Targeted gene integration for improved production of therapeutic proteins in CHO cells

Sergeeva, Daria

Publication date:
2020

Document Version
Publisher's PDF, also known as Version of record

DTU Biosustain

Targeted gene integration for improved production of therapeutic proteins in CHO cells

PhD thesis

Daria Sergeeva

The Novo Nordisk Foundation Center for Biosustainability
Technical University of Denmark, 2020

Targeted gene integration for improved production of therapeutic proteins in CHO cells

Daria Sergeeva

PhD thesis

Supervisors:

Lars Keld Nielsen (University of Queensland, Technical University of Denmark)

Lise Marie Grav (Technical University of Denmark)

Gyun Min Lee (KAIST, Technical University of Denmark)

The Novo Nordisk Foundation Center for Biosustainability

Technical University of Denmark

August 2020

Targeted gene integration for improved production of therapeutic proteins in CHO cells
PhD thesis

2020

Daria Sergeeva

Published by: DTU, Novo Nordisk Foundation Center for Biosustainability, Kemitorvet Building 220, 2800 Kgs. Lyngby, Denmark
www.dtu.dk

Preface

This thesis is written as partial fulfilment of the requirements to obtain a PhD degree at the Technical University of Denmark. The work was carried out in the CHO cell line engineering and design group at the Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, from September 2017 to August 2020. From September 2017 to July 2018 the work was supervised by senior researcher and co-PI Helene Faustrup Kildegaard and co-supervised by professor Lars Keld Nielsen, from August 2018 to August 2020 the work was supervised by professor Lars Keld Nielsen and co-supervised by postdoc Lise Marie Grav and professor Gyun Min Lee. The PhD study included research stay at the Australian Institute for Bioengineering and Nanotechnology, University of Queensland, from April to June 2018.

The funding was provided by the Novo Nordisk Foundation as a part of Copenhagen Bioscience PhD programme (NNF16CC0020908).

Kgs. Lyngby, August 2020

Daria Sergeeva

Abstract

Recombinant therapeutic proteins are crucial medicines for the treatment of human diseases including cancer, infections, inflammatory and autoimmune diseases. The production of these proteins relies on mammalian cell factories - Chinese hamster ovary (CHO) cells. For each protein, a stable CHO cell line with high productivity and protein quality must be developed. Traditionally, CHO cell line development has been long and unpredictable and required massive resources to find a single cell line with the best performance. The recent development of precise genome editing tools for mammalian cells enabled a potential paradigm shift in CHO cell line engineering, offering more rapid and predictable ways of biopharmaceuticals production. This thesis aimed to develop and optimize new CHO cell line development methods based on targeted gene integration and encourage the change to next-generation cell line development platforms.

To advance CHO cell line development, we created a targeted gene integration platform using CRISPR/Cas9 and recombinases and showed its use for the production of valuable therapeutic proteins, including a vaccine against SARS-CoV-2. We demonstrated that this platform minimizes clonal variation, which in turn reduces the need for screening of cell lines and thus shortens the timeline of CHO cell line development. This platform enables robust comparative studies of CHO cells, which was illustrated by the analysis of CHO transcriptomes. Furthermore, we increased the productivity of cell lines using multi-copy targeted integration, reaching industrially-relevant titers of therapeutic proteins. We investigated the response to increased protein production in multi-copy cell lines using RNA-seq and revealed a transcriptional limitation in protein expression that appears at high copy numbers.

Overall, the thesis proves that targeted integration is an advantageous method for CHO cell line generation that minimizes genetic heterogeneity thus making cell line generation faster, more robust and predictable. It can clearly improve CHO cell line development, reducing time, resources and cost of delivering new biopharmaceuticals to the patients.

Dansk sammenfatning

Den biofarmaceutiske industri har sit fundament i produktionen af terapeutiske proteiner i levende cellefabrikker – og den foretrukne type cellefabrik er baseret på kinesiske hamster ovarie (CHO) celler. For at skabe en ny kræftbehandling eller en ny vaccine baseret på terapeutiske proteiner, skal nye CHO-cellelinjer genereres. Den traditionelle udvikling af sådanne cellelinjer har været empirisk, og det er derfor en tid- og ressourcekrævende proces. Med de nye fremskridt inden for præcise genredigeringsværktøjer til pattedyrceller, er et paradigmeskift mod hurtigere og rationel CHO-cellelinieudvikling blevet muligt. For at fremme dette skift har vi udviklet metoder til CHO-cellelinje genredigering baseret på præcis genintegration i forudbestemte områder i genomet. Ved hjælp af CRISPR/Cas9 og rekombinase metoder for genintegration har vi genereret CHO-cellelinjer med en ensartet og konsistent ydelse, hvilket fremskynder produktionen af terapeutiske proteiner. Vi anvendte en præcis genintegrationsplatform til produktion af værdifulde biofarmaceutiske produkter, herunder en vaccine mod SARS-CoV-2, og vist de fordele platformen har for komparative studier af CHO-celler. Vi udviklede en præcis multi-kopi genintegrationsmetode for at øge produktiviteten af CHO-cellelinjer og for at nå industrielle produktionsniveauer af terapeutiske proteiner. Således beviste vi, at præcis genintegration er en fordelagtig metode til CHO-cellelinieudvikling, hvilket hjælper med at reducere tiden, ressourcerne og omkostningerne tilknyttet udviklingsprocessen og leveringen af nye biofarmaceutiske produkter til patienterne.

Acknowledgements

I would like to thank the people who helped me to climb the mountain called “PhD”, to embrace the challenges and enjoy my way to the summit. This was not only a scientific adventure, but also a personal development journey, and I am grateful to everyone who supported me.

The Center for Biosustainability felt like the place I belonged the first minute I visited it in April 2016, during the Copenhagen Bioscience PhD programme interview tour. On that sunny day in Hørsholm, I met Helene, Lise and Jae who will become my teachers in the following years and whose kindness I will admire. I still remember how I was delighted that day by Helene's caring for the people, Lise's positive energy and Jae's bright mind. This atmosphere of support, cheer and scientific excellence stayed throughout my PhD, complemented by the inspiration from the people working in CfB. In four years of my journey, many things changed in CfB, but every person I encountered helped me to grow.

First of all, I am grateful to my supervisors, who gave me the opportunity to work in CfB and encouraged my personal and scientific growth. Helene, thank you for your support and confidence in me, your diligence to build an inclusive and family-like mood inside the CHO cell line engineering group, your great management and your drive to set a place for women in science. Thank you, Gyun Min, for connecting and directing the CHO cell line engineering group and providing valuable advice to my work. Lise, thank you for being together with me throughout all my journey, there is no way that I could have finished my PhD without you. You have been great in supporting me and giving the energy to move forward. And thank you, Lars, for your open mind, holistic perspective and guidance from the over side of the globe.

