



# A pooled CRISPR/AsCpf1 screen using paired gRNAs to induce genomic deletions in Chinese hamster ovary cells

Valerie Schmieder<sup>a,b,1</sup>, Neža Novak<sup>a,b,1</sup>, Heena Dhiman<sup>a,b</sup>, Ly Ngoc Nguyen<sup>a,b</sup>, Evgenija Serafimova<sup>a,b</sup>, Gerald Klanert<sup>b</sup>, Martina Baumann<sup>b</sup>, Helene Fastrup Kildegaard<sup>c</sup>, Nicole Borth<sup>a,b,\*</sup>

<sup>a</sup> Department of Biotechnology, BOKU University of Natural Resources and Life Sciences, Muthgasse 18, Vienna, Austria

<sup>b</sup> acib GmbH, Austrian Centre of Industrial Biotechnology, Muthgasse 11, Vienna, Austria

<sup>c</sup> The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Building 220, Kemitorvet, Kgs. Lyngby, Denmark

## ARTICLE INFO

### Keywords:

Genetic screen  
CRISPR/AsCpf1  
paired gRNAs  
genomic deletion  
Chinese hamster ovary cells

## ABSTRACT

Chinese hamster ovary (CHO) cells are the most widely used host for the expression of therapeutic proteins. Recently, significant progress has been made due to advances in genome sequence and annotation quality to unravel the black box CHO. Nevertheless, in many cases the link between genotype and phenotype in the context of suspension cultivated production cell lines is still not fully understood. While frameshift approaches targeting coding genes are frequently used, the non-coding regions of the genome have received less attention with respect to such functional annotation. Importantly, for non-coding regions frameshift knock-out strategies are not feasible. In this study, we developed a CRISPR-mediated screening approach that performs full deletions of genomic regions to enable the functional study of both the translated and untranslated genome.

An *in silico* pipeline for the computational high-throughput design of paired guide RNAs (pgRNAs) directing CRISPR/AsCpf1 was established and used to generate a library tackling process-related genes and long non-coding RNAs. Next generation sequencing analysis of the plasmid library revealed a sufficient, but highly variable pgRNA composition. Recombinase-mediated cassette exchange was applied for pgRNA library integration rather than viral transduction to ensure single copy representation of pgRNAs per cell. After transient AsCpf1 expression, cells were cultivated over two sequential batches to identify pgRNAs which massively affected growth and survival. By comparing pgRNA abundance, depleted candidates were identified and individually validated to verify their effect.

## 1. Introduction

Chinese hamster ovary (CHO) cell lines are the main expression systems for the production of biopharmaceuticals due to the ability of this cell line to grow to high cell densities in suspension and to produce high yields of complex recombinant proteins harboring human-like post-

translational modifications [1,2]. Although CHO is being used for the production of recombinant biotherapeutics since 1987 [3], the link between non-coding transcripts as well as coding genes and observed phenotypes is still not fully understood [4]. One reason is the late availability of the genomic sequence for this cell line in 2011 only [5]. Nevertheless, researchers in the CHO community achieved tremendous

**Abbreviations:** AsCpf1, Cpf1 from *Acidaminococcus* sp BV3L6; Cas9, CRISPR-associated protein 9; CHO, Chinese hamster ovary; Cpf1, CRISPR-associated protein 1 from *Prevotella* and *Francisella*; CPM, counts per million reads mapped; CRISPR, Clustered Regularly Interspaced Short Palindromic Repeats; dCas9, deactivated Cas9; DE, differentially expressed; DOWN-TTS, downstream transcription termination site; DR, differentially represented; EpoFc, Erythropoietin Fc fusion protein; EV, empty vector; FACS, fluorescence activated cell sorting; FDR, false discovery rate; FC, fold change; gRNA, guide RNA; GS, glutamine synthetase; lncRNA, long non-coding RNA; ncGene, non-coding gene; NGS, next generation sequencing; NTC, no template control; oligo, oligonucleotide; PAM, protospacer adjacent motif; PCA, principal component analysis; pgRNA, paired gRNA; Qp, specific productivity; RMCE, recombinase-mediated cassette exchange; sgRNA, single guide RNA; TMM, trimmed mean of M values; UP-TSS, upstream transcription start site; VCD, viable cell density;  $\mu$ , growth rate.

\* Corresponding author.

E-mail address: [nicole.borth@boku.ac.at](mailto:nicole.borth@boku.ac.at) (N. Borth).

<sup>1</sup> These authors contributed equally to the study.

<https://doi.org/10.1016/j.btre.2021.e00649>

Received 21 December 2020; Received in revised form 6 June 2021; Accepted 16 June 2021

Available online 20 June 2021

2215-017X/© 2021 Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

progress in terms of the reference genome sequence quality in recent years [6,7,8]. These studies enabled the investigation of the genome, transcriptome as well as the epigenome [9,10,11,12,13]. So far, gene annotation and assignment of the corresponding function are based on gene homology, *de novo* prediction or expression level [9,7]. This may not adequately reflect the function of the respective gene in CHO cells that are cultured in protein free medium in stirred-suspension bioreactors. In general, genetic studies are focused on the protein-coding part of the genome [14] covering approximately 2-3% of the entire sequence in human cells, which is also a likely estimate for CHO cells [15,16]. Untranslated transcripts, such as microRNAs or long non-coding RNAs (lncRNA), have an enormous effect on the phenotype, however [17,18,19]. This demonstrates the need also for an improved understanding of the function of the non-protein coding genome in the context of recombinant protein production and bioprocessing.

Genetic screens performed using the Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) system and the CRISPR-associated protein 9 (Cas9) are the state-of-the-art approach to correlate the genome with the phenotype, as reviewed by Shalem et al. [20]. CRISPR/Cas9 screens can be conducted in array or pooled formats where the latter is preferred for genome-wide functional studies due to less complex handling and application in high-throughput screens. Therefore, a subset of proteins [21] or even the entire proteome [22] are covered by an appropriate number of guide RNAs (gRNAs) to create gene disruptions. By this approach, those genes that were responsible for the observed phenotype can be identified. Typically, gRNAs are designed to target close to the translation start site in the coding gene sequence so as to cause frameshifts or an early stop codon, thus disrupting the function of the resulting protein. Per definition, this approach is limited to coding genes, however.

Recently, Zhu et al. applied paired gRNAs (pgRNAs) with Cas9 to induce larger genomic deletions allowing the characterization of untranslated lncRNAs in a genome-scale pooled screen in human cells [23]. In this study, pgRNAs were obtained as a single-stranded oligonucleotide (oligo) pool. Additional cloning steps were required to achieve the delivery of a defined gRNA pair into a unique cell. Separate U6 promoters drove proper gRNA transcription along with the necessary scaffold sequences, which together build the single gRNA (sgRNA). Here, the alternative CRISPR-associated protein in *Prevotella* and *Francisella* (Cpf1, also known as Cas12a) from *Acidaminococcus* sp BV3L6 (AsCpf1) offers unique advantages over Cas9 as AsCpf1 requires a shorter sgRNA and has an additional RNase activity allowing the transcription of sgRNAs from a tandem array, thereby meeting high-throughput oligo pool synthesis requirements and enabling pgRNA expression from a single U6 promoter [24]. Additionally, AsCpf1 recognizes an AT-rich protospacer adjacent motif (PAM) (5'-TTTN-gRNA) [25], which could be of additional benefit when targeting the non-coding part of the genome due to its increased AT content. As state-of-the-art, CRISPR/Cas9-mediated screens in pooled format are conducted using viral transduction with a low level of multiplicity of infection to deliver a single defined gRNA into each unique cell. However, viral transduction is highly dependent on obtained virus titers and the transduction efficiency of the cell line being screened [26]. Further, the application of viral systems is subject to specific biosafety requirements, which are not given in every facility and thus can limit the use of such a screen setup [27,28]. An alternative approach is the use of recombinase mediated cassette exchange (RCME) which is typically used to target recombinant genes into highly active loci in the genome that enable stable and efficient protein production ([29]; L. [30]). If applied in library screening, the high transcriptional activity of the target locus would be of less importance, however, the ability to generate a library that has a homogeneous genetic background would reduce the occurrence of off-target effects that might be caused not by the integrated pgRNAs, but by the disruption of a gene at the integration site in some instances. In addition, due to the selection screen for replacement of the initially inserted target gene (i.e. GFP) one would

have assurance that each cell contains only a single integrant. The main drawback of the RMCE method is the low efficiency of exchange within a cell pool, which typically lies between 0.5-2% [31] and thus necessitates efficient and high throughput selection methods, such as cell sorting, to ensure covering the entire library within the cell pool isolated.

In pooled CRISPR screens, one unique gRNA is stably integrated per cell. Such stably integrated guides are essential for later discrimination of the effective gene editing events and function as barcodes for the subsequent read out. Once a desired phenotype was selected after genetic gene alteration, the introduced changes can be investigated by targeted amplification of the stably integrated gRNAs followed by next generation sequencing (NGS) and subsequent quantification of gRNA specific abundance. An important consideration in designing such a screen is the certainty that only a single pgRNA is present in each cell. Other considerations include a good representation of the library. Here, a good rule of thumb is the use of three or more redundant gRNAs per target and a high coverage within the cell pool, depending on the efficiency of integration [32,33,34,35,36].

