An analogous way of action characterizes the GS system: the GS enzyme catalyses the production of glutamine from glutamate and ammonia, allowing cells to grow properly. In this case, a methionine sulfoxamine (MSX) selection system is used to select CHO cells producing the protein of interest. MSX binds to the GS enzyme, thus inhibiting the production of glutamine, which is necessary for cells to grow. Notably, in addition to another types of selection, these two systems are able to promote transgene amplification: increased administrations of MTX or MSX induce modified cells to amplify the selection marker and therefore the gene of interest that is linked. Transgene amplifications occur at the same locus, meaning that there are many tandem copies in the insertion site which are supposed to increase

the transgene mRNA transcription and consequently therapeutic protein production. On the other hand, an exceedingly high number of copies also correlated to higher genetic instability, so the goal is to balance the protein expression and the possible drawbacks[1,3,4,9].

- *Clone screening.* Following the selection step, the recovered heterogeneous cell populations are grown, expanded and single-cell sorted to obtain clonal cell line to be further expanded[1,3]. At this step, thousands of clones are produced and evaluated for recombinant protein expression and cell growth, leading to costly and time-consuming efforts. On the other hand, screening is necessary to choose the best candidate for the Master Cell Bank (MCB) based on growth, productivity, stability, and product quality. Aliquots of the best clonal population are produced and stored in liquid nitrogen[1,3].

- *Scale-up phase and purification.* Finally, the MCB starts the scale-up phase becoming the Working Cell Bank (WCB): the cells are grown into a bioreactor to produce therapeutic proteins. In the end, the biotherapeutic protein is purified through several steps on chromatographic columns and supplements are added to obtain the Drug product[1,3].

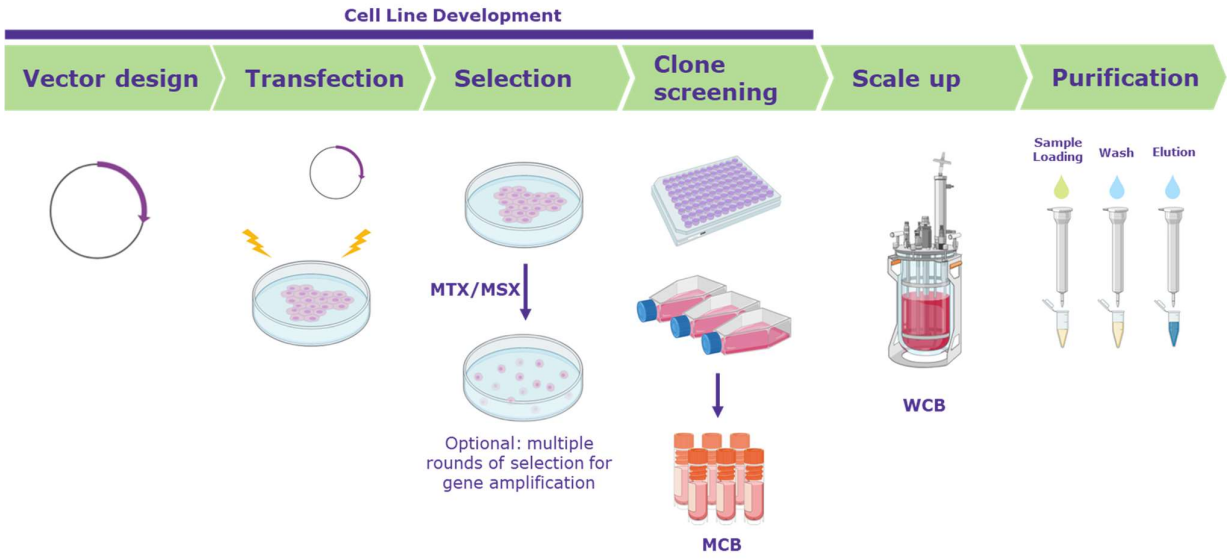


Figure 1. The manufacturing process for therapeutic protein production: the plasmid vector is transfected into the host cell line, then after some steps of selection and expansions, thousands of clones are produced. The clones are screened for recombinant protein production and growth. Finally, the best clone is selected as the MCB. The MCB is grown into bioreactors as WCB to finally produce therapeutic proteins which in the end are purified[1] (*created with BioRender.com)*.

The cell line development process described here requires from six to twelve months to select and obtain the MCB able to efficiently produce the molecule of interest. In the last years, a lot of work has been done to improve the process acting on media formulation, bioprocess optimization (e.g., feeding strategies and process parameter control) as well as on vector design. However, despite these improvements, the overall workflow has changed little, remaining time-consuming and expensive in terms of costs and labor[1,9,10]. In particular, the clone screening phase could be considered the real bottleneck in the cell line development process as a consequence of the high number of clones that need to be generated and evaluated for their growth, viability, productivity, and quality of the product.

Single-cell cloning or limiting dilution is technically necessary for the process because the productivity of individual cells in the transfected and gene-amplified cell populations is extremely variable. The heterogeneity of the clones is commonly attributed to the random integration of the transfected gene of interest into the genome[10,11]. In fact, it is well known that before and during an integration event, some modifications of both the vector and the genomic background likely occur (Figure 2). The transfected donor plasmid can be partially degraded before the integration resulting in an incomplete protein translation, or it can induce small indel events in the flanking genomic region which can also impact the transcription machinery. Moreover, it has been demonstrated that foreign DNA molecules are prone to link and form concatemers of different conformations (e.g., head-tail repeats, inverted and partial insertions), thus resulting in multiple copies of the transgene present at the same locus but with a wide range of arrays[11].

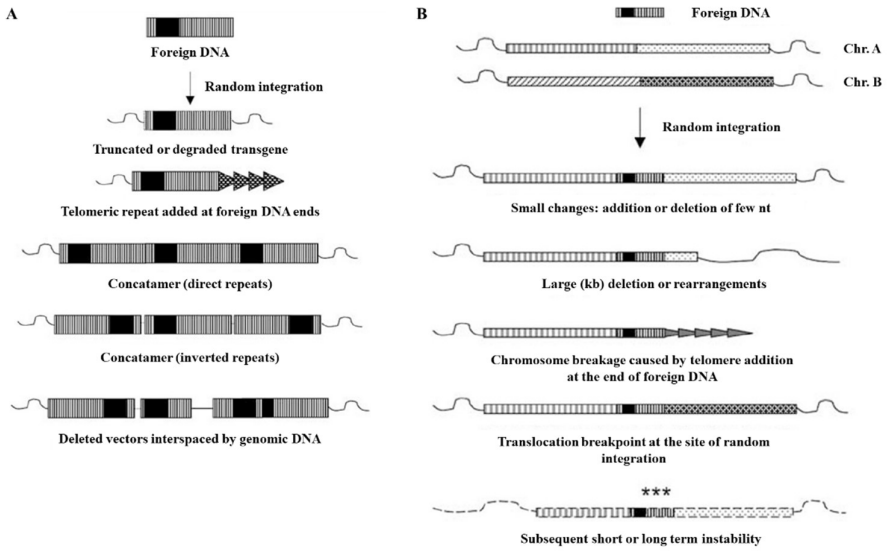


Figure 2. Examples of possible modification at the insertion site of foreign DNA by random integration. A) Some of the most common fates of integrated foreign DNA molecules. All these possibilities can be combined to give almost any imaginable and unpredictable kind of integrated array. B) Possible rearrangements of the locus targeted by the random integration which could impact the final expression of the transgene coded by the vector[11]. *(Picture readapted from Wurtele H. et al.2003)*

Furthermore, the insertion event but even more likely the gene amplification process results in large genomic rearrangements. Large deletions, duplications, and translocations could further increase heterogeneity in protein expression levels but also affect some important genes for cell homeostasis, leading to clone instability.

Finally, the chromosomal surroundings of the insertion locus exert strong influences on the promoter which in turn affects the transcription rate of the gene of interest, and this makes it difficult to obtain a homogenous level of protein expression among individual transfected cells. So, in the end, the heterogeneity caused by random gene integration and gene amplification has to be alleviated by generating monoclonal populations. The step of single-cell cloning can contribute to a more consistent bioreactor performance, product quality and, manufacturing process[8,10,11].

On the other hand, single-cell cloning necessarily results in thousands of clones to be screened and compared to identify a set of highly productive clones. This becomes even more important considering that high-producing clones are rarely generated in the heterogeneous cell populations after transfection and gene amplification[10]. Highly suitable candidate clones are selected based on high cell viability and growth, high productivity, high-quality product characteristics, and long-term expression stability. If the best candidate lacks those characteristics, it could impact future productivity in bioreactors and the quality of the final product, leading to a loss in time and cost for Pharma companies[7,10].

Advances in cell line development technologies and approaches are therefore crucial to support the rapid development of recombinant protein therapeutic products, focusing on reducing timelines and complexity of the analysis needed for clone screening. In this context, the identification of biomarkers able to predict productivity and expression stability is gaining increased attention[4,7]. In particular, researchers are now developing new molecular biology-based methods to correlate genetic parameters to the reliable performances of the clones.

Next Generation Sequencing (NGS) and Droplet Digital Polymerase Chain Reaction (ddPCR) based methods have been recently applied to the clone screening phase because of their high-throughput capability[12,13]. The ddPCR is mostly applied to the Transgene Copy Number analysis and Transgene mRNA expression, which could be directly related to final productivity[14,15]. On the other hand, NGS has a wide range of applications (e.g., genomics, transcriptomics and epigenomics) and among them, the Targeted Locus Amplification (TLA)-NGS coupled technique for the identification and characterization of the insertion site of the transgene becomes the best choice for having cost-effective and reliable analysis[16]. These techniques could be applied to characterize the generated CHO cell lines and investigate mechanisms related to productivity[6,8].

1.1.2 Characterization of protein-expressing clones

Next Generation Sequencing (NGS) and Droplet Digital Polymerase Chain Reaction (ddPCR) based methods have been recently applied to the clone screening phase because of their high-throughput capability[12,13].

Interestingly, these techniques could be also applied to investigate mechanisms related to productivity, thus predicting in early phases the future behaviour of generated CHO producing clones[6,8].

The ddPCR is mostly implemented to the Transgene Copy Number analysis and Transgene mRNA expression, that could be directly related to final productivity[14,15].

The advantage of the ddPCR relies on the absolute quantification of target nucleic acid sequences without the use of a standard curve. As illustrated in Figure 3, the ddPCR consists of different steps. This technology is based on the formation of water-oil emulsion droplets, in which the sample is randomly distributed. Subsequently, a PCR reaction takes place in each droplet using specific primers and a fluorescent probe for the target. After the target amplification, the droplets are analysed individually by reading the fluorescence generated by the different fluorophores. According to the number of droplets that are positive or negative for the target and to subsequent statistical calculations, the absolute quantification of nucleic acid expressed in copies/µL is obtained[17].

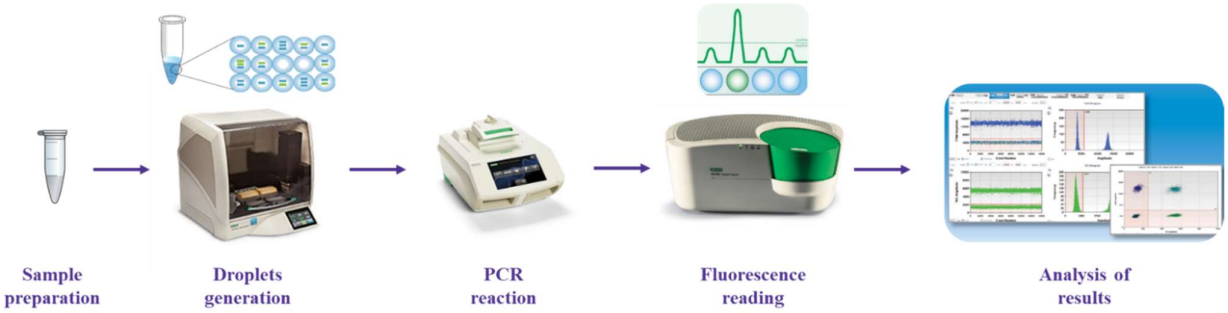


Figure 3. Schematic workflow for ddPCR experiment. The sample is added to a ddPCR master mix (primers, probes, Taq polymerase and buffer) that is partitioned in water-oil emulsion to create droplets. Target DNA in the partitions is amplified through PCR, and positive droplets are read by the Droplet Reader instrument. The ratios between positive and negative droplets are analysed by the software calculating the absolute quantification of the target sequence.

Analysis of GCN involves determining the number of copies of a given target locus concerning an invariant reference locus. In this method, the target and a housekeeping gene are amplified using duplex assays: specific primers for each target are used in combination with specific probes linked to different fluorophores. From the these two quantities, the number of copies of the target gene is calculated[17]. Similar to GCN analysis, evaluation of the target mRNA expression involves determining the number of copies of a given target RNA with respect to housekeeping RNA. From the concentration in copies/µL estimated, the ratio between the target and the reference is calculated from the software, resulting in the expression value of the target gene[11,14].

Taken together, these analyses are useful to compare and select the clones with higher GCN as well as higher mRNA expression: recent works show that high Copy Number is normally correlated to high productivity, through high levels of recombinant mRNA[18–20]. The analysis of GCN combined with mRNA data could give more information about the behaviour of the selected clones in terms of productivity. Focusing on the

production of mAbs, the cells are transfected with the sequences of both the Light Chain (LC) and the Heavy Chain (HC) that once translated pair together to form the antibody. Several studies combining GCN and RNA analysis on these mAbs-expressing cell lines showed that even if the copy number is the same for the LC and HC genes, the mRNA expression of the LC is generally higher than the expression of the HC, meaning that the expression of the latter is most likely one of the rate-limiting steps of the mAbs production[18,20]. In addition, a long-term study on clones could help in identifying clones prone to lose transgene copies during prolonged culture, which is one of the major causes of loss of productivity. Causes of loss in transgene copies could be searched for chromosomal instability that leads to translocation and re-arrangement of the construct[13]. Understanding the molecular mechanism of expression instability is extremely important to ensure that inefficient sites for transgene integration are avoided, and the comparison of GCN, transcription data coupled with deeper analysis of the genomic background of stable and unstable clones using NGS-based methods could be crucial to improve the workflow of protein production[21,22].

Indeed, NGS includes a wide range of applications (e.g., genomics, transcriptomics and epigenomics) and among them, the Targeted Locus Amplification (TLA)-NGS coupled technique for the identification and characterization of the insertion site of the transgene becomes the best choice for having cost-effective and reliable analysis[16].

The TLA technology uses the physical proximity of nucleotides within a locus of interest as the basis of selection. As a result, TLA enables the targeted enrichment and sequencing of any locus or transgene of interest without the need for detailed knowledge of the region to be investigated[23]. Therefore, this type of targeted sequencing is able to analyse the DNA segments that surround a selected site (the 'anchor') in a region of interest without loss of sequence. In Figure 4, the steps required for the TLA-seq workflow are summarised[23].

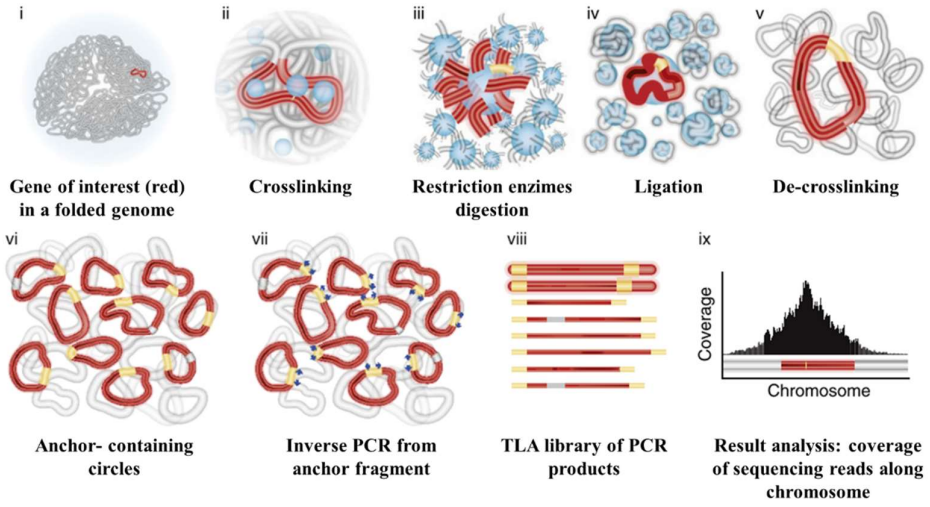


Figure 4. Summary of the steps for the TLA-NGS technology. Neighbouring sequences that form a gene or genetic locus (red) are in close spatial proximity (i) and therefore are preferentially crosslinked (ii). Digestion with a frequently

cutting enzyme (iii) and ligation (iv) results in large DNA circles composed of multiple crosslinked restriction fragments (v). Different Gene of interest (red) copies of a locus (from different cells) result in DNA circles composed of co-captured restriction fragments (vi). Fragments captured with the region of interest (the anchor sequence, yellow) are selectively PCR-amplified with anchor-specific inverse PCR primers (blue arrows) (vii). The resulting sample (viii) is highly enriched for locus-specific sequences and can be processed with standard library preparation procedures for NGS. The reads are then analysed and mapped against the reference genome (ix)[23]. *(Picture readapted from De Vree PJP. et al.2014)*

The TLA/NGS-based workflow is then able to support some of the current needs for cell line development and the clone selection process. This includes identifying clones sharing the same integration site, confirming the correct integrity of the genome-integrated plasmid sequences and identifying and eliminating clones producing aberrant protein products at an early stage[16,24].

Taken together, the correlation of insertion site information, stability of transgene copies, mRNA expression and productivity will increase the knowledge underlying instability of transgene expression providing more awareness and possibly more control on the process[21]. Indeed, the increased knowledge derived from those studies, together with the emergence of genome editing tools, provides novel opportunities for targeted genome engineering both for host cell line improvement and for replacing random integrations with all their related drawbacks. Here, the goal is to drive the system, mitigating at least some of the stochastic events that could impair the production of biopharmaceuticals.

1.2 Targeted cell line engineering

The current method of cell line development relies on the random integration of transgenes followed by laborious screening procedures to identify stable and high-producer clones. As discussed in the previous paragraph (see section 1.1.1), two major drawbacks of uncontrollable integration events include: 1) high clonal variation due to the heterogenous copy number as well as the location of transgene insertion and 2) the reduction of clonal stability over production time derived by loss of transgene copy number, chromosomal rearrangements, and epigenetic regulations. Therefore, a great investment in terms of research and optimization has been made during the last years to develop new methods based on a targeted integration of the transgene in specific loci of the CHO genome.

In particular, the transgene knock-in (KI) into transcriptionally active genomic hot spots has been favoured as a controllable and predictable CLD method[25]. Precise insertion of transgenes into the so-called genomic safe harbours ensures production consistency in long-term culture and high expression levels of the biotherapeutics. This strategy also alleviates clonal variation and results in production titers comparable with those reached by classical random integration[9,25,26]. At the same time, a high number of works focus on the understanding of effective targets to be modulated to gain better performances of mAbs-expressing cell lines taking advantage of targeted genetic manipulation. The tuning of critical cellular pathways definitely succeeded in improving CHO features acting on glycosylation genes, cellular metabolism, antiapoptotic/proapoptotic pathways, cellular productivity, cell cycle checkpoint kinases, and post-translational modifications, to report some examples[26].

In this context, the development and the optimization of programmable nucleases have boosted the field of cell line targeted engineering also affecting biopharmaceutical industries. Starting from the very first gene editing experiments, different tools have been employed, with a particular focus on zinc-finger nucleases (ZFN), transcription activator-like effector nucleases (TALEN), and Clustered Regularly Interspaced Short Palindromic Repeats/CRISPR-associated system (CRISPR/Cas) system. However, the CRISPR/Cas technology has gained more and more visibility thanks to its high targeting efficiency, cost-effectiveness, and simplicity of design, making it the first preference for genetic manipulations, and replacing almost all the other available tools.

1.2.1 CRISPR/Cas system

The CRISPR/Cas system was first discovered as a bacterial and archaeal adaptive immune system against phage infection[27,28]. The effector complex of this system is composed of an RNA molecule complementary to the target sequence, i.e., the crRNA (CRISPR-RNA), combined with an endonuclease, i.e., the Cas protein, which is able to cut the target DNA recognized by the crRNA. Once the double-stranded break (DSB) is generated, distinct types of gene editing can be carried out taking advantage of the endogenous DNA repair

pathways of the cells. If the nonhomologous end joining (NHEJ) repair occurs the target site will be mended in an inaccurate way introducing small indels (insertion/deletion) which can cause a frameshift of the targeted gene thus generating a Knock-Out (KO) (Figure 5). On the other end, it is also possible to perform the Knock-In (KI) of exogenous DNA sequence at the point of the CRISPR/Cas cleavage. This type of editing requires a donor DNA composed of the sequence of interest flanked by the so-called homology arms which need to be homologous to the genomic free-ends generated by the DSB. In this way, the homology-directed repair (HDR) pathway is able to use the donor DNA as a template to precisely repair the genome (Figure 5).

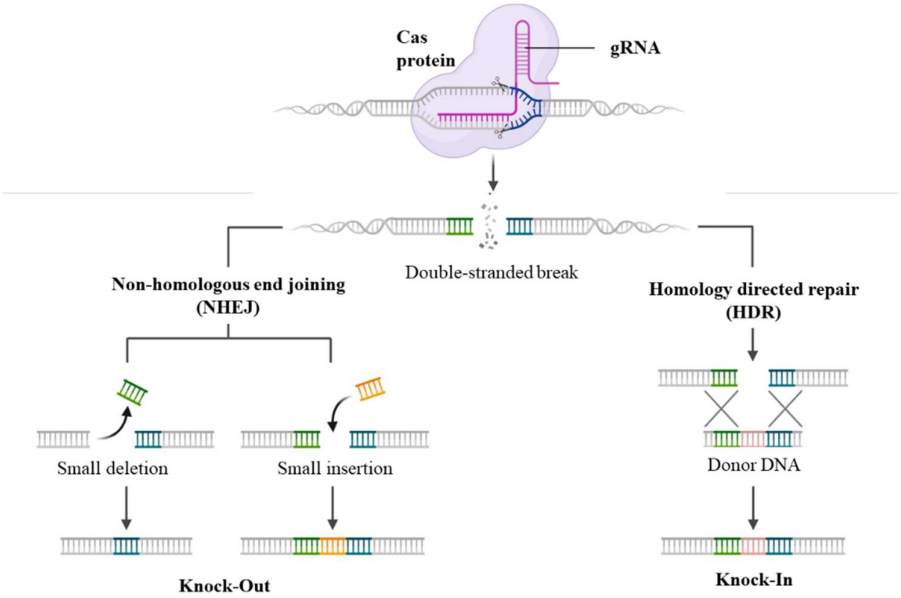


Figure 5. The mechanisms of CRISPR/Cas-mediated genomic editing are summarized. Initially, a double-strand break (DSB) is generated by the complex at a specific target site. Subsequently, KO can be generated via the non-homologous end-joining pathway which generates small deletions and/or insertions at the break site. Differently, donor DNA with long homology arms could be KI at the cleavage site via the homology-directed repair (HDR) pathway. (*Created with BioRender.com)*

CRISPR/Cas9 system is the very first complex that has been characterized and employed as a reprogrammable gene editing tool[29].

This complex is composed of the Cas9 endonuclease, which can recognize and cleave dsDNA substrates, and a guide RNA consisting of an additional small noncoding RNA, the trans-activating crRNA (tracrRNA), which base pairs with the repeat sequence in the crRNA[30]. Even in this dual-RNA structure, the target specificity is granted by the 20-nucleotides (nt) spacer of the crRNA and more specifically by the so-called seed region. In CRISPR/Cas9 and other systems of the same type, the seed region has been defined as the first 10–12 nt located at the 3' end of the spacer. Mismatches in this seed region severely impair or completely avoid target DNA

binding and cleavage, whereas close homology often leads to off-target events even with many mismatches elsewhere[29,31]. In the end, once the guide RNA pairs a complementary sequence proximal to a 5'-NGG-3' protospacer adjacent motif (PAM) which is a short, conserved motif necessary for Cas9 target recognition, the endonuclease generates a site-specific blunt-ended DSB[29] (Figure 6A).

The Cas9 protein is made of two distinct lobes, the alpha-helical recognition (REC) lobe and the nuclease (NUC) lobe. The latter contains a conserved HNH-like nuclease domain that cleaves the DNA strand complementary to the guide RNA sequence (target strand), and a RuvC-like nuclease domain responsible for cleaving the DNA strand opposite the complementary strand (nontarget strand). In addition to these, the variable C-terminal domain contains PAM-interacting sites required for PAM interrogation[32].

The Cas9 from S. pyogenes Cas9 (SpCas9) is the first Cas nuclease studied and consequently the most well-characterized type. This knowledge allowed the researchers to engineer the Cas9 protein to fit their experimental needs. Modified Cas9 nucleases are now commercially available and include the high-fidelity SpCas9 (SpCas9-HF1), which ensures better performance in terms of on-target cleavage, the SpCas9 nickase (Cas9n) which bears an inactivating mutation in one of the two nuclease domains thus creating a nick on the target sequence, and the nuclease-dead SpCas9 (dCas9) which have both nucleases domains enable to cut and could be used for gene modulation or epigenetic remodelling.

Together with the optimization of the classical Cas9, researchers explored the possibility of employing other CRISPR/Cas systems. In particular, the Cas12a (also known as Cpf1) turned out to be quite similar to the Cas9 in their functionality but with some interesting differences. The effector complex with the Cas12a needs a single crRNA molecule which is 42-44 nt in length, with the first 19-20 nt corresponding to the repeated sequence and the remaining 23-25 nt to the spacer sequence[33,34]. Regarding the Cas12a protein, the researchers have pointed out some peculiarities if compared with the Cas9. Even if the overall bilobed structure is similar, the Cas12a possesses only one nuclease site at the Ruv-C domain leading to a DSB with a 5 nt overhang on the target strand distal to the PAM site[33]. Also in this system, the PAM recognition is necessary for the correct binding and cleavage of the target DNA and Cas12a shows an efficient targeting of protospacer followed by a 5'-T reach PAM, 5'-TTTN-3' for the most employed AsCas12a and LbCas12a[34] (Figure 6B). But in the end, the distinctive feature of the CRISPR/Cas12a system is the indiscriminate ssDNAse activity upon activation. In fact, the recognition and the cleavage of the target dsDNA change the conformation of the Cas12a in a way that the catalytic pocket becomes always available for ssDNA which is degraded into single/double nucleotides[35]. This unspecific activity could be a powerful tool in addition to the classical on-target gene editing and a way to prevent undesired off-target insertion due to the prolonged presence of donor ssDNA during KI experiments.

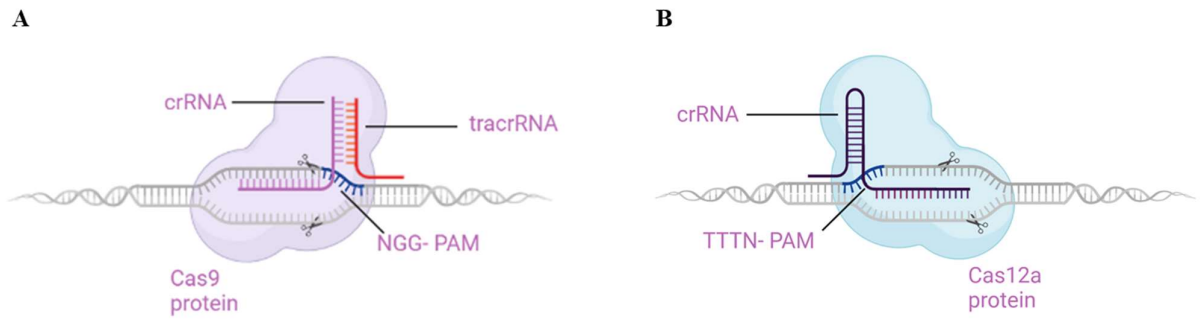


Figure 6. Schematic representation of CRISPR/Cas9 and CRISPR/Cas12a activity. (A) The duplex crRNA:tracrRNA guides the Cas9 at the target site. Upon recognition of the NGG-PAM, the Cas9 is able to cleave both strands creating a blunt cut 3 nt at the 5' of the PAM sequence. (B) The Cas12a protein requires its crRNA and a TTTN-PAM to be activated. After the binding, the endonuclease domain generates the DSB cutting the target strand (the one containing the PAM) 18-19 nt at the 3' of the PAM and the non-target strand 23nt from the PAM, resulting in a 5′ overhang of 4–5 bases. (*Created with BioRender.com)*

Considering the low complexity of the system, it becomes clear the huge interest generated by the first demonstration of its potential as a reprogrammable tool to be employed in the gene editing field [29,36]. Indeed, it is possible to theoretically target every genomic sequence by changing the sequence of the crRNA, which confers the specificity to the system. Furthermore, the engineering of an RNA molecule is extremely effortless if compared to previous nuclease-mediated DNA editing techniques such as ZFNs and TALENs in which the DNA recognition is specified by the proteins. The ZFNs complex is able to recognise 3 nt, so multiple ZFNs need to be linked together to target a longer and unique sequence[37]. However, the design and efficiencies prediction of modified zinc finger arrays result to be time-consuming and requires consolidated expertise in protein engineering[37]. This is due to the absence of a consistent ZF-base "code" and the variation in sequence recognition caused by interactions with neighbouring ZFs[38]. To the same extent, the TALEN technology requires that each targeted nt is recognized by a different TALE motif[39]. The challenges of targeting a sequence using TALENs is in the complex process for assembling the array itself and the large coding sequence needed to express the array, impacting the delivery of the TALEN-expressing vector[40]. Indeed, both ZFNs and TALENs systems require complex design which excluded these technologies from performing advanced gene editing experiments such as multiplex editing. On the other hand, the constraints of ZFNs and TALENs have profoundly boosted the application of the more flexible and easy-to-handle CRISPR/Cas system for large-scale genomic manipulation and has made the gene editing field even more accessible[32]. To date, the CRISPR/Cas system has been increasingly studied and optimized to achieve the best performances based on the different experimental demands and types of gene modification. This includes the engineering of Cas protein and RNA guides, as well as the choice of the more suitable Cas protein that the bacterial world can offer.