I was extremely lucky to be a part of the CHO program at CfB that connected brilliant people from all over the world. I would like to thank all past and present members of the CHO cell line engineering group, who built an amazing scientific community. Thank you, Manuel, Tae Kha, Che Lin, Kai, Johan, Urska, Ildze, Nachon, Nusa, Thomas, Ankita, Sara, Julie, Hooman, Saranya, Jae, Henning, Daniel, Eric, Thomas K, Michael and Kim. Thanks to all former and present members of CHO core,

who gave me excellent technical guidance: Bjørn, Sara, Patrice, Stef, Sanne, Helle, Michael, Tune, Anders, Johnny, Karen Kathrine, Karoline, Zulfiya, Marianne, Marie, Kristian, and Ivan from QMCM group. I would also like to thank Nathan Lewis from UCSD for his encouraging talks and insightful comments.

I want to thank Gert Talbo for sharing his wise words to me, and people from AIBN for welcoming me in Australia during my research stay. It was an incredible experience, I will never forget a few nights on the Great Barrier Reef that opened my eyes to new dimensions.

I am grateful for being in the NNF Copenhagen Bioscience PhD programme that built a spectacular network of talented students where I met fantastic friends. Thanks to all past and present CfB PhD students, whose enthusiasm and drive make life in CfB flourish. I am especially thankful to my Russian-speaking friends in Copenhagen who made my life in Denmark joyful. Sabina and Ivan, thank you for sharing these years with me, you are superb friends and flatmates. Thank you, Iuliia and Dmitrii, for your energy and fun activities we had together. Thank you, Ivan, for being open to deep conversations. Svetlana, Denis and Ekaterina, thanks for your care and cheer :)

Finally, I would like to thank my family who nourished my ability to learn and grow and always supported my choices.

List of publications

The overview of scientific articles and manuscripts resulting from the work of the thesis:

I. **CRISPR Toolbox for Mammalian Cell Engineering**
Daria Sergeeva, Karen Julie la Cour Karottki, Jae Seong Lee, Helene Faustrup Kildegaard. In *Cell Culture Engineering: Recombinant Protein Production (eds G.M. Lee, H. F. Kildegaard, S.Y. Lee, J. Nielsen and G. Stephanopoulos), Wiley- VCH, 2019.* https://doi.org/10.1002/9783527811410.ch8

II. **CRISPR/Cas9 as a genome editing tool for targeted gene integration in CHO cells**
Daria Sergeeva, Jose Manuel Camacho-Zaragoza, Jae Seong Lee, Helene Faustrup Kildegaard. In *CRISPR Gene Editing. Methods in Molecular Biology (eds Y.Luo), Humana Press, 2019*. https://doi.org/10.1007/978-1-4939-9170-9_13

III. **Minimizing clonal variation during mammalian cell line engineering for improved systems biology data generation**
Lise Marie Grav, Daria Sergeeva, Jae Seong Lee, Igor Marin de Mas, Nathan E Lewis, Mikael Rørdam Andersen, Lars Keld Nielsen, Gyun Min Lee, Helene Faustrup Kildegaard. *ACS synthetic biology, 2018.* https://doi.org/10.1021/acssynbio.8b00140

IV. **Multi-copy targeted integration for accelerated development of high-producing CHO cells**
Daria Sergeeva, Gyun Min Lee, Lars Keld Nielsen, Lise Marie Grav. *Accepted in ACS synthetic biology* https://doi.org/10.1021/acssynbio.0c00322

V. **Transcriptional response to recombinant protein production in isogenic multi-copy CHO cells**
Daria Sergeeva, Lise Marie Grav, Gyun Min Lee, Lars Keld Nielsen. *Manuscript in preparation*

The publication that I have contributed to during my PhD studies, but was not included in the thesis:

Using titer and titer normalized to confluence are complementary strategies for obtaining Chinese hamster ovary cell lines with high volumetric productivity of etanercept
Nuša Pristovšek, Henning Gram Hansen, Daria Sergeeva, Nicole Borth, Gyun Min Lee, Mikael Rørdam Andersen, Helene Faustrup Kildegaard. *Biotechnology journal, 2018.* https://doi.org/10.1002/biot.201700216

Table of contents

Thesis aim and structure

The aim of this thesis was to develop and optimize targeted gene integration towards improved production of therapeutic proteins in Chinese hamster ovary (CHO) cells. This thesis demonstrates different methods and applications of targeted gene integration, combining CHO genome editing, therapeutic protein expression and RNA-seq analysis.

The thesis is divided into following parts:

- **Introduction**, reviewing the field of therapeutic protein production in CHO cells and gene integration methods;
- **Chapter 1**, an overview of CRISPR/Cas tools for the editing of mammalian cells, with special focus on CHO cells;
- **Chapter 2**, a protocol for CRISPR/Cas9-mediated targeted integration in CHO cells;
- **Chapter 3**, a research paper showing the development of recombinase-mediated cassette exchange platform and its usage for the analysis of transcriptomics data;
- **Chapter 4**, a research paper describing the multi-copy targeted integration method for accelerated development of high-producing CHO cells;
- **Chapter 5**, a manuscript studying the transcriptional response to protein production in CHO cell lines generated by multi-copy targeted integration;
- **Conclusions and outlook**, providing a summary of the research and future directions;
- **Appendix**, a report demonstrating the application of targeted integration for the production of the vaccine against SARS-CoV-2 in CHO cells.

Introduction

1. Biotechnology and biopharmaceuticals

Biotechnology - the use of living organisms to make products - has served people for thousands of years. Using yeast for baking bread or brewing beer, growing fungi to make penicillin or bacteria to produce acetone: there are many examples of how biotechnology has been helping to feed, heal and fuel the world throughout the history of humankind. The bloom of biotechnology happened with the development of genetic engineering. Soon after uncovering the basic principles of molecular biology in the middle of the 20th century, scientists began to think about engineering living organisms for the benefit of people. The first experiments on recombinant DNA technology performed by Paul Berg, Herbert Boyer and Stanley Cohen in the 1970s have transformed our view on biology and marked the start of the bioengineering era. Today, with the help of genetic engineering, biotechnology creates new living organisms for the production of medicines that treat complex diseases, vaccines that train the immune system against pathogens, improved crops and foods that feed the world, and chemicals and biofuels that build sustainable industries.

One of the major applications of biotechnology is the production of biopharmaceuticals. Biopharmaceuticals in general are natural or modified products derived from living cells or the copies of biological molecules (DNA, RNA, peptides) made by chemical synthesis that are used for disease treatment and prevention. Some of the oldest forms of biopharmaceuticals are live attenuated vaccines and plasma-derived therapeutic proteins. With the development of genetic engineering, new, safer, and more potent forms of biomedicines were brought into practice, such as recombinant therapeutic proteins and recombinant vaccines, cell therapies, and gene therapies. While gene therapies (e.g. AAV-based gene therapies) and cell therapies (e.g. CAR-T cells) are at the forefront of biopharmaceuticals development in the 21st century,

recombinant therapeutic proteins continue to be the foundation of the biopharma industry. They are the major biomedical drugs for the treatment of metabolic disorders (insulin), genetic disorders (clotting factors), infections (interferons), inflammatory, autoimmune, and cancer diseases (antibodies).

