



## OPEN ACCESS

## EDITED BY

Zhanglin Lin,  
South China University of Technology,  
China

## REVIEWED BY

Wenjing Cui,  
Jiangnan University, China  
Prihardi Kahar,  
Kobe University, Japan

## \*CORRESPONDENCE

Oliver Hertel,  
oliver.hertel@uni-bielefeld.de

<sup>†</sup>These authors have contributed equally to this work and share first authorship

## SPECIALTY SECTION

This article was submitted to Synthetic Biology, a section of the journal Frontiers in Bioengineering and Biotechnology

RECEIVED 03 August 2022

ACCEPTED 30 September 2022

PUBLISHED 13 October 2022

## CITATION

Hertel O, Neuss A, Busche T, Brandt D, Kalinowski J, Bahnemann J and Noll T (2022), Enhancing stability of recombinant CHO cells by CRISPR/Cas9-mediated site-specific integration into regions with distinct histone modifications. *Front. Bioeng. Biotechnol.* 10:1010719. doi: 10.3389/fbioe.2022.1010719

## COPYRIGHT

© 2022 Hertel, Neuss, Busche, Brandt, Kalinowski, Bahnemann and Noll. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

# Enhancing stability of recombinant CHO cells by CRISPR/Cas9-mediated site-specific integration into regions with distinct histone modifications

Oliver Hertel<sup>1,2\*†</sup>, Anne Neuss<sup>3†</sup>, Tobias Busche<sup>2</sup>, David Brandt<sup>2</sup>, Jörn Kalinowski<sup>2</sup>, Janina Bahnemann<sup>4</sup> and Thomas Noll<sup>1,2</sup>

<sup>1</sup>Cell Culture Technology, Faculty of Technology, Bielefeld University, Bielefeld, Germany, <sup>2</sup>Center for Biotechnology (CeBiTec), Bielefeld University, Bielefeld, Germany, <sup>3</sup>Biochemical Engineering (AVT.BioVT), RWTH Aachen University, Aachen, Germany, <sup>4</sup>Institute of Physics, University of Augsburg, Augsburg, Germany

Chinese hamster ovary (CHO) cells are the most important platform for producing biotherapeutics. Random integration of a transgene into epigenetically instable regions of the genome results in silencing of the gene of interest and loss of productivity during upstream processing. Therefore, cost- and time-intensive long-term stability studies must be performed. Site-specific integration into safe harbors is a strategy to overcome these limitations of conventional cell line design. Recent publications predict safe harbors in CHO cells based on omics data sets or by learning from random integrations, but those predictions remain theory. In this study, we established a CRISPR/Cas9-mediated site-specific integration strategy based on ChIP-seq data to improve stability of recombinant CHO cells. Therefore, a ChIP experiment from the exponential and stationary growth phase of a fed-batch cultivation of CHO-K1 cells yielded 709 potentially stable integration sites. The reporter gene eGFP was integrated into three regions harboring specific modifications by CRISPR/Cas9. Targeted Cas9 nanopore sequencing showed site-specific integration in all 3 cell pools with a specificity between 23 and 73%. Subsequently, the cells with the three different integration sites were compared with the randomly integrated donor vector in terms of transcript level, productivity, gene copy numbers and stability. All site-specific integrations showed an increase in productivity and transcript levels of up to 7.4-fold. In a long-term cultivation over 70 generations, two of the site-specific integrations showed a stable productivity (>70%) independent of selection pressure.

## KEYWORDS

CRISPR/Cas9, CHO, cell line development, ChIP-seq, epigenetics, safe harbor, stability, histone

## 1 Introduction

Chinese hamster ovary (CHO) cells are the most important cell-based system for biopharmaceutical production, accounting for 70% of all recombinant proteins (Wurm, 2004; Walsh, 2018; Oosting, 2019). Despite decades of experience, however, many key parameters of production and cell line development—such as chromosome rearrangements, heterogeneity within cell lines, unpredictable expression levels, and gene silencing - are still not fully understood (Lee et al., 2019). The standard method for generating new CHO production cell lines for biopharmaceuticals is based on random integration of the product gene followed by selection *via* antibiotics or metabolic markers (Bebbington et al., 1992; Hamaker and Lee, 2018). Clones are then separated by single cell cloning or limiting dilution (Jayapal et al., 2007; Lai et al., 2013). This method has been continuously optimized over the years by leveraging our constantly expanding knowledge on transcription, translation, cell metabolism, signal transduction pathways and secretion (Tihanyi and Nyitray, 2021). Intuitive genetic modifications of the cells (e.g., knock-down of lactate dehydrogenase) are now standard (Zhou et al., 2011).

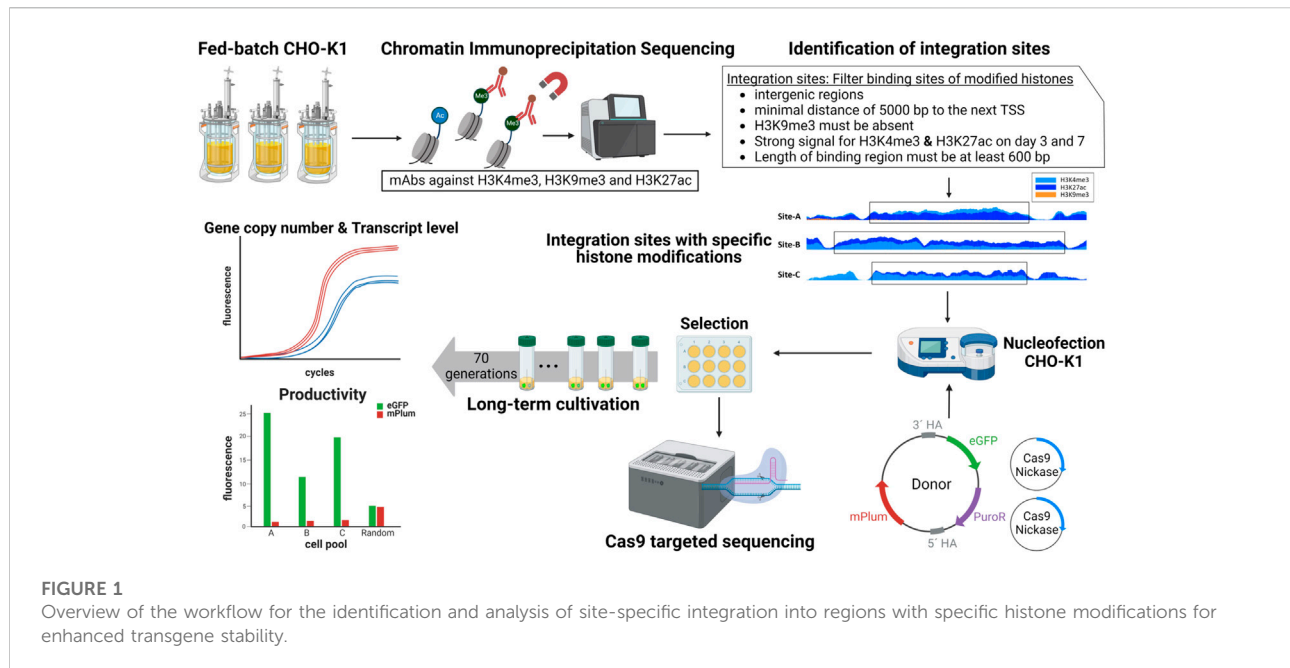
Nevertheless, despite this progress, the development of a high-producer cell line still usually takes between 6 and 12 months—making it both time-consuming as well as labor- and cost-intensive (Jayapal et al., 2007; Lai et al., 2013). This is because the random integration method still only relies on chance to generate the right production cell lines. Random integration also generates clones with different gene copy numbers and integration sites, resulting in heterogeneities in the cell population, both in terms of growth and productivity, and in terms of unpredictable expression levels that can additionally change during cultivation (Lee et al., 2019). This makes it very difficult if not impossible to obtain results that are simultaneously reproducible and predictable (Hefzi and Lewis, 2014).