1.2.2 CRISPR/Cas application for cell line development

The intrinsic ability of CRISPR/Cas to perform several types of targeted modifications together with the increased knowledge and characterization of this tool, have pushed the pharma industries to consider this system for optimizing and speeding up their cell line development workflows. In the last years, many explorative works on CHO cells have produced positive results covering all the possible needs of CLD, from the targeted integration of the transgene to the enhancement of productivity.

The random integration method employed in CLD implies that the target gene could be integrated into the heterochromatin or an unstable area of chromatin, necessitating repeated rounds of selection to acquire acceptable and stable target protein expression levels. Furthermore, proteins produced from randomly integrated cell lines have a significant degree of sequence variability, which could be attributable to the mutagenicity of the selective agents employed to select the high-producer clones[8,10,11]. Here is the opportunity to employ CRISPR/Cas systems to target the integration of transgenes into specific loci in the host cell genome. Obviously, for successful transgene KI and expression, a so-called hot spot is required, and the identification of new functional loci would be fundamental for this new cell line generation process. Although some published works applied different methodologies, such as lentivirus-based random integration, cut-and-paste procedures, along with bioinformatics tools, to identify new hot spots in CHO cells genome, the public availability of this kind of information is uncommon due to commercial value concerns[26]. Despite the limited data sharing, some works came out with promising results of targeted and stable integration in well-characterized loci such as HPRT[41,42], GRIK[42], Fer1L4[43], C12orf35[42], and ADGRL4[44] with good productivity in small-scale analysis. Of course, the KI efficiencies of the different experiments result to be highly variable depending on the experimental design and ranging from 1,4% to 41,7% with donor DNA up to 6 kb in length. Possibly, this variability likely relies on different aspects: the locus could be more or less suitable for the integration, the length of the donor DNA is known to be a key aspect as well as the design of the plasmid (length of the homology arms, in vivo linearization sites, etc.), and also the KI strategy could be optimized to best fit the experimental request[26].

Among them, the more complete and promising work on targeted integration with the CRISPR/Cas9 system is the one carried out by Zhao et al., in which three highly active sites (C12orf35, HPRT, and GRIK1) were investigated to select the best one to be used in a reliable site-specific integration strategy to develop recombinant cell lines efficiently. In this study, the donor plasmids encoding anti-PD1 monoclonal antibody (5,6kb in length) were targeted into these three loci respectively through CRISPR/Cas9 technology and stable cell lines were generated successfully after a single round of transfection and selection with an efficiency of 41,7%. All the targeted integration cell lines showed higher productivity, among which the C12orf35 locus was the most advantageous in both productivity and cell line stability. Moreover, analysis of the inserted sequence, the quality of the produced antibodies, and the check on off-target mutation confirmed the high reliability of this method with high-quality products and low off-target rates caused by CRISPR/Cas9 activity.[42]

In addition to these positive results, this work pointed out also the savings in terms of costs and time of applying the targeted method with the CRISPR/Cas system: the conventional random integration strategy requires multiple rounds of pressure selection and an extensive screening of the clones which normally takes from 6 to 12 months, whereas the targeted approach takes only 3 weeks to get targeted transfected pools and 5 weeks to grow and select integrated clones performing just a small screening[42]. Besides all these promising results, the productivity of analysed clones remains below the ones reached by current recombinant CHO cell lines grown in bioreactors[42]. However, the productivity of CRISPR/Cas-modified cell lines has been always evaluated under sub-optimal conditions and in small-scale models, so further analyses are required to comprehensively characterize the performances of these cell lines.

Together with the evaluation of CRISPR/Cas systems for the targeted integration of recombinant protein into specific loci, the researchers tried to take advantage of this precise tool for modulating endogenous genes to enhance productivity. The engineering of CHO cell lines to create a sort of super-host has always been explored with different available approaches including gene over-expression, noncoding RNA-mediated gene silencing and gene KO, but with the development of the CRISPR/Cas system, all the previous tools were almost completely replaced. In particular, KO engineering has gained importance in a high number of studies because of its rapid and easy experimental design. As mentioned above, researchers already focused on finding effective molecular targets to obtain cell lines with desirable features, so it was easy to apply the recent technologies to previous knowledge. The CRISPR/Cas KO of genes involved in various endogenous pathway results to be effective in this purpose: the disruption of genes related to amino acids catabolism enhances cell proliferation[45,46], the KO of glycosylation enzymes to add or remove specific glycosylation sites increases biopharmaceuticals efficacy[47], impaired activity of de novo methyltransferases reduces epigenetic transgene silencing[48], and ultimately the multiple-KO of host cell proteins (HCPs) increases the yield of mAbs purifications during downstream processes[49,50].
Less usual but likewise effective are the studies relying on CRISPR/Cas-mediated KIs to enhance production in CHO cells. The goal is to improve some of the well-known bottlenecks of recombinant protein production, such as protein secretion and endoplasmic reticulum (ER) stress. In this regard, the study of Wang et al. aimed to induce the unfolded protein response (UPR), a transcriptional program that promotes protein folding and secretion, but in a way to prevent the proapoptotic effect of prolonged UPR activation. The authors introduced hSurvivin and hQSOX1b genes with CRISPR/Cas, improving ER microenvironment and strengthening the antiapoptotic ability. In the end, the modified cell lines resulted to be less responsive to apoptotic signals and increased the yield of protein production by 6,4-fold to 5,55-fold[51].

Taken together, all these works demonstrate the efficacy of the employment of the CRISPR/Cas system in the CLD workflow, bringing benefits in terms of efficiency, time, and costs. Notably, in the last year most of the studies showed reliable results on real protein expression data and not only on reporter protein, providing consistent evidence to the industries. Of course, this is quite a recent field, and a lot of work could be done to optimize the system to meet the requirements of biopharmaceutical companies.

2. Aim of the project and experimental design

The cell line development workflow still relies on random integration, but industries are now exploring new strategies to overcome all the issues related to this technique. Considering the huge step forward in the field of targeted genome editing, it becomes evident why the CRISPR/Cas technology represents the best choice for driving the system in order to improve the cell line development in terms of costs, time as well as simplifying all the characterization analysis needed by regulatory authorities. Moreover, the applicability of genome editing is not only restricted to the targeted integration of molecule constructs but could be a powerful tool to engineer CHO host cell lines to improve their performances and to better fit the needs of the industrial process.

In this frame, our work aims to evaluate the application of the CRISPR/Cas system in our current workflow of cell line development starting from our in-house optimized CHO cells provided by our CLD facility (based in Vevey). We want to assess the feasibility of the CRISPR/Cas system in driving large KO and in performing targeted KI of large portions of DNA. The goal is to provide effective protocols exploring the wide variety of CRISPR/Cas tools and experimental design combinations recently highlighted by academic literature and CRISPR/Cas system providers.

For the evaluation of KO, we decided to stress the system and eliminate a very large portion of the endogenous DNA. Therefore, we started with one of our clones selected for its high productivity (i.e., super-clone) trying to eliminate the integrated sequence as a whole. The super-clone was generated by the CLD facility in Vevey using a classical random integration approach which necessarily implies genomic rearrangements involving at least the genomic locus and/or the construct itself, so we need a real picture of what happened in the integration site of this clone[11]. For this reason, the first required step was the characterization of the integration locus starting from the TLA analysis already performed during the clone screening step and completing the assessment taking advantage of long-reads Nanopore sequencing. This was extremely crucial to set up the best strategy to restore the original locus and to know the real sequence for the design of the crRNAs (Figure 7).

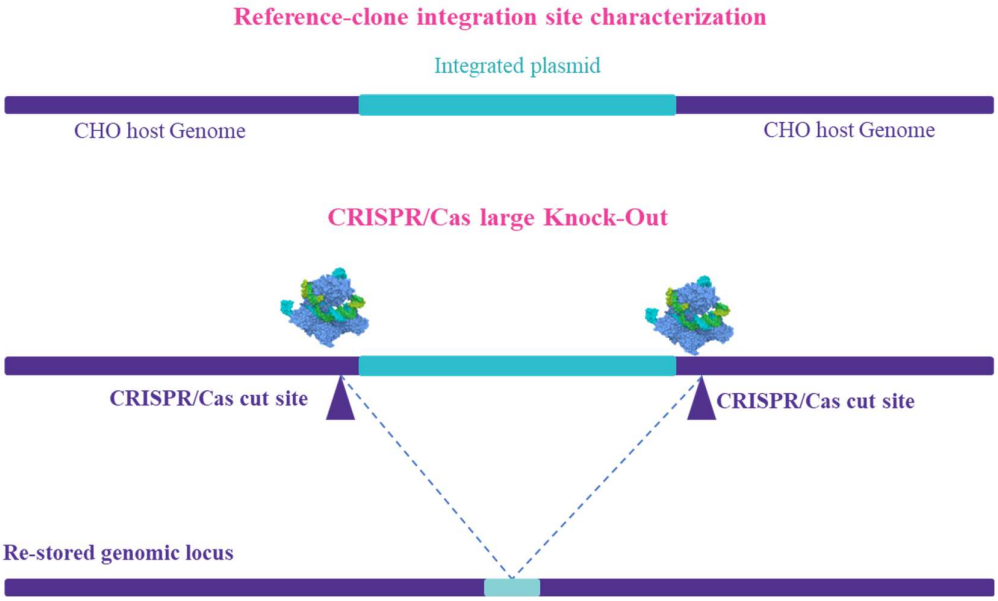


Figure 7. Schematic representation of the experimental workflow for large KO assessment. First, the sequence of the reference clone insertion site is obtained with NGS-based techniques. Then, the sequence is used to design the crRNA targeting at both sides of the integrated plasmid and all the united genetic rearrangement likely occurred for the random integration approach used to generate the clone. At the end, the free-ends are re-joined by the endogenous repair pathway of the cells.

Regarding the CRISPR/Cas tool, we explored both Cas9 and Cas12a to see which system could give the best results in our specific experimental design. In our work, the efficiency of the tested systems was first evaluated in the transfected pools with the ddPCR to see the presence/absence of the construct that we intended to delete, and then the same type of analysis was carried out on the generated monoclonal populations. Ultimately, sequencing of the edited cells was performed to uncover the restored sequence.

Regarding the KI experiments, the literature research reveals that the major issue in CRISPR/Cas integration is the decreasing of the efficiency as the donor DNA increases in length[26]. Given that the sequence coding for an antibody usually ranges from 7kb-10kb, the length of the donor could represent a problem in terms of efficiency for the CRISPR/Cas system in driving a consistent KI.

So, we applied and adapted two of the most promising strategies to see if they could fit in with our workflow. In the first one, the goal was to perform a CRISPR/Cas direct integration of a large plasmid, whereas we selected a CRISPR/Cas-recombinase combined approach for the second one. The starting point was the same for both workflows: the characterization of the target locus in the newly generated CHO host cell line by the CLD group. With Nanopore sequencing, we were able to uncover the locus identified in the KO experiments to know if it was possible to use the same best crRNA. After that, we proceeded with the donor DNA design suitable for the two selected strategies.

In the first strategy, we started with the design of donor plasmid to be directly inserted by the CRISPR/Cas system. The peculiarity of this approach was the design of the donor plasmid itself, which contained the recognition site for the CRISPR/Cas ribonucleoprotein (RNP) complex at the side of the two homology arms. This design ensures the in vivo linearization of the donor plasmid needed for the efficient and precise integration of the specific sequence of interest without the plasmid backbone[26]. As mentioned, the possible issues could be related not only to the length of the sequence which is known to decrease the efficiency but also to the type of the donor (both circular and linear dsDNA) which is more prone to random off-target events and more cytotoxic than ssDNA[52] (Figure 8A).

The second strategy consisted of a two-step integration approach. It took advantage of the so-called Landing Pads (LPs) and combined the precision of the CRISPR/Cas system with the high efficiency of recombinases in large integrations. In the first step, the integration of the Landing Pad was carried out by the CRISPR/Cas system thus ensuring the locus-specificity of this approach. The landing pad core was made up of a small, specific sequence recognized by a recombinase, but other expressing cassettes could be added such as antibiotic resistances or fluorescent markers. In this manner, the resulting cell line integrating the LP could be employed as a platform for the subsequent recombinase-based integration of the transgene of interest. With this strategy, we should avoid the issue related to the length of the donor inserted by the CRISPR/Cas system because the LP vector could be refined in a way that only 1kb is needed. The lower length of the donor also gave the possibility to synthesize the LP as ssDNA which is known to ensure lower rates of off-target and lower toxicity if compared with dsDNA. Certainly, this was a more complex strategy that would require a longer timeframe to obtain the final cell line. Nonetheless, the platform cell line could serve as an initial host for introducing various constructs in the same locus (Figure 8B).

In those approaches, the plasmid or the LP and the CRISPR/Cas RNP complexes were co-transfected in our CHO cell line. First, generated pools were monitored with the ddPCR to see the presence of the constructs and with junction PCR to have partial positioning information. These data were used to individuate the best pools to carry on for single-cell cloning. The clones were further screened with PCR and ddPCR to identify the correctly modified clones. In the end, the selected ones were sequenced with Nanopore technology to evaluate and characterize the insertions. Overall, our results would increase our expertise in handling this type of genetic modification, generating consistent protocols to be used for future projects. Moreover, the acquired knowledge could boost the implementation of the CRISPR/Cas technology in our CLD process, significantly expediting all stages of clone development and resulting in a more rapid and efficient workflow.

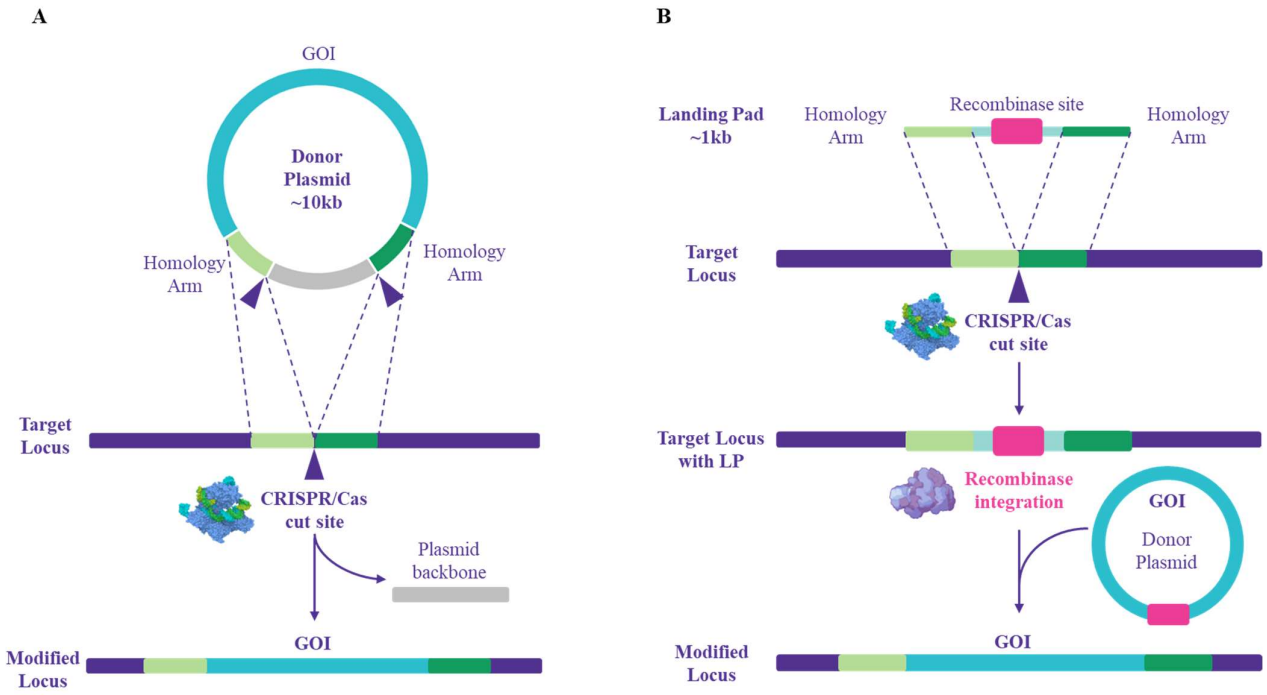


Figure 8. Schematic representation of the experimental workflow for large KI assessment. (A) The direct KI approach is summarized in the scheme. The RNP will cleave both the target locus and the donor plasmid at the side of homology arms, giving a clean insertion of the GOI (gene of interest) at the target site. (B) The Landing Pads approach is shown in the scheme. The first integration is performed by the CRISPR/Cas system to generate a master LP cell line used to perform the recombinase integration of the GOI.

3. Materials and methods

3.1 Cells and culture media

For the KO assessment, the CHO-K1 selected reference clone was used to perform experiments (supplied by CLD group in Vevey). For the KI experiments, CHO-K1 (ATCC) and CHO-K1 (ATCC) in-house optimized cell lines were employed (supplied by CLD group in Vevey). All the cell lines were cultured in an in-house developed media supplemented with glutamine 4,5mM (X0550, Biowest). Cells were incubated in 125-mL shake flasks (214-0447, VWR) at 37°C with 5% CO_2 and 115 rpm. For the clone generation, single-cell cloning media was composed of the in-house developed media supplemented with glutamine 4,5mM and 20% FBS (F7524, Sigma-Aldrich) or 10%FBS for the KO and KI cells, respectively. Transfected pools were seeded in 96-well plates in static conditions for the single-cell cloning phase: the pools were diluted to seed 0,5 cell/well for the KO experiment and 1 cell/well for the KI experiments. After outgrowth, the clones were maintained in 48-well plates during the screening phase and only selected clones were then expanded to standard shake flask culture conditions. Cell count and viability were monitored with the Vi-CELL BLU cell analyser (Beckman Coulter) based on the trypan blue dye exclusion method.

3.2 CRISPR/Cas assay and donor DNA design

The crRNAs for both Cas9 and Cas12a were designed with the online tool CHOPCHOP (http://chopchop.cbu.uib.no/) which can also provide an in-silico analysis of the possible off-targets. The design was performed by selecting the CHO-K1 GCF_003668045_CriGri-PICRH-1.0 as reference genome. The donor plasmids for the direct KI approach were designed and ordered from VectorBuilder whereas the ssDNAs employed for the Landing Pad donor were ordered from GENEWIZ (Azenta Life Science). All the transfections were performed using the AMAXA 4-D Nucleofector system (Lonza) with the SF Cell Line 4D-Nucleofecto X Kit S.

3.3 Transfection conditions

For the KO experiments, cells were seeded 3-4 days before transfection at a density of 2×10^5 viable cells/mL to have 1×10^6 viable cells for each condition to transfect. Cas9 and Cas12a RNP complexes were formed before transfection according to the manufacturer's instruction: Cas9 RNPs were generated incubating 24 µM of crRNA:tracrRNA duplex with 20,7 µM of Cas9 protein in 5 µL PBS at room temperature for 15-20 minutes (Alt-R® CRISPR-Cas9 crRNA, Alt-R® CRISPR-Cas9 tracrRNA, Alt-R™ S.p. Cas9 Nuclease V3 from IDT); Cas12a RNPs were generated by incubating 32 µM crRNA with 25,2 µM Cas12a protein in 5 µL PBS at room temperature for 15-20 minutes (Alt-R™ A.s. Cas12a crRNA, Alt-R™ A.s. Cas12a (Cpf1) V3 from IDT). The

transfection was performed on 1 × 10^6 cells resuspended in 20 µL SF solution at which 1 µL of 78 µM IDT Electroporation Enhancer (1076301, IDT) was added. After nucleofection, cells were recovered in six-well plates containing prewarmed media for 7-10 days before genetic analysis was performed.

Regarding the KI experiments, cells were seeded 3-4 days before transfection at a density of 2 × 10^5 viable cells/mL to have 1 × 10^6 viable cells for each condition to transfect. Cas12a RNP complexes were formed before transfection depending on the approach. For the direct KI Cas12a RNPs (4,8 µM in the final transfection solution) were generated by incubating 40 µM crRNA with 31,5 µM Cas12a protein in 4 µL PBS at room temperature for 15-20 minutes (Alt-R™ A.s. Cas12a crRNA, Alt-R™ A.s. Cas12a (Cpf1) V3 from IDT). The RNPs were then mixed with the donor plasmid testing 0,5µg, 1µg and 2µg (VectorBuilder). For the LP approach, three Cas12a RNPs concentrations in the final solution were tested: 1,2 µM, 2,3 µM and 4,8 µM. The complexes were generated by incubating 240 µM crRNA with 31,5 µM Cas12a protein in 4 µL PBS at room temperature for 15-20 minutes (Alt-R™ A.s. Cas12a crRNA, Alt-R™ A.s. Cas12a (Cpf1) V3 from IDT). This starting solution was then diluted 1:2 and 1:3 to obtain the conditions at 2,3 µM and 1,2 µM respectively. The RNPs were then mixed with the ssDNA for the LP at 2,5µg (GENEWIZ). For both KI protocols, the transfection was performed on 1 × 10^6 cells resuspended in 20 µL SF solution at which 1 µL of 78 µM IDT Electroporation Enhancer (1076301, IDT) was added. After nucleofection, cells were split into three wells of 96-well plate filled with pre-warmed media containing the Alt-R™ HDR Enhancer V2 (10007921, IDT) and media exchange was carried out after 16-17h pooling the three wells into a single well of a six-well plate. For the direct KI approach cells were let recover for 4 days before applying antibiotic selection with 10ug/mL of Puromycin. The cells were selected for 7 days and then maintained with 7ug/mL of puromycin excluding the phases of clone generation in which any selection was applied.

3.4 DNA extractions

The extraction of DNA was performed with different kits depending on the purpose of the consequent analysis. For the analysis of the transfected pools with ddPCR and junction PCR as well as the final evaluation of generated clones, genomic DNA is extracted from cell pellets of ~4x 10^6 cells using the QIAmp Blood Mini Kit (51104, QIAGEN Group). Each cell pellet was processed following supplier instructions. DNA derived from each column is finally eluted in 100µl of Molecular Biology water (W4502, Sigma-Aldrich) and quantified using NanoDrop instrument. All samples show 260/280 (purity respect to protein content) and 260/230 (purity respect to organic compound) NanoDrop ratios ranging from 1,8 and 2,2.

For the clone screening phase, genomic DNA was extracted with QuickExtract DNA Extraction Solution (QE09050, Biosearch Technologies) from small pellets (less than 1x 10^6 cells) taking half of the culture volume of clones grown in 48-well plates. The clones we processed following the protocol provided by the supplier decreasing the resuspension volume to 50 µl. The extracts were quantified with the Qubit 4 fluorometer (Invitrogen) using the kit dsDNA Broad Range (Q32850, Invitrogen).

The sample preparation for the Nanopore long-read sequencing was performed starting from high-molecular weight genomic DNA extraction. The DNA for the characterization of the insertion site in the reference clone was extracted using the kit QIAGEN Blood & Cell Culture DNA kit (13362, QIAGEN). The extraction was carried out following supplier instructions for freshly harvested cells. The sequencing of the new CHO host cell line and the clones generated from the KO and the KI experiments were performed starting from DNA extracted with the Puregene Cell Kit (8 x 108) (158043, QIAGEN). Cell pellets (~4x 10^6 cells) were processed following the protocol for cultured cells provided by the vendor but decreasing the elution volume to 80 µl. All the extracts were then quantified with Qubit 4 fluorometer (Invitrogen) using the kit dsDNA High Sensitivity (Q32851, Invitrogen), checked for purity with Nanodrop and checked for the degradation/size with the 4150 TapeStation System (Agilent Technologies) using the gDNA analysis (5067-5365, Agilent Technologies).

3.5 PCR conditions

3.5.1 PCR for fusion sites analysis

PCR analysis of fusion sites was conducted using the FastStart High Fidelity PCR System kit (3553400001, Roche). Specific primers (sequences not reported) for each fusion site were added at 400nM to the FastStart™ High Fidelity Reaction Buffer with 18mM of $MgCl_2$ and 1,5 U of FastStart™ High Fidelity Enzyme Blend. 100ng of extracted DNA were amplified using the C1000 Touch Thermal Cycler (Bio-Rad) following the TouchDown PCR protocol listed in Table 1.

Table 1. Cycling conditions for TouchDown PCR.

Cycling step	Temperature (°C)	Time	Number of cycles
Enzyme activation	95	10min	1
Denaturation	95	1min	10
Annealing	66-61 (0,5°C decrease at each cycle)	45sec	
Extension	72	1min	
Denaturation	95	1min	25
Annealing	61	45sec	
Extension	72	1min	
Enzyme deactivation	72	10min	1
Hold	4	Infinite	1

The PCR amplicons were then loaded on 1% agarose gel (TBE) and 120V voltage was applied for the run. The resulting fluorescent bands were visualized with the GelDoc+ System (Bio-Rad) and molecular weights were defined with the ImageLab software 6.1 (Bio-Rad).

3.5.2 PCR for Junction analysis

Long PCR for junction analysis in KI samples was conducted using the PrimeSTAR® GXL Premix (R051A, Takara) suitable for amplicons up to 30kb and GC-rich sequences adding specific primers (listed in Table 2 and Table 3, sequences on the genome are not reported) for each junction at 400nM. Then, amplification was performed using the C1000 Touch Thermal Cycler (Bio-Rad) following the PCR protocols listed in Table 2 and Table 3.

Table 2. Cycling conditions of Long PCR for the following assay.

ID primer	Sequence 5'-3'	Amplicon length
Genome5'_A fwd	Not reported	2450 bp
PGK1_rv	AGAAAGCGAAGGAGCAAAG	

Cycling step	Temperature (°C)	Time	Number of cycles
Enzyme activation	98	30sec	1
Denaturation	98	10sec	30
Annealing	59	10sec	
Extension	68	2min	
Enzyme deactivation	68	5min	1
Hold	4	Infinite	1

Table 3. Cycling conditions of Long PCR for the following assay.

ID primer	Sequence 5'-3'	Amplicon length
Genome5'_B fwd	Not reported	4650 bp
GFP1_rv	GTTGCCGTCGTCCTTGAAGAAGATG	

Cycling step	Temperature (°C)	Time	Number of cycles
Enzyme activation	98	30sec	1
Denaturation	98	10sec	30
Annealing/Extension	68	3min	
Enzyme deactivation	68	5min	1
Hold	4	Infinite	1

The PCR amplicons were analysed with gel electrophoresis. Amplicons were loaded on 1% agarose gel (TAE) and 95V voltage was applied for the run. The resulting fluorescent bands were visualized with the GelDoc Go System (Bio-Rad) and molecular weights were defined with the ImageLab software 6.1 (Bio-Rad).

3.6 ddPCR evaluation of deleted or inserted constructs

The evaluation of generated pools and selected monoclonal population were performed using the QX200 Droplet Digital PCR system (BioRad) whereas the clone screening for the KI analysis was performed with the QXONE Droplet Digital PCR system (BioRad). The ddPCR reaction mix was prepared using the ddPCR Supermix for Probes (no dUTP) as manufacturer instruction (1863024, BioRad). Each set of primers was added at the final concentration of 900nM, while TaqMan Probes were added at 250nM as final concentration or 125nM where needed to perform amplitude triplex analysis. All the analyses were performed by measuring the concentration of the target assay/assays on the corresponding construct target in relation to the reference assay designed on beta2-microglobulin (B2M) gene. The sequences of the ddPCR assays are listed in Table 4 (sequences referred to our proprietary molecules are not shown).

Table 4. 5'-3' sequence of primer and probes for ddPCR experiments.

Experiment	Assay name	ID primer	Sequence 5'-3'	Fluorescent Probe
KO and KI	B2M reference gene	B2M for	ACCCACTGCGACCGATAAAT	
		B2M rev	GACCTTGGGCTCCTTCAGAGT	
		B2M probe	TGCCTGCAGAGTTACA	VIC
KO	LC reference clone	LC for	Not reported	
		LC rev	Not reported	
		LC probe	Not reported	FAM
KO	HC reference clone	HC for	Not reported	
		HC rev	Not reported	
		HC probe	Not reported	FAM
KI	PGK1 promoter	PGK1 for	CAGGACGTGACAAATGGAAG	
		PGK1 rev	AGAAAGCGAAGGAGCAAAG	
		PGK1 probe	TGCTGTCCATCTGCACGAGACT	FAM
KI	GFP gene	GFP for	CTGACCCTGAAGTTCATCTG	
		GFP rev	GTCGTCCTTGAAGAAGATGG	
		GFP probe	ACCACATGAAGCAGCACGACTT	FAM

The DNA of each sample (total amount of 20ng) weas added to the reaction mix with 2 Units of MseI restriction enzyme (BA0525L, NEB). For each reaction mix, a No Template Control was prepared adding molecular biology water instead of sample DNA. Reaction mixes and AutoDG Oil for Probes (1864110, BioRad) were prompted to create the water-oil emulsion using the Automated Droplet Generator (Bio-Rad). Droplets were then amplified using the C1000 Touch Thermal Cycler (Bio-Rad) using the protocol in Table 5. The annealing temperature was defined after a thermal gradient PCR reaction.

Table 5. Cycling conditions for ddPCR Supermix for Probes.