Recombinant therapeutic proteins range from relatively small molecules, such as insulin, to complex and large molecules such as antibodies, coagulation factors, and cytokines (e.g. erythropoietin). Small therapeutic proteins were the first molecules to be produced by recombinant DNA technology: recombinant insulin was produced in bacteria (specifically *E.coli*) by Genentech and approved in 1982. The production of complex therapeutic proteins requires the use of more complex cells. So, the first recombinant protein synthesized in mammalian cells (specifically Chinese hamster ovary (CHO) cells) - recombinant tissue plasminogen activator (tPA) - was produced by Genentech in 1983. A few years later, recombinant erythropoietin (EPO) was produced in CHO cells by Amgen and approved for the treatment of anemia in 1989. The second breakthrough in biopharmaceuticals happened in the 1990s when humanized monoclonal antibodies were generated by recombinant DNA technology in mammalian cells. Monoclonal antibodies (mAbs) are universal proteins that can be altered to target various molecules in the body, helping to treat cancer, autoimmune and inflammatory disorders, infections, and other diseases. Today, monoclonal antibodies are the top-selling biopharmaceutical products, and more than 80% of approved mAbs are produced in CHO cells[1]. Overall, mammalian expression systems are predominant over nonmammalian systems for producing of recombinant proteins (84% of biopharmaceuticals approved in 2014-2018 were expressed in mammalian cells, the remaining were expressed in *E.coli* or yeast *S.cerevisiae*)[1].

The biopharmaceutical industry is ever-growing, and recombinant protein sales are approaching $200 billions per year[1]. The competitive biopharma market continues to drive innovation in therapeutic proteins, creating new forms: Fc-fusion proteins, antibody-drug conjugates, bispecific and multi-specific mAbs including T-cell engagers. Moreover, growing production of biosimilars makes therapeutic proteins more available for the public. Thus, the prevalence of protein-based biopharmaceuticals is likely to remain in the biopharma industry for the foreseeable future, while the new

biopharmaceutical products such as cell therapies and nucleic acid-based gene therapies are getting through the scientific, technological, manufacturing and regulatory hurdles to reach widespread adoption[1].

2. Mammalian cells for the production of therapeutic proteins

CHO cells were the first mammalian cells used for the production of complex therapeutic proteins. Fast growing, immortal and having similar molecular machinery to human cells, CHO cells have been used in the laboratories to study mammalian genetics since 1957[2]. What made CHO the first choice for the production of biopharmaceuticals in the 1980s was the development of the mutant CHO cell line that lacked the dihydrofolate reductase enzyme (DHFR). DHFR-deficient CHO cells can be used for selection and amplification of recombinant genes, which can significantly increase protein expression.

The success of the first drugs produced in CHO (tPA and EPO) helped to establish CHO as the main host for the production of complex therapeutic proteins. Although other mammalian cell lines have been used for protein production (murine NS0 and Sp2/0, baby hamster kidney (BHK21), human embryonic kidney 293 (HEK293), human fibrosarcoma HT-1080 and PER.C6 cell lines), CHO continues to be the dominant mammalian cell line. Besides the long history of approval of therapeutic proteins produced in CHO, this cell line has other key advantages such as resistance to human pathogenic viruses, ease of growth in large-scale serum-free suspension culture, and the ability to perform human-like post-translational modifications of the proteins[3].

3. Gene integration in mammalian cells

The key technique in the development of mammalian cell lines for the production of biopharmaceuticals is the establishment of a stable protein-producing cell line. The stability of the cell line ensures the robust production of therapeutic protein over a long time, from one large-scale cultivation to another. To establish a stable cell line, a gene encoding recombinant protein together with genetic elements that support its

expression is introduced into the cell and gets integrated into the genome. Thus, the gene will be passed through generations with each cell division.

The integration of exogenous DNA into the mammalian genome can be performed by various methods that can generally be divided into random integration and targeted integration. In random integration, the gene is integrated into different sites in the genome with unpredictable DNA rearrangements and copy numbers, leading to cell heterogeneity known as clonal variation. Targeted gene integration ensures precise insertion of exogenous DNA into specific sites in the genome with a control of copy number. The cells generated by targeted integration are isogenic, meaning that these cells have identical genotypes[4] (Figure 1).

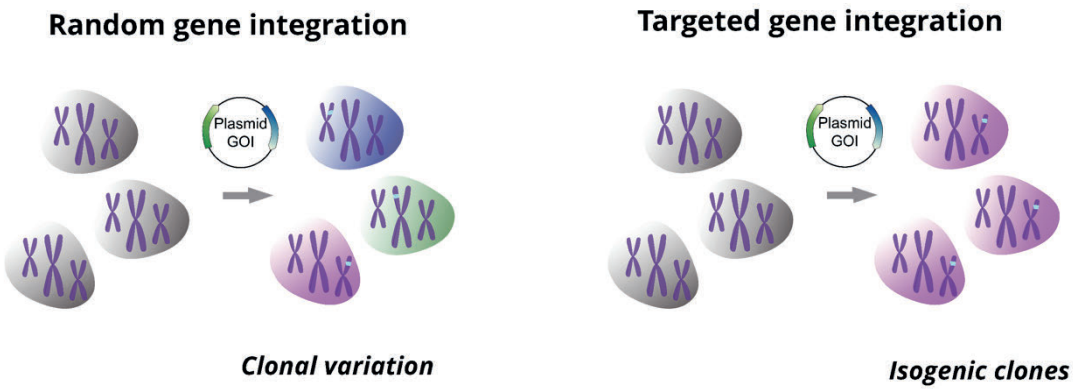


Figure 1. Gene integration in mammalian cells. Two approaches are used for the integration of exogenous DNA into the mammalian genome: random integration and targeted integration. During random gene integration, the plasmid with the gene of interest (GOI) integrates into various genomic sites with varying copy number, accompanied by DNA rearrangements. Therefore, random integration leads to high heterogeneity in the cell population known as clonal variation. During targeted gene integration, a plasmid with GOI is inserted precisely into the defined genomic site with controlled copy number, thus generating a homogenous population of isogenic cells.

3.1. Random gene integration and selection systems

When cultured mammalian cells are exposed to exogenous DNA, a small subpopulation of cells stably integrates this DNA into chromosomes. Since such events are rare, the identification of modified cells requires the use of selection markers.