Making matters even worse, adequate stability and predictability with respect to both product yield and quality are absolute prerequisites for any biopharmaceutical-grade production (Barnes et al., 2003; Noh et al., 2020). A cell line may be said to be stable if, firstly, a homogeneous cell population retains 70% or more of volumetric productivity over 70 generations and, secondly, it exhibits no “clinically meaningful differences” (according to the FDA Orange Book) compared to the reference product as determined by considering its structure, function, purity, chemical identity, and bioactivity (Bailey et al., 2012; Dahodwala and Lee, 2019). Instability within recombinant CHO cells can occur at any or all of the genome, transcriptome, or proteome level as reviewed in Dahodwala and Lee (2019). The term “position-effect” is frequently used to refer on the one hand to rearrangements of chromosomes that lead to loss or silencing of genes, and on the other hand to chromatin-related effects caused by different integration sites (Wilson et al.,

1990; Recillas-Targa, 2006). Chromosome rearrangements (Derouazi et al., 2006; Du et al., 2013; Bandyopadhyay et al., 2019) or gene copy loss (Kim et al., 2011; Beckmann et al., 2012) have both been identified as potential reasons for instability. Other studies have also indicated that instability may arise from gene silencing (Chusainow et al., 2009; Yang et al., 2010; Osterlehner et al., 2011; Marx et al., 2018), which can be caused by DNA methylations that occur at the promoter (Wippermann and Noll, 2017)—frequently involving the CMV (cytomegalovirus) promoter, specifically (Romanova and Noll, 2018). Based on this, stable clones were screened *via* DNA methylation at the CMV promoter subsequent to random integration, which, however, resulted in a non-negligible false positive rate (Osterlehner et al., 2011). In addition, histone modifications in the environment of the GOI (gene of interest) influence transcription. Some of these modifications and their relevance for gene expression are well elucidated (Kundaje et al., 2015). For example, H3K4me3 is associated with active promoter regions (Lawrence et al., 2016), H3K27ac is associated with increased activation of enhancer and promoter regions (Karličić et al., 2010) and H3K9me3 is a typical marker for constitutive, permanent heterochromatin and is associated with gene silencing (Nicetto and Zaret, 2019). Like CpG methylation, histone modifications were also analyzed to screen for stable expressing cells after random integration. Here, acetylation of H3 and also the appearance of H3K4me3 at the CMV promoter under selection pressure was shown to be a more effective indicator of stable cells than CpG methylation (Moritz et al., 2016).

In order to improve cell line development with respect to stability, reproducibility as well as product yield and quality, one of the most important research interests is on site-specific integration of product genes and the development of faster and more efficient selection and screening systems during long-term cultivation (Tihanyi and Nyitray, 2021). One optimization approach is the specific integration of product genes into what are known as “safe harbors” or “hot spots”. These are regions in the genome that are supposedly not affected by gene silencing and show increased transcriptional activity. At present, very few such safe harbor regions have been identified in CHO cells (Hamaker and Lee, 2018). Their characteristics are poorly understood and differ substantially between the relatively few known regions (Lee et al., 2019). Notwithstanding those limitations, however, it has been shown that site-specific integration into known safe harbor regions can lead to stable cell lines with high productivity (Kawabe et al., 2018; Zhao et al., 2018; Pristovšek et al., 2019).

To date, the ongoing efforts to identify additional safe harbor regions in the CHO genome has mostly been conducted *via* empirical methods such as random lentiviral integrations (Gaidukov et al., 2018; O’Brien et al., 2018; Zhou et al., 2019). Recently, Lee et al. (2019) described a pipeline using omics methods to search for safe harbors in a structured manner.



Initial approaches have also used RNA-Seq data to compare properties of stable and unstable integration sites to predict safe harbors (Pristovšek et al., 2019; Dhiman et al., 2020). Another approach sought to identify potential safe harbor regions based on the presence or absence of 16 epigenetic markers, in particular, histone modifications and RNA-Seq data (Hilliard and Lee, 2020). To date, all of these methods have remained theoretical in nature—meaning that they have not actually been tested *via* site-specific integration. Here, we establish a CRISPR/Cas9-mediated site-specific integration strategy based on chromatin immunoprecipitation (ChIP)-seq data that improves the stability of recombinant CHO cells. Figure 1 provides an overview of the identification and analysis of the integration sites with specific histone modifications. We set out in this work to establish *via* proof-of-concept that stable integration sites can be accurately predicted with epigenetic data. Therefore, a ChIP experiment was conducted against H3K4me3, H3K27ac, and H3K9me3, using samples taken from the exponential and stationary growth phase of a fed-batch cultivation of CHO-K1 cells. After genome-wide analysis of histone modifications, analysis of the observable combinations of these epigenetic markers (in broad intergenic regions) was used as a criterion for the identification of potential integration sites. To analyze the influence of histone modifications on transgene expression, the reporter gene eGFP (enhanced green fluorescent protein) was integrated by CRISPR/Cas9 into three regions harboring specific modifications. The site-specific integration was then verified and characterized by targeted Cas9 nanopore sequencing.

Subsequently, the cells with the three integration sites to be screened were compared in terms of transcript level, productivity, and stability against a randomly integrated donor vector.

## 2 Materials and methods

All primers for PCR, qPCR, and RT-qPCR, vector construction, and all guide RNAs, are listed in [Supplementary Table S1](#) and were purchased from Metabion.

### 2.1 Cell culture

The CHO-K1 cell line (strain ATCC 61-CCL) was adapted for growth in suspension and used as our model host cell line. CHO-K1 cells were cultivated in shake flasks (Triforest) or TubeSpin® Bioreactors (TPP) at 185 rpm (maximal deflection 50 mm), 37 °C, 5% CO<sub>2</sub>, and 80% humidity within either a Mytron cell culture incubator (Mettler) or in 2 L B-DCU bioreactors (Sartorius). To determine viable cell density (VCD) and viability, cells were counted using a Cedex™ automated cell counter device (Roche). The cells were passaged and diluted to 3 × 10<sup>5</sup> cells/mL every second to third day for preculture and long-term cultivation. All cultivations took place within the chemically defined TCX6D medium (Xell AG) that was supplemented with 8 mM glutamine.

## 2.2 Bioreactor operation

Bioreactor cultivation was performed as triplicates in 2 L Biostat B-DCU bioreactors (Sartorius AG), with a starting volume of 1 L and an inoculation of  $5 \times 10^5$  cells/mL. The cultivation temperature, pH-value, and dissolved oxygen concentrations were controlled at 37°C, at 7.2, and 40% of the air saturation, respectively. Stirring speed was set to 120 rpm in the beginning, before being increased to 150 rpm on day 3 and then to 180 rpm on day 6, using a Rushton turbine. The feed (1 L in total) was added, in pulses, on days 3–6 of the cultivation. Beginning with 100 ml, the feed volume was increased by 100 ml every day using CHO Basic Feed (Xell AG), supplemented with 40 mM of glutamine.

## 2.3 Chromatin immunoprecipitation

Samples of  $1 \times 10^7$  cells were harvested on days 3 and 7 after inoculation of the bioreactors. The SimpleChIP® Enzymatic Chromatin IP Kit (Cell Signaling Technology) was used for the whole ChIP workflow. All steps were performed according to the manufacturer's protocol. After resuspending fixed cell pellets, the biological replicates were pooled. To digest the chromatin, 5.64 µL micrococcal nuclease were used for each pooled sample. Lysis of nuclei was performed with the Sonifier 250 (Branson) on level 2, in 4 cycles of 5 s. Between cycles, the samples were incubated on ice. Chromatin was stored at -80°C for further analysis. DNA was extracted from 50 µL chromatin solution and the concentration was determined using Nanodrop One (ThermoFisher). The size of the digested chromatin was analyzed with a 1% agarose gel. Samples with a predominant size range between 100 and 500 bp were used for further analysis. Immunoprecipitation was performed with chromatin solution equal to 7.5 µg DNA and 2–10 µL antibody overnight. The following ChIP validated antibodies were used: Tri-Methyl-Histone H3 (Lys4) (C42D8) Rabbit mAb (#9751, Cell Signaling Technology), Acetyl-Histone H3 (Lys27) (D5E4) XP Rabbit mAb (#8173, Cell Signaling Technology), Tri-Methyl-Histone H3 (Lys9) (D4W1U) Rabbit mAb (#13969, Cell Signaling Technology), Histone H3 (D2B12) XP Rabbit mAb (ChIP Formulated) (#4620, Cell Signaling Technology) and Normal Rabbit IgG (#2729, Cell Signaling Technology). All antibodies were used according to the manufacturer's protocol. Two percent of the starting volume of chromatin were stored at -20°C for normalization of the sequencing data as input sample. ChIP eluates and the input sample were checked *via* PCR using DreamTaq DNA polymerase (Thermo Fisher Scientific), according to the manufacturer's protocol, with primers specific to *Rpl30* suited for micrococcal nuclease digested DNA.