Cycling step	Temperature (°C)	Time	Number of cycles	Ramp rate
Enzyme activation	95	10min	1	2°C/sec
Denaturation	94	30sec	40	
Annealing/Extension	59	1min		
Enzyme deactivation	98	10min	1	
Hold	4	Infinite	1	

After thermal cycling, droplets were read using the QX200 Droplet Reader using the Direct Quantification (DQ) analysis on the QX Manager software. Where the QXONE instrument was employed, all the steps from the droplet generation to the fluorescence reading were performed automatically by the QXONE system itself. The results were extrapolated from the concentrations of the targets and/or the ratios between target assay and reference assay.

3.7 Fluorescence evaluation

The GFP fluorescence after transfection and in the phase of clone generation was monitored with the EVOS M5000 microscope (Invitrogen). The quantification GFP fluorescent cells of the direct KI approach was performed using flow cytometry with FACSAria cell sorter (BD - Becton Dickinson) processing ~1 x10^6 cells for each tested condition.

3.8 Nanopore sequencing

To perform Nanopore sequencing the DNA extracts underwent size selection to decrease the presence of low molecular weight DNA fragments. The size selection was performed with the BluePippin System (Sage Science) using a pre-casted 0,75% agarose dye-free gel cassette associated with the external marker "Marker S1" (PAC20KB, Sage Science). The runs were performed by selecting the 0,75% DF Marker S1 high pass 6-10 kb vs3 protocol and loading 5 µg of samples/well. The eluates were purified with AMPure XP beads (Beckman Coulter) to substitute the BluePippin elution buffer. For each sample, two independent extractions were size selected and then pooled together at the end of the purification step. Samples were quantified with Qubit 4 fluorometer (Invitrogen) using the kit dsDNA High Sensitivity (Q32851, Invitrogen) and checked for the size with the 4150 TapeStation System (Agilent Technologies) using the gDNA analysis (5067-5365, Agilent Technologies) before proceeding with the library preparation.

For the library preparation, the SQK-LSK109 or SQK-LSK110 ligation sequencing kits (Oxford Nanopore Technologies) were used following the Oxford Nanopore protocol for gDNA, enriching for long fragments. Finally, the libraries were quantified with the Qubit 4 fluorometer (Invitrogen) using the kit dsDNA High

Sensitivity to calculate the resulting fmol based on an approximative N50 of 20kb (this value was extrapolated from other samples processed with the same protocol). For all the performed runs, flow cells R9.4.1 (FLO-MIN106 Oxford Nanopore Technologies) were loaded with ~50fmol of sample and then the sequencing was performed on the MinION (Oxford Nanopore Technologies) with the super-clone sample and on the GridION (Oxford Nanopore Technologies) with all the other samples. For the runs on the GridION the base calling was set on the super-accurate mode to ensure high-quality reads (QScore >10). The reads were filtered with in-house developed pipelines and aligned with CHO genome CriGri_PICRH-1.0 (GCA_003668045.2) or specific vectors employed in the different experiments. IGV software was employed to visualize the obtained data and perform a deep analysis of the deletion/integration events. The online tool BLAST was included in the analysis of the sequencing results to better characterize specific reads and highlight possible similarities with respect to different reference sequences.

4. Results

4.1 Knock-Out assessment

4.1.1 Characterization of the insertion site in a selected reference clone

In the feasibility study for the employment of CRISPR/Cas KO, we started from the characterization of the insertion site in one of our expressing-mAb clones selected for its high productivity, which we intended to use as reference clone. This clone was generated with the classical random integration approach and was already characterized for GCN with ddPCR as well as for the insertion locus with TLA-seq analysis. The TLA-seq identified a single integration locus genome-wide and generated five fusion sequences, that are the regions where the host genome and the mAb-expressing plasmid result to be contiguous. In a single copy integration event, two fusion sites should be detected flanking the integrated DNA. It is not unusual that in a single integration site, multiple copies of plasmid have been integrated, so multiple fusion sites could be identified. However, an odd number of fusions is not a reliable result since a fusion site should be always detected at the beginning and end of every portion of integrated DNA. So, we started the investigation of the locus by confirming the fusion sites with specific PCR to exclude false positive results of the TLA analysis and to determine the boundaries of the real integration. The results of PCR designed for the five fusion sites are shown in Figure 9. We confirmed Fusion2, Fusion3, and Fusion4 without non-specific signals in the clone or the CHO host cell line.

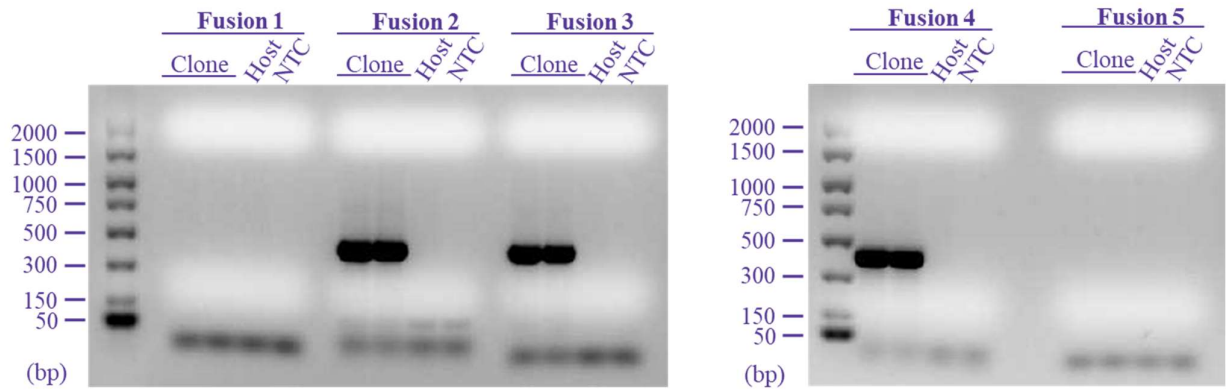


Figure 9. **Validation of the five fusion sequences by PCR**. The DNA of the clone and the CHO host cell line was amplified with specific primers designed on the five sequences provided by the TLA-seq analysis. The expected amplicon size of Fusion1, Fusion2, Fusion3, Fusion4 and Fusion5 was 348 bp, 423 bp, 405 bp, 410 bp, and 647 bp, respectively. The DNA of the reference clone was correctly amplified only with the Fusion2, Fusion3, and Fusion4 assays. All the employed primers were specific as no amplification occurred in the host CHO DNA. The no template controls (NTC) gave no amplifications.

Since the result of the PCR gave again an odd number of fusion sequences, we decided to change the approach and perform a whole-genome sequencing with the long-read Nanopore technology. Two independent extractions with the Blood & cell culture DNA kit were used for the library preparation after the size selection for high-molecular weight fragments. Then, the two libraries were loaded into distinct flow-cells and sequenced on the MinION instrument. At the end of the first run, the instrument reported that were sequenced 11,2 Gigabases (Gb) counting 521,2 k reads, thus calculating an average read length of 33,1 kb. For the second run, the output was slightly lower with 8,2 Gb and 341,7 k reads, but the reads were longer, with an average of 41,5kb. Overall, the generated output together with the length of the sequenced reads were considered enough to reconstruct the sequence of the integration site. All the obtained reads were selected for the ones that aligned to the plasmid sequence or to the genomic region identified by TLA analysis and finally used to perform a de novo assembly with a mean coverage on the target sequence of 30x. The obtained sequence was then characterized performing BLAST analysis to uncover how the plasmid, genome, and the five fusions were located within the generated assembly. The results were visualized with a Circos plot (Figure 10).

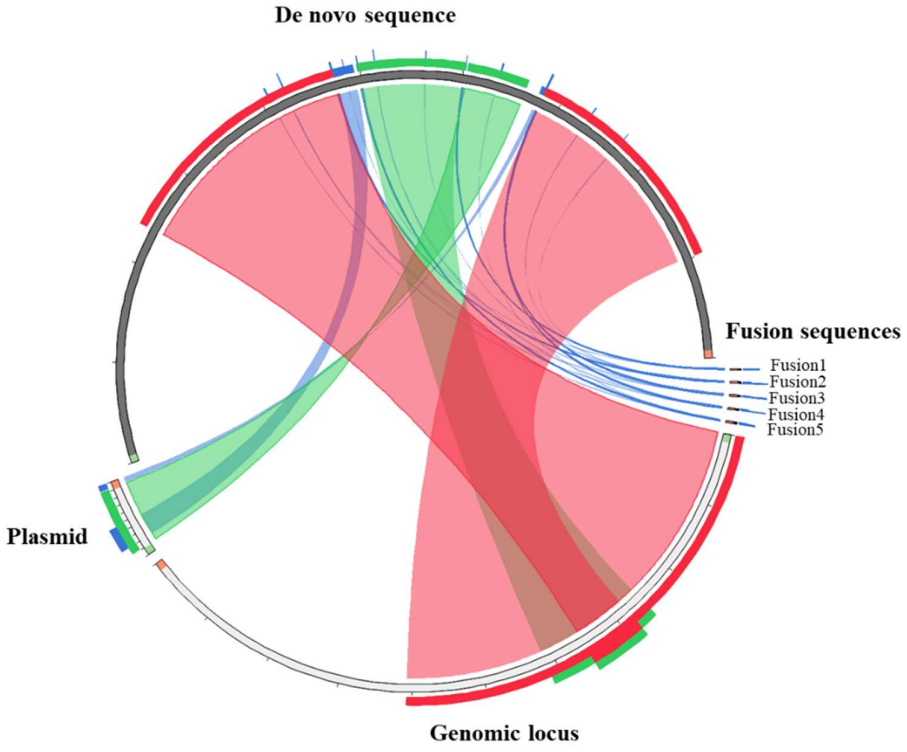


Figure 10. **De novo assembly characterization of the reference clone integration site starting from the Nanopore sequencing data.** The Circos plot (https://bat.infspire.org/circoletto) showed the alignments between the de novo assembly and the genomic locus identified with TLA-seq, the transfected plasmid and the five fusion sites.

The results of our analysis revealed a more complicated conformation of the integration site than the expected linear integration (as shown in Figure 7, paragraph 2). We identified three distinct plasmid insertions interspaced with genomic sequence. Starting from the 5' of the genomic sequence (in grey in Figure11) we spotted: a partial plasmid integration (3596 bp), a genomic region which seemed duplicated (15523 bp), a complete integration of the plasmid with an inversion (8932 bp) and a small, inverted portion of the plasmid (687 bp). The latter were separated by an unknown sequence (1995 bp) which did not give any significant similarities to any sequence in the CHO genome or plasmid. We speculated that these nucleotides were inserted during the random rearrangements of the locus (Figure 11). Moreover, we were able to explain the number and the previous results on the fusion sites. We discovered that the entire integration was delimited by Fusion2 and Fusion3 which were also pinpointed by the TLA analysis and the consequent confirmation by PCR. Although the PCR could not confirm Fusion1 and Fusion5, the de novo assembly revealed that the two sequences were located at the side of the genomic sequence between the partial and the complete plasmid integration. The failed PCR amplification could be caused by the suboptimal efficiency of the designed primers. Indeed, comparing the fusion sequences generated from the TLA-seq, used for the primer design, and the sequence obtained from the Nanopore sequencing, we observed some differences. Since the design was performed on a slightly different sequence this might have impaired the efficient amplification of the target. Regarding the Fusion4 that was confirmed by the PCR, we could not find any consistent alignment on the de novo assembly sequence, so we hypothesize the presence of a genomic region with a high grade of homology with the Fusion4 sequence identified by TLA (Figure 11).

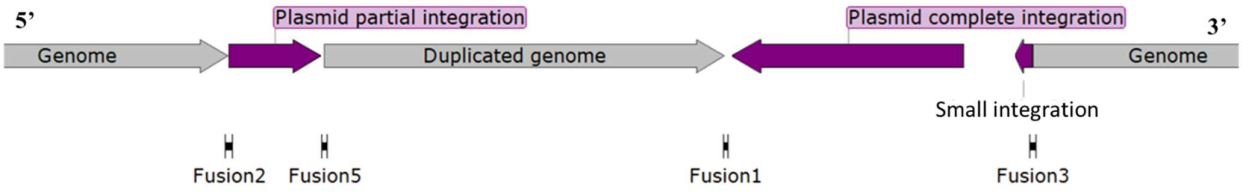


Figure 11. **Characterization of the integration site of our reference clone starting from Nanopore sequencing data.** A graphic representation of the obtained positions of plasmid sequences in the identified site. Starting from the 5' of the genomic sequence (in grey) we find a partial integration of the plasmid (in purple), a duplicated genomic sequence (in grey), a complete integration of the plasmid, and a small integration (both in purple) separated by un unknown sequence (empty space). The scheme shows also the fusion sequences highlighted in the TLA-seq analysis.

In a second instance, we conducted further analysis on the three plasmid integrations, examining the correspondences to the various plasmid components, as well as the interspaced genome sequences. Regarding the integrated constructs, we discovered that the complete integration comprised all the coding sequences, meaning that both the light and the heavy chains of the coding mAb were present with their respective promoters, as depicted in the upper part of Figure 12. The upstream partial insertion instead included the incomplete sequence of the light chain without its promoter (upper part of Figure 12). Finally, the small integration was found to be a portion of the plasmid backbone, particularly the antibiotic resistance used for the cloning step in bacteria (upper part of Figure 12).

Regarding the genomic sequence, we uncovered some large duplications of the original locus as depicted in the lower part of Figure 12. In fact, we could divide the parental genomic locus into three parts: A (2435bp), B (7357bp) and C (5728bp) as indicated in the lower part of Figure 12. Certainly, during the integration of the plasmid some rearrangements occurred resulting in a copy-and-paste of these three portions of the genome. We found that region A (deep orange in Figure 12) was present twice, at the 5' and at the 3' of the partial integration of the plasmid. The B segment (orange in Figure 12) was identified three times in the assembly: at the 5' of the partial integration, in the middle of the partial and the complete plasmid sequences and at the 3' of the small integration. The region C (yellow in Figure 12) was present twice: a the 5' of the complete integration and at the 3' of the small plasmid backbone integration.

In the end, the Nanopore sequencing combined with subsequent analysis gave us a real and accurate characterization of the plasmid integration site, resolving challenging locus conformations.

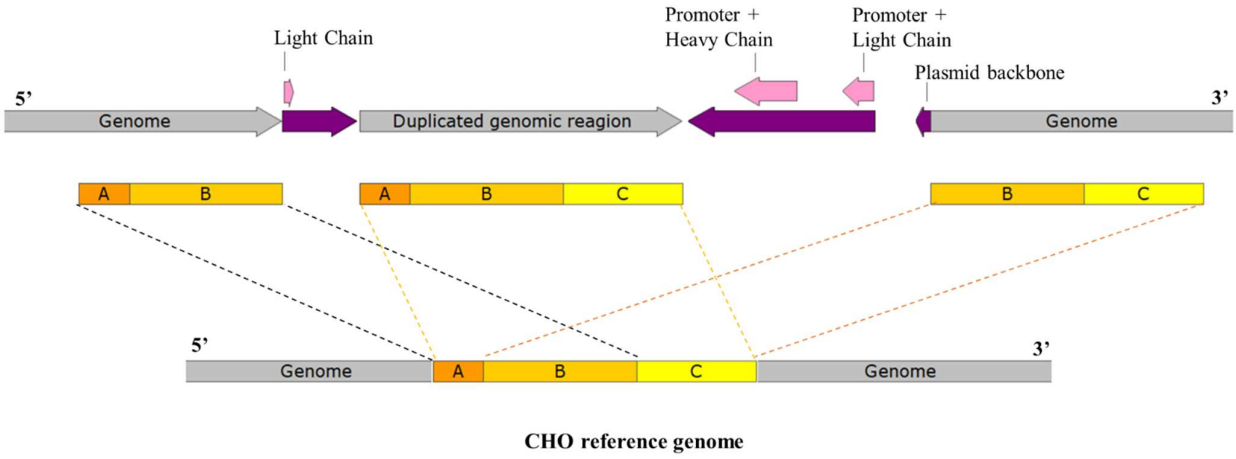


Figure 12. **A deep characterization of the genomic and plasmid sequences is depicted in the scheme.** In the upper part, the coding sequences of the plasmid are highlighted in pink, whereas in the lower part, the duplications of the original genomic locus are clarified.

4.1.2 crRNA design for the Knock-Out experiment

Considering the complexity of the integration site uncovered by the Nanopore sequencing, we wondered which was the best way to eliminate all the 38,5 kb integrated sequence considering all the existing homologies along the integration locus. We evaluated different strategies but, in the end, we found out that the most effective system to perform a clean deletion, restoring de original sequence, was to design the crRNAs on the genomic region that was found triplicated (B fragment in Figure 13). In this way, the same RNP complex could cut at both sides of the integration, enabling the re-joining of the free-ends that in the original sequence were found to be contiguous. Of course, a third cleavage could be generated on the B fragment in the middle of the integration, and its effect on the efficiency had to be assessed (Figure 13).

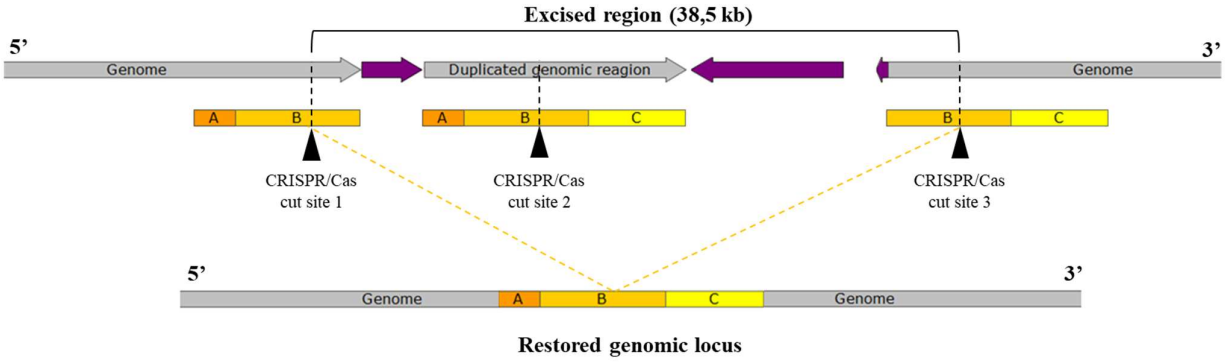


Figure 13. Representation of the KO strategy. All the designed crRNA for both Cas9 and Cas12a were designed along the B fragment to perfectly restore the genomic region after the rejoining of the free ends generated by CRISPR/Cas cleavage on sites 1 and 3. Since region B is found triplicated, an additional cut site (CRISPR/Cas cut site 2) was present in the middle of the region to be excised.

We took advantage of the online tool CHOPCHOP to design crRNAs for Cas9 and Cas12a. The tool generated a list of candidate crRNAs within the given sequence with information about CG%, self-complementarity, putative off-target (genomic regions with 0, 1, 2, or 3 mismatches), and predicted efficiency. Moreover, these parameters were used by the software itself to rank the guide-RNAs, thus helping the user in selecting the best CRISPR assays. In our study, we chose six crRNAs for the Cas9 and six for the Cas12a picking those with the highest CHOPCHOP ranks and the highest efficiency. The list of crRNAs with relative features that were tested is shown in Table 6. In general, we avoided selecting crRNAs with 1 or 2 mismatches on predicted off-targets (MM1 and MM2 in Table 6) and checked that only 1 hit was present in the predicted off-target with no mismatches (MM0 in Table 6). In fact, the target region was present in the CHO genome, so the tool recognized the on-target as a perfect match off-target. Furthermore, we decided to include some crRNAs with higher

predicted efficiencies but with more off-targets bearing 3 mismatches (MM3 in Table 6) possibly in the seed region, which actually confers the specificity to the crRNA.

Table 6. Features of selected crRNAs for Cas9 and Cas12a proteins.

crRNA	CHOPCHOP Rank	GC (%)	Self-compl	MM0	MM1	MM2	MM3	Efficiency
Cas9#1	1	35	0	1	0	0	1	61
Cas9#2	2	45	0	1	0	0	1	51
Cas9#3	6	40	0	1	0	0	3	56
Cas9#4	238	55	3	1	0	0	18	78
Cas9# 5	42	55	1	1	0	0	7	71
Cas9#6	15	40	0	1	0	0	4	66
Cas12a#1	1	46	1	1	0	0	0	90
Cas12a#2	2	54	0	1	0	0	0	81
Cas12a#3	4	54	0	1	0	0	0	80
Cas12a#4	5	46	1	1	0	0	0	80
Cas12a#5	6	50	0	1	0	0	0	78
Cas12a#6	307	38	0	1	0	0	0	98

4.1.3 Cas9 and Ca12a Ribonucleoprotein complexes transfection

Two independent transfections were carried out to test the Cas9 crRNAs and the Cas12a crRNAs. In both experiments, all the RNP complexes were transfected in duplicate with two different electroporation programs suitable for CHO cells, DT-133 and DU-158. The aim was to identify the transfection protocol with less impact on viability and cell growth while ensuring effective gene editing. Cells were let recover after electroporation and at 7 days post-transfection (DPT) they were analysed with the Vi-Cell BLU instrument to assess viability and viable cell number.

For the Cas9 RNPs, we found high viability for both programs with no significant difference as a mean of 92,7% (±2,22% considering the Standard Deviation) was calculated for the DT-133 program and a mean of 92,4% (±1,81% considering the Standard Deviation) was calculated for the DU-158. The same was for the cell counts which revealed no differences between the two protocols with $8{,}7 \times 10^6$ $\pm 2{,}1 \times 10^6$ (mean ± Standard Deviation) and $8{,}0 \times 10^6$ $\pm 3{,}0 \times 10^6$ (mean ± Standard Deviation) total viable cells respectively (Figure 14A and 14B).

Regarding the Cas12a RNPs, the statistical analysis showed a significantly higher viability of the cells transfected with DT-133 even if the obtained viabilities were close to each other, 96,6% (±0,9% considering mean ± Standard Deviation) for DT-133 and 95,2% (±0,8% considering mean ± Standard Deviation) for the DU-158. Indeed, the higher performances of the DT-133 program were confirmed by comparing the total viable cell counts: the mean of total viable cells transfected with DT-133 reached the $11{,}6 \times 10^6$ $\pm 2{,}9 \times 10^6$ (mean ± Standard Deviation) whereas the mean of those transfected with DU-158 reached the $8{,}3 \times 10^6$ $\pm 1{,}1 \times 10^6$ (mean ± Standard Deviation) (Figure 14C and 14D).

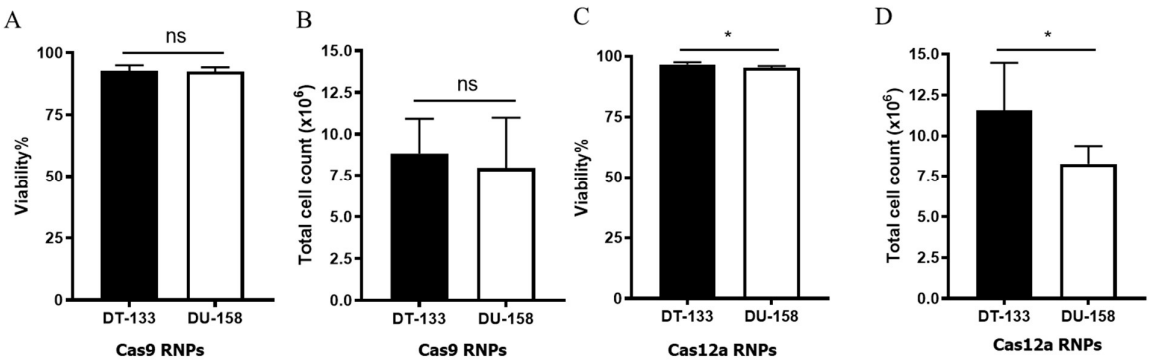


Figure 14. Viability and cell growth analysis on transfected pools with different electroporation programs 7 days after transfection. In the left part of the panel, the results for the pools transfected with Cas9 RNPs using the DT-133 or the DU-158 program are shown. The pools were analysed with the Vi cell BLU and were compared for viability (A) and the total number of viable cells (B). In the right part of the panel, the results for the pools transfected with Cas12a RNPs using the DT-133 or the DU-158 program are shown. The pools were analysed with the Vi cell BLU and were compared for viability (C) and the total number of viable cells (D). *Data show mean ± Standard Deviation (test T, ns P > 0,05; * P < 0,05).*

Then, the gene editing efficiency in the transfected pools was assessed with duplex ddPCR analysis measuring the LC or the HC together with the reference gene B2M. The resulting target/reference ratio values were compared to those of the non-transfected reference clone to evaluate the percentage of reduction for both targets. We chose to use the ratio instead of the gene copy number as a measurement parameter because we analysed heterogeneous samples with modified and non-modified cells. The data obtained for the Cas9 transfected pools (Table 7) and the data obtained for the Cas12a-transfected pools (Table 8) were employed to evaluate differences between electroporation programs and to identify the RNP complex with the best performances in terms of reduction of LC and HC signal. Regarding the Cas9 transfected pools, we observed a reduction in both LC and HC with all the crRNAs and no significant differences were highlighted between the same RNP complex transfected with the DT-133 or the DU-158 electroporation program. This data confirmed that both protocols can be efficiently used with our cells transfected with CRISPR/Cas9 assays. For the Cas12a pools, we found a strong reduction of LC and HC values with all the crRNAs except for the crRNA Cas12a#1 which showed ratio values comparable to those of the reference clone. Also in this case, the same RNP transfected with DT-133 or the DU-158 electroporation program showed similar reductions, meaning that both protocols were able to efficiently deliver the CRISPR/Cas12a complexes. Overall, we obtained good results with all conditions, but in the end, the DT-133 program seemed to have less effect on the viability and growth rate of the cells at least with the Cas12a-transfected pools.

Table 7. Ratio values of Cas9-transfected pools obtained with ddPCR. The values are compared with the non-transfected reference clone (green row) to calculate the reduction percentage of Light and Heavy Chains. In the analysis also controls of Cas9 without any crRNAs are included as negative controls (grey rows).

Transfected pool	LC/B2M ratio	HC/B2M ratio	Reduction of LC (%)	Reduction of HC (%)
DT-133 Cas9#1	1,9	1,4	48,8	39,2
DT-133 Cas9#2	2,0	1,5	45,1	36,7
DT-133 Cas9#3	1,8	1,4	51,9	41,1
DT-133 Cas9#4	1,8	1,5	50,0	39,0
DT-133 Cas9#5	1,9	1,4	48,0	40,3
DT-133 Cas9#6	1,9	1,4	49,7	40,9
DT-133 Cas9 ctrl	3,6	2,6	-	-
DU-158 Cas9#1	1,9	1,5	47,4	38,0
DU-158 Cas9#2	1,9	1,5	47,7	35,4
DU-158 Cas9#3	1,9	1,4	49,6	39,9
DU-158 Cas9#4	1,9	1,5	48,6	38,8
DU-158 Cas9#5	1,9	1,5	48,6	38,8
DU-158 Cas9#6	1,9	1,5	49,3	38,0
DU-158 Cas9 ctrl	3,6	2,4	-	-
Reference Clone	**3,7**	**2,4**	-	-

Table 8. Ratio values of Cas12a-transfected pools obtained with ddPCR. The values are compared with the non-transfected reference clone (green row) to calculate the reduction percentage of Light and Heavy Chains. In the analysis also controls of Cas12a without any crRNA are included as negative controls (grey rows).

Transfected pool	LC/B2M ratio	HC/B2M ratio	Reduction of LC (%)	Reduction of HC (%)
DT-133 Cas12a#1	3,8	2,5	7,9	8,1
DT-133 Cas12a#2	1,4	1,2	65,7	54,5
DT-133 Cas12a#3	1,6	1,3	61,3	54,1
DT-133 Cas12a#4	1,5	1,3	64,1	53,2
DT-133 Cas12a#5	1,7	1,4	58,7	49,5
DT-133 Cas12a#6	1,4	1,2	66,9	55,8
DT-133 Cas12a ctrl	3,8	2,5	-	-
DU-158 Cas12a#1	3,9	2,6	7,2	6,6
DU-158 Cas12a#2	1,5	1,3	65,3	53,9
DU-158 Cas12a#3	1,7	1,4	60,5	50,6
DU-158 Cas12a#4	1,4	1,2	67,5	56,9
DU-158 Cas12a#5	1,8	1,4	57,9	50,1
DU-158 Cas12a#6	1,3	1,2	68,0	56,9
DU-158 Cas12a ctrl	3,8	2,5	-	-
Refrence Clone	**4,1**	**2,7**	-	-

For this reason, we decided to focus our attention on samples transfected with the DT-133 electroporation program to identify the best CRISR/Cas assay. In Figure 15A, the obtained values of LC/B2M and HC/B2M ratios of Cas9 and Cas12a pools are plotted together with the ratios of the reference clone (see also Table 7 and Table 8 for the ratio values and relative reductions). We found that overall, the Cas12a RNPs (excluding the sample crRNA Cas12a#1) worked better with a mean reduction of 53,4% and 63,4% of HC and LC respectively compared to the 39,5% and 48,9% reduction of HC and LC in the Cas9 transfected pools (see

Figure 15B). Among all, the Cas12a crRNAs we identified Cas12a#6 as the best one with the highest reduction of both HC and LC.

Before proceeding with the single-cell cloning we monitored the ratios of the pool Cas12a #6 with ddPCR till 65 DPT to confirm the stability of the reduction (Figure 15C, left panel). Indeed, at 42 DPT the signal of both HC and LC decreased even more, reaching a reduction of 66% and 81% which were maintained at 65 DPT (Figure 15C, right panel). So, we could conclude that the Cas12a#6 pool was stable for the KO.