The introduction of exogenous genes into the mammalian genome became possible with the development of the first mammalian selection system in 1977. In this system, cultured mammalian cells lacking thymidine kinase (Tk^-) can be restored to Tk^+ state by the integration of herpes virus thymidine kinase gene (*HSV-tk*) provided in plasmid DNA[5]. The frequency of cells with the integrated *tk* gene in the genome was approximately one cell per million, however, by growing the population in the HAT medium, Tk^+ cells could be selected over Tk^- cells. The analysis of modified cells showed that exogenous DNA stably integrates into different chromosomal locations in multiple copies, and the integrated DNA was often segmented[6,7]. Thus, the plasmid integration in mammalian cells was shown to be random, leading to the generation of heterogeneous cell populations.

In the early 1980s, another mammalian selection system was developed, which is based on the antibiotic G418 and neomycin-resistance gene (*neoR*)[8,9]. Whereas normal non-modified cells are killed by the antibiotic G418, those that acquire and express *neoR* continue to grow in the presence of G418. The development of this system allowed the use of non-mutant recipient cells for gene integration, broadening its application to a wide variety of mammalian cells. Later, similar antibiotic selection systems using hygromycin, puromycin, zeocin, and blasticidin were brought into practice for mammalian cell engineering[10].

The first selection system that became popular in mammalian protein production is DHFR. In 1978 it was shown that mammalian cells that are resistant to increased concentrations of methotrexate (MTX) contain increased copies of *dhfr* gene[11]. This phenomenon was called gene amplification, and it was proposed that gene amplification can significantly increase protein expression. In 1982 Randal Kaufman and Phillip Sharp transfected DHFR-deficient CHO cells with a plasmid containing *dhfr* gene together with the early region of simian virus 40 (SV40). After sequential increases in MTX concentration, selected cells contained up to 1000 copies of transfected DNA and were producing increased amounts of SV40 small tumor antigen (over 10% of total protein synthesis)[12]. The characterization of clones demonstrated that plasmid DNA was randomly integrated into the CHO genome and was frequently deleted and rearranged, showing instability and variability of clones generated by gene amplification. This caveat

of the DHFR system was not considered crucial by the first biotech companies who used it for protein production since the ability to overproduce recombinant proteins by gene amplification was more important.

An alternative CHO selection system based on the glutamine synthetase (GS) was developed in 1990[13]. GS catalyzes the production of glutamine from glutamate and ammonia, which is necessary for cell growth. CHO cells that are glutamine prototrophs can be engineered to delete the endogenous GS gene, alternatively, GS activity can be inhibited by methionine sulfoximine (MSX)[14]. When transfecting recombinant DNA with GS gene into CHO, the cells with gene integration can be selected in glutamine-free media with MSX. Even a single round of MSX selection is sufficient to achieve high protein production without gene amplification[15]. The GS system became increasingly used for production of biopharmaceuticals due to a shorter timeline compared to the DHFR system. However, similar to other random integration strategies, GS-mediated random integration results in the instability and high variability of clones.

3.2. Targeted gene integration

Random gene integration has become very successful in the biotech industry in establishing mammalian cell lines for high protein production, despite frequent instability of clones, loss of productivity, heterogeneity, and unpredictability of outcomes. However, random integration is not optimal for the genetic manipulation of mammalian cells for basic genetics studies. Variability of gene expression both from chromosomal position effects and from copy number variation are hurdles when specific genes or their mutations are studied. Targeted gene integration methods have been developed to get control over gene integration position and outcomes.

The discovery of homologous recombination in mammalian cells in the middle 1980s allowed to target exogenous DNA into a specific location in the genome and precisely introduce genes or mutations[16,17], which became the first method of targeted integration in mammalian cells. To modify a predetermined genomic site by homologous recombination, donor DNA should be flanked by sequences homologous to the integration site (homology arms). Although the frequency of homology-directed targeted integration is low (about one in a thousand cells), the edited cells can be efficiently

enriched by positive-negative selection using *neoR* and *HSV-tk* markers[18]. The method of gene targeting by homologous recombination became routine for the engineering of mammalian organisms, resulting in the creation of a collection of knockout mouse models for biomedical studies, and was recognized by a Nobel Prize in 2007[19].

Homologous recombination relies on the homology-directed DNA repair machinery of the cell, however, its efficiency in mammalian cells is low. To enhance homology-directed repair, a double-stranded break (DSB) can be introduced in the target locus by site-specific endonucleases[20]. The sites of DSB generation can be controlled by the use of programmable endonucleases. In the 2000s it was shown that by co-delivering programmable zinc-finger nuclease (ZFN) together with a donor plasmid bearing homology arms, a gene can be integrated precisely into the genome with high frequency (5-15%) without selection[21]. Soon after this discovery, a toolbox of site-specific programmable nucleases was expanded and included not only ZFN but also transcription activator-like effector nucleases (TALENs) and CRISPR/Cas systems (clustered regularly interspaced short palindromic repeats/CRISPR-associated protein)[22]. While ZFNs and TALENs are programmed by customizing the amino acid sequence of their DNA-binding domains, Cas proteins are programmed by short guide RNA, providing great flexibility and simplicity to precise gene editing by CRISPR/Cas systems. With the explosive development of CRISPR technology for mammalian genome editing, a range of techniques for CRISPR-mediated gene integration and modification appeared, revolutionizing the whole field of genetic engineering[23].

DSB in the mammalian genome can be repaired by multiple DNA repair mechanisms, which can be divided into two main arms: homology-directed repair (HDR) and end-joining (EJ). HDR leads to precise editing of the target site, but it is generally less active in mammalian cells compared to EJ. End-joining (non-homologous end joining (NHEJ) or microhomology-mediated end joining (MMEJ)) leads to uncontrolled and less predictable insertions or deletions (indels), but it is typically more efficient than HDR[23]. By using programmable nucleases for DSB generation, a gene can be inserted into a specific genomic site by either of these pathways (Figure 2a). Recent studies have shown that NHEJ and MMEJ can be used for gene integration in mammalian cells with efficiency up to 40-60% using double-stranded DNA (dsDNA) for CRISPR/Cas9-mediated

homology-independent integration in HEK293 cells[24]. However, EJ-mediated integration is accompanied by mutations in the target site and donor DNA, so it is less preferable for precise genome editing. HDR can be used to precisely integrate genes into mammalian cells with efficiency up to 50% without selection, using single-stranded DNA (ssDNA) in T cells[25] or dsDNA in HEK293 cells[26] by CRISPR/Cas9-mediated gene integration. However, even when donor DNA is designed for HDR-mediated integration, outcomes of repair might be imprecise due to EJ-mediated repair happening at the same time[26]. Thus, although CRISPR/Cas is a flexible tool for gene integration, the reliance on DNA repair machinery of the cell dictates the efficiency and precision of gene integration, which varies depending on the cell type, availability of DNA repair factors, design of donor DNA and location of the integration site.