In this study, we developed a protocol that enables the study of the entire coding and non-coding CHO genome by generating and validating a small-scale deletion approach with an alternative AsCpf1 pgRNAs pooled screening strategy. We established an *in silico* pipeline for the high-throughput computational design of pgRNAs guiding CRISPR/AsCpf1 and designed an approach for high-throughput cloning and delivery of the pgRNA library into cells at single copy frequency, using RCME. An initial small-scale pgRNA library was created including 2,348 pgRNAs. pgRNAs targeting 500 previously identified differentially expressed (DE) lncRNAs [18], 45 published bioprocess-related protein-coding genes ([37,38,39,40]b; [41,42,43,44]), the recombinant Erythropoietin FC fusion protein (EpoFc) and 164 randomers as non-targeting controls were designed. Effects of library synthesis, cloning and delivery into cells were monitored by NGS to evaluate variations on pgRNA representation and integrity. Recombinase-mediated cassette exchange (RMCE) was applied to maintain unique pgRNA integration into the host cell genome. After generation of genomic deletions and cultivation, Illumina sequencing of pgRNAs from genomic DNA was performed to determine the abundance of specific pgRNAs that had an immediate impact on growth. Ten pgRNAs were identified as possible growth reducing hits and were validated by transiently transfecting those pairs individually into CHO along with AsCpf1 to analyze changes in CHO growth behavior.

## 2. Material and methods

### 2.1. Computational high-throughput design of pgRNAs

An *in silico* pipeline was established for the high-throughput design of CHO-specific pgRNAs navigating AsCpf1. For each target, the 2 kb flanking region upstream and downstream was evaluated for possible guides by using 27 nt windows starting at each consecutive position. The sequences with PAM sites (TTT[AGC]) at the 5'-end were selected as guides and mapped across the whole genome [8] to identify off-targets using Bowtie alignment software (v1.2.2). Alignment was performed with options for prioritizing finding all possible alignments over speed (-h) and report all alignments per each read (-a) without allowing any mismatch (-m=0) in a seed length of 18 nt (-l=18). The guides with non-unique alignment or individual GC content not within 30-70% or with GC content of the complete fragment not within 41±3% were filtered out. The guide sequences containing recognition sites for restriction enzymes *BsmBI* or *BbsI* (CGTCTC, GAAGAC) or repetition of the same nucleotide more than 5 times were removed from the list as well. Non-redundant guides were further ranked based on the cumulative scores that were computed based on the values found in Table S1. Preference was given to 'TTTA' sequence at the PAM site for both mates of the gRNA pair, alignment with zero mismatch and GC content of the fragment being equal to 0.41 [45]. Each target was then reported with

top four scoring pgRNA sequences.

## 2.2. Construction of CRISPR/AsCpf1 pgRNA plasmid library

The CRISPR/AsCpf1 pgRNA deletion library targeting non- as well as coding genes consisted of 2,348 different sequences. Each oligo had a length of 133 bp and the library was ordered as a ~12,000 single-stranded oligo pool (CustomArray). Phusion High-Fidelity DNA Polymerase (Thermo Fisher Scientific) was used to generate the double-stranded oligo pool. Therefore, the following reaction was performed in quadruplicates. Per replicate, a 50  $\mu$ L PCR reaction was mixed according to the manual using 3.5 ng original single-stranded oligos as template. The reaction was run with 98°C 30 sec; 10 cycles of 98°C 5 sec, 60°C 10 sec, 72°C 15 sec; 72°C 5 min. PCR products were pooled and cleaned by DNA Clean & Concentrator®-5 (Zymo Research). Amplification was checked on a 2% Midori Green Advance agarose (Biozym) TAE gel. Next, the double-stranded oligo pool was cloned into pRMCE-EpoFc-UP or -DOWN (Fig. S1) via *BsmBI* (*Esp3I*) FastDigest™ (Thermo Fisher Scientific) by Golden Gate assembly as described previously [46]. After heat-shock transformation of NEB® 10-beta Competent *Escherichia coli* cells (NEB) (5  $\mu$ L Golden Gate assembly mix (in theory 10 ng) added to 50  $\mu$ L competent cells), a serial dilution of the transformation reactions was prepared to investigate cloning and transformation efficiencies for both plasmid versions. A cut-off of a pgRNA representation of at least 250-fold with a reasonable low re-ligation rate was set. The rest of the transformation reactions was used to inoculate bacterial overnight cultures to obtain plasmid DNA pools (UP- and DOWN-pgRNA) using the EndoFree Plasmid Maxi Kit (Qiagen) as described in the manual.

## 2.3. Plasmid library quality assessment

Library quality was assessed by pgRNA sequence rescue followed by in-depth sequencing. 5  $\mu$ g of UP-pgRNAs and DOWN-pgRNAs plasmid pool DNA were double digested with *NdeI* (NEB) and *PciI* (NEB) to isolate pgRNA DNA fragments (loss of 39 pgRNAs due to restriction recognition sites within oligo sequences, which served as restriction control). After restriction, samples were run on a 2% EtBr agarose (Biozyme) TAE gel and 164 nt long fragments of interest were extracted from the gel. DNA was extracted from the gel slices using the Hi Yield® Gel/PCR DNA Fragment Extraction Kit (Süd-Laborbedarf Gauting) followed by a clean-up step using DNA Clean & Concentrator®-5 (Zymo Research). Samples were sequenced by Vienna Biocenter Core Facilities GmbH with a HiSeq2500 instrument (Illumina) paired end 125 bp (V4) in rapid mode as spike-in.

## 2.4. Processing sequencing results of plasmid library

Raw reads were trimmed using Trimmomatic (v. 0.36) [47] to remove low quality reads and sequencing adapters. To remove all possible reads with only backbone plasmid sequences, reads were mapped to the pRMCE-EpoFc backbone plasmids using Bowtie2 (v. 2.26) [48] and only unmapped reads were used in subsequent steps. Cutadapt (v. 1.16) [49] and Trimmomatic (v. 0.36) were used to remove ends of reads belonging to plasmid backbones. Additionally, forward sequencing reads were excluded due to incomplete pgRNA spanning by trimming the backbone sequence with Trimmomatic and setting the minimal length to 85 bp. Reads were mapped to pgRNA sequences used for oligo library synthesis with BMap (v. 38.25) [50] (minid=0.95, maxindel=10). Variants were called with FreeBayes [51] (v. 1.2.0) by setting the ploidy to 1 (-p), minimal alternate fraction (-F) to 0.15 and minimal alternate count (-C) to 5. Called variants were split into four groups - insertions, deletions, substitutions and complex, which were defined as mutations spanning > 3 nt on the reference and having multiple alternative sequences. Share of variants was calculated as percent of each variant group of the total variant count. Variant

frequency was calculated as share of the reads supporting the variant, excluding perfect reads and sequencing errors on the specific position. Reads not supporting variants were extracted using VariantBam [52] and together with reads without any mismatches were counted per pgRNA. The counts were imported into R (v. 3.6.0.) and normalized using the edgeR (v. 3.26.8) trimmed mean of M values (TMM) normalization method [53]. The  $R^2$  between normalized counts of UP-pgRNAs and DOWN-pgRNAs was calculated using the cor() function with 'Spearman's rank correlation coefficient'. pgRNA representation was shown as percent of pgRNA containing variants, not containing variants and not represented at all in the plasmid pool.

## 2.5. Generation of pre- and post-screening CHO pools

RMCE to generate the pre-screening CHO pools – stable integration of one pgRNA per cell – was performed as described previously [54]. After antibiotic enrichment of exchanged cells (exchange of CD4 with EpoFc along with pgRNA), cold capture surface staining against CD4 and EpoFc was applied in combination with fluorescence-activated cell sorting (FACS) to enrich for CD4-/EpoFc+ cells in three subsequent rounds according to published protocols [55]. Copy number determination via qPCR against the U6 promoter, which is driving pgRNA transcription, was applied using FUT8 as an internal reference. Additionally, gene expression analysis to determine fold change (FC) in CD4 and EpoFc expression was done to monitor successful cassette exchange. Pre-screening pgRNA CHO pools were applied to plasmid-based transfection for transient AsCpf1 or Cas9 enzyme expression. 7 days post transfection altered cell pools were used to inoculate a 1<sup>st</sup> batch in spin bioreactor tubes. The 2<sup>nd</sup> batch was inoculated using cells from day 5 of the 1<sup>st</sup> batch. Growth and productivity were measured daily and were used to calculate specific productivities (Qp) and growth rates ( $\mu$ ) ([40] b). Detailed protocols for each step can be obtained from the supplementary information.

## 2.6. Cell library sample preparation for sequencing

Stably integrated pgRNAs were amplified using gDNA sample pools and checked for pgRNA abundance by NGS. For each sample, 4×50  $\mu$ L PCR reactions were performed using 2 U Phusion High-Fidelity DNA Polymerase (Thermo Fisher Scientific), the stable\_pgRNA\_pool primer pair and 825 ng gDNA as template. Remaining PCR reagents were added according to the manual. The following PCR cycling conditions were applied: 98°C 30 sec; 25 cycles of 98°C 5 sec, 65°C 10 sec, 72°C 15 sec; 72°C 5 min. PCR reactions were pooled per sample prior clean up using the DNA Clean & Concentrator®-5 kit (Zymo Research) as described in the manual. Quality of 137 bp pgRNA amplicons was monitored by NanoDrop (Thermo Fisher Scientific). NGS was performed on a HiSeq2500 instrument (Illumina) spike-in as paired end 125 bp (V4).

## 2.7. Computational analysis of screen results

Low quality reads, sequencing adapters and PCR primers were removed from the raw reads using Trimmomatic (v. 0.36). Reads were mapped to pgRNA sequences used for oligo library synthesis with BMap (v. 38.25); minid=0.95, maxindel=10) and variants were analyzed using FreeBayes (v. 1.2.0; -F 0.15 -C 5 -p 1). Reads without variants were counted per pgRNA using a custom script. Counts were normalized using the edgeR (v. 3.26.8) TMM method and used for principal component analysis (PCA). To estimate common dispersion, the estimateDisp() function was used with all the non-targeting pgRNA counts. The results of the differential representation (DR) were acquired using the glmFit() and glmLRT() functions [53]. Depleted pgRNAs between AsCpf1 and Cas9 samples were defined by false discovery rate (FDR) value < 0.25 and log<sub>2</sub>FC < 0. Enriched pgRNAs were defined by the same FDR value and log<sub>2</sub>FC > 0. Intersections of depleted pgRNA in different datasets were analyzed using the UpSetR package (v. 1.4.0)

[56]. To check for potential off-targets and additional target sites of selected pgRNAs, BLAST command line application was used against the latest Chinese hamster genome assembly (CriGri-PICR) [8] as database with the following parameters: -evalue 1000 -word\_size 7 -perc\_identity 80 -ungapped. The results were additionally filtered for alignment length > 22 and identity > 90%.