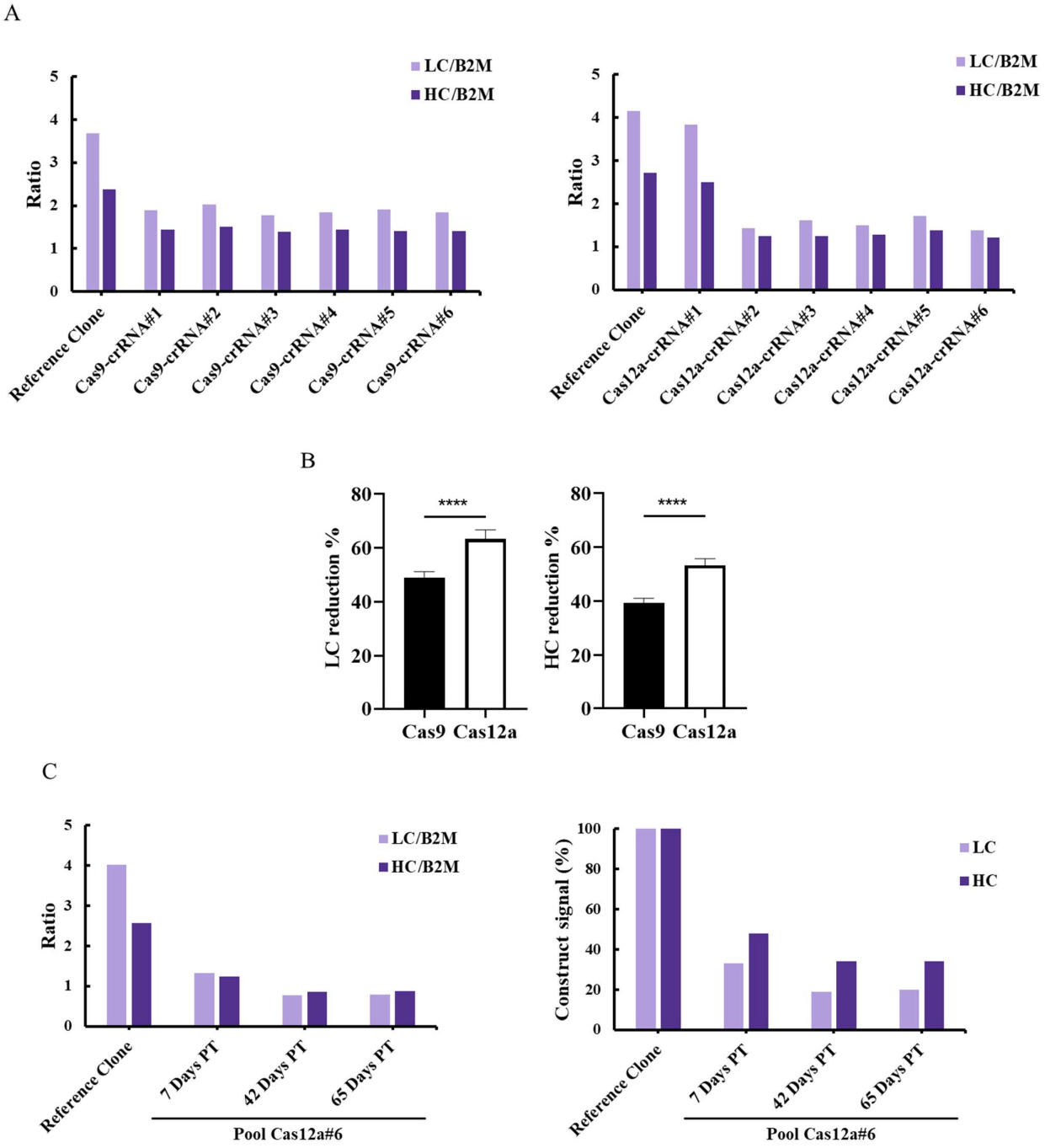


Figure 15. ddPCR analysis on transfected pools. (A) DNA from all the pools was extracted at 7DPT and analysed with ddPCR. LC/B2M and HC/B2M ratios of all the Cas9 and Cas12a RNP complexes are shown in comparison to the reference clone. (B) The mean reduction percentage of Cas9 and Cas12a assays are shown for both LC and HC. Data are shown as mean *mean ± Standard Deviation (test T, **** P < 0,0001)*. (C) The selected Cas12a #6 pool was monitored at different time points: 7 DPT, 42 DPT, and 65 DPT. The LC and HC ratios obtained during time are plotted with the reference clone. In the right panel, the results are plotted considering the ratios of the reference clone as 100% to evaluate the LC and HC reduction expressed as a percentage.

4.1.4 Knock-Out clone generation and screening

Since the pool Cas12a #6 showed a stable reduction of both LC and HC we proceeded with the single-cell cloning phase to generate monoclonal populations. The cloning conditions for our cells were optimized by testing two types of dilution methods. We performed four rounds of single-cell cloning and clones generated from each one were expanded and then screened with ddPCR. In Round 1 we tested the array dilution method that implies 1:2 dilutions of the input cells directly in the 96-well plates in a way that wells on the plate diagonal carry just one cell. This method succeeded in generating viable clones, but the total number of clones per plate was too low to be used for a consistent screening (7 clones in one 96-well plate). So, we decided to test a limiting dilution approach for the Round 2 and Round 3 diluting the cell suspension in a way that 1 cell/well was plated. With this approach, we were able to increase the number of clones (~20 clones in one 96-well plate), but a high number of wells contained 2 or more distinct colonies which were not suitable for the screening. Moreover, we could not exclude that some of the analysed clones were derived from two or more cells that were too close to be seen as distinct clusters. For this reason, in Round 4 we decided to dilute the cell suspension to 0,5 cell/well and increase the number of plates to have a reasonable number of clones. Around 15 clones were generated for each plate with very few wells bearing 2 or more clusters, so we decided to select this condition as the one with the best outcome.

We let the clones grow and once they were plated in 24-well plates the DNA was extracted with the QuickExtract solution to be analysed with ddPCR to identify clones 100% KO for both LC and HC. Since we were comparing monoclonal populations, we considered the values of copy number of LC and HC taking as reference gene B2M. In Figure 16 the copy number of the two targets in each of the tested clones is reported. Overall, 185 clones were tested, and we identified 4 KO clones that showed no signal for both LC and HC: R1_C5, R4_P1B4, R4_P2D1 and R4_P7A1.

Interestingly, we noticed that the KO clones were generated in the cloning Round 1 and Round 4 that are the ones with a cloning protocol that gave higher confidence in having just one cell/well.

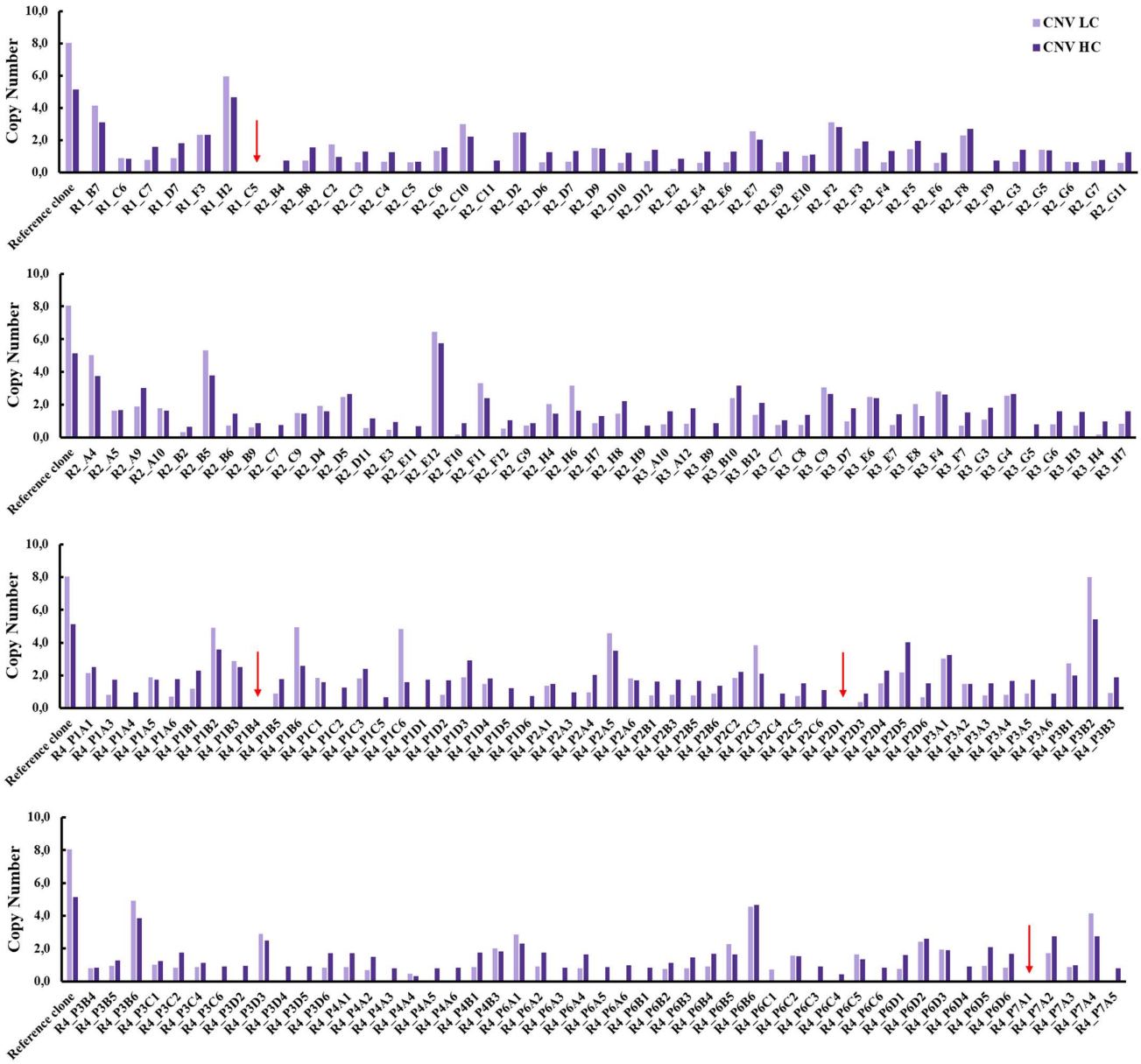


Figure 16. ddPCR analysis on generated clones. The copy number for LC and HC are reported for every clone together with the reference clone. Clones are named after their relative round of cloning as R1, R2, R3 or R4. The clones 100% KO are marked with a red arrow.

4.1.5 Characterization of R1_C5 KO clone

We decided to better characterize one of the identified KO clones in order to demonstrate the real deletion of the constructs. We chose the R1_C5 and we performed the MSX resistance test as further evidence of the deletion after the expansion in shaking flasks. Actually, the construct inserted in the reference clone coded also for an extra copy of the glutamine synthetase (GS), conferring to the cells the resistance to the MSX which is able to block the activity of the endogenous glutamine synthetase. In this way, we expected that the MSX treatment in the KO clone would lead to an almost complete death of cells. Then, we treated the clone R1_C5 and the non-transfected reference clone with a final concentration of MSX equal to 25µM. The cells were evaluated for their growth and viability for 14 days together with the relative untreated controls. As shown in Figure 17A, the viability of the C5 KO clone treated with MSX started decreasing from day 2 reaching 7% of viability at day 14. On the contrary, the reference clone subjected to the same treatment maintained the same values as the non-treated controls (>95%). The MSX treatment had also an important effect on the cell growth of the C5 KO clone: the cells were seeded at 0,5 x10^6 cells/mL and did not grow during the entire window of treatment. On the contrary, the reference clone treated with MSX showed a normal growth profile similar to the controls (Figure 17B). These data demonstrated that the R1_C5 clone lost the resistance to the MSX implying that the construct coding for the exogenous GS was no more present in the genome of cells.

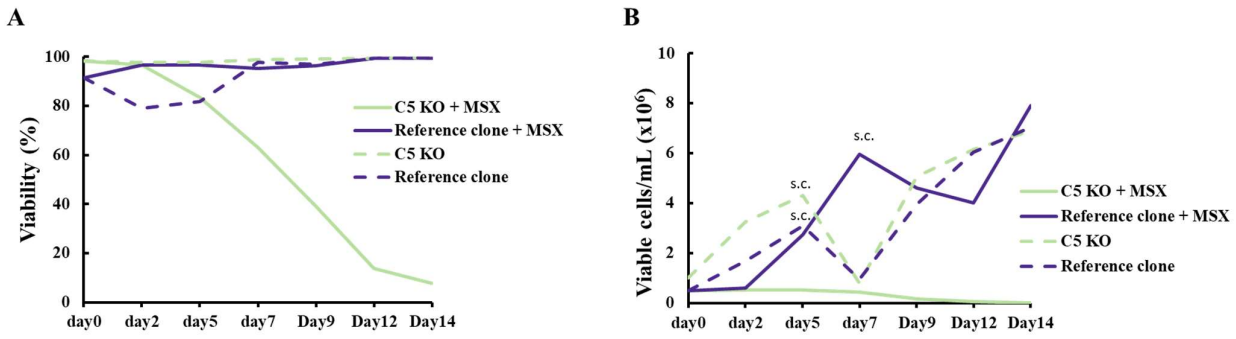


Figure 17. MSX resistance experiment. (A) The viability of cells was evaluated every 2 days. The controls without the MSX treatment (dashed green line for the C5 KO clone and violet dashed line for the reference clone) showed a viability that was stable and higher than 95%. The reference clone treated with MSX (violet line) showed a slight decrease in viability after 2 days, but it returned stably over 95% after 7 days, meaning that the cells were resistant to MSX. On the contrary, the viability of the C5 KO clone (green line) treated with MSX dropped to 7% after 14 days. (B) The viable cell concentration was monitored every 2 days. Both non-treated controls (dashed green line for the C5 KO clone and violet dashed line for the reference clone) needed a sub-culturing (s.c. in the graph) after 5 days and the same was for the MSX-treated reference clone (violet line) after 7 days. The C5 KO clone treated with MSX (green line) did not show any increase in the cell count and after 7 days almost all the cells died.

However, sequencing analysis remains the gold standard to really prove the success of a genetic modification. For this reason, we decided to take advantage of the long-read Nanopore technology to confirm that all the inserted constructs were deleted from the target locus and to see the rearrangement of the sequence after the deletion.

The DNA of the C5 KO clone was extracted with the Puregene Cell Kit and, after the size selection step, the library was prepared. The library was loaded on the GridION to perform a single run setting the basecalling as "super-accurate", meaning that the reads were immediately filtered for the Quality score (>10). At the end of the run, the instrument reported that 3,3 Gigabases (Gb) were sequenced counting 335,4 k reads, thus calculating an average read length of 15,8 kb.

The generated reads were aligned to the sequence of the parental reference clone and the same locus in the *Cricetulus Griseus* genome (CriGri-PICRH-1.0). Starting from the analysis with the reference clone locus, we found that there was a drop in the coverage (5x on average along the entire sequence) exactly at the regions spanning the two partial and the complete integration of the plasmid (Figure 18). We highlighted some small reads (shorter than 400bp) in these gaps which were further investigated (red asterisks in Figure 18).

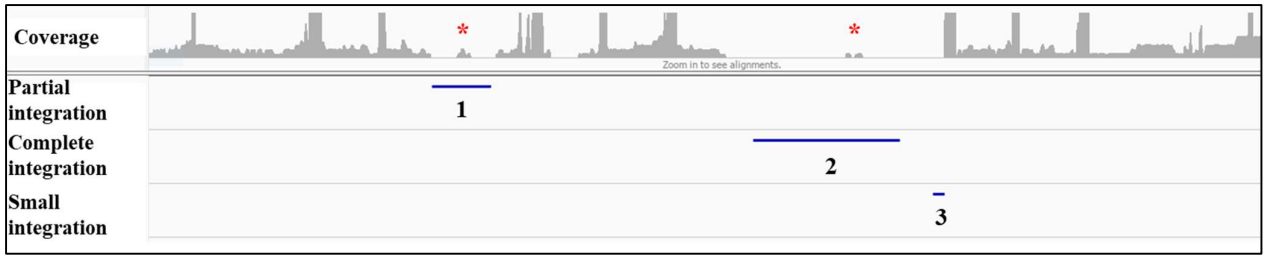


Figure18. IGV visualization of the sequencing data aligned on the reference clone locus. The image was taken from IGV software in which the sequence of the reference clone was loaded as reference genome. The coverage plot showed two significant drops. From right to left, represented with blue lines in the plot, we had a drop at the partial integration site (1) and a drop spanning all the region between the complete (2) and the small integration (3). From this analysis were also visible some hits mapping on plasmid sequences (pointed out by red asterisks).

We checked the reads sequences on our plasmid, and we discovered that the reads mapped on the GS transgene. So, we carried out a NCBI-BLAST analysis using the selected reads and limiting the field "organism" to *Cricetulus Griseus* genome. Reads showed a high percentage of identity (>95%) to the endogenous GS gene (Table 9), so we could conclude that those reads were just an artifact of the alignment due to the homologies between the endogenous GS and the GS gene on the plasmid. In the end, our analysis pointed out that all plasmid sequences were eliminated from the targeted locus.

Table 9. BLAST results of reads mapping on plasmid insertions.

Read ID: 4308b9df-e834-4d4a-8c52-930f6cc93e87 (373bp)		
Description	**Query Cover**	**Per. ident**
Cricetulus griseus Glul pseudogene (LOC100689337) on chromosome 1	100%	97,4
Cricetulus griseus glutamine synthetase mRNA complete cds	100%	96,8
Cricetulus griseus glutamine synthetase (GLUL) mRNA complete cds	100%	95,2
Cricetulus griseus glutamate-ammonia ligase (Glul) mRNA	100%	95,2
PREDICTED: Cricetulus griseus glutamate-ammonia ligase (Glul) transcript variant X3 mRNA	100%	95,2
Read ID: 95f09f35-b63f-4d83-8fa3-67c4ce125d37 (293bp)		
Description	**Query Cover**	**Per. ident**
Cricetulus griseus Glul pseudogene (LOC100689337) on chromosome 1	100%	97,0
Cricetulus griseus glutamine synthetase mRNA complete cds	100%	97,0
Cricetulus griseus glutamine synthetase (GLUL) mRNA complete cds	100%	95,0
Cricetulus griseus glutamate-ammonia ligase (Glul) mRNA	100%	95,0
PREDICTED: Cricetulus griseus glutamate-ammonia ligase (Glul) transcript variant X3 mRNA	100%	95,0

As mentioned before, the reads generated from the Nanopore sequencing were also aligned on the targeted genomic locus in the reference *Cricetulus Griseus* genome, which was supposed to be similar to the re-stored sequence of the KO. We visualized the output on IGV software, and we focused our attention on the re-joining site, corresponding to the Cas12a crRNA#6 sequence (8x coverage). We noticed that all the reads mapped across the cut site, without any interruption. This means that there were not reads mapping only for a portion on that specific genomic region and continued with a sequence which was not present in the reference (e.g., the plasmid integrations), confirming that the re-joining of the predicted free-ends occurred in the R1_C5 KO clone and that the constructs inserted were completely deleted (Figure 19).

From this analysis we also had information about the sequence of the re-joining. Indeed, we found in all the reads some single-nucleotide mutations and some small insertions exactly at the cutting site of the employed RNP complex. Moreover, a more consistent base insertion was found at the 5' of the Cas12a crRNA#6 sequence (Figure 19).

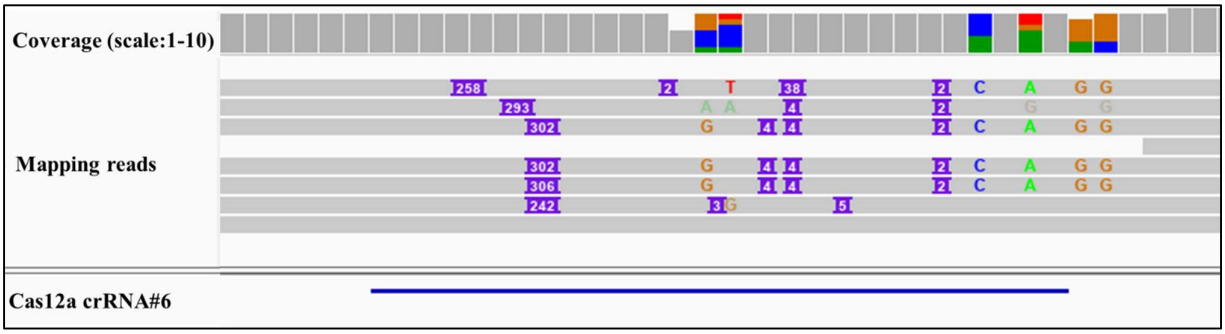


Figure 19. IGV visualization of the sequencing data aligned on the putative restored sequence. A zoomed window on IGV displayed the region targeted by the RNP complex Cas12a#6 (blue line). Empty grey reads showed perfect

matches against the sequence of the reference genome whereas purple boxes and coloured nucleotides indicate bases insertions (the number inside the box referred to the number of inserted bases) and nucleotide mutation respectively.

We further analysed these insertions and we found that the sequence of 300bp was almost the same in all reads, meaning that they were not an artifact of the sequencing. Then, we aligned these sequences to all the reference clone integration locus to see if it could be a residue region that was not eliminated by the KO. In the end, this analysis did not reveal any homologies between the read insertions and the plasmid of the reference clone, but there were high similarities (83-99% identities) to a genomic region at the 5' of all the locus involved in the reference clone integration. For this reason, we can speculate that during the re-joining this region was used by the repair machinery as a template to mend the break in the genome.

Overall, the Nanopore sequencing of the R1_C5 KO clone and the subsequent analysis showed that all the inserted plasmid sequences together with the duplicated genomic regions were correctly eliminated at the targeted locus which showed the expected re-joining even if some recurrent mutations occurred.

4.2 Knock-In assessment – Direct KI approach

4.2.1 Sequencing of new host cell line (in-house optimized CHO-K1)

Considering the reliable results that we obtained with the Cas12a #6 RNP for the KO assessment, we decided to use the same crRNA-Cas12a for the KI part. Since we changed the host CHO cell line for the KI experiments employing our in-house optimized CHO-K1 host cells, we proceeded with the Nanopore sequencing of the new cells to check whether the target sequence of crRNA#6 remained the same. For this reason, the DNA of progenitor CHO-K1 (ATCC) cells was collected by our colleagues in Vevey and directly used for the library preparation for long fragments sequencing. Here, the size selection with the BluePippin was not performed because we were interested in a small region of the genome (less than 1 kb around the crRNA#6 sequence), and we preferred to maintain smaller fragments to increase the depth of coverage. The library was loaded on the GridION to perform a single run setting the base calling as "super-accurate" (Q score >10). At the end of the run, the instrument reported that were sequenced 8,6 Gb counting 1,1 M reads, thus calculating an average read length of 20,2 kb. At this point, the reads were aligned to the genomic locus identified in the KO experiments, focusing on the crRNA#6 target sequence. The analysis with IGV showed a total coverage of 6x and even if some reads showed some mutations, these were not recurrent in the other reads, suggesting that they could be considered errors of the sequencing itself (Figure 20). Therefore, we concluded that the RNP Cas12a#6 was suitable for our KI experiments with the optimized CHO-K1 cell line.

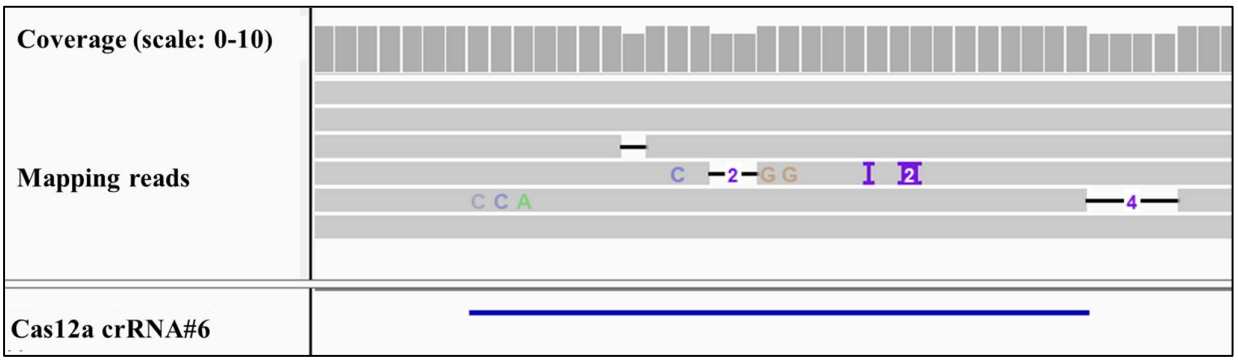


Figure 20. IGV visualization of the sequencing data at the Cas12a crRNA#6 site. A zoomed window on IGV software was used to see the region cleaved by the RNP complex Cas12a#6 (blue line). Empty grey reads show that the sequence of the read perfectly matches the sequence of the reference genome whereas purple boxes, black lines and coloured nucleotides indicate base insertions (the number inside the box refers to the number of inserted bases), deletion, and nucleotide mutation respectively.

4.2.2 Large donor plasmid design

Once RNP Cas12a#6 was confirmed to be suitable, we proceeded with the design of the donor DNA to be employed for the direct KI approach. For the design of the Homology Arms (HAs), we followed the guidelines provided by IDT together with information available in the literature[53]. It is demonstrated that Cas12a shows preferences in the donor template strand, with higher performances when the donor is designed as the non-targeting strand (strand containing the PAM sequence)[53]. Moreover, Cas12a protein gives better KI results when the homology arms are designed starting from positions 12-16 from the 3' of the PAM. So, we designed the HAs for our large donor starting from position 14 of the non-targeting strand in order to increase as much as possible the HDR efficiency (Figure 21)[53]. Moreover, it is known that also the length of HAs could affect the efficiency of KI depending on the length of the donor. We chose long homology arms as suggested by the plasmid vendor, including 900 nt at the 5' of the starting position for the left homology arm and 900 nt at the 3' for the right homology arm.

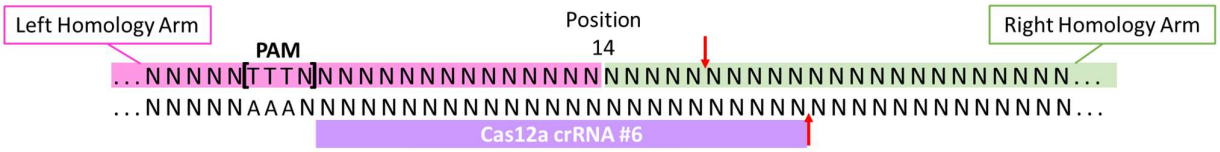


Figure 21. Scheme of the homology arms design strategy. In the image, the TTN-PAM site is put in brackets on the non-targeting strand while the crRNA#6 is paired with the specific site on the targeting strand. As suggested by the Cas12a vendor, the starting point to design the homology arms should be set around position 14 from the PAM of the non-targeting strand regardless of the cleavage sites (red arrows) of the RNP complex. Here the Left homology arm is highlighted in pink whereas the right homology arm is depicted in green. Both the homology arms covered 900nt.

As discussed in the experimental design section (paragraph 2), we added the recognition site of RNP Cas12a #6 flanking the homology arms to obtain the in vivo linearization of the plasmid, possibly integrating the coding sequences without the plasmid backbone (Figure 22). As coding sequence, we decided to add genes useful for the characterization and visualization of correctly modified cells trying to reach an overall length similar to those needed for the expression of a monoclonal antibody, comprehensive of a light chain and a heavy chain with their respective promoters. Namely, from 5' to 3' we added: the PGK1 promoter upstream the Puromycin resistance gene (PuroR in Figure 22), the CMV enhancer and CMV promoter upstream the EGFP reporter gene, the EF-1α core promoter upstream the Hygromycin resistance gene (HygR in Figure 22), and the PGK1 promoter coupled with a recombinase cassette. In the end, the entire plasmid reached a length of 10'837 bp with a sequence to be used as donor DNA of 8'758 bp (including coding sequences and the homology arms).

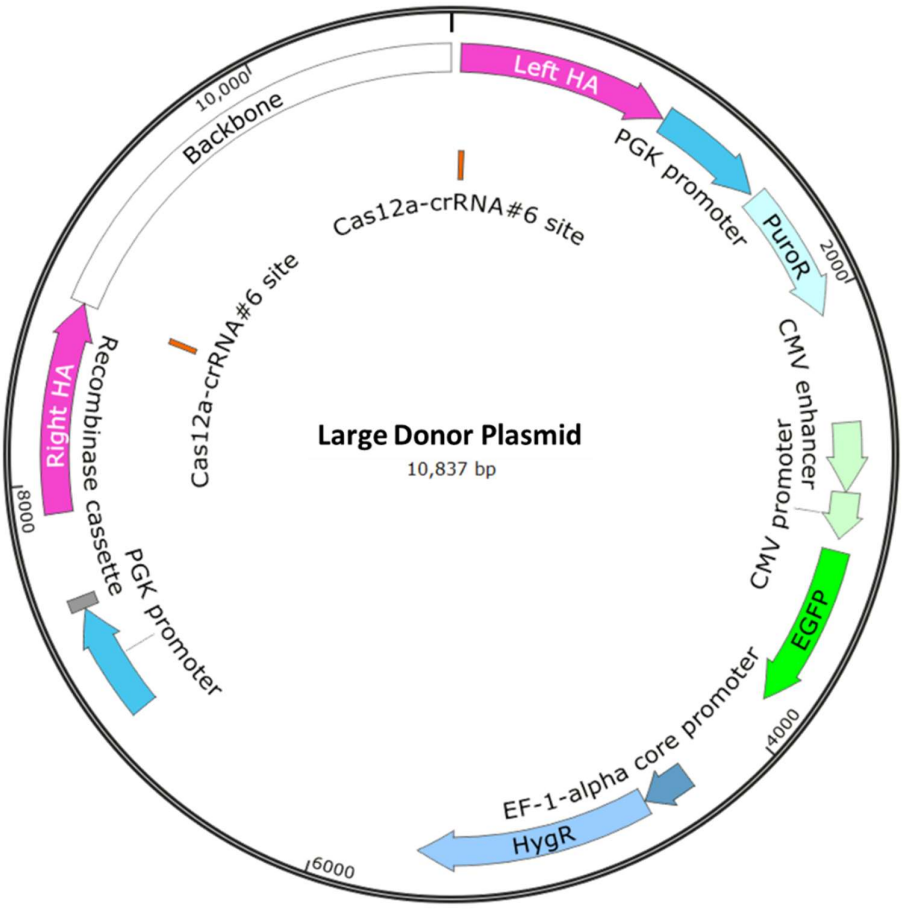


Figure 22. Snapgene map of the designed large donor plasmid. The 900 bp homology arms (in pink) are flanked by a CRISPR/Cas12a #6 recognition site (in orange) to perform the integration of the coding sequences in between. This conformation would avoid the integration of the no more useful plasmid backbone (in white).