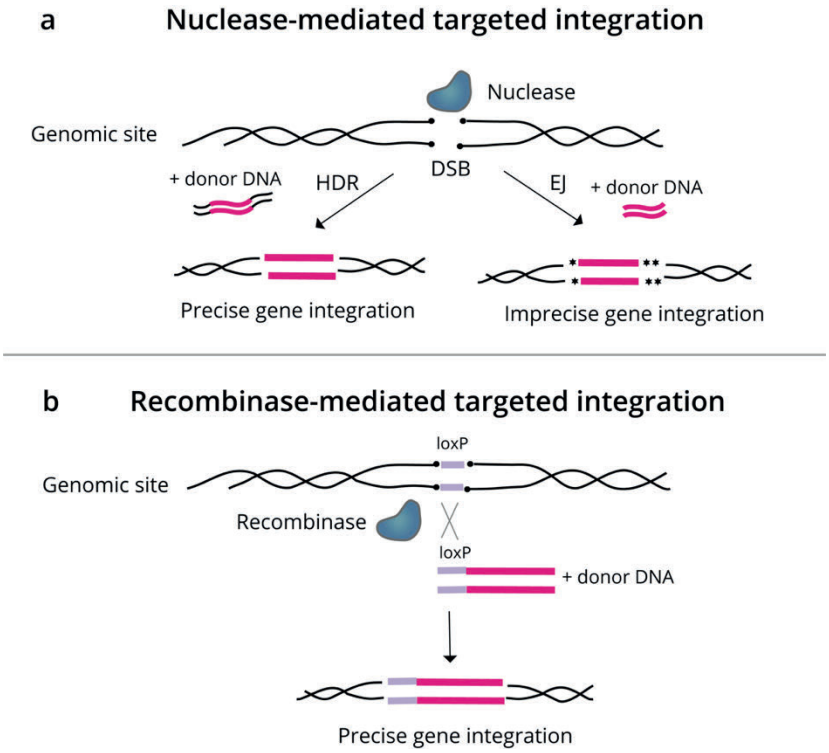


Figure 2. Main strategies for targeted gene integration. Two main strategies are currently used for targeted gene integration: nuclease-mediated (a) and recombinase-mediated (b). **a.** Nuclease facilitates targeted integration by introduction of a double-stranded break (DSB). DSB is repaired by cellular DNA repair machinery via homology-directed repair (HDR) or end-joining (EJ), leading to the integration of donor DNA. **a.** In recombinase-mediated targeted integration, DSB is not required, and gene integration is directed by site-specific recombinase via autonomous recombination between recombination sites (loxP) located on genomic DNA and donor DNA.

Alternatives to gene integration mediated by nucleases is gene integration mediated by site-specific recombinases and transposases. Recombinases and transposases do not rely on cellular DNA repair machinery and can ligate DNA autonomously, thus, they should be functional in nearly any cell type independent of cell state and lead to more predictable outcomes.

Transposons are mobile genetic elements located in the genome, which can be divided into two large groups: retrotransposons that transpose DNA by "copy-and-paste" mechanism (they are first transcribed to RNA, then reverse transcribed back into DNA and integrated into a new genomic site) and DNA transposons that transpose by "cut-and-paste" mechanism (they are excised from one site in the genome and inserted into the new site)[27]. Sleeping Beauty (SB) is the first "cut-and-paste" transposon that was used for the genetic modification of mammalian cells. SB transposase can promote the integration of donor DNA flanked by terminal repeat sequences into TA-rich regions throughout the genome. Similar transposases such as piggyBac, Tol2 and Leap-In have been later used in mammalian cells[28,29]. Although transposase-mediated integration is not programmable and semi-random[30], there have been attempts to fuse transposase to DNA-binding domains or Cas proteins to direct the integration into a specific genomic site in mammalian cells[31–34]. Transposase fusions showed biased activity at their intended genomic sites, however, significant off-target integration was observed. The development of programmable fusions of retrotransposases for gene integration in mammalian cells has also been reported[35]. Recently, CRISPR-associated transposases were discovered in bacteria and were used for RNA-guided targeted gene integration, however, off-target effects were also detected[36,37]. In principle, when optimized for mammalian cells and improved for on-target integration, programmable transposases could become a promising tool for gene integration, targeting any site in the genome and promoting the integration of donor DNA without introduction of DSB.

Site-specific recombinases are proteins that enable predictable and targeted gene integration while avoiding DSB induction (Figure 2b). During site-specific recombination, two short 30-50 bp DNA sequences (recombination sites) at separate locations are recognized and rejoined by recombinase without DNA synthesis, degradation, or cofactors. As a result, DNA can be integrated, inverted or excised depending on the

position and orientation of recombination sites[38]. Recombinases have high specificity to recombination sites, and the sequences of recombination sites are constrained, thus the off-target activity of recombinases is low. Since the first application of site-specific recombinases Cre and FLP in mammalian cells in the early 1990s[39,40], recombinases have become widespread in mammalian cell line and mammalian organism engineering[38,41,42]. Prior to the recombinase-mediated integration of the desired gene, recombination sites must be placed in the genome by other integration methods (random or targeted). Although this is an extra step compared to other targeted integration strategies, the cell with recombination sites is universal and can be used as a platform for precise recombinase-mediated genome engineering. Site-specific recombinases could become close to ideal tools for targeted gene integration due to their independence of cellular DNA repair and high specificity, however, they have limitations. While *in vitro* activity of recombinases can approach 100%[38], *in vivo* activity in mammalian cells is much lower and integration efficiency rarely exceeds 20%[43], perhaps due to complex chromatin states and inefficient DNA delivery into the cell. Moreover, recombinases are not programmable and are constrained on the recombination site sequences, making the first step of recombination site placement in the genome unavoidable. The dream about designer recombinases that can be configured to recombine sequences of any choice in any target site has been around for many years. However, the attempts of directed evolution of recombinases, fusions of recombinases to DNA-binding domains or designing structure-specific recombination sites have not yet resulted in the development of more efficient and programmable recombinases for mammalian cell engineering[38,44–46].

To summarize, targeted gene integration methods are crucial for the precise genetic engineering of mammalian cells, and new techniques continue to develop. Nevertheless, there is not yet a single method that would fit the “wish list”. The ideal targeted integration tool should be: i) highly efficient on-target and show no off-target integration, ii) fully and easily programmable to modify any site in the genome, iii) controllable to introduce only desirable outcomes, iv) applicable to any cell type.

4. CHO cell line development using targeted integration

Both random integration and targeted integration have been used for the generation of protein-producing CHO cell lines, however, random gene integration is currently the prevalent method for the industrial production of biopharmaceuticals. Its popularity can be explained by the historical legacy, simplicity of gene introduction, and, most importantly, the ability to generate high-producing cells that have multiple copies of GOI inserted in the genome. The drawbacks of random integration - heterogeneity and instability of clones - oblige scientists to do a time-consuming and labor-intensive screening of clones, bioprocess optimization and stability testing each time a new protein has to be produced. While lab automation can help to streamline this process, random integration remains unpredictable, long and resource-intensive.