## 2.8. pgRNA hit validation

To validate the hits from the screen, respective pgRNAs were cloned individually into pY010(AsCpf1) and transfected as mentioned above into CHO-K1 glutamine synthetase (GS)<sup>-</sup> Herceptin producing cells. Growth after transfection in a 5 day batch experiment was analyzed by daily measurements on ViCELL XR Cell Counter (Beckman Coulter) and an in-house developed R package ViCellR ([40]b). gDNA was isolated on day 4 as described above and deletion PCR was performed using the following cycling conditions: 98°C 2 min; 35 cycles of 98°C 10 sec, 62°C 30 sec, 72°C 1 min 30 sec; 72°C 10 min.

## 3. Results

### 3.1. High-throughput design of CRISPR/AsCpf1 pgRNA

An *in silico* pipeline was prepared for high-throughput design of CRISPR/AsCpf1 pgRNAs considering the availability of an efficient CHO- as well as an AsCpf1-specific tool. Per target, the 2 kb region upstream of the transcription start site (UP-TSS) as well as the 2 kb downstream region of the transcription termination site (DOWN-TTS) were screened for any 5'-TTTN PAM sequence (Fig. S2). The 23 nt sequence downstream of a PAM was defined as the gRNA sequence. UP-TSS and DOWN-TTS gRNAs without off-target activity were paired in the oligo frame for pool synthesis while avoiding redundancy. In the absence of gRNAs with preferred parameters for off-target activity, those with mismatches outside the seed were used. Each frame with non-redundant gRNAs was further scored based on the criteria suggested by Kim et al. for a high-throughput profiling of CRISPR/Cpf1 activity [45]. The scoring focus was set on low off-target activity, giving preference to TTTA sequence at the PAM site and an optimal GC content of 41% (Table S1). Genome editing capacity of pgRNAs designed by the *in silico* tool was evaluated in wet lab experiments for the top three pairs for FUT8, RAD21 and CHD4 as protein coding genes as well as the highest scoring pair for three non-coding genes (ncGene) after transfection in CHO-K1 cells (Fig. S2).

This tool was applied to design a small-scale CHO-specific AsCpf1 pgRNA library containing four pgRNAs per target. Hence, the resulting oligo library consisted of 2,348 pgRNA sequences (of which 2,328 are unique and were used in the downstream bioinformatic analysis) targeting 250 up- and down DE intergenic lncRNAs, 45 process-related coding genes, and the recombinant EpoFc model protein. As non-targeting controls, 164 random pairs not binding anywhere in the CHO genome were used (Table S2). While cumulative scores of the top four pgRNAs range from 1.85 to 3.00 for all coding genes, those for lncRNAs range from as low as 0.05 to 3.00 (Table 1, Table S3-5).

**Table 1**

Summary of CRISPR/AsCpf1 pgRNA scores against coding and non-coding targets.

Score	pgRNAs against DE lncRNAs		pgRNAs against coding genes	
	Counts	Percent of total	Counts	Percent of total
3.0	547	27.4	59	32.8
2.5 – 3.0	999	50.0	104	57.8
2.0 – 2.5	358	17.9	14	7.8
< 2.0	96	4.8	3	1.7
Total	2000	100.0	180	100.0

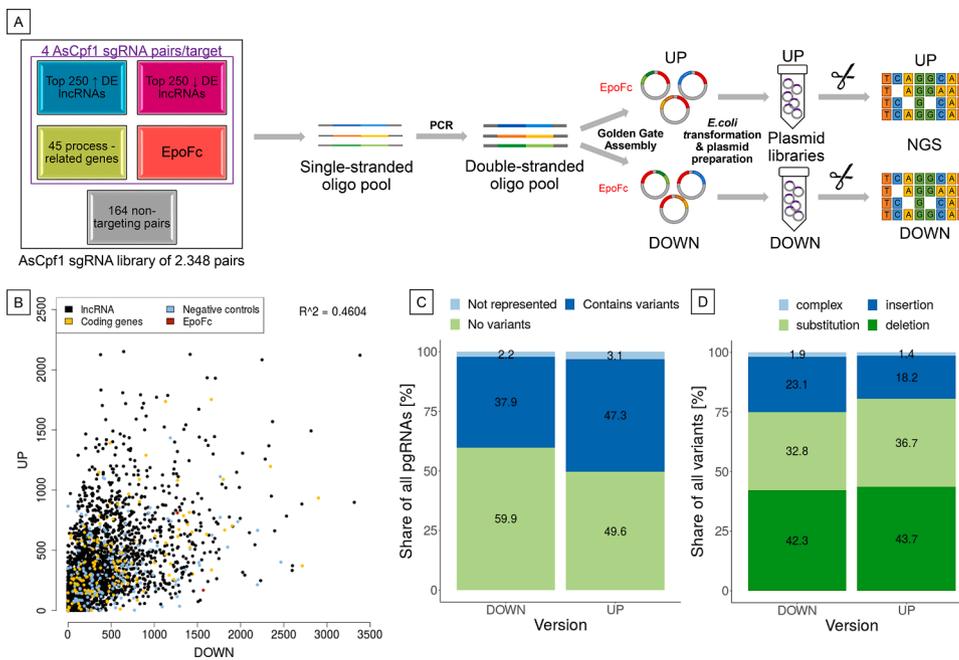
### 3.2. Plasmid library cloning and quality assessment

The pgRNA library was received from CustomArray Inc. as a 12,000 single-stranded oligo pool (5-fold representation per pgRNA), PCR amplified with a low number of cycles and further cloned into two versions of the delivery plasmid backbone pRMCE-EpoFc either up- or downstream of the model protein EpoFc yielding the two plasmid libraries UP-pgRNAs and DOWN-pgRNAs. Cloning and transformation efficiencies were monitored during the plasmid pool preparation resulting in a 250- (UP-pgRNAs) and 300-fold (DOWN-pgRNAs) representation per pgRNA with a backbone re-ligation rate of < 1% for both plasmid pools. An extensive analysis on the prepared plasmid pools was performed to assess pgRNA sequence quality as well as pgRNA representation per sequence/target using rescued pgRNA DNA fragments (Fig. 1a). After NGS raw data processing, obtained reads ( $1.97 \times 10^6$  for UP-pgRNAs and  $1.27 \times 10^6$  for DOWN-pgRNAs) were mapped to the sequences used for the oligo pool synthesis and sequence errors that were present in more than 15% of the reads were called as variants. The reads containing the variants were removed while the rest of the mapping reads (perfect matching and those with only sequencing errors) were counted per pgRNA. Next, the correlation of the two library versions (Fig. 1b) and the representation of the pgRNAs (Fig. 1c) were analyzed. In UP-pgRNAs, 3.1% of the total 2,328 pgRNAs were not present at all in the pool and for DOWN-pgRNAs 2.2% of the pairs were missing only. For the remaining pgRNAs, normalized counts were ranging from 0.9 to 3382.0 showing high fluctuation in pgRNA representation. The normalized counts of both library versions were plotted against each other to estimate the similarity, however, no correlation was detected ( $R^2 = 0.4245$ ). Furthermore, the sequence quality was assessed by analyzing the presence as well as type of variants in the plasmid pool. Per variant, frequency of reads supporting each variant was calculated as ratio between reads supporting the variant and total reads at that position. The frequency of the reads that supported each variant had a wide range and was skewed towards the lower values (Fig. S3a). Variants were defined as insertions, deletions, substitutions or complex variants. The latter had to have multiple alternative sequences or span more than 3 nt. For UP-pgRNAs 47.3% and for DOWN-pgRNAs 37.9% of pgRNAs had at least one and up to four variants in the plasmid pool (Fig. 1c, Fig. S3b, Table S6). However, 97.4% of these pgRNAs in UP-pgRNAs and 98.2% in DOWN-pgRNAs had reads without variants in the sequence pool. Deletions occurred the most often with 42.3% and 43.7% of all variants for UP-pgRNAs and DOWN-pgRNAs, respectively. The second most frequent variant type were substitutions (32.8% and 36.7%), insertions (23.1% and 18.2%) and complex variants were less abundant (1.9% and 1.4%) (Fig. 1d). In summary, variation in pgRNA representation and sequence variants that accumulated during the preparation of the plasmid pools were observed.

### 3.3. Generation of the pre- and post-screening CHO pools

Next, UP-pgRNAs and DOWN-pgRNAs plasmid pools were used to generate pre-screening CHO cell pools containing stably integrated pgRNAs based on a previously published landing pad cell line [54]. For RMCE, antibiotic selection, FACS, and all cell cultivation steps an appropriate number of cells were used to maintain a pgRNA representation of at least 100-fold (Table S7). In summary, two replicates of cell pools carrying the UP-pgRNA or DOWN-pgRNA library were generated (UP and DOWN), respectively. pgRNA and EpoFc co-integration was maintained by RCME and enrichment of CD4-/EpoFc+ was conducted via FACS. Successful cassette exchange and single copy integration per cell were verified via copy number determination targeting the U6 promoter and gene expression analysis against CD4 as well as EpoFc (Fig. S4).

Generated UP and DOWN pre-screening CHO cell pools underwent transient plasmid based AsCpf1 expression to initiate the formation of genomic deletions. In parallel, UP and DOWN cell pools were treated



**Fig. 1. Generation of AsCpf1 pgRNA library targeting lncRNAs and assessment of plasmid pool quality.** (A) Graphical depiction of AsCpf1 pgRNA library composition, pgRNA library generation in pRMCE-EpoFc-UP and DOWN version, and plasmid pool quality assessment by next-generation (NGS) sequencing. (B) Correlation plot of normalized pgRNA counts of the two different plasmid versions. (C) Bar plot of pgRNA sequence quality in the plasmid pools. Identification of pgRNAs with correct sequences, sequences containing variants or pgRNAs not present in the library version. (D) Bar plot of determined pgRNA sequence variants in plasmid libraries.