4.2.3 Large donor plasmids-Cas12a RNP transfection and pools evaluation

The CHO-K1 optimized host cell line was transfected with Cas12a RNP #6 and the large donor plasmid. We tested three different amounts of plasmid (0,5µg, 1µg, and 2µg), to identify the condition with low cytotoxic effect and the best rate of integration. In particular, two independent transfections for each condition were analysed. To evaluate the cytotoxicity, the cells were monitored for 17 days after transfection assessing both viability and cell growth. In the left panel of Figure 23, the viability of the three pools is plotted over time and, besides an initial decrease of the viability of the sample transfected with 2µg of plasmid (Day 4 after transfection), all the conditions maintained the cell viability over 90%. In the right panel of Figure 23, the viable cell count is shown till 11 days after transfection, when a subculturing step was needed. Also in this case, the sample transfected with 2µg of plasmid showed a slower growth in the first 4 days after transfection with respect to the other two samples. After 7 DPT, all pools started growing at a similar rate. These data revealed that the amount of transfected DNA in the tested range had a non-significant impact on cell viability and growth.

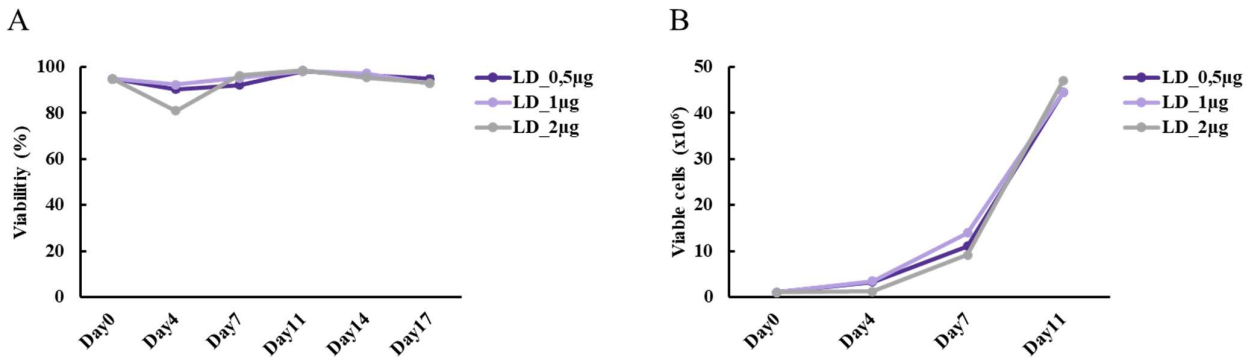


Figure 23. Evaluation of plasmid DNA cytotoxic effect on transfected pools. (A) The viability of cells was monitored till 17 DPT for all three transfected pools. (B) The total number of viable cells is plotted for the three pools till Day11 when a subculturing step was needed. *Data show mean ± Standard Deviation.*

The integration efficiency on the pools was assessed with FACS analysis for the expression of the GFP protein in combination with ddPCR to evaluate the presence of the GFP gene. Moreover, the DNA of the cells was analysed with a 5' junction PCR to confirm the integration at the target site. Since the large donor coded for antibiotic resistance genes, the cells were treated with Puromycin for 7 days and then, both selected and non-selected controls were tested with FACS to evaluate the percentage of cells integrating the plasmid as well as the enrichment of GFP+ cells after the antibiotic selection. As shown in Figure 24, all the analysed samples resulted in a mean of 0,5% GFP+ cells before the selection, giving a measurement of the plasmid integration rate without applying any type of enrichment. The analysis performed on the corresponding selected samples

revealed that the treatment with puromycin succeeded in selecting GFP+ cells with >90% of purity even if the percentage of stably modified cells was not high at the beginning.

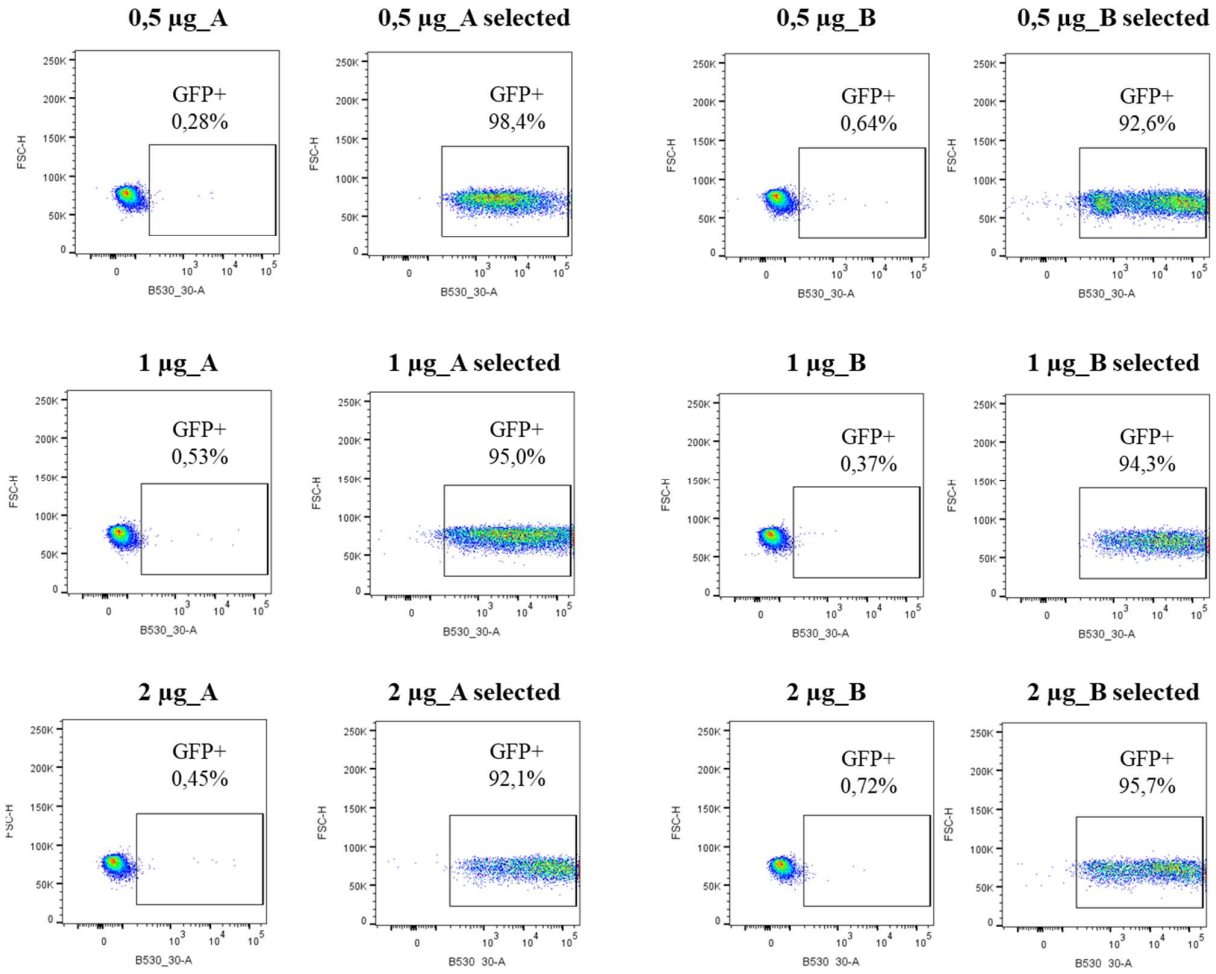


Figure 24. FACS analysis for the GFP expression in all the puromycin-selected pools and their related non-treated control. The result of the experiment is reported in each plot as a percentage of GFP+ cells in tested samples. In this analysis, the pools transfected with 0,5 µg (upper panel), 1 µg (middle panel), or 2 µg (bottom panel) of donor plasmid treated with puromycin were tested with their respective non-treated control sample. Since each transfection was performed in duplicates, both were tested and the two pools with the same transfection condition are indicated as A and B in the figure.

On the same day (24DPT) in which the GFP fluorescence was assessed, we performed the analysis with ddPCR to confirm the FACS data from the genetic point of view. We evaluated the same pools with duplex assays calculating the percentage of the GFP gene signal over the reference gene B2M. The values expressed as percentages of the non-selected pools (control samples) are shown in the left part of Table 10 in which also the results of FACS analysis are reported. The outcomes of the two analyses were comparable: the ddPCR showed that a mean of 0,7% of the genomes contained the GFP gene and the FACS analysis showed 0,5% of GFP+ cells on average. To the same extent, the analysis performed on the selected samples showed 100% of the DNA bearing the GFP gene confirming the high percentage of GFP+ cells obtained with the fluorescence analysis. The calculated percentages were always higher than 100% which could be due to potential multiple insertions in the genome of some cells. Indeed, we used a reference gene present in a single copy in the genome, so if some cells had integrated multiple copies of the plasmid the ratio GFP/B2M inevitably would increase to two or more.

Table 10. The results of ddPCR analysis for the GFP gene are summarized and compared to the FACS evaluation of GFP fluorescent cells. In the left part, non-selected samples are listed; in the right part, the puromycin-treated samples are shown.

	ddPCR		FACS
Control sample	GFP/B2M ratio	GFP gene percentage	Percentage GFP+ cells
LD_0,5µg A	0,0042	0,42%	0,28%
LD_0,5µg B	0,0047	0,47%	0,64%
LD_1µg A	0,0055	0,55%	0,53%
LD_1µg B	0,0056	0,56%	0,37%
LD_2µg A	0,0079	0,79%	0,45%
LD_2µg A	0,0129	1,29%	0,72%

	ddPCR		FACS
Selected sample	GFP/B2M ratio	GFP gene percentage	Percentage GFP+ cells
LD_0,5µg A	1,93	193	98,40%
LD_0,5µg B	1,88	188	92,60%
LD_1µg A	3,08	308	95,00%
LD_1µg B	1,41	141	94,30%
LD_2µg A	2,83	283	92,10%
LD_2µg A	3,73	373	95,70%

Altogether these data demonstrated that the overall efficiency of integration does not change significantly between the different amounts of donor plasmid employed in the transfection. As expected for such a large donor DNA, the percentage of modified cells is not high and a selection step with antibiotics, in this case puromycin, is recommended to enrich the population bearing the integration.

In addition, a junction PCR analysis was performed to assess if the construct was inserted at the targeted site. The design of the primer of the junction PCR is shown in Figure 25A: the Assay1 was designed to span the genomic region flanking the left Homology Arm of the construct and the beginning of the coding sequence, on the PGK1 promoter; the Assay2 spanned a wider region from the 5' genomic region to the GFP gene, approximately in the middle of the inserted construct. The DNA of all the transfected pools treated with puromycin as well as the non-treated, was preliminary tested with the Assay1. As shown in Figure 25B, all the

tested samples gave specific amplification with a prominent amplification band in the pool LD_1µg A selected with puromycin.

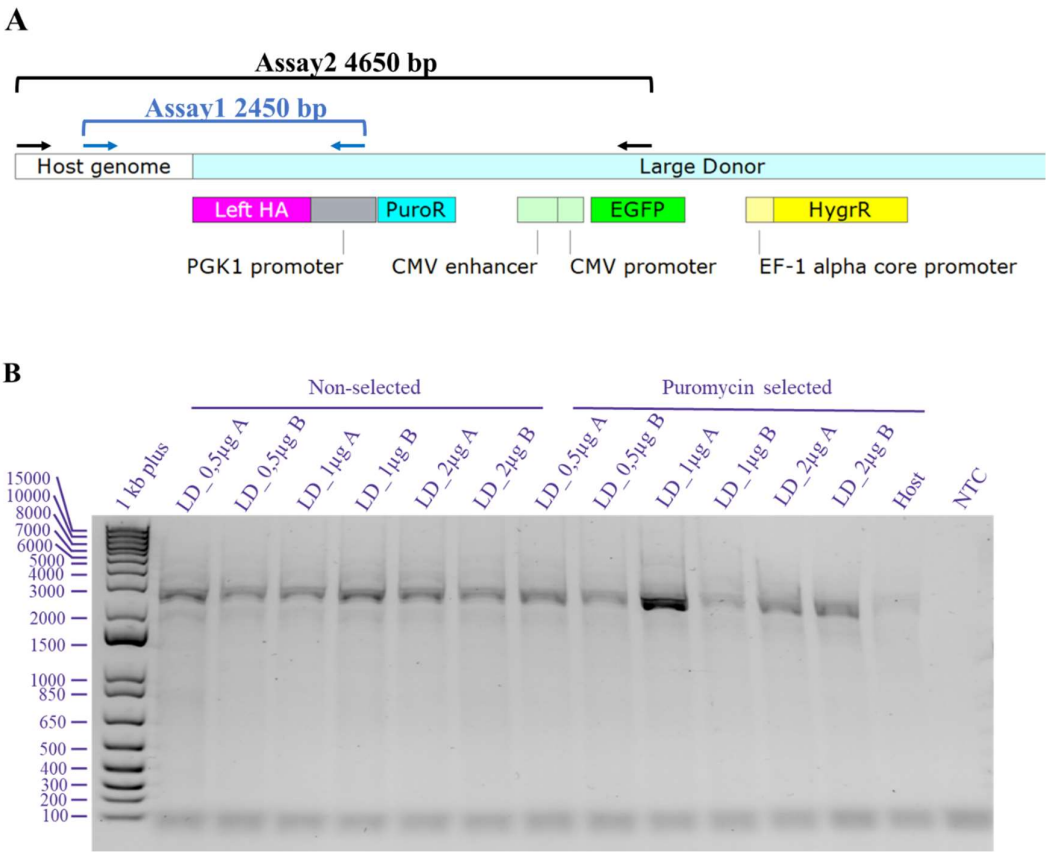


Figure 25. **Junction PCR analysis on transfected pools**. (A) The left part of the donor is shown next to the targeted genomic region. The Assay1 spanned the host genome and the PGK1 promoter with an amplicon length of 2450 bp (primers are indicated with blue arrows). The Assay2 spanned the host genome and the GFP gene with an amplicon length of 4650 bp (primers are indicated with black arrows). (B) The same amount of DNA derived from the 6 pools was analysed with the Assay1. The amplification occurred in all the pools, both in the selected and the non-selected ones at the expected amplicon size (± 10% of 2450 bp). The no template controls (NTC) gave no amplifications.

4.2.4 Knock-In clone generation and screening

Since the pool A transfected with the RNP Cas12a #6 and 1µg of large donor plasmid resulted to be the best sample, with a consistent amplification with junction PCR as well as a good percentage of GFP+ cells, we proceeded with the single cell cloning phase to generate monoclonal population. The cloning conditions were optimized starting from the KO experiments. We diluted the cell suspension to 1 cell/well and seven 96-well plates were generated to have a reasonable number of clones to be screened. After 14 days the GFP+ clones

were plated in 24-well plates and then DNA was extracted with the QuickExtract solution to be analysed with junction PCR. The first screening was performed on all the 225 generated clones, using the Assay1 and 16 clones were found to be positive for the amplification (data not shown in this thesis). These 16 clones were then re-tested with the Assay2 to be more confident of the complete integration of the plasmid. From this analysis, only the clones 5_A2 and 6_A6 resulted to be positive also for the second set of primers (Figure 26). In the end, the efficiency for the on-target direct integration of 8'758 bp with CRISPR/Cas12a was 0,89%.

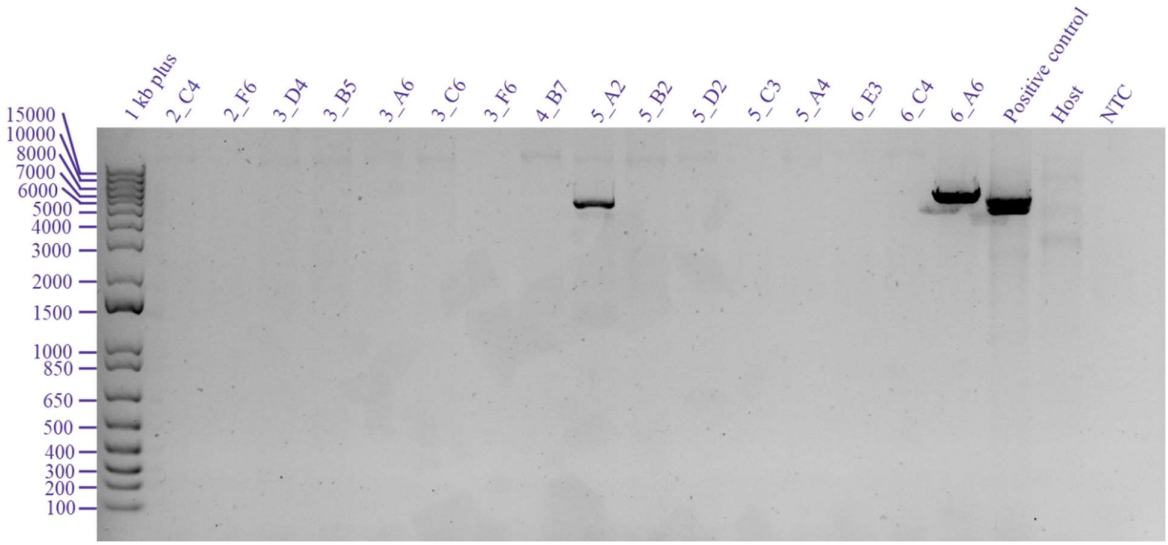


Figure 26. **Final junction PCR analysis on clones.** The DNA of the 16 clones positive for the Assay1 was tested with the Assay2, spanning half of the donor plasmid. The amplification occurred in 5_A2 and 6_A6 clones as well as in the positive control (pool LD_1µg A) with the expected amplicon size (4650 bp). The primers were specific as no amplification occurred in the host CHO DNA. The no template control (NTC) gave no amplifications.

4.2.5 Characterization of 5_A2 direct KI clone

After the screening with the two junction PCRs, we decided to deeply characterize the genetic modification that occurred in one of the two selected clones to really assess the on-target integration as well as the presence of possible off-targets that are not detectable with the previous analysis. To do this, we implemented the Nanopore sequencing starting from the protocol used for the characterization of the KO clone. The extraction of 5_A2 clone DNA was carried out with the Puregene Cell Kit and after the size selection step the library was prepared. The library was loaded on the GridION to perform the run setting the base calling as "super-accurate", filtering for Quality score >10. At the end of the first sequencing, the instrument reported that 7,46 Gigabases (Gb) were sequenced counting 531,99 k reads, thus calculating an average read length of 20,73 kb.

Distinct types of analysis were carried out to assess: 1) the integrity of the donor DNA, 2) the clean integration of the GOI without the plasmid backbone and 3) the correct on-target integration at the CRISPR/Cas12a #6 target site 4) without off-target integration events.

To check the plasmid integrity (1), all the generated reads were aligned directly to the sequence of the expected integration sequence, comprising the coding sequence and the homology arms of the donor flanked by the targeted genomic sequence. In Figure 27 the IGV visualization of this type of analysis is shown with evidence that no gaps in the sequence coverage were present (with at least 6x coverage). In this visualization, some indels and base mismatches were depicted as coloured boxes, but a deep analysis of the sequence revealed that all these alterations were not recurrent within all the reads and they were located at highly repeated regions, suggesting that they were caused by errors during the sequencing.

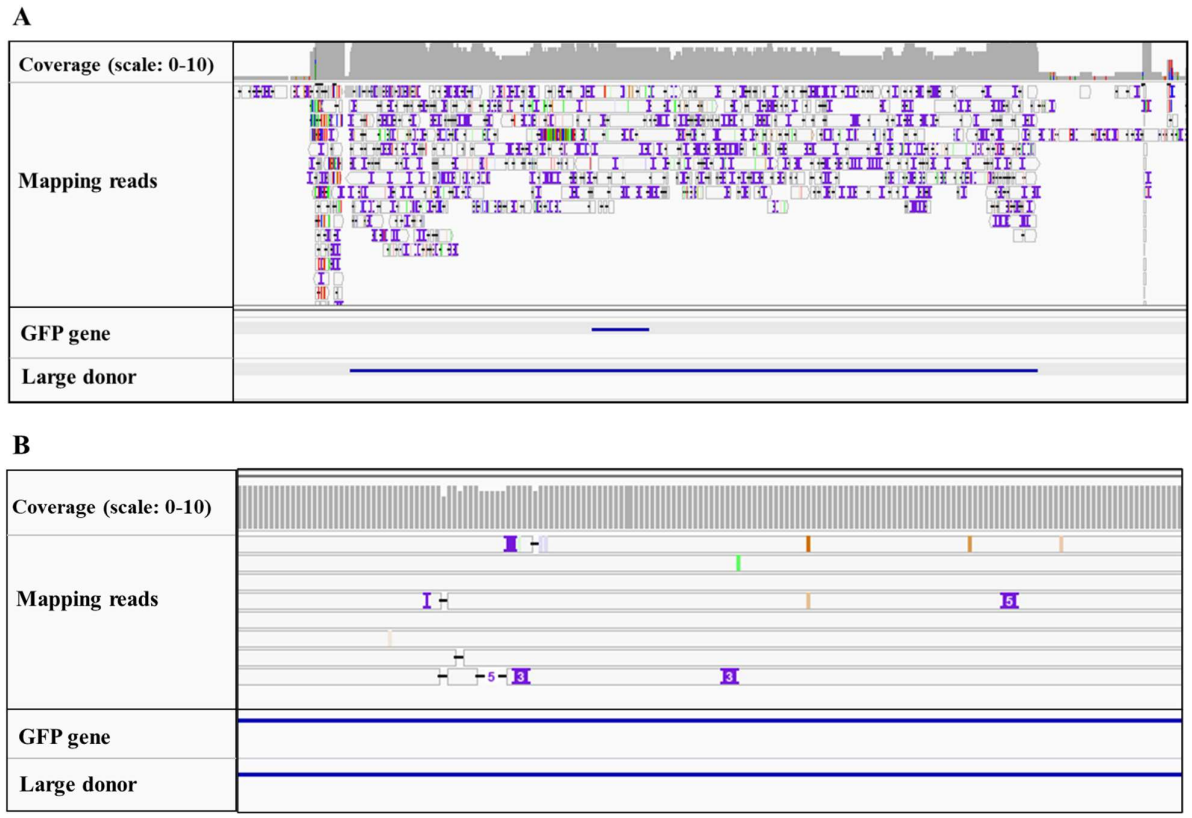


Figure 27. IGV visualization for the analysis of plasmid integrity. From top to bottom: coverage depth, mapping reads (white rectangles), GFP gene sequence and large donor sequence (coding sequence and the homology arms). (A) A wide window on IGV software was used to visualize the complete donor sequence. The coverage reported in the upper part of the image showed that the entire donor sequence was covered by the sequenced reads, confirming that the donor was integrated without deletions. (B) The sequence was checked also for potential mutations in the donor sequence. Empty white reads show that the sequence of the read perfectly matches the sequence of the reference genome whereas purple boxes, black lines and coloured rectangles indicate base insertions, deletions, or base changes from the reference sequence. Here a zoomed portion of the GFP gene was reported to show that the obtained mutations were not recurrent along the mapping reads.

The absence of the plasmid backbone (2) and the correct on-target integration (3) were assessed in the same analysis. All the generated reads were filtered for the ones matching at least a part of the plasmid sequence and then these reads were aligned on the entire CHO reference genome GCA_003668045.2_CriGri-PICRH-1.0.

The results were visualized on IGV software showing the soft clipped bases, meaning that we could see the region of the reads matching the genome (showed in grey or white) and how they continue in the non-matching part (coloured segments of the read meaning that the bases are different from the reference). We started from the chromosome targeted by our designed CRISPR/Cas12a RNP and we identified an integration signal exactly at the cut site of our RNP (Figure 28). As shown in the IGV visualization, there were reads matching the genome sequence till the crRNA#6 target site (the grey part of the reads), at the point selected for the design of the homology arms (in Figure 28B, where left HA and Right HA starts) and continued with a different sequence from the reference genome, e.g. the integrated DNA (coloured part of the reads). We deeply analysed the reads with this type of conformation revealing that the soft-clipped bases corresponded to the adjacent coding sequence of the donor plasmid. Interestingly, we did not identify any reads with the plasmid backbone, proving that at the target site, the integration of the plasmid occurred as we designed.

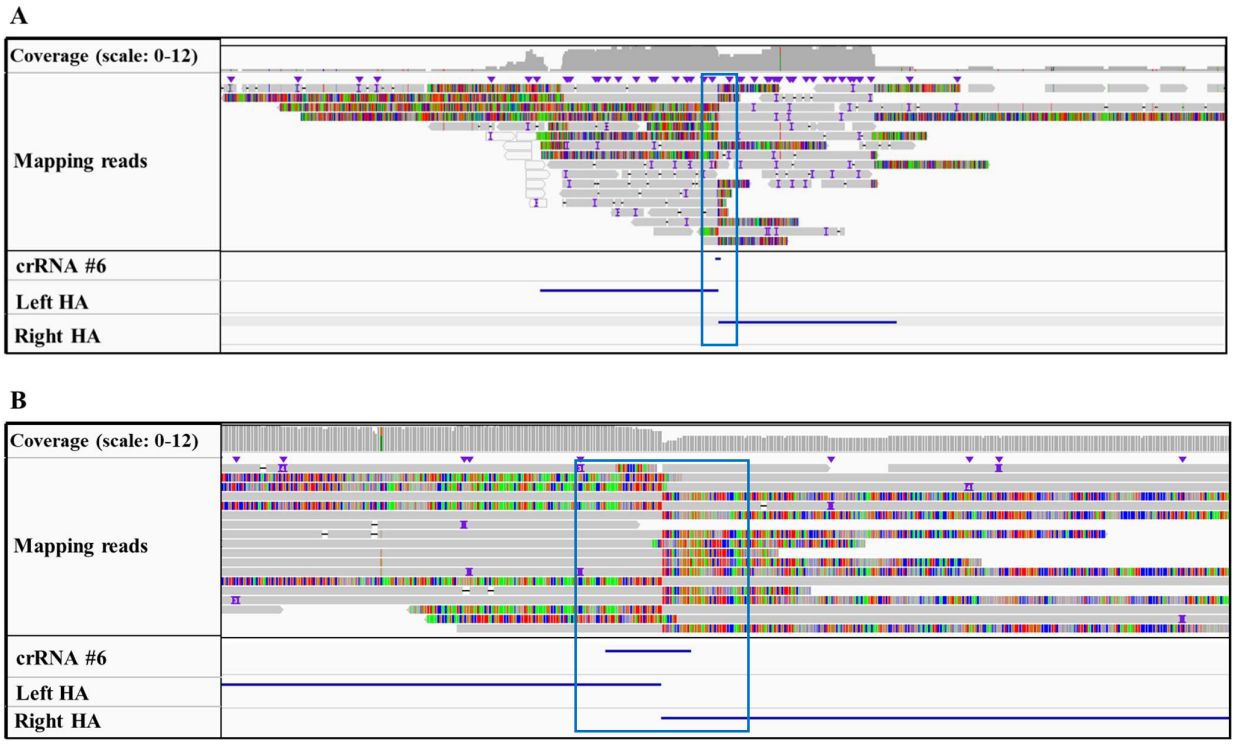


Figure 28. IGV visualization at the targeted integration site. In all figures, from top to bottom: coverage depth, mapping reads (grey and white rectangles), crRNA#6 sequence, Left Homology arm sequence, Right Homology arm sequence. (A) A wide IGV window was used to visualize the entire signal at the on-target site. Here, the soft clipped bases (coloured portions of the mapping reads) were shown to identify the breakpoint of the integration (highlighted by the blue box). (B) A zoomed window at the identified breakpoint site highlighted how the insertion occurred at the targeted site, where the homology arms were designed. Moreover, the soft-clipped bases showed the same colour

pattern, meaning that the inserted sequence is the same in all the reads and consequently they are not an artifact of the sequencing.

To assess if there were off-target integrations (4) along the genome, we investigate the metrics of the identified picks along the genome to find signals similar to the one at the on-target site, revealing other insertions. We examined the values reported in terms of depth of coverage and length of the region covered by the reads, sorting the hits for the metrics, and checking the first 100 picks on IGV. In Table 11 the first 10 hits for the higher coverage depth are reported and in Figure 29 the picks of the 3 best hits are shown. Indeed, we conclude that the mean depth of coverage did not give useful information since the sites with the higher coverage were most of the time small regions (always below 2kb) with high repetitions of nucleotides, and the soft-clipped bases of the reads were just the prosecution of the repeated stretch of nucleotide. Of note, the on-target site was not ranked in the first 100 hits (ranked as #619, data not shown).

Table 11. Best hits sorted for the higher coverage depth.