## 2.4 ChIP-sequencing and data analysis

Library preparation of the ChIP eluates and input samples was performed using TruSeq ChIP Library Preparation Kit—Set A (Illumina). Sequencing was carried out using one NextSeq 500/550 High Output Kit v2.5 (75 cycles, Illumina). Reads with a minimal length of 25 nt after trimming of the sequencing adapters were used for further analysis. Quality control of raw reads was conducted by FastQC v0.11.8 (Andrews 2010). Reads were aligned to the *Cricetulus griseus* Genbank assembly GCA\_003668045.1 using Bowtie v1.2.3 (Langmead et al., 2009), with the parameters `--best --strata -m 1`. Alignment statistics were checked using samtools v1.10.2 (Li et al., 2009), and cross-correlation analysis was performed using the `run_spp.R` script from Phantompeakqualtools package (Kharchenko et al., 2008; Landt et al., 2012). Peaks were called using MACS2 v2.2.6. (Zhang et al., 2008; Feng et al., 2012), with the parameters `callpeak -g mm --extsize 147 --nomodel -B`, and with the additional parameter `--broad` added for H3K9me3. Differential peak analysis between days 3 and 7 was performed using MACS2 v2.2.6 (Zhang et al., 2008; Feng et al., 2012), with the parameters `bdgdiff -C 11` for H3K4me3 and H3K27ac data, and with `-C 4` added for H3K9me3. The parameter `-C` was used to keep only the strongest peaks for the identification of integration sites. The common peaks were annotated using HOMER 3.12 (Heinz et al., 2010) with the gtf file from the Genbank assembly used for alignment.

## 2.5 Identification of integration sites

The annotated outputs of ChIP-seq data analysis—which contained the strongest common peaks of both times—were used as input for the identification of possible integration sites. These were filtered to keep only intergenic peaks with a minimal distance of 5000 bp to the next TSS (transcription start site). These peak lists were converted to bed format. To keep only regions with both activating histone marks (H3K4me3 and H3K27ac), and no repressive histone marks (H3K9me3), Bedtools intersect v2.27.1 (Quinlan and Hall 2010) was used to first join peaks of activating histone marks and then to perform an anti-join with all peaks of H3K9me3. Overlapping peaks or neighboring peaks with a distance less than 150 bp were merged using Bedtools merge v2.27.1 (Quinlan and Hall, 2010). Possible integration sites shorter than 600 bp were removed. The integration sites bed file was converted to fasta format using the `getfasta` module from Bedtools v2.27.1 (Quinlan and Hall, 2010). Designing gRNAs for the identified possible safe harbor regions with the CHOPCHOP website (<https://chopchop.rc.fas.harvard.edu/>), the parameters Chinese hamster ovary, knock-in and CRISPR-Cas9 nickase were used (Labun et al., 2019). The gRNAs with the highest efficiency score were chosen (Table 1).

TABLE 1 gRNAs for site-specific integration into three regions with specific histone modifications using Cas9 nickases.

Site-A	<u>CCAGCTTAGGGCCACGCTGCTGC</u>	GCGAGCATCTTGAAGGGCGGGG
Site-B	TGACAACAACCGTTGCCAGAGG <u>AGG</u>	<u>CCACGCTATTAGAAAACCTGCTC</u>
Site-C	<u>CCCATGATAGAGGTTCCGACCT</u>	GCTTCATTGGCTGGCTAGAAAGG

Six gRNAs were used in this study (underlined are the PAM sequences):

Vectors carrying the gRNAs, and eGFP-Cas were designed for all three integration sites. Single strand oligonucleotides with *BbsI* overhangs were synthesized by Metabion, phosphorylated and annealed. The gRNAs, were ligated to the digested (*BbsI*) plasmid pX461 (Ran et al., 2013).

## 2.6 Donor plasmid construction

To investigate the potential of the integration sites to serve as potential safe harbor regions, the reporter gene eGFP was introduced to the identified regions for analysis of expression and stability. The donor plasmids consist of site-specific 5' and 3' homology arms (about 800–850 bp each), protein expression cassette [hCMV-promoter, eGFP, bGH Poly(A)], puromycin expression cassette [SV40-promoter, puromycin, bGH Poly(A)], mPlum expression cassette [hCMV-promoter, mPlum, bGH Poly(A)], and an Ampicillin/ori backbone. The plasmids were cloned *via* classical restriction and insert cloning into an in-house plasmid. The components of the plasmids were then amplified using Q5 or Phusion High Fidelity DNA Polymerase (both NEB), according to the manufacturer's protocol. Finally, the plasmids were purified using the Miniprep Kit NucleoSpin Plasmid or the Midiprep Kit NucleoSnap Plasmid (both Macherey-Nagel), and sequenced at the Sequencing Core Facility (CeBiTec, Bielefeld University).

## 2.7 Generation of stable cell pools

For the generation of stable cell pools,  $2 \times 10^6$  CHO-K1 cells were co-transfected with the corresponding CRISPR- and donor-plasmids (a total amount of 6  $\mu$ g DNA, equally divided to each plasmid) with the Lonza Nucleofector 2b system (Lonza), using the Amaxa Cell Line Nucleofector Kit V (Lonza) and program U-023, according to manufacturer's instruction. Two days after transfection, the cells were treated with 4  $\mu$ g/ml puromycin for selection. The transfected cell pools were monitored by flow cytometry (BioRad S3e Cell Sorter) to determine the eGFP and/or mPlum positive cell populations during the selection phase.

## 2.8 Isolation of genomic DNA

Genomic DNA was isolated from  $5 \times 10^6$  cells and purified using the Wizard Genomic DNA Purification Kit (Promega), according to the manufacturer's protocol for Tissue Culture Cells and Animal Tissue.

To isolate high molecular weight DNA from  $5 \cdot 10^6$  cells for nanopore sequencing, the NucleoBond HMW DNA Kit (Macherey-Nagel) was used as outlined in the manufacturer's instructions.

## 2.9 Targeted Cas9 sequencing and data analysis

For targeted Cas9 sequencing the Cas9 Sequencing Kit (SQK-CS9109) (Oxford Nanopore Technologies) was used according to the manufacturer's instructions. For each sample, 5  $\mu$ g genomic DNA were used as input. The DNA was then dephosphorylated to block untargeted adapter ligation and cleaved at the eGFP target site using Cas9 and custom crRNAs (Integrated DNA Technologies). Blunt ends resulting from Cas9 cleavage were then dA-tailed and ligated to sequencing adapters. Sequencing was carried out on the GridION MK1 using one R9.4.1 flow cell per sample without barcoding.

Fastq files from each flowcell containing reads that passed the quality filtering were concatenated. To identify reads that originate from the target sequence, these were mapped to the eGFP sequence using Minimap2 v2.17 (Li, 2018) with the parameter *-ax map-ont*. Alignment statistics were checked, and mapped reads were extracted, using samtools v1.10.2 (Li et al., 2009). Sam files with extracted reads were converted to fastq format using the SamToFastq module of Picard Toolkit (Broad Institute, 2019). These targeted reads were aligned to the *Cricetulus griseus* Genbank assembly GCA\_003668045.1 using Minimap2 v2.17 (Li, 2018) with the parameter *-ax map-ont*. Alignments marked as supplementary or with MAPQ score smaller than 20 were discarded. The enrichment of eGFP integrations achieved by targeted sequencing compared to the theoretical expected number of eGFP integrations considering the sequencing depth and gene copy numbers was calculated.

$$\text{fold enrichment} = \frac{\text{reads mapping to eGFP integration} / 2}{\text{gene copies eGFP} \times \left( \frac{\text{bases sequenced}}{\text{bases CHO genome}} \right)}$$



The start of the genomic alignments was compared to the intended integration site and then classified as either site-specific or random to calculate the integration specificity. This was done using a custom R script.