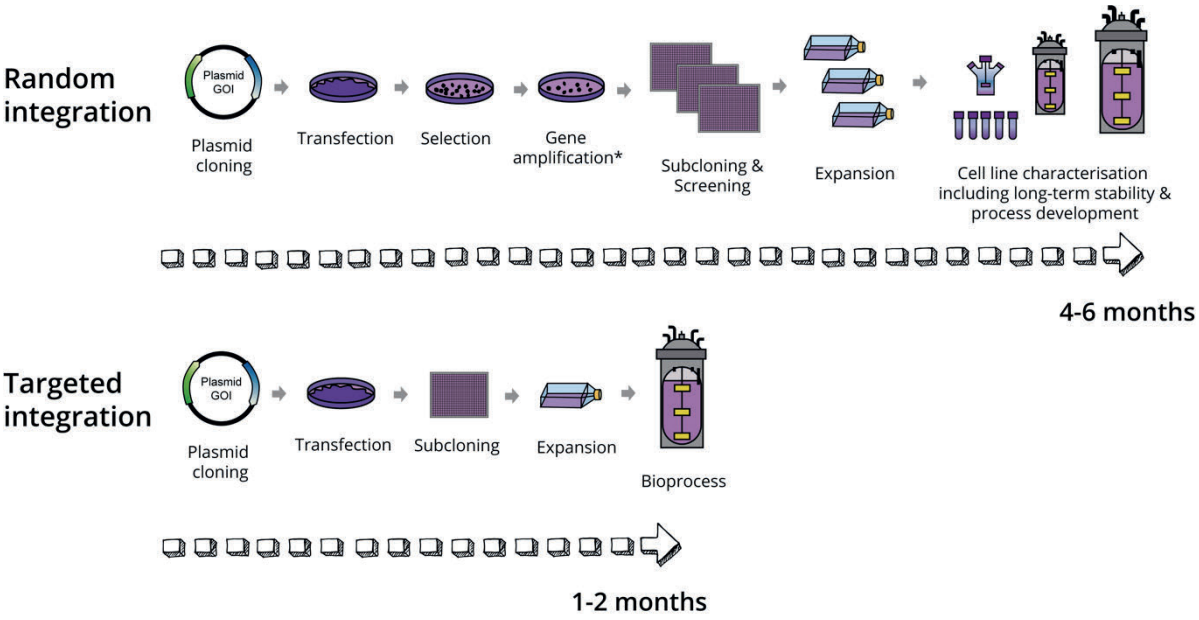


Figure 3. Stable CHO cell line generation for the production of therapeutic proteins. Traditional methods for cell line generation using random integration and DHFR or GS selection systems last about 4-6 months and require extensive cell line screening (*GS system does not require gene amplification step). Targeted integration approach reduces the time of cell line development and ensures predictable cell performance.

Targeted integration can accelerate cell line development and save resources[47]. Isogenic clones with predictable and stable performance do not require tedious screening, thus the cell line development process can be shortened to 1-2 months instead of 4-6 months (Figure 3). Nevertheless, industrial use of targeted integration has

been limited up until now, mainly due to the low productivity of clones that had only a few copies of GOI integrated into the genome. Besides that, the obstacles with finding optimal genomic sites and fights over CRISPR IP have hindered the adoption of targeted integration technologies in the biopharma industry.

In recent years, the interest in targeted integration has been rising in the biopharmaceutical industry, and several companies have published papers[48–53] and patents[54–57] describing targeted integration methods for CHO cell line generation. All these methods utilize site-specific recombinases such as Cre, Flp and Bxb1 for single-site recombination or recombinase-mediated cassette exchange (RMCE), with the latter being more prevalent. For the introduction of the "landing pad" with recombination sites into the CHO genome, random integration or CRISPR/Cas9-mediated targeted integration is used. By screening the sites of integration, several transcriptionally active "safe harbor" genomic sites in CHO cells have been identified[58,59]. The productivity of clones generated by targeted integration has been improving over the years, primarily by the increase of GOI copy number. In such a way, Genentech team demonstrated in 2013 that by insertion of two mAb cassettes by Cre-mediated RMCE into a single genomic site, specific productivity (q_P) can be doubled and reach 10 pg/cell/day (pcd)[49]. Similarly, a linear increase in antibody expression was demonstrated by the Pfizer/MIT team in 2018, when multiple copies of mAbs cassettes were integrated into multiple sites in the CHO genome[53]. In 2020, Genentech improved mAbs q_P up to 20-50 pcd by the integration of multiple copies of heavy and light chains into a single genomic site in the RMCE system[52], thus reaching productivity normally expected from random integration clones. These studies show that upon optimization, targeted integration can become an efficient method for CHO cell line development, offering accelerated and predictable strategies for high therapeutic protein production. However, the commercial nature means that these studies lack the necessary details on the design of the expression systems and do not report any limitations of the methods, making them difficult to reproduce. To make targeted integration more accessible for the scientific community and more popular for biopharmaceutical production, an open, systematic and detailed analysis of multi-copy targeted integration is required.

5. CHO cell line engineering: building next-generation cell factories

CHO cells have not evolved to be a secretory cell growing at high density in suspension culture: this cell line originates from a fibroblast cell in ovary tissue, so it has natural limitations with regard to protein modification and secretion capacity. In the past 30 years, most of the improvements in therapeutic protein production in CHO cells were achieved by optimization of cultivation conditions (media, feeding, bioprocess). Although numerous, the attempts to improve CHO secretory capacity and protein quality by genetic engineering (knockout or overexpression of genes) have achieved moderate success compared to bioprocess optimization (reviewed in [60,61]). There have been several obstacles that limited CHO cell engineering: lack of efficient genome editing tools and genetic elements, missing information about the CHO genome, fragmented understanding of molecular mechanisms underlying efficient protein production. Although these obstacles are now surmountable due to the development of CRISPR/Cas9 editing tools[62] and sequencing of CHO genome[63], one additional obstacle continues to persist - clonal variation. Clonal variation caused by random gene integration makes CHO cell engineering unpredictable and often non-reproducible[61]. Moreover, it hinders understanding of CHO biology using omics technologies, causing a substantial variation in global expression due to random gene integration[64,65]. Targeted integration can advance CHO cell engineering and analysis of omics data by reducing clonal variation during cell line development. Moreover, it can be used for predictable and rational engineering of cells. For example, targeted integration can be applied to control post-translational modifications of therapeutic proteins[66], to determine the causes of low protein expression[67] or to reduce apoptosis by controlled overexpression of effector genes[68].