Chromosome	start	end	coverage depth	coverage lenght
Chromosome A	40624969	40625301	53,8	332
Chromosome B	67078302	67078495	32,2	193
Chromosome B	17790648	17791395	32,1	747
Chromosome C	235151525	235151651	31,2	126
Chromosome D	22673745	22674611	29,6	866
Chromosome E	10601299	10601497	26,4	198
Chromosome F	62680723	62680980	19,5	257
Chromosome G	12234942	12235488	18,9	546
Chromosome H	80066312	80066514	17,8	202
Target chromosome	78673575	78673832	17,3	257

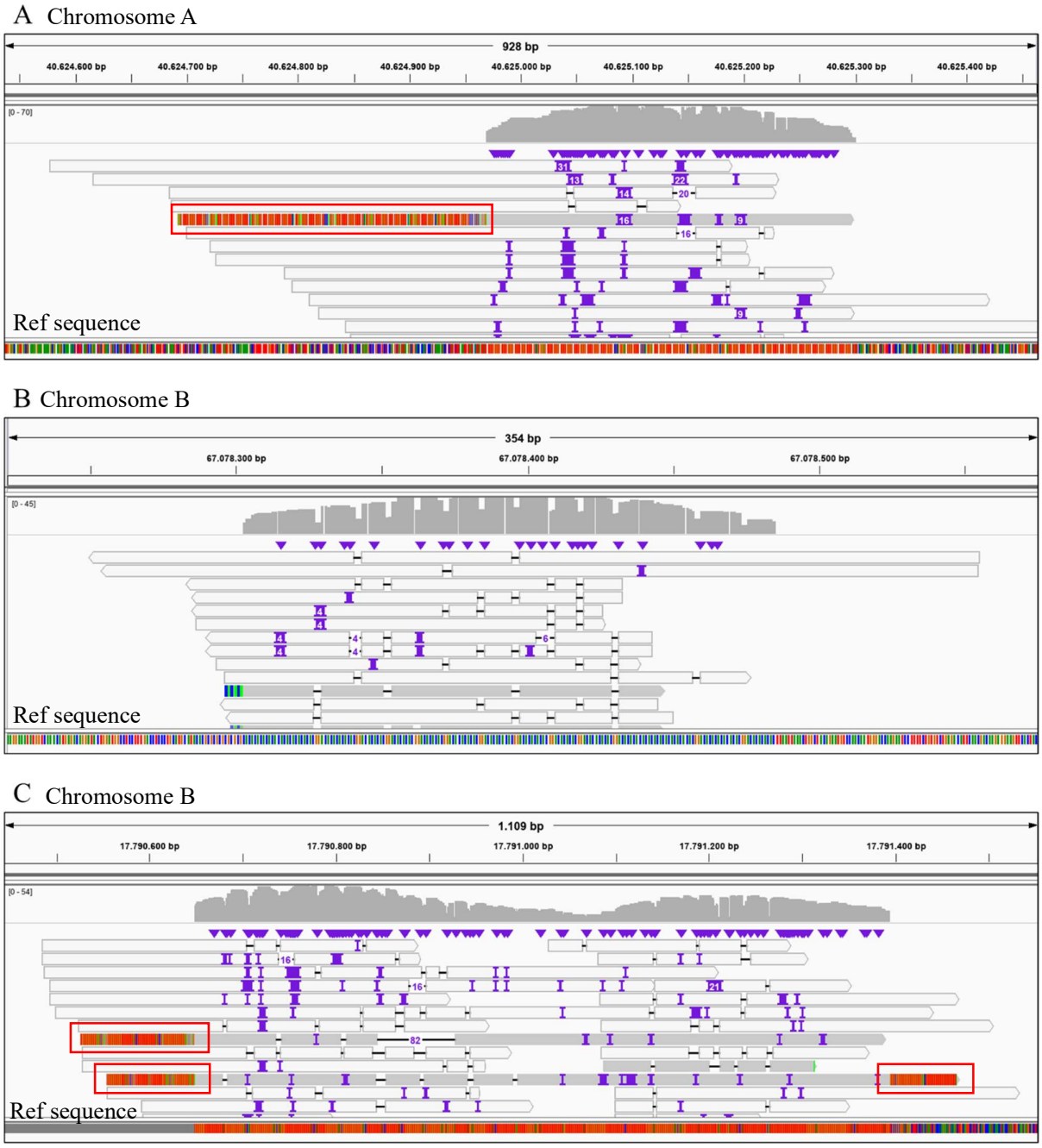


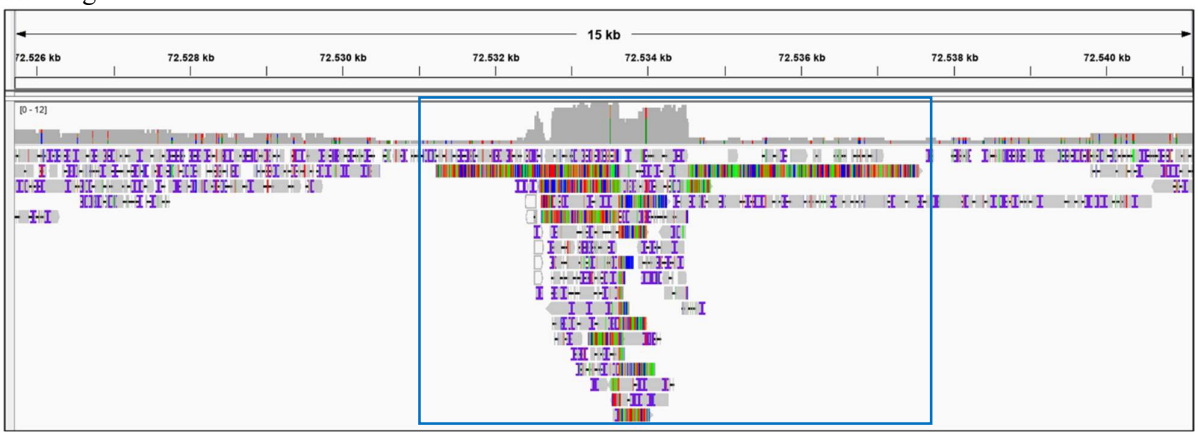
Figure 29. The first 3 best hits sorted for the deeper coverage length shown on IGV. In all figures, from top to bottom: length of genomic region spanned by the reads, coverage depth, mapping reads (grey and white rectangles) and reference sequence with coloured nucleotides (same colour represents the same base). (A) The first hit shows reads matching highly repeated regions (a repetitive pattern of colours in the reference sequence at the bottom) and with the same motif in the soft-clipped bases (in the red box). (B) The second hit shows a short coverage length (354 bp) and no soft-clipped bases (no coloured bases are shown at the side of the matching reads). (C) The third hit shows again reads covering highly repeated regions which continued in the soft-clipped bases (in the red boxes).

On the other hand, filtering the reads for the length of the region covered on the genome gave different results. Indeed, the on-target site was ranked as the first best hit suggesting that this metric could discriminate better a real integration pick with respect to the depth of coverage. In Table 12 are reported the best 10 hits and in Figure 30 the 3 best hits are shown as examples. All the examined loci did not resemble the on-target integration pick because no soft clipped bases were present at the side of the reads, meaning that the reads matched exclusively the genome. This was probably due to some similarities between the plasmid sequence and the reference genome itself. Moreover, in most of the cases, the region was covered by short reads in highly repeated regions (Figure 30C). Altogether, this data confirmed that no off-target integration events occurred in the selected 5_A2 clone.

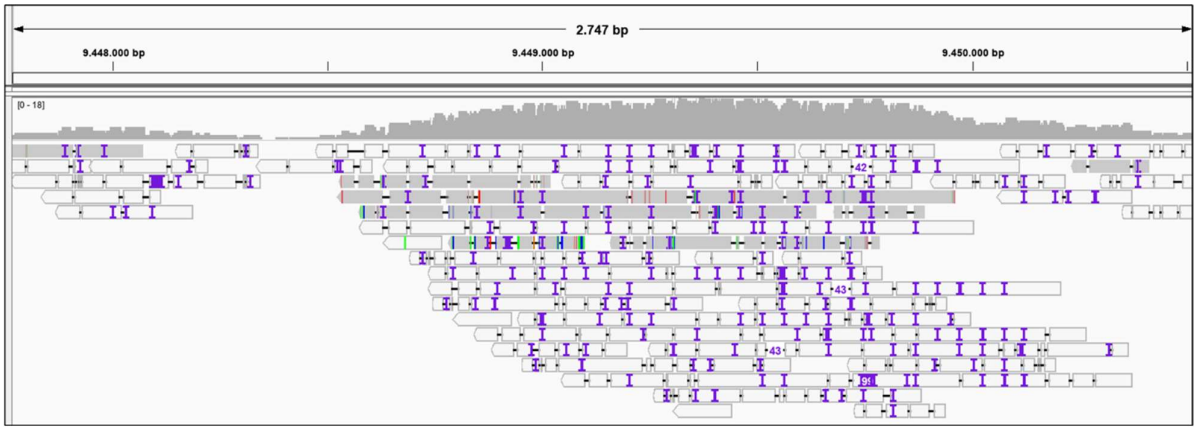
Table 12. Best hits sorted for the higher coverage length. The on-target integration site is reported in yellow.

Chromosome	Start	end	coverage depth	coverage lenght
Target chromosome	72525704	72541229	3,1	15525
Chromosome I	9447766	9450532	8,8	2766
Chromosome I	5072795	5075302	14,3	2507
Chromosome B	279525733	279527501	1,8	1768
Chromosome I	9453748	9455439	4,2	1691
Chromosome F	231222353	231223947	10,5	1594
Chromosome I	5070620	5072170	8,0	1550
Chromosome I	56245809	56247247	2,4	1438
Chromosome I	5075871	5077240	11,4	1369
Chromosome H	129888559	129889816	1,1	1257

A Target chromosome



B Chromosome I



C Chromosome I

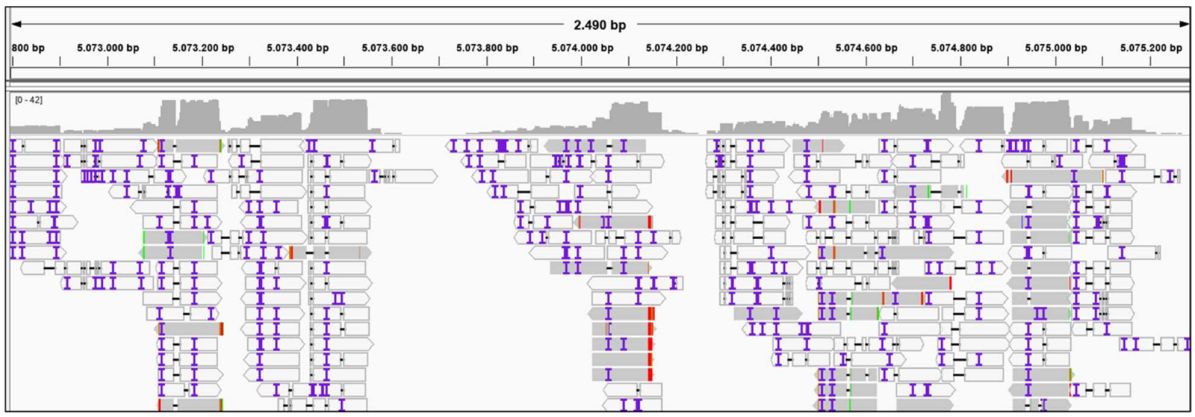


Figure 30. First 3 best hits sorted for the higher coverage length shown on IGV. In all figures, from top to bottom: length of genomic region spanned by the reads, coverage depth, mapping reads (grey and white rectangles). (A) The first hit corresponded to the on-target integration site (highlighted in the blue box), spanning a 15 kb region. (B) The second hit showed a shorter coverage length (2747 bp) and no soft-clipped bases at the side of grey/white matching reads. (C) The third hit showed again short length coverage (2490 bp) and very small reads (approximately 200 bp of grey/white matching reads) and no soft-clipped bases were shown at the side of the reads.

4.3 Knock-In assessment – Landing Pads approach

4.3.1 Landing Pad and recombinase-large donor design

In parallel to the direct KI approach, we also evaluated the two-step integration strategy creating our platform cell line with the CRISPR/Cas integration of the landing pads (LPs). For the design of the ssDNA LP donor, we proceeded with the HAs construction. Since, we would employ the same RNP Cas12a#6 we decided to maintain the same starting point selected for the large donor, so to design the HAs at the 5' or 3' of position 14 on the non-targeting strand (see Figure 21, paragraph 4.2.2). For the length of HAs, we followed the guidelines given by IDT, the provider of the CRISPR/Cas reagents and the ssDNA. The suggested range of length for ssDNA donors to be used for the HAs is 100-500 nt, so we chose HAs of 250 nt. Then we designed the core of the LP as discussed in the aim and experimental design (see paragraph 2). Figure 31 shows the schematic representation of the LP: the recombinase recognition site (in orange) is added downstream of the PGK1 promoter which is used after the recombinase-mediated integration to select cells integrating the large donor at the LP site. In the end, the entire ssDNA reached a length of 1'247 nt.

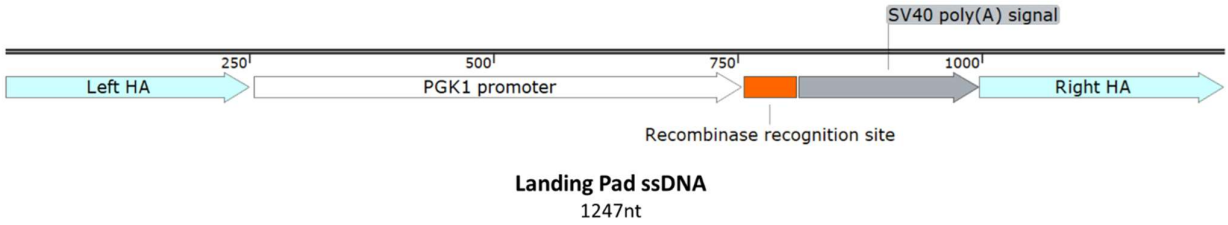


Figure 31. Snapgene map of the designed ssDNA Landing Pad. The 250 nt homology arms (in light blue) flanked the PGK1 promoter (in white) and the recombinase recognition site (in orange).

For the second integration step with the recombinase enzyme, we needed the recombinase-expressing plasmid and the large donor plasmid to be inserted by the recombinase at the LPs site.

Both plasmids were provided by our colleagues in Vevey, who performed the eGFP gene cloning into the proper donor plasmid backbone (hereafter it will be called eGFP donor).In Figure 32, the map of the eGFP donor plasmid is reported to highlight the point that is recognized by the recombinase enzyme (in orange): in this position, the plasmid is opened and inserted at the recombinase recognition site of the LP. In this way, the Puromycin resistance gene is inserted downstream the PGK1 promoter, allowing the selection of the cells inserting the eGFP gene at the LP site.

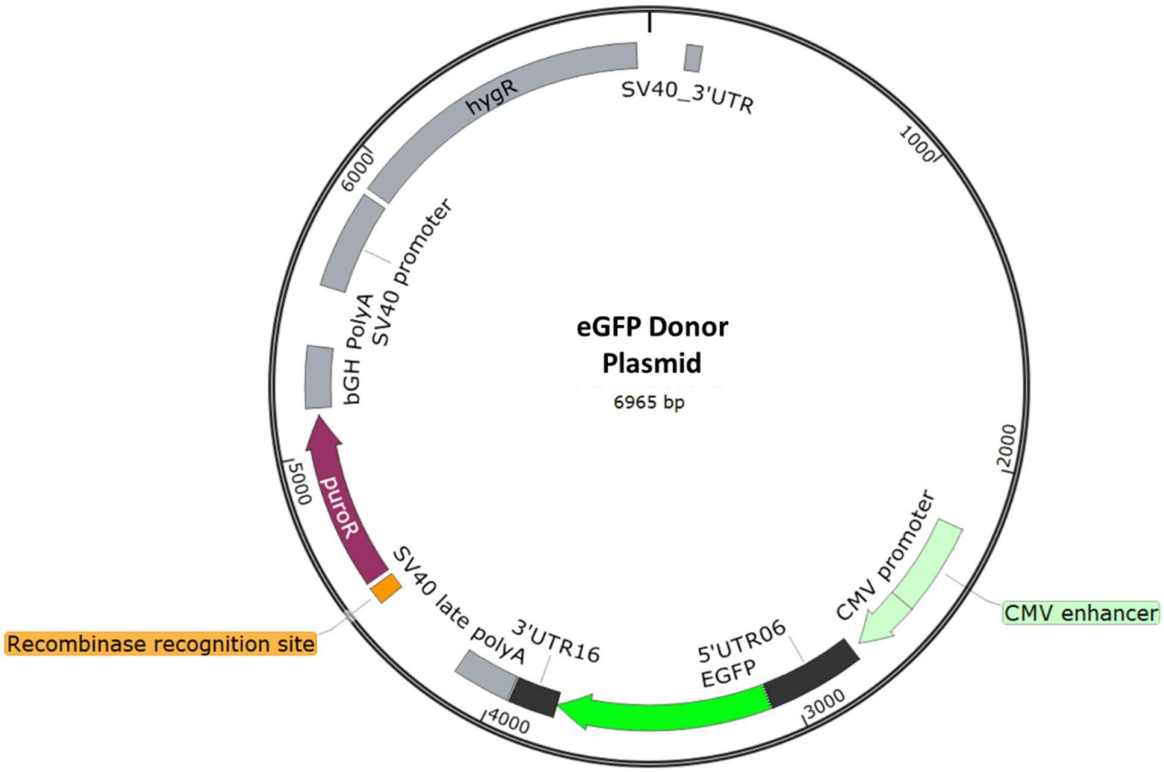


Figure 32. Snapgene map of the eGFP donor plasmid. The components of the donor plasmid for the recombinase-mediated integration at the LP are shown. In green, the eGFP gene with its CMV promoters will confer the green fluorescence to all the cells integrating the plasmid. The recombinase recognition site is highlighted in orange just at the 5' of the puromycin resistance gene (puroR) which will confer antibiotic resistance only if the plasmid is integrated at the LP site, with the PGK1 promoter expressing the resistance gene.

4.3.2 Landing pad - CRISPR/Cas12a RNPs transfection and pool evaluation

The setup for the correct amount of both CRISPR/Cas12a RNP #6 and donor ssDNA was discussed with IDT specialists to find the best transfection conditions. Preliminary tests were conducted on the ATCC CHO-K1 progenitor cell line since the optimized cells were not ready at the time.

The ssDNA is known to give less cytotoxicity, so we decided to transfect the maximum suggested amount of ssDNA: 2,5 µg of LP donor. On the other hand, the amount of RNP complexes in this specific case could have been more critical. Indeed, Cas12a is known for its ssDNase activity upon target recognition and cleavage, so higher amounts of RNP complexes could lead to unwanted degradation of ssDNA donors before its integration. For this reason, we tested three different RNP concentrations starting from the amount used for the KO and direct KI experiments, 4,8 µM, and decreasing it to 2,3 µM and 1,2 µM in the final transfection reaction. After the transfection, we implemented the duplex ddPCR analysis to assess differences in the donor integration rates. We monitored the three different pools for 30 days after transfection, using the specific assay designed on the PGK1 promoter together with the reference gene B2M. In Table 13 the ratio values obtained with the two targets at 30 DPT are shown. We also expressed the percentage to have a preliminary quantification of the integrated sequence with respect to the analysed genomes of the pools. This type of analysis revealed that the lowest concentration of RNP complex provided the best results in terms of ssDNA donor. For this reason, we decided to use the 1,2 µM concentration in further experiments for modifying our in-house optimized CHO-K1 cell line.

Table 13. Results of the ddPCR analysis of the three transfected pools with different concentrations of Cas12a RNP complexes. Values are reported as the mean value obtained with the duplicates as described in material and methods paragraph.

	ddPCR analysis	
Sample	PGK1/B2M ratio	PGK1 percentage
LP_1,2 µM	0,0707	7,07%
LP_2,3 µM	0,0139	1,39%
LP_4,8 µM	0,0071	0,71%

The in-house optimized CHO-K1 cell lines were used for the final evaluation of the landing pad strategy and three independent transfections were conducted using 1,2 µM of Cas12a #6 RNP complexes and 2,5 µg of LP ssDNA as donor template. The cells were monitored for the first 13 DPT to check eventual negative effects on viability and cell growth. As shown in Figure33, both the viability and the total viable cells were not affected by the transfection: the viability remained for most of the time over 90% for all the monitored windows and after 10 days the cells were ready for the scale-up phase in shaking flasks, reaching over 20 x10^6 cells in 3 days (13 DPT).

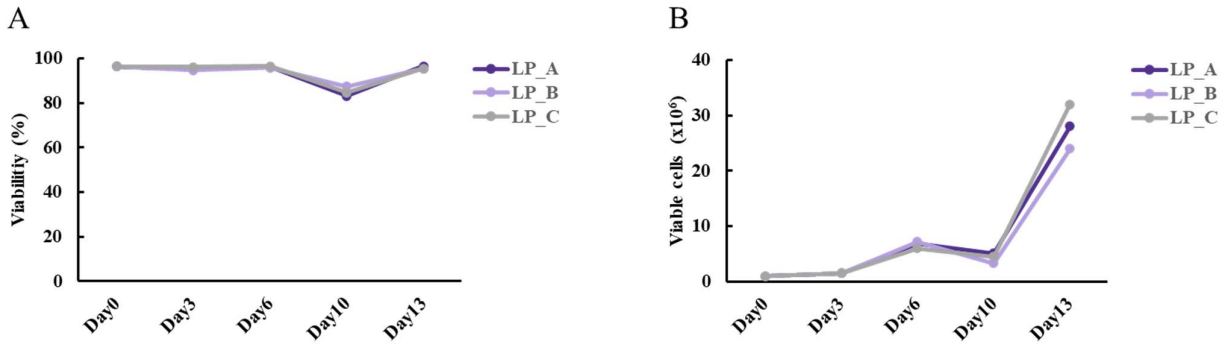


Figure 33. Evaluation of ssDNA cytotoxic effect on the three transfected pools: LP_A, LP_B, LP_C.
(A) The viability of cells was monitored till 13 DPT for all three transfected pools. (B) The total number of viable cells is plotted for the three pools till Day 13.

With the LP approach, we decided to reduce as much as possible the length of the donor DNA, so in the designed LP ssDNA no antibiotic resistance gene was inserted, preventing the use of any type of selection. For this reason, the transfected pools were directly tested with the ddPCR and junction PCR to have preliminary data on the integration rate and the presence of on-target integration respectively.
As in the set-up phase, the cells were monitored with ddPCR for 41 days till the values obtained were stabilized. We employed duplex assays evaluating the ratio of the PGK1 promoter concentration with the reference gene B2M. The values expressed as PGK1/B2M ratio and percentage are shown in Table 14.

Table 14. Results of the ddPCR analysis of the three transfected pools with the same LP donor. Values are reported as the mean of the tested duplicates.

	ddPCR analysis	
Sample	PGK1/B2M ratio	PGK1 percentage
LP_A	0,0600	6,00%
LP_B	0,0193	1,93%
LP_C	0,0199	1,99%

A preliminary junction PCR analysis was performed on the LP pools to assess if the ssDNA was inserted at the targeted site. The assay for the junction PCR is shown in Figure 34A, where the primers employed for the direct KI were adapted to be used for the LP-transfected samples. The forward primer on the genomic region flanking the left Homology Arm and the reverse primer on the PGK1 promoter would generate an amplicon of 2447 bp. So, we analysed the DNA collected on day 27 post-transfection, as the ddPCR results were already stable. As shown in Figure 34B, all the tested samples gave specific amplification with the Assay1 confirming the on-target integration of the LP in our cell pools.

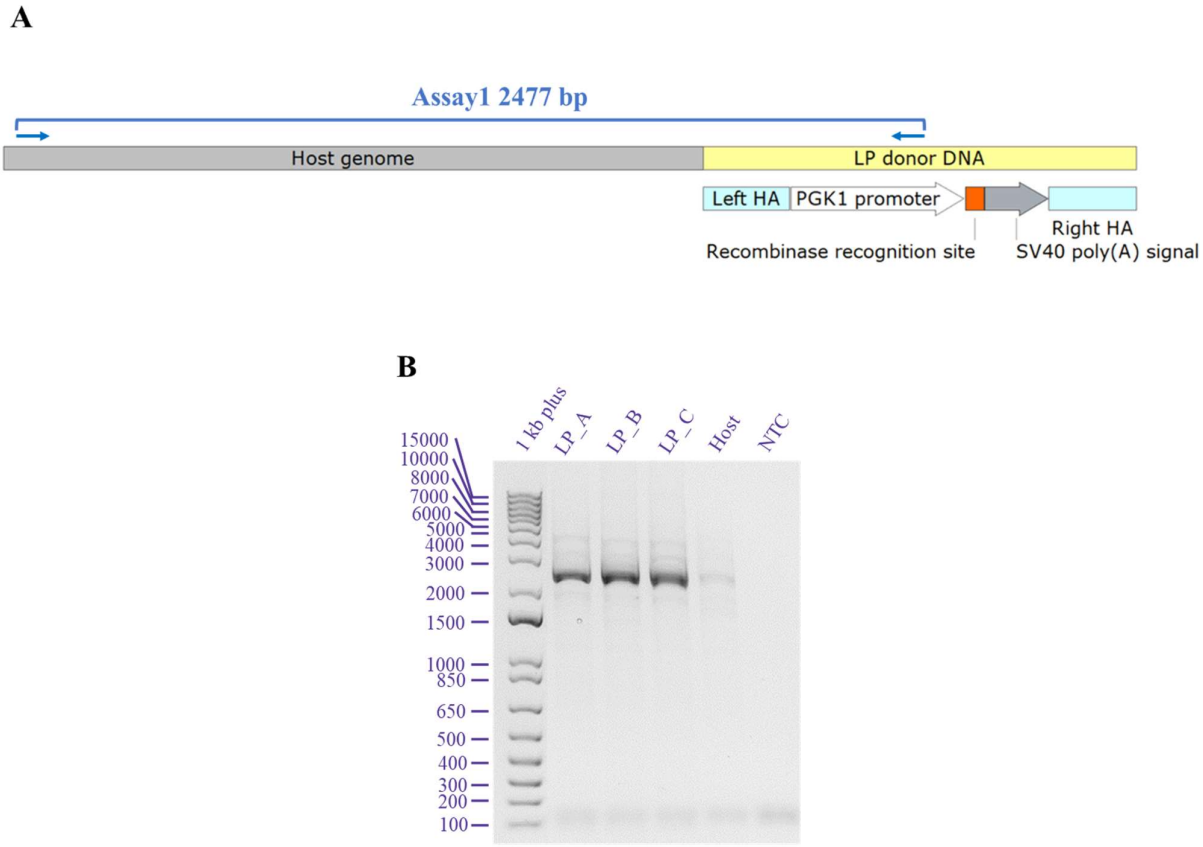


Figure 34. **Junction PCR analysis on transfected pools**. (A) The LP donor is shown next to the targeted genomic region. The Assay1 spanned the host genome and the PGK1 promoter with an amplicon length of 2477 bp (primers are indicated with blue arrows). (B) The same amount of DNA derived from the 3 pools was analysed with Assay1. The amplification occurred in all the pools with the expected amplicon size (2477 bp). The no template controls (NTC) gave no amplifications as well.

4.3.3 Generation of platform cell line: LP clone generation and screening

We decided to use only the LP pool_A to perform single-cell cloning because it was the one with the higher integration rate based on the ddPCR analysis on PGK1, and no relevant differences between the pools were highlighted by the junction PCR.

The cloning conditions were maintained as the ones employed for the direct CRISPR/Cas12a KI. We diluted the cell suspension to 1 cell/well and seven 96-well plates were generated to obtain a reasonable number of clones to be screened. Since no antibiotic selection was applied to the pool, we decided to first screen the generated clones with ddPCR for the cells at least integrating the LP in their genome. 254 clones were tested in duplex reaction with the PGK1 promoter and B2M assay using the new QXONE platform which is able to process automatically up to five 96-well plates from the droplets' generation to the reading of fluorescence signals.

From this screening, we found 16 positive clones (data not shown) which were then tested with junction PCR to identify clones with on-target integration of the LP. We used the Assay1 to perform the screening which in the end revealed that the clones 5_E5 and 5_A7 integrated the LP at the targeted locus (Figure 35). The on-target integration of a 1'247 nt ssDNA with the CRISPR/Cas12a system in our in-house optimized cell line reached the 12,5% calculated on the total modified cells.

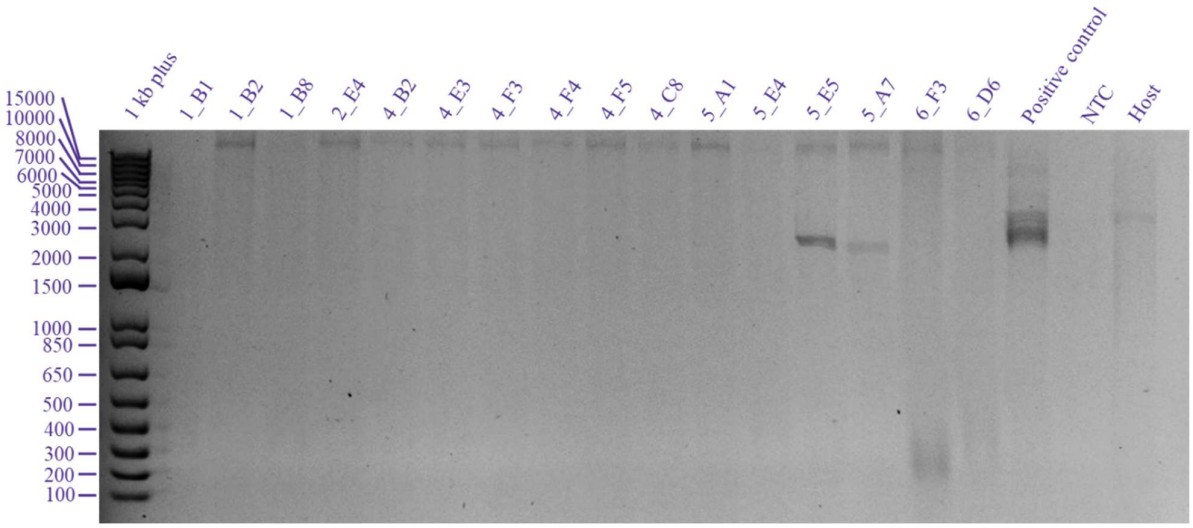


Figure 35. **Junction PCR analysis on LP clones.** The DNA of the 16 clones positive for the ddPCR were tested with junction PCR to see on-target integration of the LP. The amplification occurred in 5_E5 and 5_A7 clones as well as in the positive control (pool LP_A) with the expected amplicon size (2477 bp). The primers were specific as no amplification occurred in the host CHO DNA. The no template controls (NTC) gave no amplifications.