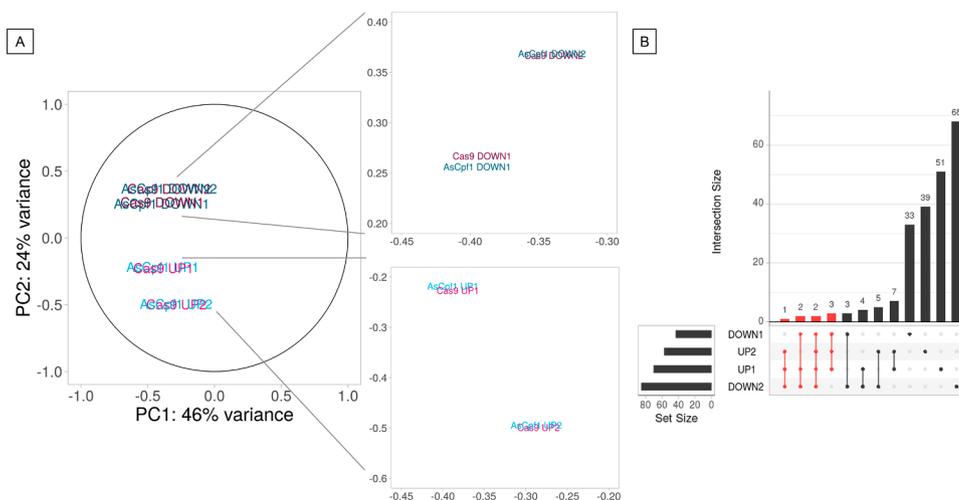
with a Cas9 expressing plasmid to serve as an unmodified head-to-head control, which underwent the same process steps as AsCpf1 edited cells.

As the lncRNAs targeted in the library were differentially expressed during batch culture, we hypothesized that some of them would have a direct effect on growth rate – either enhancing or inhibiting. We therefore performed a rapid screen for strong effects on growth, by subjecting the pools to two subsequent rounds of batch cultivation in duplicate spin tubes. No obvious phenotypic change between AsCp1 and Cas9 treated cell pools was observed (Fig. S5). To prove functionality of the alternative CRISPR/AsCpf1 pgRNA deletion screen strategy, the entire workflow was evaluated using a single gRNA pair targeting FUT8 as a model gene. Stably FUT8 pgRNA expressing pre-screening CHO pools were generated and characterized. Further, cell pools were applied to transient AsCpf1 and Cas9 expression and analyzed with regards to fucosylation over cultivation (17d). The summarized results are provided in the supplement (Fig. S6).

### 3.4. Analysis of NGS data revealing changes in pgRNA abundance

Characterization of pgRNA presence in the genome and changes in

abundance of AsCpf1 treated cell pools relative to Cas9 samples were performed via NGS of pgRNAs. Therefore, gDNA from cell pools of the exponential phase of the 2<sup>nd</sup> batch (day 5) was isolated from a number of cells that ensured a coverage of 280-fold for each pgRNA and used as template for pgRNA-targeted amplification via PCR. Sequencing data were analyzed as described for plasmid NGS samples and reads without variants were counted per pgRNA. Counts were transformed to counts per million (CPM) reads and pgRNAs were used for further analysis in case pgRNA CPM values were greater than 5 in at least three of the eight sample sets. 1,945 pgRNAs were left after this step, which represents 83.5% of the original library and shows a lower pgRNA representation than in the plasmid pools. Log<sub>2</sub> CPM counts were used for PCA to generate a plot indicating that samples were clustering according to their parental pre-screening CHO pool. CRISPR enzyme treated post-screening CHO pools were clustering more closely than biological duplicates from different CHO pre-screening pools. Results indicated a high variability in the representation of the pgRNAs (Fig. 2a). The data were split into four sets according to the parental cell pool (DOWN1, DOWN2, UP1, UP2) and filtered for CPM > 5 in at least one of the two samples (AsCpf1 or Cas9). In summary, 1,579 – 1,806 (67.8% - 77.6%) pgRNAs



**Fig. 2. Assessment of differentially represented (DR) pgRNAs in cell pools after genomic deletion and phenotypic screening.** (A) Principal component analysis (PCA) plot of log<sub>2</sub> counts per million (CPM) in pgRNA abundance of UP (light) and DOWN (dark) cell pools applied to transient AsCpf1 (pink) or Cas9 (blue) expression. (B) Intersections of significantly depleted pgRNAs between AsCpf1 and Cas9 treated samples in four analyzed datasets. The bottom left bar plot shows the number of significantly depleted pgRNAs (logFC < 0, false discovery rate (FDR) < 0.25) in each dataset. The bars on the right show the number of depleted pgRNAs in each intersection or those found in one of the datasets only. pgRNAs present in at least three of the datasets or lncRNAs targeted by at least two pgRNAs within minimum two datasets were defined as potential hits (red).

were left after this filtering step. Between 43 as the lowest (DOWN1) and 85 as the maximum (DOWN2) hits of depleted pgRNAs with a FDR < 0.25 were found in each set (Table S8-11). To improve the statistical power of potential hits, intersections of depleted pgRNAs in sets were analyzed (Fig. 2b). This showed two pgRNAs were depleted in all four sets (lnc4 P4 and lnc7 P3) and six were depleted in three of the sets (lnc1 P1, lnc2 P4, lnc3 P1, lnc4 P3, lnc5 P4 and lnc6 P3). pgRNAs that were only depleted in two sets were also selected as hits if a different pair targeting the same lncRNA was depleted in three or four sets (lnc6 P4, lnc7 P4). This way, ten candidate pgRNAs – all targeting lncRNAs only – were identified and the corresponding reads from Cas9 treated control pools were analyzed in more detail. Candidate pgRNAs not identified in all four datasets were not detectable in neither the Cas9 nor AsCpf1 missing dataset(s) or represented in the sample set(s) (Table S12). Additionally, normalized counts of identified candidate pgRNAs were re-evaluated in the UP and DOWN plasmid pool sequencing results showing no bias due to under- or overrepresentation of those particular pgRNAs in generated plasmid libraries (Fig. S7). However, between 25 (DOWN1) and 60 (DOWN2) significant hits of enriched pgRNAs were identified in each set. Three pgRNA hits build intersections in two datasets having no additional pair tackling the same target in the remaining datasets, respectively (Fig. S8). Thus, no enriched candidate for further hit validation was selected.

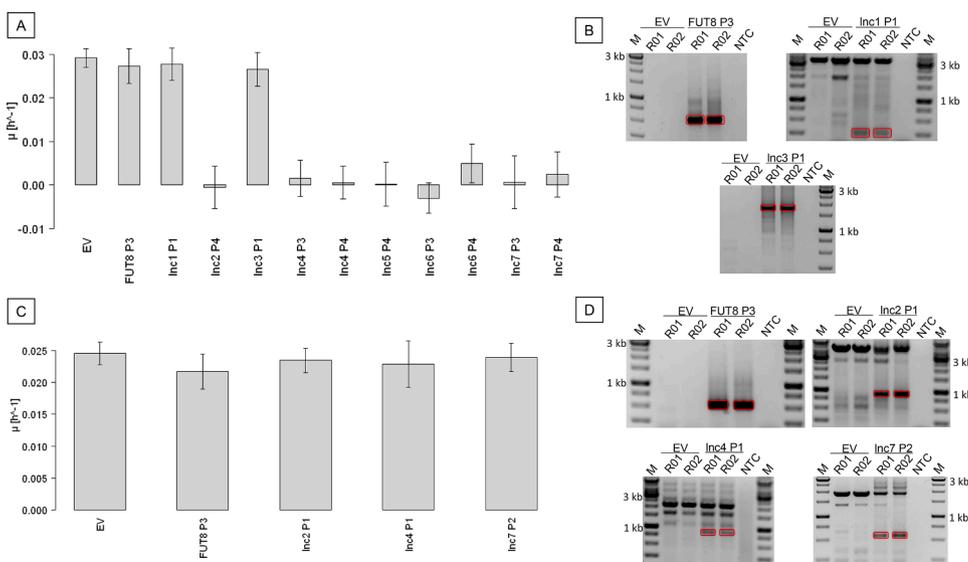
### 3.5. Validation of growth perturbing properties of candidate pgRNAs

To validate the effect on growth of the ten depleted candidate pgRNAs, each of the pairs was individually cloned into the pY010 (AsCpf1) vector (Fig. S9) and transfected into a CHO-K1  $GS^-$  Herceptin producer cell line. VCD and viability were measured daily to calculate  $\mu$  during batch cultivation over 5 days (Fig. 3a, Fig. S10) and 4 days after transfection, deletion PCR was performed using gDNA as template (Fig. 3b). FUT8 P3 was used as a control in addition to an empty vector (EV) control, to assess the effect of stress from cell treatment on growth and to serve as a positive control for deletion PCR. For lnc1 P1 and lnc3 P1, amplicons of the appropriate size were observed, however, no significant effect on growth was detectable for edited cell pools from those samples. All other pgRNAs showed a significant effect on growth after transfection of the individual pgRNAs compared to the controls. For lnc2 P4, lnc4 P3, lnc4 P4, lnc5 P4, lnc6 P4, lnc7 P3, lnc7 P4 reduced growth rates were obtained. In case of lnc6 P3, negative growth rates were calculated. Deletion of the corresponding lncRNAs could not be

confirmed on DNA level, as no bands of the appropriate size were visible after deletion PCR. All pgRNAs strongly affecting growth were undertaken a follow-up BLAST alignment resulting in multiple hits per tested pgRNA against the latest Chinese hamster genome assembly and transcriptome [8] (Table S13-16). To investigate whether the observed effect on growth triggered by eight of the ten hits was due to specific lncRNA deletion or a result of tackling the additional pgRNA target sites, further pgRNAs targeting the respective lncRNA were tested in a similar experiment (Fig. 3c, Fig. S11). No significant effect on cell growth was observed for newly designed pgRNAs, although deletion PCR confirmed efficient lncRNA removal from a larger part of the cell populations (Fig. 3d).