## 2.10 RNA Isolation

For total RNA extraction, cell suspension equal  $5 \times 10^6$  cells were centrifuged for 7 min at 800 g, 4°C. Cell pellets were then resuspended in 600  $\mu$ l TRI Reagent (Zymo Research), and stored at -80°C. The lysed samples were mixed with 120  $\mu$ l chloroform and were centrifuged at 12,000 g, 4°C for 15 min. The aqueous phase was extracted again with 300  $\mu$ l TRI Reagent and 120  $\mu$ l chloroform. RNA was precipitated from the aqueous phase by incubation with the same volume isopropanol and centrifuged at 12,000 g, 4°C for 10 min. The RNA pellet was washed with 600  $\mu$ l 75% ethanol twice. After drying, the pellet was resuspended in 50  $\mu$ l RNase-free water.

## 2.11 Analysis of relative gene copy numbers and mRNA levels

The relative *eGFP* transgene copy numbers and mRNA levels were determined *via* qPCR analysis. *eGFP*, glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*),  $\beta$ -2-microglobulin (*B2m*),  $\beta$ -actin (*ACTB*), and vezatin (*VEZT*) primers were designed and tested for specificity. To determine gene copy numbers of *eGFP* relative to *GAPDH* and *B2m*, 1000 ng isolated DNA was combined with 15  $\mu$ l GoTaq qPCR Master Mix (Promega), 0.6  $\mu$ l forward and 0.6  $\mu$ l reverse primer (10  $\mu$ M), and then replenished with water up to a total volume of 20  $\mu$ l. The analysis was carried out with the LightCycler<sup>®</sup> 480 (Roche) in triplicates for each sample. The following thermal cycling parameters were applied: 10 min at 95°C followed by 40 cycles of 15 s at 95°C, 30 s at 60°C and 30 s at 72°C. A melting-curve was generated after the last cycles from 50°C to 97°C. For mRNA level detection relative to *ACTB* and *VEZT*, the Luna Universal One-Step RT-qPCR Kit (NEB) was used according to manufacturer's instructions with 500 ng of RNA for each sample. The measurement was done in triplicates with the LightCycler<sup>®</sup> 480 (Roche).

The relative quantifications were determined by the analysis of crossing point (CP)-values with the second derivative maximum method. All results were normalized to an inter-run calibrator for possible variation in input amount and quality of DNA or RNA between samples. Relative amounts were calculated by pairing the target gene with both reference genes (pairing rule: all to mean). Afterwards, both the mean and standard deviations were calculated.

## 2.12 Determination of fluorescence signal

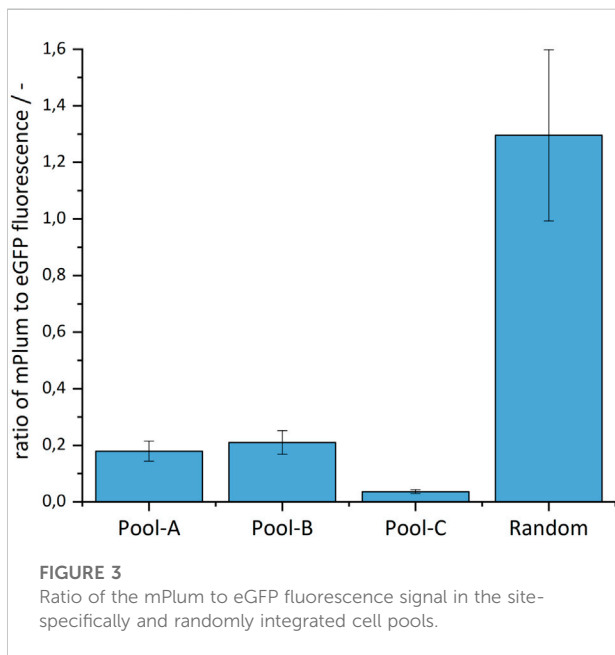
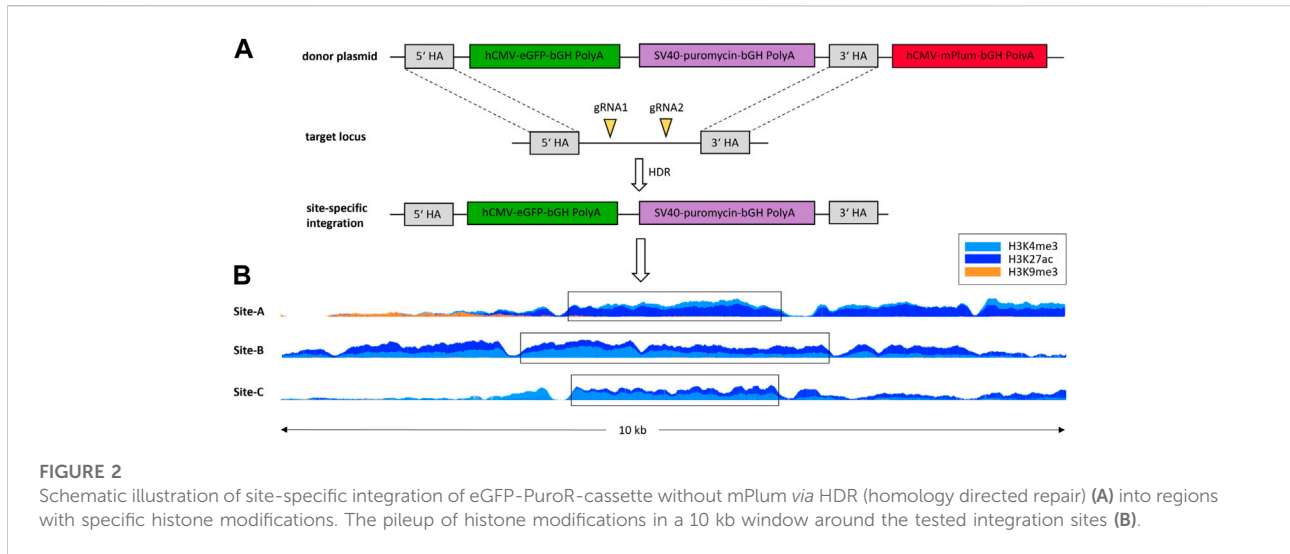
For the determination of the fluorescent signal,  $5 \times 10^6$  cells were first centrifuged for 5 min at 200 g, 4°C, then washed, and then centrifuged again for 7 min at 400 g, 4°C. The cell pellet was then frozen at -80°C. The frozen pellet was then lysed with 1 ml of lysis buffer (Supplementary Table S2) supplemented with PMSF (Phenylmethylsulfonylfluoride), cooled down on ice, and homogenized by ultrasonic treatment (Branson 1210 Ultrasonic Cleaner, Emerson Electric) for 5 min. After being cooled on ice for 30 min, the samples were centrifuged for 20 min at 15,000 g, 4°C. The fluorescence signal was analyzed in triplicates with 100  $\mu$ l of the supernatant each using the TECAN-Reader Infinite 200 Pro MNano+ and corrected by the autofluorescence of the cells. The excitation of eGFP (emission 507 nm) was measured at 488 nm, the excitation of mPlum (emission 590 nm) at 649 nm.

## 3 Results

### 3.1 Genome-wide analysis of histone modifications for the identification of integration sites

The host cell line CHO-K1 was cultivated in a chemically defined medium in 2 L bioreactors and fed with a commercial feed solution. Samples for ChIP-Seq were cross-linked and frozen during the exponential growth phase on day 3 (viability >95%), before feeding and the stationary growth phase (92% viability) 1 day after the last feed pulse (Supplementary Figure S1). ChIP-Seq was performed for three of the six core histone modifications according to the International Human Epigenetics Consortium (Kundaje et al., 2015). Sequencing yielded 34.9–58.0 M uniquely mapping reads, with a mean quality of 34.4 and a length of 83 bp. Cross-correlation analysis resulted in a normalized strand cross-correlation coefficient (NSC) > 1.05 and a relative strand cross-correlation coefficient (RSC) > 1.8 for narrow peaks which corresponds to a quality tag (Phantompeakqualtools) of two. The quality tag for broad peaks was one.