4.3.4 Characterization of 5_E5 LP clone

We applied the same approach used for the direct KI clone characterization, thus taking advantage of the long-reads Nanopore sequencing. The extraction of 5_E5 clone DNA was carried out with the Puregene Cell Kit and after the size selection step, the two libraries were prepared. The libraries were loaded on the GridION to perform two runs setting the base calling as "super-accurate. At the end of the first run, the instrument reported that 9,1 Gigabases (Gb) were sequenced counting 882,69 k reads, thus calculating an average read length of 13,9 kb. Similarly, the second sequencing run generated 7,48 Gb and 781,29 k reads, so with an estimated mean length of 13,73 kb.

We proceeded with the same analysis performed for the characterization of the direct KI-selected clone. Here the goal was to verify: 1) the integrity of the donor DNA, 2) the absence of any type of mutation in the Recombinase recognition site and, 3) the correct on-target integration at the CRISPR/Cas12a #6 target site without off-target integration events (4).

To check ssDNA donor integrity (1) and potential mutations (2), all the generated reads from the two sequencing runs were aligned directly to the sequence of the expected integration sequence, comprising the PGK1 promoter sequence, the recombinase recognition sequence and the homology arms of the donor flanked by targeted genomic sequence. In Figure 36A the IGV visualization of this type of analysis is shown with evidence that no gaps in the sequence coverage were present (with at least 4x coverage summing the reads of the two runs). In this visualization, some indels and base mismatches were depicted as coloured boxes, but a deep analysis of the sequence revealed that all these alterations were artifacts of the sequencing itself since they were not recurrent within all the reads and were located at highly repeated regions. Moreover, the zoomed window on the entire recombinase recognition site revealed that the sequence did not bear any mutations that could impair the further recombinase integration of large donors (Figure 36B). Interestingly, from this analysis we could also suggest that in the 5_E5 clone, a heterozygous insertion occurred: the presence of reads with a "deletion" spanning the 747 bp LP core (Figure 36A, grey reads spanning the Left and Right HAs connected by a black line are highlighted with red asterisks) means that these reads perfectly matched the non-modified locus, so our ssDNA was likely integrated in only one of the two alleles.

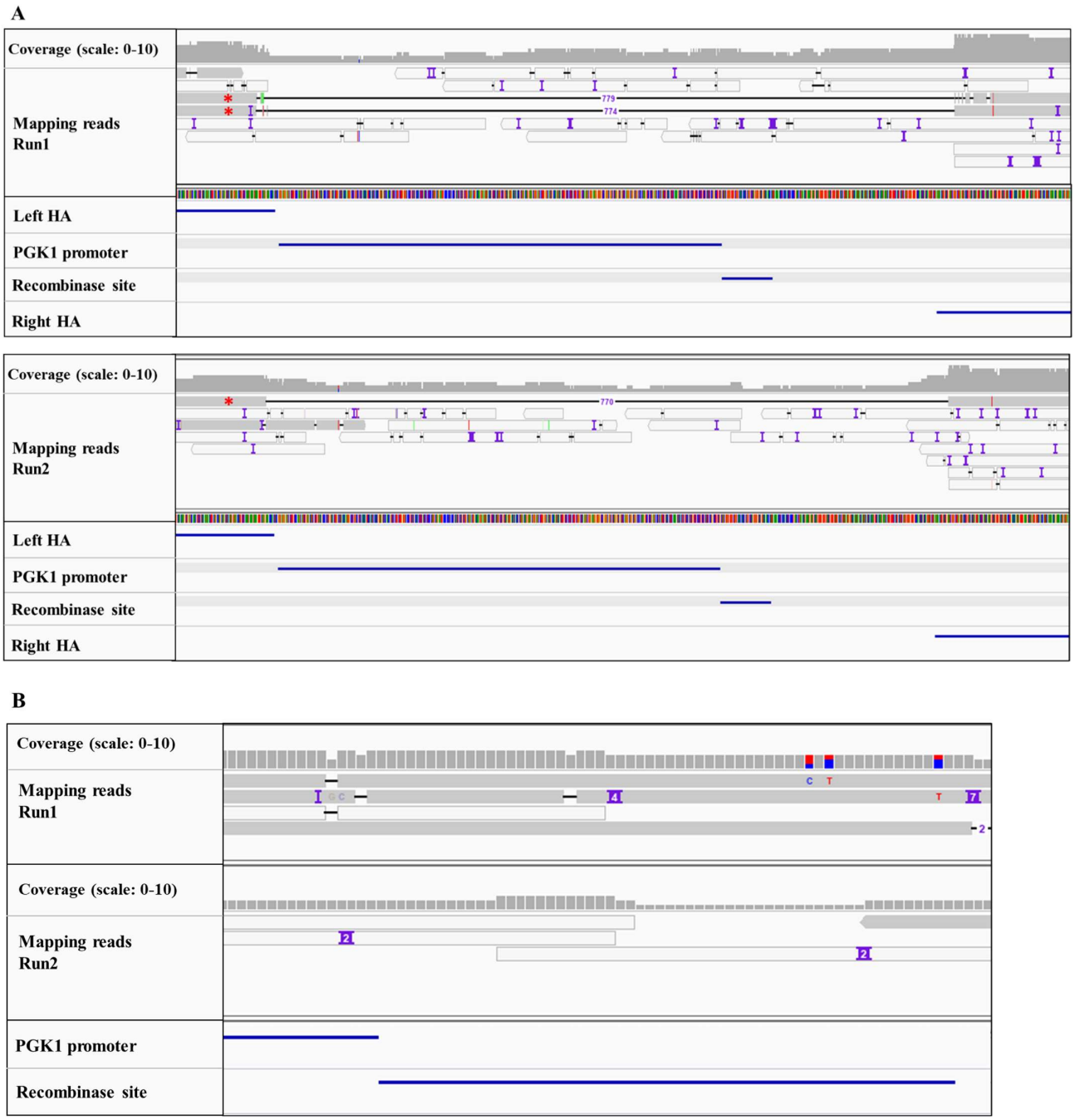


Figure 36. IGV visualization for the analysis of LP integrity. (A) The current window on IGV software was used to see all the donor insertions considering both the homology arms, PGK1 promoter and recombinase recognition site (blue lines in the lower part of the panel). The reads (grey and white rectangles) of the two runs are shown separately with their respective coverage. Empty white or grey reads show that the sequence of the read perfectly matches the reference sequence whereas purple boxes, black lines and coloured nucleotides indicate base insertions, deletions, and different nucleotides with respect to the reference. The reads derived from the allele in which the LP was not integrated were highlighted with red asterisks. (B) A zoomed window on the recombinase recognition site revealed that no recurrent mutations were introduced in the sequence crucial for the next recombinase integration.

The correct on-target integration check (3) was conducted exactly as for the direct KI clone characterization. All the generated reads were filtered for the ones matching the LP sequence and then aligned to the entire CHO reference genome GCA_003668045.2_CriGri-PICRH-1.0. In the end, the results were visualized on IGV software showing the soft clipped bases, to better identify potential integration events of our donor DNA.

We first investigated the chromosome targeted by our designed CRISPR/Cas12a RNP and we identified an integration signal exactly at the cut site of our RNP (Figure 37). As shown in the IGV visualization, there were reads matching the genome sequence till the crRNA#6 target site, at position 14 from the PAM site (e.g., the point selected for the design of the homology arms) and then the software highlighted insertions of 743 bases. The light-yellow boxes in Figure 37 are the pop-out windows of the selected purple triangle or purple boxes on the relative reads (identified by the red asterisks in Figure 37).

We deeply analysed the reads with those insertions revealing that they corresponded to the LP sequence with PGK1 promoter, recombinase recognition site and SV40 poly(A) signal. This analysis confirmed the heterozygous insertion. The presence of reads with just small mutations at the CRISPR/Cas12a cut site (the one not highlighted with the asterisks) suggested that in one allele our LP was not inserted.

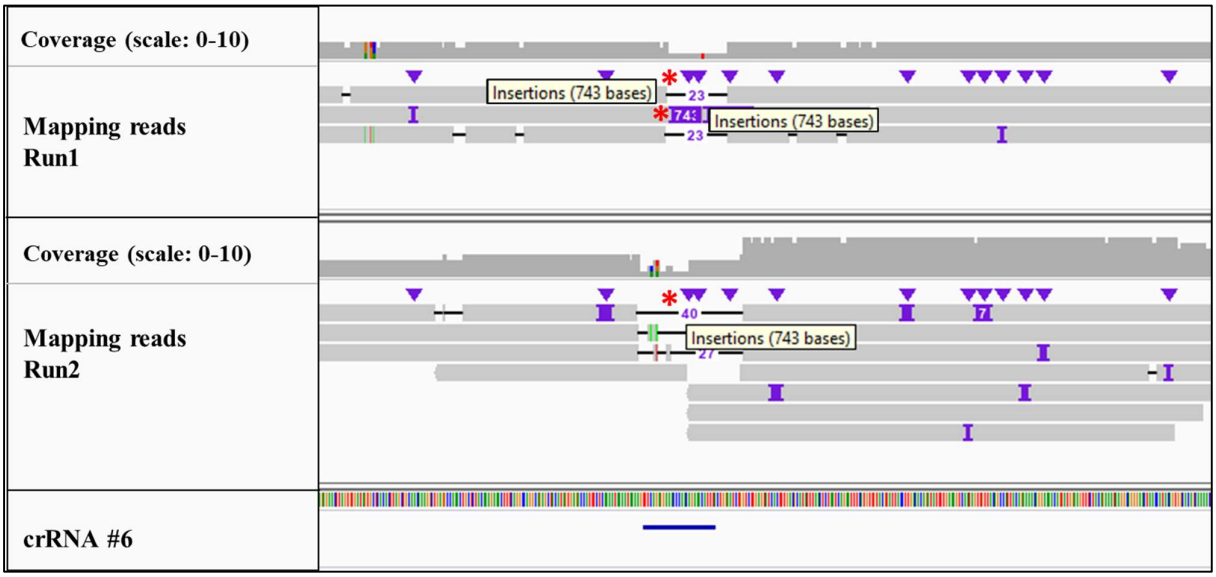


Figure 37. IGV visualization for the analysis of LP on-target integration. A zoomed window on IGV software was used to explore the integration signal at the crRNA #6 target site. Both runs revealed two different types of reads: the ones marked with the red asterisks contained an insertion of 743 bp (the light yellow pop-up windows are reported to see the information of the selected mutational event) corresponding to the LP sequence; the others revealed some little mutations at the CRISPR/Cas12a cut site, but no integrations were present. This conformation suggested that probably the LP integration occurred only in one of the two alleles.

To assess if there were off-target integrations (4) along the genome, we investigate the metrics of the identified picks along the genome to find signals resembling the on-target site insertional event. Considering the results obtained with the previously analysed KI clone, we decided to examine the length of the region covered by the reads. Since the previous analysis revealed that the signals with a coverage length of less than 400 bp were just highly repeated regions with very short reads, we decided to further filter the picks eliminating the smaller ones (less than 400 bp). Indeed, a single pick was reported from Run1, corresponding to the target site, and for the Run2 only two hits were generated, with the on-target site as the first hit (Table 15).

We analysed also the possible off-target picks highlighted by the analysis on Run2, but it resulted to be just one read covering a 477bp region with high similarity with respect to our LP (Figure 38). In conclusion, this data confirmed that our LP was integrated at the target site with no off-target integration events in the selected 5_E5 clone.

Table 15. Best hits sorted for the higher coverage length.

Chromosome	start	end	coverage lenght
Run1			
Target chromosome	72524143	72560031	35888
Run2			
Target chromosome	72515095	72552034	36939
Chromosome F	64560810	64561287	477

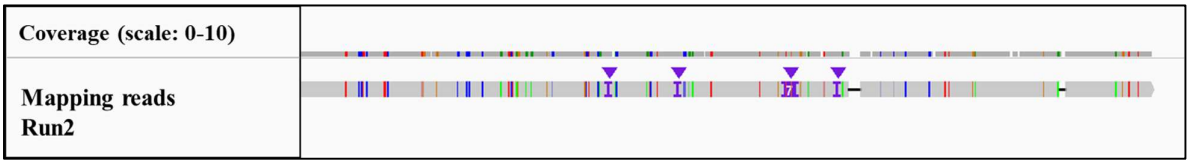


Figure 38. IGV visualization for the unique putative off-target event. The analysis revealed that at the highlighted site just one read was present with a consistent number of mutations with respect to the reference genome and without any type of huge sequence insertion. These features together with the fact that this pick was discovered only in one of the two runs, lead us to exclude the presence of off-target integration events of the LP.

4.3.5 Recombinase integration of a large plasmid

To verify the functionality of the CRISPR/Cas inserted LP, we decided to evaluate the recombinase integration with a large plasmid containing the reporter GFP gene (the eGFP donor). The plasmid shown in Figure 32 was designed in order to have the puromycin resistance gene without promoter next to the breakpoint generated by the recombinase itself.

We transfected the 5_E5 clone with both the recombinase-expressing plasmid and the eGFP donor plasmid. We followed the indications provided by our colleagues of the CLD group which performed preliminary experiments with this technology, performing the transfection with plasmids ratio 1:1 for a total amount of 1 µg of DNA. After that, we optimized the antibiotic selection workflow to find the best conditions in terms of puromycin concentration and administration timing. As a result of this preliminary setup phase, we developed a well-established protocol. The cells were let recover for 10 days after the transfection to have enough available cells for the selection and then we started the selection in 6-well plates at a cell density of 1 $x10^6$ cells/mL adding 10µg/mL of puromycin. The cells were let recover for 2 more weeks and then we implemented a constant puromycin selection: we plated the cells at 1 $x10^6$ cells/mL in 6-well plates adding 7µg/mL of puromycin to the replaced fresh medium twice a week. The administration protocol is shown in Figure 39A. During the experiment, the cells were monitored for their viability and growth to appreciate the efficacy of the selection. In the end, after 25 days of continuous selection, we were confident that all the viable cells were bearing the GFP donor plasmid at the LP site since the growth of the cells was not impaired and the viability reached 84% (Figure 38B). To further confirm the insertion of the GFP donor plasmid, the cells were also observed at the microscope to check the GFP fluorescence (Figure 39C) along all the selection window and indeed, at day50 post-transfection the cells showed a spread expression of the GFP if compared to the early phase of the selection protocol. In conclusion, the platform cell line generated with the CRISPR/Cas technology was suitable to express a gene of interest with an efficient protocol leading to ready selected cells in 50 days.

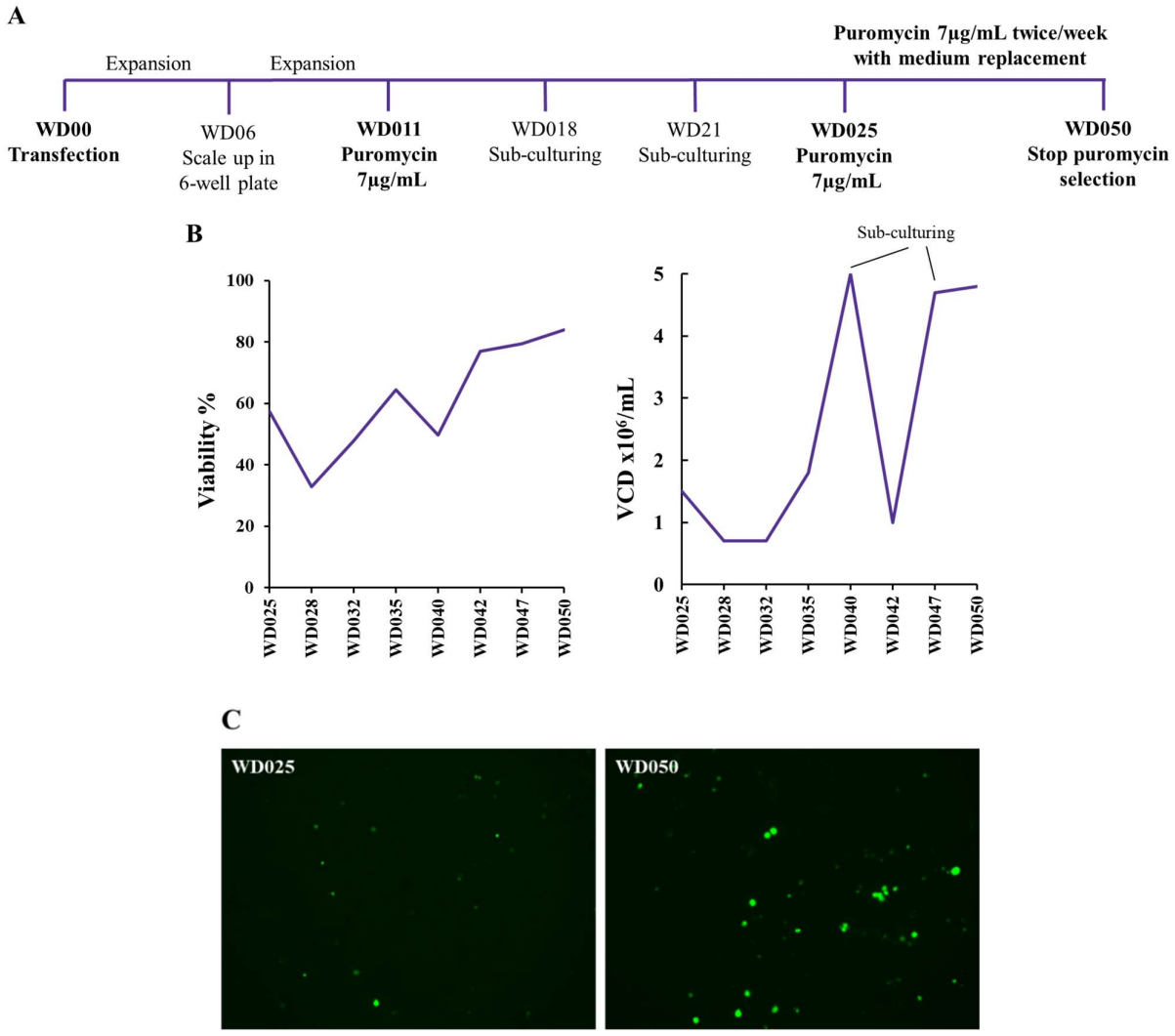


Figure 39. Selection experiment on 5_E5 LP clone transfected with GFP donor. (A) the scheme of selection is shown in the timeline with the steps followed on the different working days (WD). (B) In the panel below the viability and the viable cell density are reported throughout the more stringent selection window (from WD025 to WD050). (C) The GFP fluorescence was assessed at the microscope during the selection phase and from WD025 to WD050 an increase of GFP+ cells was detected.

5. Discussion and future perspectives

Despite all the improvements in the classical cell line development approach, the currently employed workflow remains time-consuming and quite expensive, due to all the uncontrollable insertion and modification events occurring during the establishment of cell lines[22]. Therefore, a significant investment in terms of research and optimization has been made during the last decades to develop new methods based on targeted cell line engineering, in order to reduce timelines and complexity of the analysis needed for the clone characterization[25]. A real boost in the cell line engineering field raised together with the development and optimization of programmable genome editing technologies and in particular after the discovery of the CRISPR/Cas system as a precise, efficient and versatile tool[29,36]. Indeed, the CRISPR/Cas technology started being explored also in the industrial field and the first preliminary studies have been published exploring a wide range of experimental strategies and final goals, from the integration of biopharmaceutical protein genes (e.g., monoclonal antibodies) to the optimization of the performances of the host CHO cell lines for being used as platform for protein production[26].

In this context, our work wanted to investigate the application of the CRISPR/Cas system with our in-house optimized host CHO cell line, exploring large KO and targeted KI of large portions of DNA. The goal was to provide straightforward protocols and effective strategies to modify our CHO cell lines, increasing the overall final protein production as well as making the CLD process more efficient and reproducible.

For the KO assessment, we tried to stress the CRISPR/Cas system and decided to eliminate all the integrations in a reference clone generated from our random integration-based approach. Unfortunately, the analysis employed to characterize the clones, e.g., GCN evaluation with ddPCR and insertion site identification with TLA-seq, were not adequate to really uncover the conformation of the insertion site. For this reason, we decided to apply the Oxford Nanopore technology to analyze the reference clone. This third-generation sequencing technique is able to process and sequence ultra-long reads (up to megabases)[54] which allows the resolution of complex genomic structures, such as concatemers, typically present in random transfected cells[55]. Thanks to the Nanopore sequencing, we obtained the first real picture of the integration site in one of our candidate clones, uncovering the complexity of the genetic rearrangements occurring after the current CLD workflow. In particular, our analysis revealed a 38,5Kb region comprising some partial integration of the expressing vector as well as duplications of the nearby genomic locus. It would be interesting to know if those large mutations could at least in part contribute to the high-producer behavior of this specific clone, for example duplicating some regulatory elements promoting a higher expression rate. Unfortunately, the locus was not annotated at all, so we were not able to correlate the conformation of the integration site to the performances of the clone.

Subsequently, we took advantage of the obtained sequence to properly design the CRISPR/Cas assays to delete the integrated expressing vector and all the related genetic rearrangements. We tested different crRNAs in combination with the Cas9 or Cas12a nucleases. Indeed, Cas9 has been used extensively in the last years for a

wide range of applications, including CHO cell engineering[26], while Cas12a is a comparatively recent addition to the genome editing portfolio, though its structure and mechanism of action have been carefully studied[35]. These two systems have common features but some peculiarities in the Cas12a way of action could give an effective improvement in our workflows. As mentioned in the introduction (see paragraph 1.2.1), Cas12a targets T-rich PAM sites, thus giving the possibility to target loci previously inaccessible to Cas9[34]. Additionally, Cas12a generates staggered DSBs distal from its PAM site, potentially permitting multiple rounds of targeting, unlike Cas9 which creates blunt-end cuts proximal to its PAM site, often causing the PAM to be destroyed during NHEJ[56]. Finally, while Cas12a and Cas9 have shown similar activity in CHO-K1 cells, Cas12a is reported to have a higher specificity than Cas9, decreasing the risk of unwanted off-target effects[57,58]

We analyzed the transfected pools with ddPCR to appreciate the modification rate and identified the most promising complex to use: the crRNA#6 for the Cas12a. Interestingly, crRNA#6 was identified with the online tool CHOPCHOP as the one with the highest predicted efficiency even if the CHOPCHOP ranking resulted to be low due to the low CG content. On the contrary, the highest crRNA ranked by the tool (crRNA#1 for Cas12a) was not able at all to modify our cells. This suggested that the in-silico analysis of the crRNA is a valid starting point, but the efficiency must be always assessed on the transfected cells. The application of ddPCR analysis could prove to be beneficial in acquiring a prompt and reliable preliminary assessment of the activity of the RNP complexes.

Moreover, the ddPCR could be applied for the clone screening phase to allow a quick and high-throughput analysis of clones' DNA. Taking as reference only the optimized round of single-cell cloning (see clone screening Round 4 in paragraph 4.1.4), we were able to obtain 3 KO clones from a small-scale screening of 99 samples processed in less than 4 weeks starting from the single-cell cloning (20 days considering also the expansion of the clones) to the DNA extraction and ddPCR analysis (3-4 days).

We achieved an overall large KO efficiency of 3% which is low if compared with some published results, but we must consider the peculiarity of our KO experiment. In fact, the conformation of the locus was extremely complex and besides we were able to use a unique CRISPR assay to perform both cleavage flanking the target region, we had a third predicted cut site in the middle of the target region which for sure increased the probability of partial deletion of the entire region. Furthermore, the presence of homologous sequences created by the random integration likely prevented a correct and clean restoration of the locus. In fact, the clone screening revealed that we did not generate only KO or non-modified clones, on the contrary, most of the clones showed an intermediate situation with a partial decrease in one or both chains, meaning that the CRISPR/Cas cleavages occurred, but the homologous sequences increased the probability of other aberrant re-joining of the detached portions of DNA.

To have a complete picture of our KO experiment, we decided to characterize one of the obtained KO clones with the long reads of Nanopore sequencing. We performed a deep analysis of the generated sequencing data, and we were able to confirm the deletion of the 38,5Kb target region. Moreover, our analysis showed the

presence of some random mutations at the re-joining site as a natural consequence of the error-prone NHEJ repair pathway. Of course, the ideal situation would be to avoid this type of random mutation and it could be possible by coupling the deletion to a small donor KI. For instance, we can use some "bridging oligo", very small ssDNA donors with a known sequence flanked by small homology arms, to be inserted at the re-joining site. This would help the junction of the correct ends and would give more control over the sequence of the re-joining site which could also be used as a genetic reporter to be detected by a ddPCR assay.

Through transfection optimization, clone screening implementation, and analysis, we have achieved the establishment of a highly efficient protocol for the deletion of large DNA segments within the CHO genome. Moreover, cutting-edge technology such as ddPCR and Nanopore sequencing, were efficiently implemented to obtain a precise and rapid characterization of the modified cells. A visual representation of the optimized workflow can be found in Figure 40.

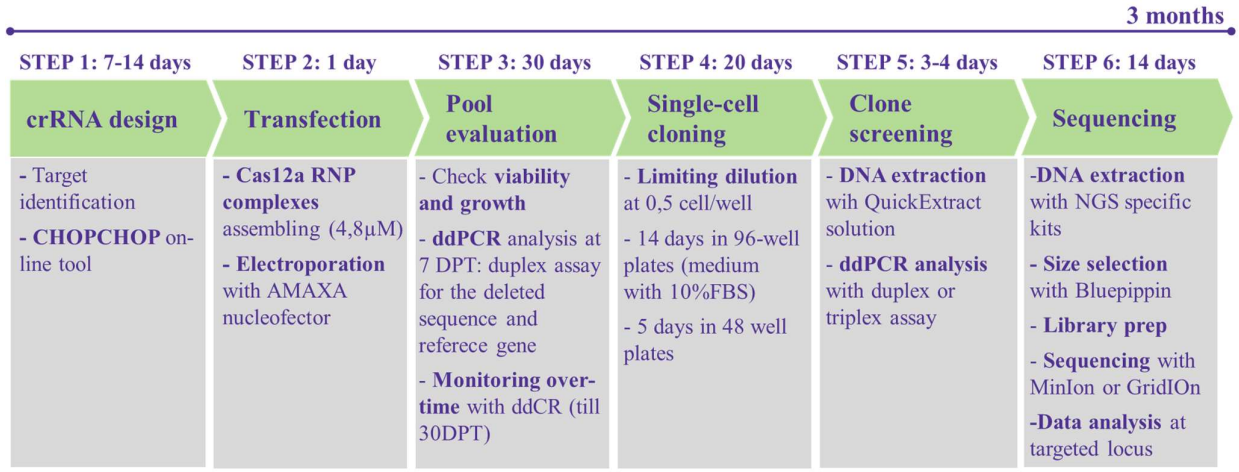


Figure 40. Schematic summary and timelines of the workflow for large KO experiments. (STEP 1) The crRNA design is conducted taking advantage of the online tool CHOPCHOP which gives the possibility to select the crRNAs for their predicted efficiency or an overall ranking. (STEP 2) The cells are transfected with the Cas12a RNP complexes using the AMXA nucleofector. (STEP 3) The generated pools are then monitored for viability, cell growth and for the target sequence with ddPCR-specific assays to appreciate signal decrease. The ddPCR is conducted over a period of 30 days to assess the stability of the modification. (STEP 4) Single-cell cloning is needed to obtain correctly modified cells, so limiting dilution is applied in 96-well plates. The cells are let recover and grow to be then scaled up in 48-well plates. (STEP 5) A small-scale clone screening is performed with the same ddPCR employed for the pool evaluation to identify clones with no signal for the KO targeted sequence. (STEP 6) Nanopore long-reads sequencing is performed, and the resulting data are analysed focusing on the target locus.

Regarding the CRISPR/Cas KI assessment we decided to follow in parallel two strategies that could be both useful in our process.

The first one was to directly integrate, with the CRISPR/Cas system, a 10'837 pb plasmid similar to those usually employed for random integration. We decided to test a peculiar design of the donor, which enables the in vivo linearization of the plasmid itself. We added the recognition site of the CRISPR/Cas12a RNP complex (the same used to target the genome) at both sides of the coding sequence of interest (8'758 bp), just at the end of the homology arms. In this way, we exclusively integrated the sequence of interest without all the plasmid backbone which is normally co-integrated with the transgene in the random integration approach[26].

Since the direct integration of a large donor could have some throwbacks in terms of efficiency, we tested a two-step integration strategy that could bypass this issue[26]. In detail, this second approach includes a first step consisting of the integration of a Landing Pad with the CRISPR/Cas system at the selected target site, thus generating a platform cell line. In this step the donor is quite smaller than the plasmid used in the first strategy, allowing us to obtain an integration with higher efficiency. This engineered cell line is then used to perform the recombinase-based integrations of a large donor with the sequence of interest. Since enzymes such as recombinases or transposases have fewer difficulties in integrating a large portion of exogenous DNA compared to the CRISPR/Cas system, with this second step we avoided donor length-related issues of low efficiency integration[59,60].