## 4. Discussion

Functional genetic screens based on CRISPR/Cas9 are powerful tools to study the correlation between the genome and the observed phenotype. Here, we present an alternative CRISPR/AsCpf1 screening strategy, which targets both the coding and non-coding genome by the formation of genomic deletions using paired gRNAs. Other than in the traditional frameshift approach, the application of gRNA pairs and the full deletion of the intermediary genomic sequence allows to study also the untranslated part of the genome. Even in the case of protein-coding targets, genes are completely removed from the cell system reducing the risk for potential cell stress as a result of the expression of truncated or nonsense proteins [57]. Additionally, CRISPR/AsCpf1 pgRNA delivery was performed by RMCE to avoid the need for viral transduction as biosafety requirements have to be fulfilled when working with viruses. Moreover, RMCE offers the opportunity to co-transport larger DNA fragments encoding for instance model molecules or even CRISPR enzymes. In this study, RMCE was used to stably co-integrate the genetic information for a recombinant product to mimic conditions CHO cells are facing in biopharmaceutical industry more closely [5]. Nowadays, numerous CRISPR libraries for Cas9 or dCas9 are commercially available and online tools are accessible to design and generate customized gRNA libraries. However, this is not the case for AsCpf1 and for pgRNAs to perform deletions. CHO-specific libraries are not listed for Cas9 and AsCpf1, respectively, although design tools for CHO and Cas9 are available [58]. Therefore, pgRNAs meeting the requirements described by Kim et al. [45] were designed by the in-house established *in silico* pipeline. Testing the pipeline with three coding genes and three non-coding transcript targets showed different editing efficiencies



**Fig. 3. pgRNA hit validation from screening.** (A) Determination of growth rates ( $\mu$ ) from batch cultivation (days 1-5) after transient transfection of AsCpf1 along with candidate pgRNAs selected from screening. (B) Deletion PCR performed on day 4 of batch cultivation for lnc1 P1 and lnc3 P1 treated samples on genomic level, which had no effect on growth. (C) Determination of growth rates ( $\mu$ ) from batch cultivation (days 1-8) after transient transfection of AsCpf1 along with additional pgRNAs with minimal off-targets tackling lncRNA verified to be growth perturbing in (A) by the application of less scored pgRNAs. (D) Deletion PCR performed on day 4 of batch cultivation for lnc2 P1, lnc4 P1 and lnc7 P2 treated samples on genomic level, which had no effect on growth. Batch experiments and deletion PCRs performed in biological replicates (R01/R02). Transfection with empty vector (EV) and FUT8 served as controls without change in growth phenotype. Successful target deletion indicated by red box. M – ladder; NTC – no template control.

depending on the pgRNA being used. No obvious correlation between genome editing capacity and *in silico* score was observed for pgRNAs generated from the in-house pipeline, however, only a small number of pairs was tested. Further, there is a chance that InDels or other random mutations can occur due to the high genomic variability of CHO cells, therefore making the altered sequence inaccessible for the formation of a genomic deletion. However, these specific mutations that typically occur only in a subpopulation of cells, would have to be in precisely the same cells that received those specific pgRNAs to be of consequence. Nevertheless, this highlighted the need for multiple pgRNAs per gene (or genomic region) in the later screen to increase the certainty of target alteration. As published previously, genomic deletions of up to 150 kb were demonstrated to be size-independent and are rather influenced by individual gRNA on-target efficiency [24]. For the small-scale oligo library, 2,348 different CRISPR/AsCpf1 pgRNAs were designed tackling 45 process-related coding genes, 500 untranslated intergenic lncRNAs, one model gene - each with four pairs per target - as well as 164 non-targeting pgRNAs. Overall, designed pgRNAs reached high *in silico* scoring values. Variations in the GC content and the presence of 5'-TTTS PAM sequences are the primary reasons for reduced scoring values. Additionally, a tendency towards gRNA selection with mismatch mapping outside the seed sequence of lncRNA pgRNAs was observed. lncRNA gene bodies are found in intronic regions of coding genes as well as intergenic [59,60], whereas repetitive sequences are more common [61,62,5], which makes it more difficult to identify unique gRNAs.

Plasmid library quality assessment via NGS analysis - which is not often done for genetic screens - showed sufficient representation of pgRNA sequences as only a small percentage of pgRNAs were missing. By isolating pgRNA containing plasmid fragments rather than PCR amplifying additional polymerase-mediated sequence errors were avoided. Interestingly, no correlation in pgRNA representation was observed between the two generated plasmid versions emphasizing that variability accumulated during plasmid pool preparations. The two versions were separated after PCR amplification of the single-stranded oligo pool, hence the fluctuation most likely developed during cloning into delivery plasmids and amplification in *E.coli*. Additionally, the quality was affected by variants in the sequences, which were found for approximately 40% of pgRNAs and were mostly deletions, but also substitutions, insertions and complex variants. The latter occurred very rarely showing that detected mutations seem to be rather simple without larger rearrangements. Sequence errors in the customized single-stranded oligo pool or generated during PCR amplification - although a high-fidelity Phusion polymerase was used with minimal replication cycles - might be the root cause for the observed sequence variants. An error rate of 1 out of 150 bases to 1 out of 200 bases was stated by the oligo library supplier, which accumulates over the course of oligo synthesis. By the application of AsCpf1 instead of Cas9, defined gRNA pairs - significantly shorter in length - were synthesized as a single array which was processed intracellularly by AsCpf1 RNase activity prior genome editing, therefore avoiding the accumulation of even more sequence errors during plasmid library preparation by additional cloning steps. However, the majority of pgRNAs were present with high coverage and good sequence quality in the prepared plasmid pools.

As the state-of-the-art library delivery method, viral transduction is applied with a low level of multiplicity of infection assuming the integration of one gRNA copy per cell. However, this presumption is based on statistically equal gRNA distribution within the so generated pre-screening cell pool. In contrast, pgRNA integration by RMCE is more controlled. In theory, pgRNAs are integrated at the same spot in the landing pad cell line by exchange with a previously introduced marker gene leading to comparable pgRNA transcription. However, a risk of additional random integration of the pgRNAs cannot be excluded based on the experiments performed here. Still, overall, RMCE offers a number of advantages over viral based delivery methods, including higher certainty of single integration events, less genomic background and straightforward analysis enrichment/depletion by PCR.

Usually, transgenic Cas9-expressing cell lines are used as a starting point for pooled CRISPR-associated screens increasing the chance for successful genome editing [63]. At the same time, continuous CRISPR enzyme presence in the cells triggers the risk of unwanted off-target effects, which was prevented in this study by transient CRISPR protein addition. Introduction of genomic alterations is followed by a screening step, which is typically based on a desired phenotypic trait and can be either conducted by positive or negative selection [32,64]. In this experiment, a viability-based negative selection screen was applied by cultivation of edited cells under standard batch bioprocess conditions for 17 days to identify targets with an immediate effect on growth. Treatment of pre-screening cell pools with transient Cas9 expression, which should not result in genome editing, served as a head-to-head control in this study. Furthermore, proof-of-principle of the alternative AsCpf1 deletion screen strategy was confirmed by targeting FUT8 by one individual pgRNA.

NGS results of pgRNA abundance in post-screening cell pools comparing AsCpf1 samples with the Cas9 controls revealed high variability in pgRNA representation between samples even when looking at genetically unmodified controls. In addition to the two plasmid libraries used, UP and DOWN cell pools were split into two replicates after transfection of pgRNA plasmid pools for RMCE and further treated separately. The high heterogeneity is reflected in the generated PCA plot. Here, replicates are clustering further apart than AsCpf1 and Cas9 edited samples within the same replicate. Thus, the implementation and representation of the library seems to have a stronger effect on clustering than the deletion of coding genes and non-coding transcripts. Moreover, a standard differential representation analysis, whereby replicates are used to estimate gene-wise dispersion, was not possible due to high fluctuation between biological replicates. To overcome this, non-targeting pgRNAs were used to estimate the common dispersion between samples and DR pgRNAs were called in each replicate separately. In the screened cells, representation of pgRNAs was overall lower than in the initial plasmid libraries. This was observed for AsCpf1 altered as well as Cas9 unmodified cell pools. The loss of representation was most likely triggered by the low efficiency of genome integration of the library. The application of replicates as well as maintaining a high pgRNA coverage of at least 100-fold aimed to reduce these effects, but may benefit from more stringent coverage targets in future screen. Based on pgRNA abundance analysis between AsCpf1 and Cas9, only depleted pgRNAs were identified as intersections between multiple datasets. As described in the literature, the improvement of CHO growth characteristics is difficult [65] and negative selection screens based on growth reveal a higher number of depleted hits [66,64]. To identify a pgRNA with an effect on growth, a longer screen would have been necessary, to enable cells to increase or to decrease in numbers sufficiently to be detectable by the bioinformatic analysis. Accordingly, the ten pgRNAs targeting seven different lncRNAs that were determined in the screen were all depleted. The counts of the candidates were very low or zero in the AsCpf1 treated samples implying that cells receiving these pgRNAs were almost completely removed from the pool after genomic alteration and cultivation over 17 days under batch conditions. No pgRNA tackling coding genes was tracked as significantly depleted presumably because no essential gene was targeted in our library.