To build next-generation CHO cell factories for the efficient production of complex biopharmaceuticals, a holistic approach should be taken. This includes the design and optimization of expression cassettes by synthetic biology, host cell engineering by genome editing and unraveling protein production mechanisms by systems biology. Advanced host cell lines in combination with high-performing vector systems, a rapid cell line generation process, and efficient bioprocess platforms form a basis for improved production of therapeutic proteins in CHO cells[60]. Targeted integration can

help to progress biopharmaceuticals production in all these areas: by building a robust system for gene expression, modifying the CHO genome for efficient protein secretion by effector gene overexpression, and accelerating cell line and bioprocess development by reducing clonal variation. The presented PhD thesis sought to encourage this progress by developing and optimizing targeted integration methods for CHO cell engineering.

Chapter 1 - CRISPR toolbox for mammalian cell engineering

This chapter provides an overview of CRISPR technologies for mammalian cell engineering that facilitate efficient knockout, knockin, activation, and repression of genes as well as epigenetic modifications. Specific attention is devoted to the applications of CRISPR for the editing of CHO cells towards improved production of biopharmaceuticals.

The chapter is reprinted from: Sergeeva, D., la Cour Karottki, K. J., Lee, J. S., and Kildegaard, H. F. (2019) "CRISPR Toolbox for Mammalian Cell Engineering", Cell Culture Engineering: Recombinant Protein Production, (eds G.M. Lee, H. F. Kildegaard, S.Y. Lee, J. Nielsen and G. Stephanopoulos), Wiley-VCH, 2019.

8

CRISPR Toolbox for Mammalian Cell Engineering

Daria Sergeeva[1], Karen Julie la Cour Karottki[1], Jae Seong Lee[1,2], and Helene Faustrup Kildegaard[1]

[1] The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Building 220, Kemitorvet, 2800 Kgs. Lyngby, Denmark
[2] Department of Molecular Science and Technology, Ajou University, 206 Worldcup-ro, Yeongtong-gu, Suwon 16499, Republic of Korea

8.1 Introduction

The conversion of bacterial CRISPR/Cas9 (clustered regularly interspaced short palindromic repeats/CRISPR-associated protein 9) immune system into a simple and versatile genome editing tool has revolutionized biological research. CRISPR/Cas9 has been rapidly adapted for a vast range of applications in diverse organisms. Especially, CRISPR technology has transformed the engineering of mammalian cells, providing tools for precise and efficient genome manipulations including gene knockout and knock-in, transcriptional activation and repression, and epigenetic modifications. Nowadays, CRISPR has been implemented in many research groups to study cell biology, establish human disease models, develop new therapeutic methods, and build complex synthetic gene circuits in mammalian cells [1–4]. Moreover, functional studies have been facilitated by genome-scale CRISPR screens, where knockout or transcriptional modulation approaches were used to elucidate gene functions [5].

Advances in CRISPR/Cas9 offer new opportunities in biotechnology toward the development of cell factories producing chemicals and drugs. The biotechnological potential of CRISPR/Cas9 has been demonstrated by metabolic engineering of microbial cell factories (bacterial and yeasts cells) for bio-based production of chemicals and fuels [6]. In the case of mammalian cell factories, the potential of CRISPR/Cas9 can be illustrated by optimizing Chinese hamster ovary (CHO) cells, the most commonly used host cells for production of therapeutic glycoproteins. This chapter will describe recent developments of CRISPR technology for mammalian cell engineering and will discuss how the technology can be applied in CHO cell engineering toward improved production of biopharmaceuticals.

Cell Culture Engineering: Recombinant Protein Production,
First Edition. Edited by Gyun Min Lee and Helene Faustrup Kildegaard.

8.2 Mechanism of CRISPR/Cas9 Genome Editing

CRISPR/Cas systems have evolved in prokaryotes as a defense mechanism against foreign genetic elements such as viruses [7]. The most investigated CRISPR/Cas9 system (class 2 type II; described in Section 8.3.1) is composed of three main components: Cas9 endonuclease, CRISPR RNA (crRNA), and trans-activating crRNA (tracrRNA) (Figure 8.1). crRNAs are processed from CRISPR arrays that are clusters of repeat sequences interspaced by variable sequences (spacers) homologous to foreign genetic elements (protospacers). Transcribed crRNA and tracrRNA hybridize to a RNA duplex known as guide RNA (gRNA) that binds to the Cas9 protein and form an active ribonucleoprotein

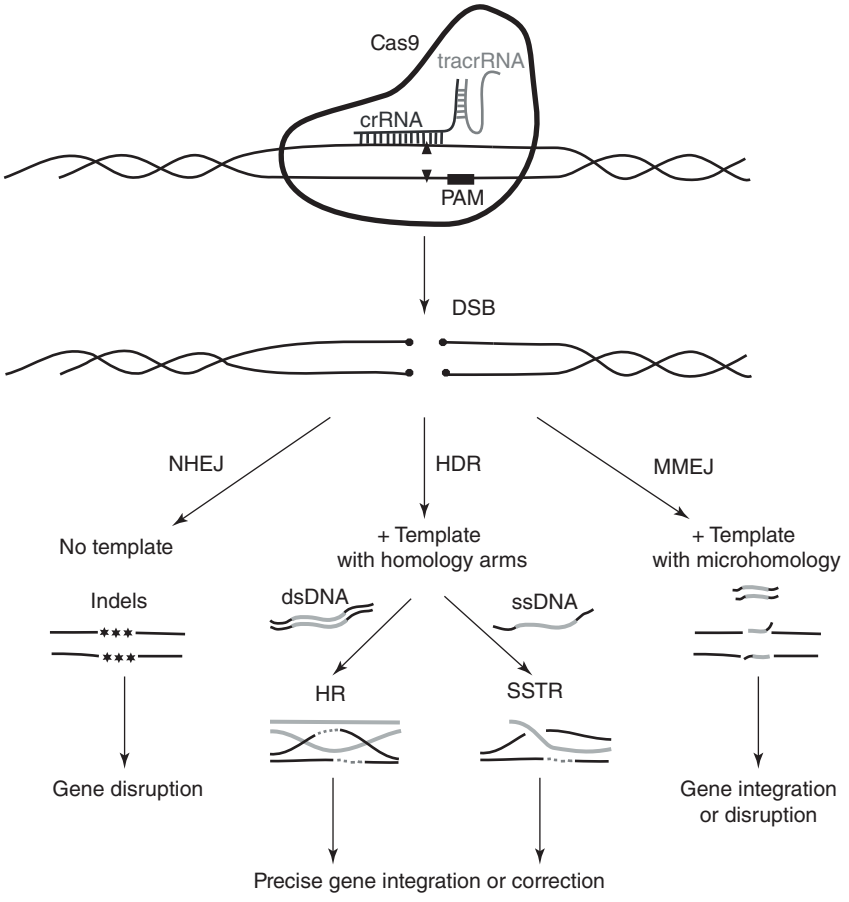


Figure 8.1 *CRISPR/Cas9-mediated genome editing*. gRNA-directed Cas9 induces a double-stranded break (DSB) in DNA, which can then be repaired by cellular DNA repair mechanisms. Error-prone NHEJ-mediated repair can introduce indels of variable length at the site of the DSB, resulting in gene disruption. HDR-mediated repair can lead to precise gene integration or correction when double-stranded or single-stranded DNA donor template with homology arms is provided. Alternatively, gene integration or disruption can result from MMEJ-mediated repair with an assistance of short homologous sequences in a donor template.