Our CHO in-house optimized cell lines were transfected with the selected Cas12a RNP complexes and the respective donor DNA: the 10'837 bp plasmid for the direct KI and the ssDNA LP (1'247nt) for the CRISPR/Cas-recombinase approach. We evaluated the pools with both ddPCR and junction PCR to have preliminary data on integration rate and on-target integration respectively. The ddPCR analysis revealed that without any type of selection, the direct KI had an integration rate of 0,5% whereas the LP integration resulted in a 6% integration rate. Moreover, both transfections gave positive amplification for the junction PCR, meaning that at least some cells integrated the donor at the target locus. After that we were able to apply an antibiotic selection step on the direct KI cells, increasing the number of correctly modified and thus improving the step of single-cell cloning.

We succeeded in performing small-scale cloning and applying different approaches for the clone screening depending on the two types of strategies. For the direct KI cells, which were selected with a purity >95%, we proceeded with a first junction PCR screening on 225 clones and then, positive clones were re-tested with a second junction PCR assay to confirm the on-target integration.

In the end, we obtained two positive clones integrating 8'758 bp donor DNA at the target site, calculating an on-target integration efficiency of 0,89%.

For the LP approach, we decided to start with a ddPCR screening in other to make a preliminary selection for the edited clones. Here 254 clones were tested with the QXONE instrument which ensured an even higher throughput and speed of sample processing. The 16 positive clones identified with ddPCR were then tested with the junction PCR to isolate the clones bearing the on-target integration. From this analysis we identified

two KI clones, thus calculating a 12,5% efficiency for the on-target integration over the overall edited cells. After the screening, we selected one clone generated from each workflow for the characterization with the Nanopore long-read sequencing and subsequent data analysis.

In both clones we confirmed the correct and complete integration of the respective donors at the target site and excluded other off-target integration. We demonstrated the precision of the CRISPR/Cas system and highlighted how the long-reads Nanopore sequencing could be a valuable tool for the deep characterization of CRISPR/Cas-based editing. Of course, we set up an analysis workflow which aimed to reliably confirm the intended modification in an effective and forthright way, excluding the presence of off-target insertions. However, in the future, more detailed characterizations could be carried out. Starting from the same sequencing raw data generated by our sequencing workflow, a de novo assembly could be performed to reconstruct from scratch the integration locus. This type of analysis requires more time and effort for its implementation, but it could provide even more accurate information on our modified cells, ensuring the resolution of possible tandem copies generated during the insertion.

In addition to the genetic characterization, the selected LP clone was then expanded to create our platform cell line which was tested for its suitability in performing the recombinase-mediated integration. Encouragingly, the integration of a reporter vector with the recombinase enzyme yielded promising results. Notably, the antibiotic selection was employed to isolate cells integrating the donor at the specific LP site, and following a 50-day treatment period, the cells exhibited complete recovery from the selection process.

Together with the successful integrations, also timelines had to be considered. Following protocol optimization, we succeeded in generating our direct KI clones in 3 months from the transfection to the sequencing of the selected clone. On the other hand, the LP approach needed 4-5 months from the transfection of the LP donor with the CRISPR/Cas system to the selected cells integrating the final large donor with the recombinase enzyme. Of course, the advantage of the LP approach is that once the platform cell lines are created, we can start from those cells to integrate any sequence of interest without any limitations of length and obtain the selected cells in just 1-2 months. The summary of the optimized protocols for both strategies is illustrated in Figures 41 and 42.

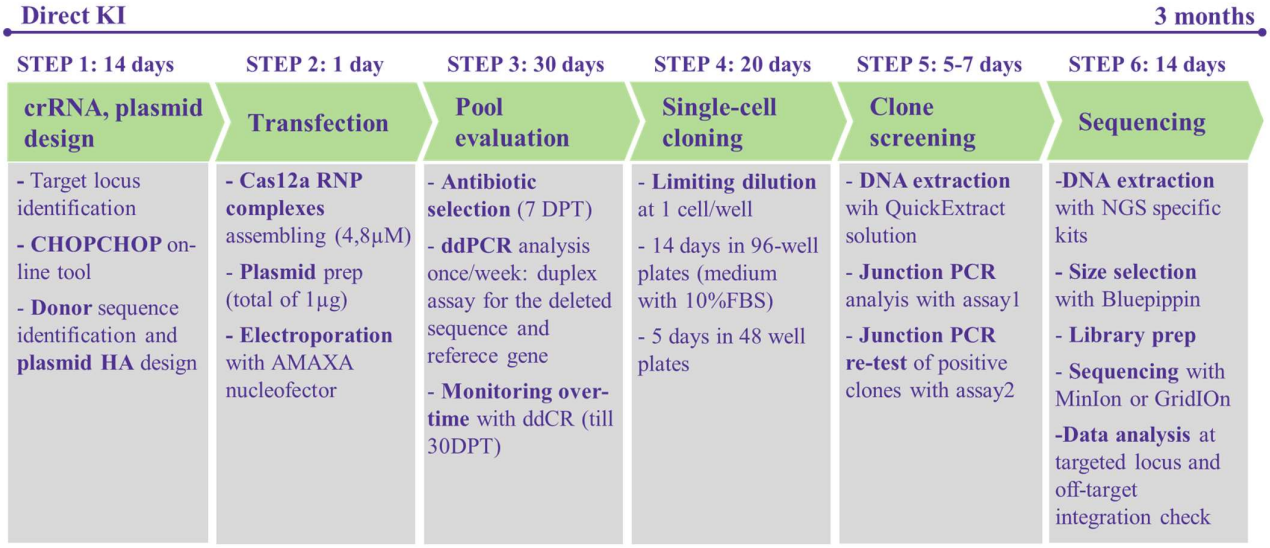


Figure 41. Schematic summary and timelines of the workflow for large direct Ki experiments. (STEP 1) The crRNA design is conducted taking advantage of the online tool CHOPCHOP and once selected, the Homology Arms are properly designed to maximize the efficiency. (STEP 2) The cells are transfected with the Cas12a RNP complexes and the donor plasmid using the AMXA nucleofector. (STEP 3) The generated pools are treated with antibiotics to select only modified cells and the donor sequence is monitored over time with ddPCR. (STEP 4) Single-cell cloning is needed to obtain modified cells at the targeted site, so limiting dilution is applied in 96-well plates. The cells are let recover and grow to be then scaled up in 48-well plates. (STEP 5) A small-scale clone screening is carried out to identify the on-target modified clones. A first junction PCR test is performed checking one of the two sides of the inserted sequence and then, the obtained positive clones are re-tested with a different junction PCR to be more confident of the integrity of the large donor plasmid. (STEP 6) Nanopore long-reads sequencing is performed, and the resulting data are analysed focusing on the target locus and possible off-target insertions.

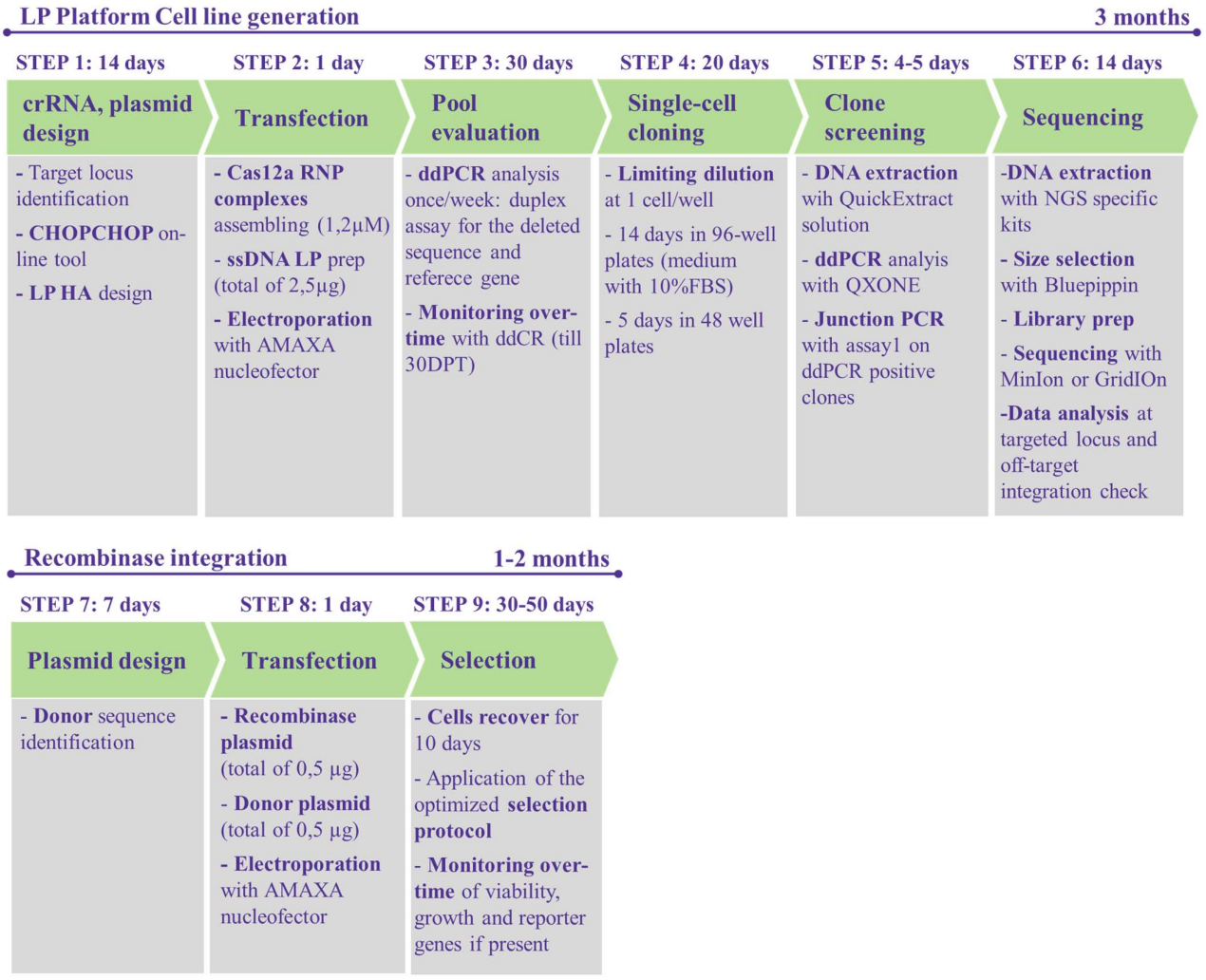


Figure 42. Schematic summary and timelines of the workflow for the CRISPR/Cas-recombinase KI experiments. (STEP 1) The crRNA design is conducted taking advantage of the online tool CHOPCHOP and once selected, the Homology Arms of the LP are properly designed to maximize the efficiency. (STEP 2) The cells are transfected with the Cas12a RNP complexes and the LP ssDNA to be inserted using the AMXA nucleofector. (STEP 3) The generated pools are monitored over time with ddPCR using specific assays for the LP sequence. (STEP 4) Single-cell cloning is needed to obtain modified cells at the targeted site, so limiting dilution is applied in 96-well plates. The cells are let recover and grow to be then scaled up in 48-well plates. (STEP 5) A small-scale clone screening is carried out to identify the on-target modified clones. A first screening is performed with ddPCR in order to select only the cells bearing the construct in their genome. Then, the obtained positive clones are re-tested with junction PCR to identify the cells with the on-target integration of the LP. (STEP 6) Nanopore long-reads sequencing is performed, and the resulting data are analysed focusing on the target locus and possible off-target insertions. This step is needed to characterize the generated LP platform cell line. (STEP 7) The donor sequence that we want to insert at the LP site is designed according to the experimental demand. (STEP 8) Both the plasmids for the recombinase enzyme and the donor sequence are transfected with AMAXA nucleofector into the platform cell line. (STEP 9) the antibiotic selection is applied to obtain the cells with the donor plasmid inserted at the single LP site. The cells are monitored for viability, cell growth and for the reporter genes if present in the donor construct.

Both strategies could be efficiently applied in our workflows and the choice to use one instead of the other relies on the final aim of the engineering. Indeed, the LP could be placed in one of the well-known hotspots in the CHO genome such as HPRT[41,42], GRIK[42], Fer1L4[43], C12orf35[42], and ADGRL4[44] thus ensuring a constant and high expression of the gene of interest. But if we need to frequently change the site of integration or need an ad hoc modification of specific loci, it could be more efficient in terms of costs and time to proceed with a direct integration of the desired sequence.

Taken together, our experimental results showcase the remarkable potential of CRISPR/Cas technology in our process, demonstrating efficiencies that align with those reported in the current literature for large deletions and integrations[26]. By exploiting the cutting-edge advancements in CRISPR/Cas toolkit, such as employing the Cas12a enzyme and the RNP complex, we have successfully optimized highly efficient protocols using our in-house CHO cell line. These protocols can now be seamlessly integrated into our CLD workflows, offering diverse and adaptable applications tailored to the specific needs of our Company. The integration of our molecules with CRISPR/Cas-based approach in well-characterized hotspots could mitigate the risk of generating non-productive clones. Furthermore, this approach streamlines the steps involved in cell characterization and selection of modified clones. Equally, these precise tools can be leveraged to enhance the performance of host cells themselves. The literature gives a wide range of studies wherein the modulation of productivity-related genes has successfully increased the yield of therapeutic proteins. Targets for the KO could be various and focused on company needs: glycosylation enzymes could be deleted to remove or gain specific glycosylation patterns on our molecules[47], genes involved in the epigenetic machinery could be impaired to avoid transgene silencing[48], and genes of protein negatively impacting the purification of our molecules, such as HCPs, could be eliminated to increase the yield in the final purification steps[49,50]. On the other hand, CRISPR/Cas KI could be adapted to enhance the expression of genes positively impacting protein production. The strategies to adopt could be different and adjustable on the intended target: regulatory elements such as strong promoters or enhancers can be inserted upstream of the target gene or candidate genes exogenous copies can be directly integrated into highly-expressing loci for increasing the overall mRNA production. The targets can be identified in specific cellular pathways which can alleviate the endoplasmic reticulum (ER) stress upon massive protein translation[51], or in genes involved at different stages of the secretory pathway, passing from the correct processing of secretory proteins in the ER[61] to the lipidic metabolism which impacts the fluidity of the membranes and consequently the secretion capabilities [62].

In conclusion, the vast range of available opportunities presents exciting prospects for the customization of our CHO cell lines through the implementation of our CRISPR/Cas-optimized protocols. By exploring and implementing CRISPR-based strategies, we can enhance the production of our biopharmaceuticals, thereby proactively responding to evolving market demands. In addition, this approach allows us to enhance the overall efficiency and reproducibility of the entire process, ensuring that we remain at the forefront of innovation and market responsiveness.

6. References

1. Matasci M, Hacker DL, Baldi L, Wurm FM. Recombinant therapeutic protein production in cultivated mammalian cells: current status and future prospects. *Drug Discov Today Technol*. 2008;5(2-3). doi:10.1016/J.DDTEC.2008.12.003

2. Walsh G. Biopharmaceutical benchmarks 2014. *Nat Biotechnol 2014 3210*. 2014;32(10):992-1000. doi:10.1038/nbt.3040

3. Wurm FM. Production of recombinant protein therapeutics in cultivated mammalian cells. *Nat Biotechnol*. 2004;22(11):1393-1398. doi:10.1038/NBT1026

4. Richelle A, Lewis NE. Improvements in protein production in mammalian cells from targeted metabolic engineering. *Curr Opin Syst Biol*. 2017;6:1-6. doi:10.1016/j.coisb.2017.05.019

5. Fischer S, Otte K. CHO Cell Engineering for Improved Process Performance and Product Quality . *Cell Cult Eng*. Published online 2019:207-250. doi:10.1002/9783527811410.ch9

6. De Jesus M, Wurm FM. Manufacturing recombinant proteins in kg-ton quantities using animal cells in bioreactors. *Eur J Pharm Biopharm*. 2011;78(2):184-188. doi:10.1016/j.ejpb.2011.01.005

7. Datta P, Linhardt RJ, Sharfstein ST. An 'omics approach towards CHO cell engineering. *Biotechnol Bioeng*. 2013;110(5):1255-1271. doi:10.1002/BIT.24841

8. Stolfa G, Smonskey MT, Boniface R, et al. CHO-Omics Review: The Impact of Current and Emerging Technologies on Chinese Hamster Ovary Based Bioproduction. *Biotechnol J*. 2018;13(3):1700227. doi:10.1002/BIOT.201700227

9. Tihanyi B, Nyitray L. Recent advances in CHO cell line development for recombinant protein production. *Drug Discov Today Technol*. 2020;38:25-34. doi:10.1016/J.DDTEC.2021.02.003

10. Lai T, Yang Y, Ng SK. Advances in Mammalian Cell Line Development Technologies for Recombinant Protein Production. *Pharmaceuticals*. 2013;6(5):579. doi:10.3390/PH6050579

11. Würtele H, Little KCE, Chartrand P. Illegitimate DNA integration in mammalian cells. *Gene Ther 2003 1021*. 2003;10(21):1791-1799. doi:10.1038/sj.gt.3302074

12. Lattenmayer C, Trummer E, Schriebl K, et al. Characterisation of recombinant CHO cell lines by investigation of protein productivities and genetic parameters. *J Biotechnol*. 2007;128(4):716-725. doi:10.1016/J.JBIOTEC.2006.12.016

13. He L, Winterrowd C, Kadura I, Frye C. Transgene copy number distribution profiles in recombinant CHO cell lines revealed by single cell analyses. *Biotechnol Bioeng*. 2012;109(7):1713-1722. doi:10.1002/BIT.24428

14. Lin HT, Okumura T, Yatsuda Y, Ito S, Nakauchi H, Otsu M. Application of Droplet Digital PCR for Estimating Vector Copy Number States in Stem Cell Gene Therapy. *Hum Gene Ther Methods*. 2016;27(5):197. doi:10.1089/HGTB.2016.059

15. Berman J, Heredia N, Regan J, et al. *Droplet Digital TM PCR : High-Resolution Copy Number Variation Analysis*. Accessed January 21, 2023. https://www.bio-rad.com/webroot/web/pdf/lsr/literature/Bulletin_6475.pdf

16. Aeschlimann SH, Graf C, Mayilo D, et al. Enhanced CHO Clone Screening: Application of Targeted Locus Amplification and Next-Generation Sequencing Technologies for Cell Line Development. *Biotechnol J*. 2019;14(7):1-11. doi:10.1002/biot.201800371

17. Biorad. Droplet Digital ™ PCR Applications Guide. *Biorad*. Published online 2018:145. http://www.bio-rad.com/webroot/web/pdf/lsr/literature/Bulletin_6407.pdf

18. Dorai H, Csirke B, Scallon B, Ganguly S. Correlation of Heavy and Light Chain mRNA Copy Numbers to Antibody Productivity in Mouse Myeloma Production Cell Lines. *http://www.liebertpub.com/hyb*. 2006;25(1):1-9. doi:10.1089/HYB.2006.25.1

19. Veith N, Ziehr H, MacLeod RAF, Reamon-Buettner SM. Mechanisms underlying epigenetic and transcriptional heterogeneity in Chinese hamster ovary (CHO) cell lines. *BMC Biotechnol*. 2016;16(1). doi:10.1186/S12896-016-0238-0

20. Lao-Gonzalez T, Bueno-Soler A, Duran-Hernandez A, et al. Screening and selection strategy for the establishment of biosimilar to trastuzumab-expressing CHO-K1 cell lines. *AMB Express*. 2021;11(1):1. doi:10.1186/S13568-020-01157-6

21. Dhiman H, Campbell M, Melcher M, Smith KD, Borth N. Predicting favorable landing pads for targeted integrations in Chinese hamster ovary cell lines by learning stability characteristics from random transgene integrations. *Comput Struct Biotechnol J*. 2020;18:3632. doi:10.1016/J.CSBJ.2020.11.008

22. Zhang JH, Shan LL, Liang F, Du CY, Li JJ. Strategies and Considerations for Improving Recombinant Antibody Production and Quality in Chinese Hamster Ovary Cells. *Front Bioeng Biotechnol*. 2022;10:321. doi:10.3389/FBIOE.2022.856049/BIBTEX

23. De Vree PJP, De Wit E, Yilmaz M, et al. Targeted sequencing by proximity ligation for comprehensive variant detection and local haplotyping. *Nat Biotechnol 2014 3210*. 2014;32(10):1019-1025. doi:10.1038/nbt.2959

24. Stadermann A, Gamer M, Fieder J, et al. Structural analysis of random transgene integration in CHO manufacturing cell lines by targeted sequencing. *Biotechnol Bioeng*. 2022;119(3):868-880.

doi:10.1002/BIT.28012

25. Shin SW, Lee JS. Optimized CRISPR/Cas9 strategy for homology-directed multiple targeted integration of transgenes in CHO cells. *Biotechnol Bioeng*. 2020;117(6):1895-1903. doi:10.1002/bit.27315

26. Amiri S, Adibzadeh S, Ghanbari S, et al. CRISPR-interceded CHO cell line development approaches. *Biotechnol Bioeng*. Published online 2023. doi:10.1002/BIT.28329

27. Mojica FJM, Díez-Villaseñor C, García-Martínez J, Soria E. Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J Mol Evol*. 2005;60(2):174-182. doi:10.1007/S00239-004-0046-3

28. Mojica FJM, Díez-Villaseñor C, García-Martínez J, Almendros C. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology*. 2009;155(Pt 3):733-740. doi:10.1099/MIC.0.023960-0

29. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science (80-)*. 2012;337(6096):816-821. doi:10.1126/science.1225829

30. Deltcheva E, Chylinski K, Sharma CM, et al. CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature*. 2011;471(7340):602-607. doi:10.1038/NATURE09886

31. Pattanayak V, Lin S, Guilinger JP, Ma E, Doudna JA, Liu DR. High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nat Biotechnol*. 2013;31(9):839-843. doi:10.1038/NBT.2673

32. Jiang F, Doudna JA. CRISPR–Cas9 Structures and Mechanisms. *https://doi.org/101146/annurev-biophys-062215-010822*. 2017;46:505-529. doi:10.1146/ANNUREV-BIOPHYS-062215-010822

33. Zetsche B, Gootenberg JS, Abudayyeh OO, et al. Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System. *Cell*. 2015;163(3):759-771. doi:10.1016/J.CELL.2015.09.038

34. Paul B, Montoya G. CRISPR-Cas12a: Functional overview and applications. *Biomed J*. 2020;43(1):8-17. doi:10.1016/J.BJ.2019.10.005

35. Stella S, Mesa P, Thomsen J, et al. Conformational Activation Promotes CRISPR-Cas12a Catalysis and Resetting of the Endonuclease Activity. *Cell*. 2018;175(7):1856-1871.e21. doi:10.1016/J.CELL.2018.10.045

36. Cong L, Ran FA, Cox D, et al. Multiplex genome engineering using CRISPR/Cas systems. *Science (80-)*. 2013;339(6121):819-823. doi:10.1126/SCIENCE.1231143/SUPPL_FILE/PAPV2.PDF

37. Handel EM, Cathomen T. Zinc-finger nuclease based genome surgery: it's all about specificity. *Curr Gene Ther*. 2011;11(1):28-37. doi:10.2174/156652311794520120

38. Porter SN, Levine RM, Pruett-Miller SM. A Practical Guide to Genome Editing Using Targeted Nuclease Technologies. *Compr Physiol*. 2019;9(2):665-714. doi:10.1002/CPHY.C180022

39. Castro NG, Bjelic J, Malhotra G, Huang C, Hasan Alsaffar S. Molecular Sciences Comparison of the Feasibility, Efficiency, and Safety of Genome Editing Technologies. Published online 2021. doi:10.3390/ijms221910355

40. Deng P, Carter S, Fink K. Design, Construction, and Application of Transcription Activation-Like Effectors. *Methods Mol Biol*. 2019;1937:47-58. doi:10.1007/978-1-4939-9065-8_3

41. Kawabe Y, Komatsu S, Komatsu S, et al. Targeted knock-in of an scFv-Fc antibody gene into the hprt locus of Chinese hamster ovary cells using CRISPR/Cas9 and CRIS-PITCh systems. *J Biosci Bioeng*. 2018;125(5):599-605. doi:10.1016/J.JBIOSC.2017.12.003

42. Zhao M, Wang J, Luo M, et al. Rapid development of stable transgene CHO cell lines by CRISPR/Cas9-mediated site-specific integration into C12orf35. *Appl Microbiol Biotechnol*. 2018;102(14):6105-6117. doi:10.1007/s00253-018-9021-6

43. FEARY M, YOUNG RJ, MOFFAT M, CASPERSON GF, JONES HL, ZHANG L. MULTI-SITE SPECIFIC INTEGRATION CELLS FOR DIFFICULT TO EXPRESS PROTEINS. Published online August 23, 2018.

44. Baumann M, Gludovacz E, Sealover N, et al. Preselection of recombinant gene integration sites enabling high transcription rates in CHO cells using alternate start codons and recombinase mediated cassette exchange. *Biotechnol Bioeng*. 2017;114(11):2616-2627. doi:10.1002/BIT.26388

45. Mulukutla BC, Mitchell J, Geoffroy P, et al. Metabolic engineering of Chinese hamster ovary cells towards reduced biosynthesis and accumulation of novel growth inhibitors in fed-batch cultures. *Metab Eng*. 2019;54:54-68. doi:10.1016/J.YMBEN.2019.03.001

46. Ley D, Pereira S, Pedersen LE, et al. Reprogramming AA catabolism in CHO cells with CRISPR/Cas9 genome editing improves cell growth and reduces byproduct secretion. *Metab Eng*. 2019;56:120-129. doi:10.1016/J.YMBEN.2019.09.005

47. Ha TK, Hansen AH, Kildegaard HF, Lee GM. Knockout of sialidase and pro-apoptotic genes in Chinese hamster ovary cells enables the production of recombinant human erythropoietin in fed-batch cultures. *Metab Eng*. 2020;57:182-192. doi:10.1016/J.YMBEN.2019.11.008

48. Jia YL, Guo X, Lu JT, Wang XY, Qiu L Le, Wang TY. CRISPR/Cas9-mediated gene knockout for

DNA methyltransferase Dnmt3a in CHO cells displays enhanced transgenic expression and long-term stability. *J Cell Mol Med*. 2018;22(9):4106-4116. doi:10.1111/jcmm.13687

49. Kol S, Ley D, Wulff T, et al. Multiplex secretome engineering enhances recombinant protein production and purity. *Nat Commun 2020 111*. 2020;11(1):1-10. doi:10.1038/s41467-020-15866-w

50. Dovgan T, Golghalyani V, Zurlo F, et al. Targeted CHO cell engineering approaches can reduce HCP-related enzymatic degradation and improve mAb product quality. *Biotechnol Bioeng*. 2021;118(10):3821-3831. doi:10.1002/BIT.27857

51. Wang W, Zheng W, Hu F, et al. Enhanced Biosynthesis Performance of Heterologous Proteins in CHO-K1 Cells Using CRISPR-Cas9. *ACS Synth Biol*. 2018;7(5):1259-1268. doi:10.1021/ACSSYNBIO.7B00375/SUPPL_FILE/SB7B00375_SI_001.PDF

52. Roth TL, Puig-Saus C, Yu R, et al. Reprogramming human T cell function and specificity with non-viral genome targeting. *Nat 2018 5597714*. 2018;559(7714):405-409. doi:10.1038/s41586-018-0326-5

53. Schubert MS, Thommandru B, Woodley J, et al. Optimized design parameters for CRISPR Cas9 and Cas12a homology-directed repair. *Sci Rep*. 2021;11(1):1-15. doi:10.1038/s41598-021-98965-y

54. Jain M, Koren S, Miga KH, et al. Nanopore sequencing and assembly of a human genome with ultra-long reads. *Nat Biotechnol 2018 364*. 2018;36(4):338-345. doi:10.1038/nbt.4060

55. Clappier C, Böttner D, Heinzelmann D, et al. Deciphering integration loci of CHO manufacturing cell lines using long read nanopore sequencing. *N Biotechnol*. 2023;75(December 2022):31-39. doi:10.1016/j.nbt.2023.03.003

56. Swarts DC, Jinek M. Cas9 versus Cas12a/Cpf1: Structure–function comparisons and implications for genome editing. *Wiley Interdiscip Rev RNA*. 2018;9(5):e1481. doi:10.1002/WRNA.1481

57. Strohkendl I, Saifuddin FA, Rybarski JR, Finkelstein IJ, Russell R. Kinetic Basis for DNA Target Specificity of CRISPR-Cas12a. *Mol Cell*. 2018;71(5):816-824.e3. doi:10.1016/J.MOLCEL.2018.06.043

58. Kim D, Kim J, Hur JK, Been KW, Yoon SH, Kim JS. Genome-wide analysis reveals specificities of Cpf1 endonucleases in human cells. *Nat Biotechnol 2016 348*. 2016;34(8):863-868. doi:10.1038/nbt.3609

59. Pallarès-Masmitjà M, Ivančić D, Mir-Pedrol J, et al. Find and cut-and-transfer (FiCAT) mammalian genome engineering. *Nat Commun*. 2021;12(1):1-9. doi:10.1038/s41467-021-27183-x

60. Low BE, Hosur V, Lesbirel S, Wiles M V. Efficient targeted transgenesis of large donor DNA into

multiple mouse genetic backgrounds using bacteriophage Bxb1 integrase. *Sci Reports 2022 121*. 2022;12(1):1-16. doi:10.1038/s41598-022-09445-w

61. Le Fourn V, Girod PA, Buceta M, Regamey A, Mermod N. CHO cell engineering to prevent polypeptide aggregation and improve therapeutic protein secretion. *Metab Eng*. 2014;21:91-102. doi:10.1016/j.ymben.2012.12.003

62. Budge JD, Knight TJ, Povey J, et al. Engineering of Chinese hamster ovary cell lipid metabolism results in an expanded ER and enhanced recombinant biotherapeutic protein production. *Metab Eng*. 2020;57(October 2019):203-216. doi:10.1016/j.ymben.2019.11.007