Separate wet lab experiments of computationally identified hits validated eight out of ten pgRNAs to have a significantly negative effect on growth in a CHO Herceptin producer cell line. However, pgRNAs, which have additional target sites for one or both gRNAs, were confirmed to be the root cause for growth-perturbing properties in CHO rather than the specific deletion of the underlying DE lncRNA targets. In case of a lack of four gRNAs per target site during library design, which are unique as well as perfectly mapping or mapping with mismatches outside the seed sequence, the tool allowed to include gRNAs with non-unique target sites for pairing. lncRNAs are often located in repetitive genomic regions making it more complicated to design unique gRNAs, thus increasing the risk to create unwanted effects. Hence, growth-

perturbing properties from validated hits were most likely caused by fractionation of the genome due to thousands of additional gRNA editing sites in CHO. A different kind of bias was observed in a previous CRISPR screen targeting lncRNA splice sites [67], where some of the identified lncRNA hits were found in regions with copy number amplification or to be hosted by essential protein-coding genes. Cleavage of these regions led to higher rates of DNA damage translating into decreased cell viability and loss-of-function of the lncRNA hosting protein [68]. Such false positives are difficult to avoid when studying the less explored non-coding part of the genome, however, controlling and recognizing them is important and different strategies depending on the screen type have been proposed [69].

## 5. Conclusions

In this study, a small-scale CRISPR/AsCpf1 deletion screen in CHO cells was established that targets both coding and non-coding regions of the genome. The overall feasibility of the proposed workflow was proven by the performance and validation of hits from a first small-scale deletion screening. Based on our experience, an initial screening experiment to get hands on experience and to identify potential pitfalls is highly recommended by the authors before moving forward to a genome-wide study. The following critical steps were identified during the preparation, execution and analysis, which deserve special attention in the design and performance of genome-scale screens:

- Both positive and negative controls, that affect the desired phenotype that will be screened for, need to be included in the screen design. Additionally, validated controls, which are specific for the cell type, are of advantage.
- The plasmid pool library must be of high quality, which should be ensured by good oligo pool synthesis, use of high proof-reading capacity polymerase and low number of cycles for PCR. Quality control by NGS sequencing is recommended before proceeding with the screen.
- Representative library delivery is a major challenge. More efficient methods are needed for this step which can provide high library coverage. Furthermore, each pre-screening cell pool that is used as starting material for the study needs to be sequenced, so that screened cells can be directly compared against their starting material rather than against the designed library.
- To ensure full representation of the library, multiple transfections and the generation of pools with high coverage is recommended. Therefore, cell line specific characteristics, the library size and the screening strategy have to be taken into account. In general, a higher coverage will increase the robustness and statistical power of the screen.
- Introduction of genomic deletions is followed by a screening step for phenotypes of interest. Here rare events need to be found, which is difficult especially in view of the high technical variability originating from preparation steps. During screen design, this should be considered and the use of highly effective methods to enrich for cells with the desired phenotype, such as cell sorting, is recommended.

## Availability of data and materials

All CRISPR/AsCpf1 gRNA sequences as well as sequencing, PCR, and qPCR primer sequences are listed in Table S17. CRISPR/AsCpf1 pgRNAs oligo library sequence information is available in Table S2. Detailed bioinformatic scripts can be found on GitHub (<https://github.com/NezaNo/CRISPRscreen>) and NGS raw data at ENA under accession number **PRJEB35876**.

## Authors' contributions

NN and VS contributed equally to the study. NN and VS were

responsible for experimental design. HD conducted the *in silico* design of AsCpf1 pgRNA. VS performed the wet lab part of the screen and NN the bioinformatic analysis. ES & NN conducted the lab validation of screen hits. LNN and MB provided support for generating the stable cell line. GK, MB, HFK and NB conceived and supervised the study. The manuscript has been read and approved by all authors.

## Declaration of Competing Interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgements

The authors thank Krishna Motheramgari for language editing and Thomas Hackl and Ivana Buric for technical help. Next Generation Sequencing was performed by the NGS Facility at the Vienna Biocenter Core Facilities (VBCF), member of the Vienna Biocenter (VBC), Austria. This work was supported by the "eCHO Systems" ITN Ph.D. Program funded by the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No 642663; and by the Austrian Center of Industrial Biotechnology, a COMET K2 competence center of the Austrian Research Promotion Agency FFG. The COMET center acib: Next Generation Bioproduction is funded by BMVIT, BMDW, SFG, Standortagentur Tirol, Government of Lower Austria und Vienna Business Agency in the framework of COMET - Competence Centers for Excellent Technologies. The COMET-Funding Program is managed by the Austrian Research Promotion Agency FFG.

## Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:[10.1016/j.btre.2021.e00649](https://doi.org/10.1016/j.btre.2021.e00649).

## References

- [1] T. Amann, V. Schmieder, H. Fastrup Kildegaard, N. Borth, M.R. Andersen, Genetic engineering approaches to improve posttranslational modification of biopharmaceuticals in different production platforms, *Biotechnol. Bioeng.* 116 (2019) 2778–2796, <https://doi.org/10.1002/bit.27101>.
- [2] G. Walsh, Biopharmaceutical benchmarks 2014, *Nat Biotech* 32 (2014) 992–1000, <https://doi.org/10.1038/nbt.3040>.
- [3] J.B. Griffiths, A. Electricwala, Production of tissue plasminogen activators from animal cells, in: *Vertebrate Cell Culture I, Advances in Biochemical Engineering/Biotechnology*. Springer Berlin Heidelberg (1987) 147–166.
- [4] G. Stolfa, M.T. Smonskey, R. Boniface, A.-B. Hachmann, P. Gulde, A.D. Joshi, A. P. Pierce, S.J. Jacobia, A. Campbell, CHO-Omics Review: The Impact of Current and Emerging Technologies on Chinese Hamster Ovary Based Bioproduction, *Biotechnol J* 13 (2018), 1700227, <https://doi.org/10.1002/biot.201700227> e.
- [5] X. Xu, H. Nagarajan, N.E. Lewis, S. Pan, Z. Cai, X. Liu, W. Chen, M. Xie, W. Wang, S. Hammond, M.R. Andersen, N. Neff, B. Passarelli, W. Koh, H.C. Fan, Jianbin Wang, Y. Gui, K.H. Lee, M.J. Betenbaugh, S.R. Quake, I. Famili, B. O. Palsson, Jun Wang, The genomic sequence of the Chinese hamster ovary (CHO)-K1 cell line, *Nat. Biotechnol.* 29 (2011) 735–741, <https://doi.org/10.1038/nbt.1932>.
- [6] K. Brinkrolf, O. Rupp, H. Laux, F. Kollin, W. Ernst, B. Linke, R. Kofler, S. Romand, F. Hesse, W.E. Budach, S. Galosy, D. Müller, T. Noll, J. Wienberg, T. Jostock, M. Leonard, J. Grillari, A. Tauch, A. Goesmann, B. Helk, J.E. Mott, A. Pühler, N. Borth, Chinese hamster genome sequenced from sorted chromosomes, *Nat Biotech* 31 (2013) 694–695, <https://doi.org/10.1038/nbt.2645>.
- [7] N.E. Lewis, X. Liu, Y. Li, H. Nagarajan, G. Yerganian, E. O'Brien, A. Bordbar, A. M. Roth, J. Rosenbloom, C. Bian, M. Xie, W. Chen, N. Li, D. Baycin-Hizal, H. Latif, J. Forster, M.J. Betenbaugh, I. Famili, X. Xu, J. Wang, B.O. Palsson, Genomic landscapes of Chinese hamster ovary cell lines as revealed by the *Cricetulus griseus* draft genome, *Nat. Biotechnol.* 31 (2013) 759–765, <https://doi.org/10.1038/nbt.2624>.
- [8] O. Rupp, M.L. MacDonald, S. Li, H. Dhiman, S. Polson, S. Griep, K. Heffner, I. Hernandez, K. Brinkrolf, V. Jadhav, M. Samoudi, H. Hao, B. Kingham, A. Goesmann, M.J. Betenbaugh, N.E. Lewis, N. Borth, K.H. Lee, A reference genome of the Chinese hamster based on a hybrid assembly strategy, *Biotechnol. Bioeng.* 115 (2018) 2087–2100, <https://doi.org/10.1002/bit.26722>.
- [9] J. Becker, M. Hackl, O. Rupp, T. Jakobi, J. Schneider, R. Szczepanowski, T. Bekel, N. Borth, A. Goesmann, J. Grillari, C. Kaltschmidt, T. Noll, A. Pühler, A. Tauch, K. Brinkrolf, Unraveling the Chinese hamster ovary cell line transcriptome by next-