Regions with favorable histone modifications for the stable expression of a transgene should meet the following criteria: high levels of both H3K4me3 and H3K27ac, and low levels of H3K9me3, at both time points in a region with a minimum width of 600 bp. These regions were not allowed to overlap loci expressing coding RNAs or regions 5000 bp upstream of the associated TSS. Application of this criteria resulted in the identification of 709 possible integration sites with specific histone modifications. The pileup at the three tested integration sites is shown in Figure 2. In all integration sites, H3K9me3 is absent. Integration site A has H3K4me3 as dominant histone modification, while sites B and C have a



higher occupancy of H3K27ac. The integration sites are all between 2600 and 3900 bp wide.

### 3.2 Identification of site-specific integration and verification using targeted Cas9 sequencing

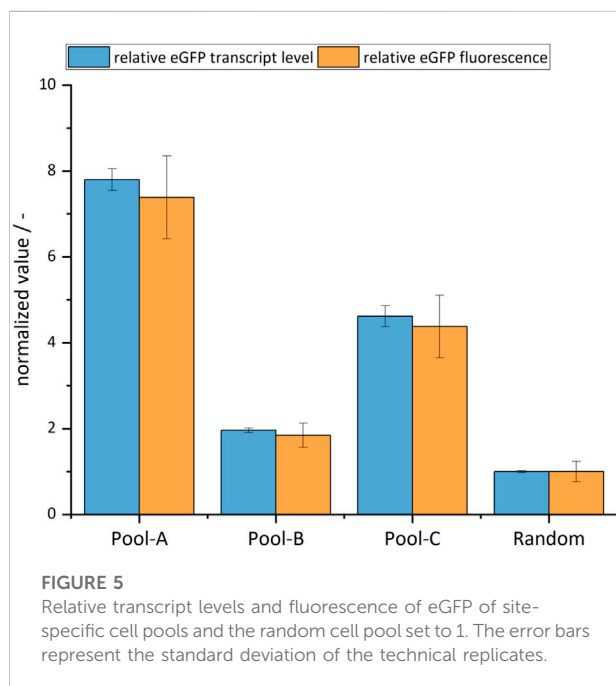
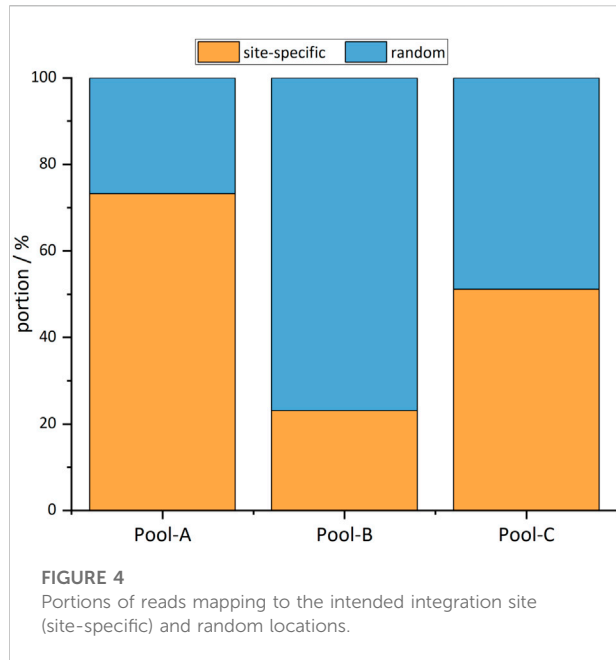
Following Sergeeva et al. (2019), a sequence for an additional fluorescent protein - the mPlum - was located on the plasmid outside the homologous arms. This served as a

marker for the type of integration. Through random integration, both eGFP and mPlum should integrate (resulting in green/red fluorescent cells), whereas in site-specific integration, only the eGFP should be stably integrated into the genome (resulting in only fluorescent green cells, Figure 2A). As shown in Figure 3, the ratio of the mPlum to eGFP fluorescence signal in the site-specifically integrated cell pools was between 0.05 and 0.2 at the end of the selection phase. The ratio of the random cell pool was approximately 1.3. This ratio provided a first indication of the specificity of the CRISPR experiment.

The ratio of mPlum to eGFP fluorescence provides a strong preliminary signal for the type of integration—but sequencing is necessary to provide confirmation. In addition, the specificity of the integration must also be assessed. Since PCR errors should be avoided and long reads are needed to span the insert, homologous arms and integration site, Targeted Cas9 Sequencing from Oxford Nanopore Technologies was used. Compared to other Cas9-based target-enrichment sequencing protocols, no capture step is required and read length is limited solely by the fragment length of the DNA (Slesarev et al., 2019; van Haasteren et al., 2021).

From each sample, 33500–56300 mappable reads (with an average length of ~5 kb) were sequenced. This is equal to 6–17% of the CHO genome. The integration site must be enriched, in order to facilitate detection, because only 26–57 reads are assigned to both the insert and the genome (which corresponds to an on-target rate of 0.15–0.33% at base level). Compared to the theoretically expected number of inserts in the sequencing runs—normalized to the gene copy number—targeted reads are accordingly enriched 86–244x with an average length of ~12.6 kb.

The comparison of the alignments with the intended integration sites shows that site-specific integration took place



in all cell pools. However, random integration also occurred to a different extent in all pools (Figure 4). The proportion of site-specific integration is identical with the specificity: Pool-A displayed an integration specificity of 73%, Pool-B of 23%, and Pool-C of 51%. In most cases, site-specific integration occurred at exactly one location; however, in all pools, there were integrations in the same regions but with an offset of up to 900 bp. In the following measurements, cell pools with a mixture

of site-specific and random integration are considered. The unspecific integration seems to be truly random and should therefore perform like the random cell pool.

### 3.3 Analysis of transcript levels and productivity

After successful integration of eGFP into regions with specific histone modifications, the first step was to compare productivity and transcript levels with random integration. Cell pools were sub-cultured with selection pressure for 3 weeks, and the eGFP fluorescence intensity was then measured. The transcript level of eGFP was analyzed at the same time point *via* RT-qPCR, in order to ensure that any increased productivity could be correctly attributed to transcription rather than translation. The results are shown in Figure 5.

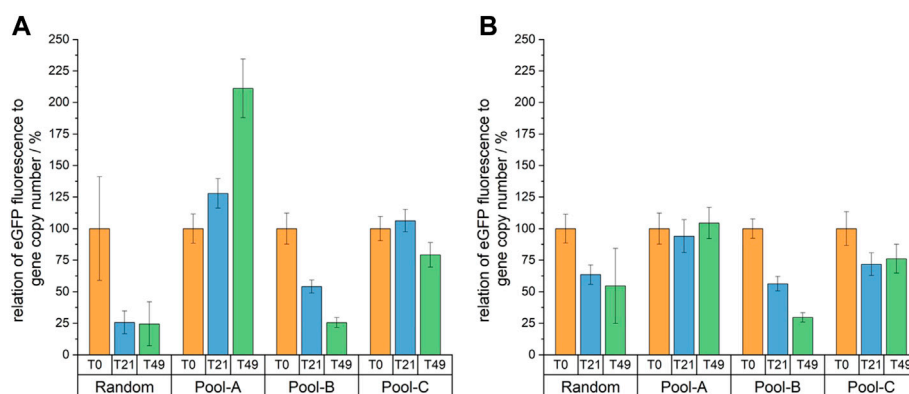
It was found that all site-specifically integrated cell pools showed a higher fluorescence intensity when compared with random integration. Specifically, Pool-A showed an increase by a factor of 7.4 ( $\pm 0.97$ ), Pool-B by a factor of 1.9 ( $\pm 0.97$ ), and Pool-C by a factor of 4.4 ( $\pm 0.73$ ). The transcript levels also showed almost the same fold changes compared to the random integration. There was a strong correlation between transcript levels and eGFP fluorescence (Pearson  $r=0.99$ ). Compared to the specificity of integration, the productivity data indicated a similar relationship. This suggests that productivity in monoclonal cell lines would be even higher, and at a similar level, in all site-specifically integrated cell pools. In summary, both the productivity and transcript levels of all site-specifically integrated cell pools were higher than those of the randomly integrated one.