- generation sequencing, *J. Biotechnol.* 156 (2011) 227–235, <https://doi.org/10.1016/j.jbiotec.2011.09.014>.
- [10] J. Feichtinger, I. Hernández, C. Fischer, M. Hanscho, N. Auer, M. Hackl, V. Jadhav, M. Baumann, P.M. Krempf, C. Schmidl, M. Farlik, M. Schuster, A. Merkel, A. Sommer, S. Heath, D. Rico, C. Bock, G.G. Thallinger, N. Borth, Comprehensive genome and epigenome characterization of CHO cells in response to evolutionary pressures and over time, *Biotechnol Bioeng* 113 (2016) 2241–2253, <https://doi.org/10.1002/bit.25990>.
- [11] O. Rupp, J. Becker, K. Brinkrolf, C. Timmermann, N. Borth, A. Pühler, T. Noll, A. Goemann, Construction of a Public CHO Cell Line Transcript Database Using Versatile Bioinformatics Analysis Pipelines, *PLoS One* 9 (2014), <https://doi.org/10.1371/journal.pone.0085568>.
- [12] A. Singh, H.F. Kildegaard, M.R. Andersen, An Online Compendium of CHO RNA-Seq Data Allows Identification of CHO Cell Line-Specific Transcriptomic Signatures, *Biotechnol. J.* 13 (2018), 1800070, <https://doi.org/10.1002/biot.201800070>.
- [13] A. Wippermann, S. Klausning, O. Rupp, S.P. Albaum, H. Büntemeyer, T. Noll, R. Hoffrogge, Establishment of a CpG island microarray for analyses of genome-wide DNA methylation in Chinese hamster ovary cells, *Appl Microbiol Biotechnol* 98 (2014) 579–589, <https://doi.org/10.1007/s00253-013-5282-2>.
- [14] E. Hartenian, J.G. Doench, Genetic screens and functional genomics using CRISPR/Cas9 technology, *FEBS J.* 282 (2015) 1383–1393, <https://doi.org/10.1111/febs.13248>.
- [15] T.R. Gregory, Synergy between sequence and size in Large-scale genomics, *Nat Rev Genet* 6 (2005) 699–708, <https://doi.org/10.1038/nrg1674>.
- [16] International Human Genome Sequencing Consortium, Initial sequencing and analysis of the human genome, *Nature* 409 (2001) 860–921, <https://doi.org/10.1038/35057062>.
- [17] S. Fischer, T. Buck, A. Wagner, C. Ehrhart, J. Giancaterino, S. Mang, M. Schäd, S. Mathias, A. Aschrafi, R. Handrick, C. Otte, A functional high-content miRNA screen identifies miR-30 family to boost recombinant protein production in CHO cells, *Biotechnol. J.* 9 (2014) 1279–1292, <https://doi.org/10.1002/biot.201400306>.
- [18] I. Hernandez, H. Dhiman, G. Klanert, V. Jadhav, N. Auer, M. Hanscho, M. Baumann, A. Esteve-Codina, M. Dabad, J. Gómez, T. Alioto, A. Merkel, E. Raineri, S. Heath, D. Rico, N. Borth, Epigenetic regulation of gene expression in Chinese Hamster Ovary cells in response to the changing environment of a batch culture, *Biotechnol. Bioeng.* 116 (2019) 677–692, <https://doi.org/10.1002/bit.26891>.
- [19] D. Vito, C.M. Smales, The Long Non-Coding RNA Transcriptome Landscape in CHO Cells Under Batch and Fed-Batch Conditions, *Biotechnol. J.* 13 (2018), 1800122, <https://doi.org/10.1002/biot.201800122>.
- [20] O. Shalem, N.E. Sanjana, E. Hartenian, X. Shi, D.A. Scott, T.S. Mikkelsen, D. Heckl, B.L. Ebert, D.E. Root, J.G. Doench, F. Zhang, Genome-Scale CRISPR-Cas9 Knockout Screening in Human Cells, *Science* 343 (2014) 84–87, <https://doi.org/10.1126/science.1247005>.
- [21] R. Han, L. Li, A.P. Ugalde, A. Tal, Z. Manber, E.P. Barbera, V.D. Chiara, R. Elkon, R. Agami, Functional CRISPR screen identifies AP1-associated enhancer regulating FOXF1 to modulate oncogene-induced senescence, *Genome Biol* 19 (2018) 118, <https://doi.org/10.1186/s13059-018-1494-1>.
- [22] E. Shifrut, J. Carnevale, V. Tobin, T.L. Roth, J.M. Woo, C.T. Bui, P.J. Li, M. E. Diolaiti, A. Ashworth, A. Marson, Genome-wide CRISPR Screens in Primary Human T Cells Reveal Key Regulators of Immune Function, *Cell* 175 (2018) 1958–1971, <https://doi.org/10.1016/j.cell.2018.10.024>, e15.
- [23] S. Zhu, W. Li, J. Liu, C.-H. Chen, Q. Liao, P. Xu, H. Xu, T. Xiao, Z. Cao, J. Peng, P. Yuan, M. Brown, X.S. Liu, W. Wei, Genome-scale deletion screening of human long non-coding RNAs using a paired-guide RNA CRISPR library, *Nat Biotechnol* 34 (2016) 1279–1286, <https://doi.org/10.1038/nbt.3715>.
- [24] V. Schmieder, N. Bydlinksi, R. Strasser, M. Baumann, H.F. Kildegaard, V. Jadhav, N. Borth, Enhanced Genome Editing Tools For Multi-Gene Deletion Knock-Out Approaches Using Paired CRISPR sgRNAs in CHO Cells, *Biotechnol. J.* 13 (2018), 1700211, <https://doi.org/10.1002/biot.201700211>.
- [25] B. Zetsche, J.S. Gootenberg, O.O. Abudayyeh, I.M. Slaymaker, K.S. Makarova, P. Essletzbichler, S.E. Volz, J. Joung, J. van der Oost, A. Regev, E.V. Koonin, F. Zhang, Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System, *Cell* 163 (2015) 759–771, <https://doi.org/10.1016/j.cell.2015.09.038>.
- [26] L.A. Miles, R.J. Garippa, J.T. Poirier, Design, execution, and analysis of pooled in vitro CRISPR/Cas9 screens, *FEBS J.* 283 (2016) 3170–3180, <https://doi.org/10.1111/febs.13770>.
- [27] Z. Debyser, Biosafety of lentiviral vectors, *Curr Gene Ther* 3 (2003) 517–525, <https://doi.org/10.2174/1566523034578177>.
- [28] P.L. Sinn, S.L. Sauter, P.B. McCray, Gene therapy progress and prospects: development of improved lentiviral and retroviral vectors—design, biosafety, and production, *Gene Ther* 12 (2005) 1089–1098, <https://doi.org/10.1038/sj.gt.3302570>.
- [29] N.K. Hamaker, K.H. Lee, Site-specific integration ushers in a new era of precise CHO cell line engineering, *Curr. Opin. Chem. Eng.* 22 (2018) 152–160, <https://doi.org/10.1016/j.coche.2018.09.011>.
- [30] L. Zhang, M.C. Inniss, S. Han, M. Moffat, H. Jones, B. Zhang, W.L. Cox, J.R. Rance, R.J. Young, Recombinase-mediated cassette exchange (RMCE) for monoclonal antibody expression in the commercially relevant CHOK1SV cell line, *Biotechnol. Prog.* 31 (2015) 1645–1656, <https://doi.org/10.1002/btpr.2175>.
- [31] Q.V. Phan, J. Contzen, P. Seemann, M. Gossen, Site-specific chromosomal gene insertion: Flp recombinase versus Cas9 nuclease, *Sci. Rep.* 7 (2017) 17771, <https://doi.org/10.1038/s41598-017-17651-0>.
- [32] J.G. Doench, Am I ready for CRISPR? A user's guide to genetic screens, *Nat. Rev. Genet.* 19 (2018) 67–80, <https://doi.org/10.1038/nrg.2017.97>.
- [33] Hart, T., Tong, A.H.Y., Chan, K., Van Leeuwen, J., Seetharaman, A., Aregger, M., Chandrashekar, M., Hustedt, N., Seth, S., Noonan, A., Habsid, A., Sizova, O., Nedyalkova, L., Climie, R., Tworzynski, L., Lawson, K., Sartori, M.A., Alibeh, S., Tieu, D., Masud, S., Mero, P., Weiss, A., Brown, K.R., Usaj, M., Billmann, M., Rahman, M., Costanzo, M., Myers, C.L., Andrews, B.J., Boone, C., Durocher, D., Moffat, J., 2017. Evaluation and Design of Genome-Wide CRISPR/SpCas9 Knockout Screens. *G3 (Bethesda)* 7, 2719–2727. <https://doi.org/10.1534/g3.117.041277>.
- [34] T. Wang, K. Birsoy, N.W. Hughes, K.M. Krupczak, Y. Post, J.J. Wei, E.S. Lander, D. M. Sabatini, Identification and characterization of essential genes in the human genome, *Science* 350 (2015) 1096–1101, <https://doi.org/10.1126/science.aac7041>.
- [35] L. Wei, D. Lee, C.-T. Law, M.S. Zhang, J. Shen, D.W.-C. Chin, A. Zhang, F.H.-C. Tsang, C.L.-S. Wong, I.O.-L. Ng, C.C.-L. Wong, C.-M. Woone, Genome-wide CRISPR/Cas9 library screening identified PHGDH as a critical driver for Sorafenib resistance in HCC, *Nat. Commun.* 10 (2019) 4681, <https://doi.org/10.1038/s41467-019-12606-7>.
- [36] S. Zhu, Z. Cao, Z. Liu, Y. He, Y. Wang, P. Yuan, W. Li, F. Tian, Y. Bao, W. Wei, Guide RNAs with embedded barcodes boost CRISPR-pooled screens, *Genome Biol* 20 (2019) 20, <https://doi.org/10.1186/s13059-019-1628-0>.
- [37] P. Doolan, N. Barron, P. Kinsella, C. Clarke, P. Meleady, F. O'Sullivan, M. Melville, M. Leonard, M. Clynes, Microarray expression profiling identifies genes regulating sustained cell specific productivity (S-Qp) in CHO K1 production cell lines, *Biotechnol J* 7 (2012) 516–526, <https://doi.org/10.1002/biot.201100255>.
- [38] D. Fomina-Yadlin, M. Mujacic, K. Maggiora, G. Quesnell, R. Saleem, J.T. McGrew, Transcriptome analysis of a CHO cell line expressing a recombinant therapeutic protein treated with inducers of protein expression, *J. Biotechnol.* 212 (2015) 106–115, <https://doi.org/10.1016/j.jbiotec.2015.08.025>.
- [39] E. Harreither, M. Hackl, J. Pichler, S. Shridhar, N. Auer, P.P. Labaj, M. Scheideler, M. Karbiener, J. Grillari, D.P. Kreil, N. Borth, Microarray profiling of preselected CHO host cell subclones identifies gene expression patterns associated with increased production capacity, *Biotechnol J* 10 (2015) 1625–1638, <https://doi.org/10.1002/biot.201400857>.
- [40] G. Klanert, D.J. Fernandez, M. Weinguny, P. Eisenhut, E. Bühler, M. Melcher, S. A. Titus, A.B. Diendorfer, E. Gludovacz, V. Jadhav, S. Xiao, B. Stern, M. Lal, J. Shiloach, N. Borth, A cross-species whole genome siRNA screen in suspension-cultured Chinese hamster ovary cells identifies novel engineering targets, *Sci Rep* 9 (2019) 8689, <https://doi.org/10.1038/s41598-019-45159-2>.
- [41] P. Meleady, P. Doolan, M. Henry, N. Barron, J. Keenan, F. O'Sullivan, C. Clarke, P. Gammell, M.W. Melville, M. Leonard, M. Clynes, Sustained productivity in recombinant Chinese Hamster Ovary (CHO) cell lines: proteome analysis of the molecular basis for a process-related phenotype, *BMC Biotechnol* 11 (2011) 78, <https://doi.org/10.1186/1472-6750-11-78>.
- [42] P.M. Nissom, A. Sanny, Y.J. Kok, Y.T. Hiang, S.H. Chuah, T.K. Shing, Y.Y. Lee, K.T. K. Wong, W.-S. Hu, M.Y.G. Sim, R. Philip, Transcriptome and proteome profiling to understand the biology of high productivity CHO cells, *Mol. Biotechnol.* 34 (2006) 125–140, <https://doi.org/10.1385/mb:34:2:125>.
- [43] N. Vishwanathan, A. Yongky, K.C. Johnson, H.-Y. Fu, N.M. Jacob, H. Le, F.N. K. Yusufi, D.Y. Lee, W.-S. Hu, Global insights into the Chinese hamster and CHO cell transcriptomes, *Biotechnol. Bioeng.* 112 (2015) 965–976, <https://doi.org/10.1002/bit.25513>.
- [44] J.C. Yee, Z.P. Gerdzen, W.-S. Hu, Comparative transcriptome analysis to unveil genes affecting recombinant protein productivity in mammalian cells, *Biotechnol. Bioeng.* 102 (2009) 246–263, <https://doi.org/10.1002/bit.22039>.
- [45] H.K. Kim, M. Song, J. Lee, A.V. Menon, S. Jung, Y.-M. Kang, J.W. Choi, E. Woo, H. C. Koh, J.-W. Nam, H. Kim, In vivo high-throughput profiling of CRISPR-Cpf1 activity, *Nat. Methods* 14 (2017) 153–159, <https://doi.org/10.1038/nmeth.4104>.
- [46] D.E. Bauer, M.C. Canver, S.H. Orkin, Generation of Genomic Deletions in Mammalian Cell Lines via CRISPR/Cas9, *J Vis Exp* (2015), <https://doi.org/10.3791/52118>.
- [47] A.M. Bolger, M. Lohse, B. Usadel, Trimmomatic: a flexible trimmer for Illumina sequence data, *Bioinformatics* 30 (2014) 2114–2120, <https://doi.org/10.1093/bioinformatics/btu170>.
- [48] B. Langmead, S.L. Salzberg, Fast gapped-read alignment with Bowtie 2, *Nat. Methods* 9 (2012) 357–359, <https://doi.org/10.1038/nmeth.1923>.
- [49] M. Martin, Cutadapt removes adapter sequences from high-throughput sequencing reads, *EMBnet.journal* 17 (2011) 10–12, <https://doi.org/10.14806/ej.17.1.200>.
- [50] B. Bushnell, BMAP: A Fast, Accurate, Splice-Aware Aligner (No. LBNL-7065E), Lawrence Berkeley National Lab. (LBNL), Berkeley, CA United States, 2014.
- [51] Garrison, E., Marth, G., 2012. Haplotype-based variant detection from short-read sequencing 20.
- [52] J. Wala, C.-Z. Zhang, M. Meyerson, R. Beroukhim, VariantBam: filtering and profiling of next-generation sequencing data using region-specific rules, *Bioinformatics* 32 (2016) 2029–2031, <https://doi.org/10.1093/bioinformatics/btw111>.
- [53] M.D. Robinson, D.J. McCarthy, G.K. Smyth, edgeR: a Bioconductor package for differential expression analysis of digital gene expression data, *Bioinformatics* 26 (2010) 139–140, <https://doi.org/10.1093/bioinformatics/btp616>.
- [54] M. Baumann, E. Gludovacz, N. Sealover, S. Bahr, H. George, N. Lin, K. Kayser, N. Borth, Preselection of recombinant gene integration sites enabling high transcription rates in CHO cells using alternate start codons and recombinase mediated cassette exchange, *Biotechnol. Bioeng.* 114 (2017) 2616–2627, <https://doi.org/10.1002/bit.26388>.