### 3.4 Enhanced Transgene Expression Stability

To evaluate the stability of eGFP production, the three site-specifically integrated cell pools and the random cell pool were cultivated for 49 days with (4  $\mu\text{g/ml}$  puromycin) and without selection pressure. Fluorescence intensities and gene copy numbers were analyzed at the beginning (T0), in the middle (T21) and at the end (T49) of this long-term cultivation period.

In all long-term cultivations performed, the viability of the cultures was above 95% over the entire cultivation period, regardless of the addition of selection pressure. No differences were observed in the cell densities and growth rates that were achieved between any cultures at the same time. An increase in the specific growth rate  $\mu$  was observed equally in all cultures over the cultivation period. This rate was assessed once a week, and observed over a period of 3 days. The specific growth rate  $\mu$  initially averaged 0.90  $\text{days}^{-1}$ , and increased to an average of about 1.05  $\text{days}^{-1}$  by the end of cultivation. Similar observations of increased growth rates for CHO cells with increasing number





**FIGURE 6**

Relative eGFP fluorescence normalized to the gene copy number of days 0, 21 and 49 of a long-term cultivation with (A) and without (B) selection pressure. All values are shown relative to T0.

of passages were also noticed in other long-term cultivations (Kaneko et al., 2010; Beckmann et al., 2012).

To determine the stability of eGFP production, fluorescence intensities were normalized to gene copy numbers. The results are shown in Figure 6. In the progress of cultivation clear differences began to manifest due to the decreasing fluorescence intensities of the random cell pool—and by day 21, an increased fluorescence intensity could be detected for all site-specifically integrated cell pools compared to the randomly integrated cell pool. After 49 cultivation days, however, this could no longer be observed for cell pool B, as the fluorescence intensity of this cell pool decreased noticeably. For Pool-C, only a slight decrease in normalized fluorescence intensity was observed. For Pool-A, the intensity remained constant at about 120 without selection pressure and increased from about 175 to about double with selection pressure, presumably due to a selection on a subpopulation with strong fluorescence.

The criteria for the stability of a homogeneous cell line are the retention of 70% of the productivity over 70 generations, as well as the maintenance of “clinically relevant” properties such as structure, function, purity, chemical identity, and bioactivity (Dahodwala and Lee, 2019). As eGFP is not a clinically relevant product, this aspect cannot be assessed for this experimental approach. In addition, it should be noted that the present cultures are all cell pools, rather than homogeneous cell lines.

Nevertheless, it was observed that Pool-A remained stable over time in relation to productivity without selection pressure ( $97 \pm 13\%$  of productivity is maintained) and that productivity even increased under selection pressure ( $211 \pm 31\%$ ), due to selection to high fluorescent cells. This translates to a 15-fold increase in productivity after 49 days compared to the random integration. For Pool-C,  $79 \pm 14\%$  with puromycin and  $72 \pm 14\%$

without puromycin of the fluorescence intensity were preserved. If the standard deviations are not included, then this cell pool also fulfilled the stability criterion—although it should be noted that only about 51% of the cells corresponded to the desired site-specifically integrated cells. In a homogeneous cell line, the maintenance of productivity would have been possibly even higher, as the decrease was probably largely due to silencing in the random integration sites. The same considerations should also be kept in mind when looking at the percentage fluorescence intensity decrease of Pool-B. After 70 generations, the fluorescence intensity was only  $26 \pm 5\%$  and  $31 \pm 4\%$  of the initial value. Pool-B can thus not be described as stable. However, since only about 23% of the cells were site-specifically integrated, it is possible that a monoclonal cell line with this integration site could also be stable.

As expected, the fluorescence intensity of the random cell pool decreased from day 0 to day 49. This confirmed the typical problem of instability over time for CHO cell lines. Since the relative gene copies were determined by qPCR, and did not change over time, a loss of gene copy number can be excluded as a reason for the instability in this case.

## 4 Discussion

In terms of biopharmaceutical production, CHO cells represent the most prevalent platform—and despite decades of experiments, many parameters of cell line development (like expression levels and gene silencing) still remain very difficult to predict. Nevertheless, this study represents a proof-of-principle that safe harbors can be predicted using ChIP-Seq with a minimal set of histone modifications.

For this purpose, three histone modifications were analyzed genome-wide from a CHO-K1 bioreactor fed-batch cultivation.

The criterion for possible integration sites was defined as the intergenic occurrence of the modifications H3K4me3 and H3K27ac, and the absence of H3K9me3. Using this method, 709 potential integration sites with these specific histone modifications were identified. Compared to other safe harbor predictions based on a combination of multiple omics datasets (Dhiman et al., 2020; Hilliard and Lee, 2020), the method presented here can be much more readily adapted to individual host cell lines, as a single cultivation run with sequencing on a single flowcell is sufficient. Three integration sites were examined for expression strength and stability, only one (B) of which matched the landing pads predicted by Dhiman et al. (2020).

For integration, the CRISPR/Cas9 system was used. A donor vector coding for eGFP was site-specifically integrated. Outside the homology arms was the gene coding for mPlum, a second fluorescent protein, located. Thus, cells with site-specific integration displayed only green fluorescence, while cells with random integration displayed both green and red fluorescence. This double-fluorescence system was particularly useful during selection, showing an mPlum:eGFP ratio of 0.05–0.2 for site-specifically integrated cell pools and 1.3 for random integration. However, the system did not provide quantitative information about the specificity and selectivity of integration—so Cas9 targeted nanopore sequencing (Gilpatrick et al., 2020) was used first-in-CHO. To be able to identify the integration-sites, the method was adapted by inverting the crRNAs. With this technique, no PCR amplification and capture step is needed, and the read length is limited solely by the length of the DNA fragments. An enrichment of 86–244-fold was achieved with a mean read length of 12.6 kb. Sequencing showed that site-specific integration was achieved in all cell pools, but the specificity differed. Pool-A had a specificity of 73%, Pool-B 23% and Pool-C 51%. Thus, after limiting dilution, in the worst case about every fourth cell would show the desired integration, which would make the search for strongly and stably expressing cells unnecessary. By this, weeks could be saved in cell line generation, and it would no longer rely on chance. Surprisingly, in Pool-A an integration of mPlum into the desired integration site could also be detected in a few reads.

Until now, predictions of safe harbors from omics data sets have remained theoretical. Therefore, the described cell pools were directly used for the analysis of transcription, productivity and stability. An initial evaluation of the eGFP transcript levels and fluorescence showed an increase of 1.9–7.4-fold compared to the random integration without taking into account the specificity of the integration. Both analyses showed a high correlation for all cell pools. Even more important in this context is maintaining productivity of at least 70% over 70 generations (Dahodwala and Lee, 2019). To investigate this, the cell pools were sub-cultivated with and without selection pressure for 49 days (equivalent to 70 generations). In this period the specific growth rate increased from 0.9 days<sup>-1</sup>

to 1.05 days<sup>-1</sup> in all cultures. The same effect was previously observed in other CHO long-term cultivations (Kaneko et al., 2010; Beckmann et al., 2012). The eGFP fluorescence was measured over time and normalized to the gene copy number to analyze the silencing effect independent from the loss of gene copies. Random integration and Pool-B, which had the lowest integration specificity, are both unstable over the period analyzed. Pool-A and -C, on the other hand, maintain their productivity both with and without selection pressure. Under selection pressure, the productivity of Pool-A even doubles over time. Part of the effect could be explained by the fact that resistance to puromycin in cells with undirected integration is epigenetically silenced over time and they die. A similar increase in productivity over time was observed by Zhao et al. (2018) after site-specific integration into C12orf35 locus.

The strategy to integrate transgenes into regions with high levels of H3K4me3, H3K27ac and low levels of H3K9me3 in intergenic regions improved the development of stable cell lines significantly without altering the growth behavior. Constructing a stable recombinant CHO cell line takes between 6 and 12 months using conventional random integration (Lai et al., 2013). With the method presented in this study, targeted transfected cell pools with mostly intended integration-sites can be created in 3 weeks. After limiting dilution, just a handful of cells must be screened to obtain a stable cell line. Moreover, no laborious and time-consuming screening for integration sites by random integration is necessary for the identification of safe harbors (Zhou et al., 2019), but only a single ChIP sequencing experiment from one cultivation run.

## Data availability statement

The sequencing data presented in this study are deposited in the NCBI SRA, NCBI BioProject PRJNA865478.

## Author contributions

Conceptualization: OH; Methodology: OH, AN, TB, DB; Formal analysis and investigation: OH, AN; Writing—original draft preparation: OH, AN; Writing—review and editing: OH, AN, JB, TN, JK, DB; Resources: JK, TN; Supervision: JK, TN.

## Funding

The authors gratefully acknowledge support provided by the “European Regional Development Fund (EFRE)” through project “Cluster Industrial Biotechnology (CLIB) Kompetenzzentrum Biotechnologie (CKB)” (34. EFRE-0300095/1703FI04) for Oliver Hertel, Tobias Busche and David Brandt.

## Acknowledgments

The authors gratefully thank Xell AG for providing the medium for all cultivations, Biofidus AG for providing the fluorescence reader, and Larissa Leßmann (Bielefeld University) for her advice and help in the laboratory (RNA extraction and RT-qPCR).

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## References

- Andrews, S. (2010). FastQC: A Quality Control Tool for High Throughput Sequence Data [Online]. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>
- Bailey, L. A., Hatton, D., Field, R., and Dickson, A. J. (2012). Determination of Chinese hamster ovary cell line stability and recombinant antibody expression during long-term culture. *Biotechnol. Bioeng.* 109, 2093–2103. doi:10.1002/bit.24485
- Bandyopadhyay, A. A., O'Brien, S. A., Zhao, L., Fu, H.-Y., Vishwanathan, N., and Hu, W.-S. (2019). Recurring genomic structural variation leads to clonal instability and loss of productivity. *Biotechnol. Bioeng.* 116, 41–53. doi:10.1002/bit.26823
- Barnes, L. M., Bentley, C. M., and Dickson, A. J. (2003). Stability of protein production from recombinant mammalian cells. *Biotechnol. Bioeng.* 81, 631–639. doi:10.1002/bit.10517
- Bebbington, C. R., Renner, G., Thomson, S., King, D., Abrams, D., and Yarranton, G. T. (1992). High-level expression of a recombinant antibody from myeloma cells using a glutamine synthetase gene as an amplifiable selectable marker. *Nat. Biotechnol.* 10, 169–175. doi:10.1038/nbt0292-169
- Beckmann, T. F., Krämer, O., Klausning, S., Heinrich, C., Thüte, T., Büntemeyer, H., et al. (2012). Effects of high passage cultivation on CHO cells: A global analysis. *Appl. Microbiol. Biotechnol.* 94, 659–671. doi:10.1007/s00253-011-3806-1
- Broad Institute (2019). *Picard Toolkit*. Available at: <http://broadinstitute.github.io/picard/>. 12 July 2021.
- Chusainow, J., Yang, Y. S., Yeo, J. H. M., Toh, P. C., Asvadi, P., Wong, N. S. C., et al. (2009). A study of monoclonal antibody-producing CHO cell lines: What makes a stable high producer? *Biotechnol. Bioeng.* 102, 1182–1196. doi:10.1002/bit.22158
- Dahodwala, H., and Lee, K. H. (2019). The fickle CHO: A review of the causes, implications, and potential alleviation of the CHO cell line instability problem. *Curr. Opin. Biotechnol.* 60, 128–137. doi:10.1016/j.copbio.2019.01.011
- Derouazi, M., Martinet, D., Besuchet Schmutz, N., Flaction, R., Wicht, M., Bertschinger, M., et al. (2006). Genetic characterization of CHO production host DG44 and derivative recombinant cell lines. *Biochem. Biophys. Res. Commun.* 340, 1069–1077. doi:10.1016/j.bbrc.2005.12.111
- Dhiman, H., Campbell, M., Melcher, M., Smith, K. D., and Borth, N. (2020). 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.* 18, 3632–3648. doi:10.1016/j.csbj.2020.11.008
- Du, Z., Mujacic, M., Le, K., Caspary, G., Nunn, H., Heath, C., et al. (2013). Analysis of heterogeneity and instability of stable mAb-expressing CHO cells. *Biotechnol. Bioprocess Eng.* 18, 419–429. doi:10.1007/s12257-012-0577-1
- Feng, J., Liu, T., Qin, B., Zhang, Y., and Liu, X. S. (2012). Identifying ChIP-seq enrichment using MACS. *Nat. Protoc.* 7, 1728–1740. doi:10.1038/nprot.2012.101
- Gaidukov, L., Wroblewska, L., Teague, B., Nelson, T., Zhang, X., Liu, Y., et al. (2018). A multi-landing pad DNA integration platform for mammalian cell engineering. *Nucleic Acids Res.* 46, 4072–4086. doi:10.1093/nar/gky216
- Gilpatrick, T., Lee, I., Graham, J. E., Raimondeau, E., Bowen, R., Heron, A., et al. (2020). Targeted nanopore sequencing with Cas9-guided adapter ligation. *Nat. Biotechnol.* 38, 433–438. doi:10.1038/s41587-020-0407-5
- Hamaker, N. K., and Lee, K. H. (2018). Site-specific integration ushers in a new era of precise CHO cell line engineering. *Curr. Opin. Chem. Eng.* 22, 152–160. doi:10.1016/j.coche.2018.09.011
- Hefzi, H., and Lewis, N. E. (2014). From random mutagenesis to systems biology in metabolic engineering of mammalian cells. *Pharm. Bioprocess.* 2, 355–358. doi:10.4155/pbp.14.36
- Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y. C., Laslo, P., et al. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell.* 38, 576–589. doi:10.1016/j.molcel.2010.05.004
- Hilliard, W., and Lee, K. H. (2020). Systematic identification of safe harbor regions in the CHO genome through a comprehensive epigenome analysis. *Biotechnol. Bioeng.* 118, 659–675. doi:10.1002/bit.27599
- Jayapal, K. P., Wlaschin, K. F., Hu, W.-S., and Yap, M. G. S. (2007). "Recombinant protein therapeutics from CHO cells - 20 Years and counting," in *Chemical engineering progress*, 40–47.
- Kaneko, Y., Sato, R., and Aoyagi, H. (2010). Evaluation of Chinese hamster ovary cell stability during repeated batch culture for large-scale antibody production. *J. Biosci. Bioeng.* 109, 274–280. doi:10.1016/j.jbiosc.2009.09.044
- Karlič, R., Chung, H.-R., Lasserre, J., Vlahovicek, K., and Vingron, M. (2010). Histone modification levels are predictive for gene expression. *Proc. Natl. Acad. Sci. U. S. A.* 107, 2926–2931. doi:10.1073/pnas.0909344107
- Kawabe, Y., Komatsu, S., Komatsu, S., Murakami, M., Ito, A., Sakuma, T., et al. (2018). 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.* 125, 599–605. doi:10.1016/j.jbiosc.2017.12.003
- Kharchenko, P. V., Tolstorukov, M. Y., and Park, P. J. (2008). Design and analysis of ChIP-seq experiments for DNA-binding proteins. *Nat. Biotechnol.* 26, 1351–1359. doi:10.1038/nbt.1508
- Kim, M., O'Callaghan, P. M., Droms, K. A., and James, D. C. (2011). A mechanistic understanding of production instability in CHO cell lines expressing recombinant monoclonal antibodies. *Biotechnol. Bioeng.* 108, 2434–2446. doi:10.1002/bit.23189
- Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-Moussavi, A., et al. (2015). Integrative analysis of 111 reference human epigenomes. *Nature* 518, 317–330. doi:10.1038/nature14248
- Labun, K., Montague, T. G., Krause, M., Torres Cleuren, Y. N., Tjeldnes, H., and Valen, E. (2019). CHOPCHOP v3: Expanding the CRISPR web toolbox beyond genome editing. *Nucleic Acids Res.* 47, W171–W174. doi:10.1093/nar/gkz365
- Lai, T., Yang, Y., and Ng, S. K. (2013). Advances in Mammalian cell line development technologies for recombinant protein production. *Pharmaceuticals* 6, 579–603. doi:10.3390/ph6050579