- [55] J. Pichler, F. Hesse, M. Wieser, R. Kunert, S.S. Galosy, J.E. Mott, N. Borth, A study on the temperature dependency and time course of the cold capture antibody secretion assay, *J. Biotechnol.* 141 (2009) 80–83, <https://doi.org/10.1016/j.jbiotec.2009.03.001>.
- [56] J.R. Conway, A. Lex, N. Gehlenborg, UpSetR: an R package for the visualization of intersecting sets and their properties, *Bioinformatics* 33 (2017) 2938–2940, <https://doi.org/10.1093/bioinformatics/btx364>.
- [57] S. Reber, J. Mechttersheimer, S. Nasif, J.A. Benitez, M. Colombo, M. Domanski, D. Jutzi, E. Hedlund, M.-D. Ruepp, CRISPR-Trap: a clean approach for the generation of gene knockouts and gene replacements in human cells, *MBoC* 29 (2018) 75–83, <https://doi.org/10.1091/mbc.E17-05-0288>.
- [58] C. Ronda, L.E. Pedersen, H.G. Hansen, T.B. Kallehauge, M.J. Betenbaugh, A. T. Nielsen, H.F. Kildegaard, Accelerating genome editing in CHO cells using CRISPR Cas9 and CRISPy, a web-based target finding tool, *Biotechnol. Bioeng.* 111 (2014) 1604–1616, <https://doi.org/10.1002/bit.25233>.
- [59] K. Motheramgari, R.V.-B. Curell, I. Tzani, C. Gallagher, M.C. Rivadeneyra, L. Zhang, N. Barron, C. Clarke, Expanding the Chinese hamster ovary cell long non-coding RNA transcriptome using RNASeq, *Biotechnol. Bioeng.* (2020), <https://doi.org/10.1002/bit.27467>.
- [60] J.D. Ransohoff, Y. Wei, P.A. Khavari, The functions and unique features of long intergenic non-coding RNA, *Nat. Rev. Mol. Cell Biol.* 19 (2018) 143–157, <https://doi.org/10.1038/nrm.2017.104>.
- [61] A. Kapusta, C. Feschotte, Volatile evolution of long noncoding RNA repertoires: mechanisms and biological implications, *Trends Genet.* 30 (2014) 439–452, <https://doi.org/10.1016/j.tig.2014.08.004>.
- [62] H. Lee, Z. Zhang, H.M. Krause, Long Noncoding RNAs and Repetitive Elements: Junk or Intimate Evolutionary Partners? *Trends Genet.* 35 (2019) 892–902, <https://doi.org/10.1016/j.tig.2019.09.006>.
- [63] X.-H. Zhang, L.Y. Tee, X.-G. Wang, Q.-S. Huang, S.-H. Yang, Off-target Effects in CRISPR/Cas9-mediated Genome Engineering, *Molecular Therapy - Nucleic Acids* 4 (2015) 264, <https://doi.org/10.1038/mtna.2015.37>, e.
- [64] S. Sharma, E. Petsalaki, Application of CRISPR-Cas9 Based Genome-Wide Screening Approaches to Study Cellular Signalling Mechanisms, *Int J Mol Sci* 19 (2018), <https://doi.org/10.3390/ijms19040933>.
- [65] G. Klanert, N. Bydlinski, P. Agu, A.B. Diendorfer, M. Hackl, M. Hanscho, M. Melcher, M. Baumann, J. Grillari, N. Borth, Transient manipulation of the expression level of selected growth rate correlating microRNAs does not increase growth rate in CHO-K1 cells, *J. Biotechnol.* 295 (2019) 63–70, <https://doi.org/10.1016/j.jbiotec.2019.02.011>.
- [66] S.J. Liu, M.A. Horlbeck, S.W. Cho, H.S. Birk, M. Malatesta, D. He, F.J. Attenello, J. E. Villalta, M.Y. Cho, Y. Chen, M.A. Mandegar, M.P. Olvera, L.A. Gilbert, B. R. Conklin, H.Y. Chang, J.S. Weissman, D.A. Lim, CRISPRi-based genome-scale identification of functional long noncoding RNA loci in human cells, *Science* 355 (2017), <https://doi.org/10.1126/science.aah7111>.
- [67] Y. Liu, Z. Cao, Y. Wang, Y. Guo, P. Xu, P. Yuan, Z. Liu, Y. He, W. Wei, Genome-wide screening for functional long noncoding RNAs in human cells by Cas9 targeting of splice sites, *Nat. Biotechnol.* 36 (2018) 1203–1210, <https://doi.org/10.1038/nbt.4283>.
- [68] Horlbeck, M.A., Liu, S.J., Chang, H.Y., Lim, D.A., Weissman, J.S., 2020. Fitness effects of CRISPR/Cas9-targeting of long noncoding RNA genes. *Nat. Biotechnol.* <https://doi.org/10.1038/s41587-020-0428-0>.
- [69] D.M. Munoz, P.J. Cassiani, L. Li, E. Billy, J.M. Korn, M.D. Jones, J. Golji, D. A. Ruddy, K. Yu, G. McAllister, A. DeWeck, D. Abramowski, J. Wan, M.D. Shirley, S.Y. Neshat, D. Rakiec, R. de Beaumont, O. Weber, A. Kauffmann, E.R. McDonald, N. Keen, F. Hofmann, W.R. Sellers, T. Schmelzle, F. Stegmeier, M.R. Schlabach, CRISPR screens provide a comprehensive assessment of cancer vulnerabilities but generate false-positive hits for highly amplified genomic regions, *Cancer Discov.* 6 (2016) 900–913, <https://doi.org/10.1158/2159-8290.CD-16-0178>.