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fbioe.2022.1010719/full#supplementary-material>

- Landt, S. G., Marinov, G. K., Kundaje, A., Kheradpour, P., Pauli, F., Batzoglou, S., et al. (2012). ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome Res.* 22, 1813–1831. doi:10.1101/gr.136184.111
- Langmead, B., Trapnell, C., Pop, M., and Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 10, R25. doi:10.1186/gb-2009-10-3-r25
- Lawrence, M., Daujat, S., and Schneider, R. (2016). Lateral thinking: How histone modifications regulate gene expression. *Trends Genet.* 32, 42–56. doi:10.1016/j.tig.2015.10.007
- Lee, J. S., Kildegaard, H. F., Lewis, N. E., and Lee, G. M. (2019). Mitigating clonal variation in recombinant mammalian cell lines. *Trends Biotechnol.* 37, 931–942. doi:10.1016/j.tibtech.2019.02.007
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078–2079. doi:10.1093/bioinformatics/btp352
- Li, H. (2018). Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics* 34, 3094–3100. doi:10.1093/bioinformatics/bty191
- Marx, N., Grünwald-Gruber, C., Bydlinki, N., Dhiman, H., Ngoc Nguyen, L., Klanert, G., et al. (2018). CRISPR-based targeted epigenetic editing enables gene expression modulation of the silenced beta-galactoside alpha-2, 6-sialyltransferase 1 in CHO cells. *Biotechnol. J.* 13, 1700217. doi:10.1002/biot.201700217
- Moritz, B., Woltering, L., Becker, P. B., and Göpfert, U. (2016). High levels of histone H3 acetylation at the CMV promoter are predictive of stable expression in Chinese hamster ovary cells. *Biotechnol. Prog.* 32, 776–786. doi:10.1002/btpr.2271
- Nicetto, D., and Zaret, K. S. (2019). Role of H3K9me3 heterochromatin in cell identity establishment and maintenance. *Curr. Opin. Genet. Dev.* 55, 1–10. doi:10.1016/j.gde.2019.04.013
- Noh, S. M., Shin, S., and Lee, G. M. (2020). “Cell line development for therapeutic protein production,” in *Cell culture engineering: Recombinant protein production*. Editors G. M. Lee and H. Fastrup Kildegaard (Weinheim: Wiley VCH), 23–47.
- O'Brien, S. A., Lee, K., Fu, H.-Y., Lee, Z., Le, T. S., Stach, C. S., et al. (2018). Single copy transgene integration in a transcriptionally active site for recombinant protein synthesis. *Biotechnol. J.* 13, 1800226. doi:10.1002/biot.201800226
- Oosting, R. S. (2019). “Molecular Biotechnology: From DNA sequence to therapeutic protein,” in *Pharmaceutical Biotechnology: Fundamentals and applications*. Editors D. J. A. Crommelin, R. Sindelar, and B. Meibohm 5th Edition. Cham, Switzerland: Springer Nature.
- Osterlehner, A., Simmeth, S., and Göpfert, U. (2011). Promoter methylation and transgene copy numbers predict unstable protein production in recombinant Chinese hamster ovary cell lines. *Biotechnol. Bioeng.* 108, 2670–2681. doi:10.1002/bit.23216
- Pristovšek, N., Nallapareddy, S., Grav, L. M., Hefzi, H., Lewis, N. E., Rugbjerg, P., et al. (2019). Systematic evaluation of site-specific recombinant gene expression for programmable mammalian cell engineering. *ACS Synth. Biol.* 8, 758–774. doi:10.1021/acssynbio.8b00453
- Quinlan, A. R., and Hall, I. M. (2010). BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841–842. doi:10.1093/bioinformatics/btq033
- Ran, F. A., Hsu, P. D., Wright, J., Agarwala, V., Scott, D. A., and Zhang, F. (2013). Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.* 8, 2281–2308. doi:10.1038/nprot.2013.143
- Recillas-Targa, F. (2006). Multiple strategies for gene transfer, expression, knockdown, and chromatin influence in mammalian cell lines and transgenic animals. *Mol. Biotechnol.* 34, 337–354. doi:10.1385/mb:34:3:337
- Romanova, N., and Noll, T. (2018). Engineered and natural promoters and chromatin-modifying elements for recombinant protein expression in CHO cells. *Biotechnol. J.* 13, 1700232. doi:10.1002/biot.201700232
- Sergeeva, D., Camacho-Zaragoza, J. M., Lee, J. S., and Kildegaard, H. F. (2019). CRISPR/Cas9 as a genome editing tool for targeted gene integration in CHO cells. *Methods Mol. Biol.* 1961, 213–232. doi:10.1007/978-1-4939-9170-9\_13
- Slesarev, A., Viswanathan, L., Tang, Y., Borgschulte, T., Achtien, K., Razafsky, D., et al. (2019). CRISPR/CAS9 targeted CAPTURE of mammalian genomic regions for characterization by NGS. *Sci. Rep.* 9, 3587. doi:10.1038/s41598-019-39667-4
- Tihanyi, B., and Nyitray, L. (2021). Recent advances in CHO cell line development for recombinant protein production. *Drug Discov. Today Technol.* 38, 25–34. doi:10.1016/j.ddtec.2021.02.003
- van Haasteren, J., Munis, A. M., Gill, D. R., and Hyde, S. C. (2021). Genome-wide integration site detection using Cas9 enriched amplification-free long-range sequencing. *Nucleic Acids Res.* 49, e16. doi:10.1093/nar/gkaa1152
- Walsh, G. (2018). Biopharmaceutical benchmarks 2018. *Nat. Biotechnol.* 36, 1136–1145. doi:10.1038/nbt.4305
- Wilson, C., Bellen, H. J., and Gehring, W. J. (1990). Position effects on eukaryotic gene expression. *Annu. Rev. Cell Biol.* 6, 679–714. doi:10.1146/annurev.cb.06.110190.003335
- Wippermann, A., and Noll, T. (2017). DNA methylation in CHO cells. *J. Biotechnol.* 258, 206–210. doi:10.1016/j.jbiotec.2017.08.005
- Wurm, F. M. (2004). Production of recombinant protein therapeutics in cultivated mammalian cells. *Nat. Biotechnol.* 22, 1393–1398. doi:10.1038/nbt1026
- Yang, Y., MariatiChusainow, J., and Yap, M. G. (2010). DNA methylation contributes to loss in productivity of monoclonal antibody-producing CHO cell lines. *J. Biotechnol.* 147, 180–185. doi:10.1016/j.jbiotec.2010.04.004
- Zhang, Y., Liu, T., Meyer, C. A., Eeckhoute, J., Johnson, D. S., Bernstein, B. E., et al. (2008). Model-based analysis of ChIP-seq (MACS). *Genome Biol.* 9, R137. doi:10.1186/gb-2008-9-9-r137
- Zhao, M., Wang, J., Luo, M., Luo, H., Zhao, M., Han, L., et al. (2018). Rapid development of stable transgene CHO cell lines by CRISPR/Cas9-mediated site-specific integration into C12orf35. *Appl. Microbiol. Biotechnol.* 102, 6105–6117. doi:10.1007/s00253-018-9021-6
- Zhou, M., Crawford, Y., Ng, D., Tung, J., Pynn, A. F., Meier, A., et al. (2011). Decreasing lactate level and increasing antibody production in Chinese Hamster Ovary cells (CHO) by reducing the expression of lactate dehydrogenase and pyruvate dehydrogenase kinases. *J. Biotechnol.* 153, 27–34. doi:10.1016/j.jbiotec.2011.03.003
- Zhou, S., Ding, X., Yang, L., Chen, Y., Gong, X., Jin, J., et al. (2019). Discovery of a stable expression hot spot in the genome of Chinese hamster ovary cells using lentivirus-based random integration. *Biotechnol. Equip.* 33, 605–612. doi:10.1080/13102818.2019.1601998