(RNP) complex. With the guidance of the gRNA, this complex searches for complementary sequences in foreign DNA and directs its cleavage by two nuclease domains of Cas9. Each nuclease domain cleaves one strand of DNA; hence, Cas9 activity leads to a double-stranded break (DSB) of DNA [8].

The breakthrough in the field happened, when it was shown that the bacterial CRISPR/Cas9 system could be used for genome editing in eukaryotic cells, especially in mammalian cells [9, 10]. To simplify the system, the crRNA:tracrRNA duplex was fused into a chimeric single guide RNA (sgRNA) [9]. gRNA binds to target DNA through an approximately 20-nucleotide (nt) region that is adjacent to the protospacer-adjacent motif (PAM), which is recognized by the PAM-interacting domain of Cas9. Thus, by customizing a 20-nt region of the gRNA to pair with the DNA sequence of interest, Cas9 can be targeted to any genomic locus containing a PAM sequence, making it an easily programmable platform for genome editing.

DSBs generated by Cas9 activate the intrinsic cellular DNA repair mechanisms in mammalian cells, such as non-homologous end joining (NHEJ) and homology-directed repair (HDR). These DNA repair pathways predominate at different cell cycle phases and recruit various molecular factors to perform the repair [11]. Mammalian cells may generate random insertion/deletion mutations (indels) at the site of DSB via NHEJ, leading to the potential disruption of genes and functional gene knockout. In contrast, HDR allows precise targeted gene integration, gene replacement, or correction. HDR can be exploited in the presence of a double-stranded DNA (dsDNA) or single-stranded DNA (ssDNA) template with homologous regions spanning the DSB. Depending on the DNA used as a template, homologous recombination (HR) or single-stranded template repair (SSTR) occurs [12]. Besides NHEJ and HDR, an alternative repair mechanism called microhomology-mediated end joining (MMEJ) can occur at the site of DSB. MMEJ requires short homologous sequences (5–20 nt) for DSB repair and harnesses HDR- and NHEJ-independent DNA repair machinery. All these naturally occurring DNA repair pathways are utilized in CRISPR/Cas9-mediated genome editing for different genome engineering purposes.

8.3 Variants of CRISPR-RNA-guided Endonucleases

8.3.1 Diversity of CRISPR/Cas Systems

CRISPR/Cas systems display a wide evolutionary diversity in bacteria and archaea. Based on the differences in their components, CRISPR/Cas systems have been divided into two classes: class 1 systems (types I, III, and IV) that rely on multi-subunit protein complexes and class 2 systems (types II, V, and VI) that utilize single-effector proteins [7].

The widely studied DNA-targeting CRISPR/Cas9 system belongs to class 2 type II and comprises the single-effector protein Cas9, which contains RuvC and HNH nuclease domains. The most commonly used Cas9 has been adapted from *Streptococcus pyogenes* (SpCas9). The SpCas9-mediated DNA recognition requires a 20-nt target complementary sequence in the crRNA and 5′-NGG-3′ PAM in the

target DNA. SpCas9 gene has a relatively large size (4.2 kb), which can hamper its efficient delivery into mammalian cells. Smaller Cas9 orthologs with different PAM requirements have been described and developed as genome editing endonucleases, for example, Cas9 from *Staphylococcus aureus* (SaCas9), which requires a 5′-NNGRRT-3′ PAM sequence [13].

Another CRISPR system – class 2 type V CRISPR/Cpf1 (also known as Cas12a) expands the options for mammalian genome editing [14]. Cpf1 appears to have increased specificity and facilitates targeting of AT-rich sequences [15, 16]. Containing only one RuvC endonuclease domain, Cpf1 creates a staggered DNA DSB and mediates DNA cleavage by recognition of a short 5′-TTTN-3′ PAM. In contrast to Cas9s, Cpf1 is guided by a single short crRNA and does not require a tracrRNA. As both Cpf1 and its gRNAs are smaller than SpCas9 counterparts, it overcomes some limitations of CRISPR delivery into mammalian cells. Moreover, Cpf1 has an ability to process its own crRNA, which can be used to simplify multiplexed genome editing [17].

In contrast to type II and type V, recently characterized class 2 type VI effector proteins Cas13 mediate single-stranded RNA (ssRNA) cleavage [18]. Cas13 binding is determined by a crRNA with a 28 nt sequence complementary to RNA protospacer, which should be flanked by A, U, or C. The discovery of RNA-targeted CRISPR systems opens the door for the development of new RNA editing tools for mammalian cells (described in Section 8.5.3).

8.3.2 Engineered Cas9 Variants

Since the first applications of CRISPR/Cas9 as a genome editing tool, engineered Cas9 variants have appeared. One of the first developments was mutating one of the two nuclease domains in Cas9, creating a "nickase" Cas9 (nCas9) [10, 19]. nCas9 cleaves only one strand of DNA leading to nicks, which are predominantly repaired with higher fidelity than DSBs. Inactivation of both nuclease domains of Cas9 results in "dead" Cas9 (dCas9), which lacks nuclease activity but retains RNA-guided DNA-binding ability [19]. Fusions of dCas9 with various effector domains mediate site-specific transcriptional or epigenetic regulation without cleaving target DNA (described in Section 8.5.2).

To improve Cas9 specificity and alter PAM recognition, different protein engineering efforts have been performed. Cas9s with enhanced specificity were constructed by mutating residues, forming unspecific bonds with DNA strand [20, 21]. Using structural information and molecular evolution, PAM specificities of SpCas9 and SaCas9 were expanded for broadening the Cas9 targeting range [22, 23]. Furthermore, to allow inducible control of genome modifications, Cas9 was split into two fragments, which can be brought back together upon chemical or light induction [24].

8.4 Experimental Design for CRISPR-mediated Genome Editing

Overall, the generation of a mammalian cell line with CRISPR/Cas9-mediated modification, for example, cell lines with gene knockouts, can be achieved within

four weeks [25]. The process includes the selection of target sites, construction of reagents, transfection of Cas9 and gRNA, single-cell isolation, expansion of clones, and analysis of modifications (Figure 8.2). The initial design of experiments for CRISPR-mediated genome editing requires several considerations at the following steps: target site selection, design of gRNA, and choice of delivery methods for CRISPR/Cas9 components. If the goal of experiment is targeted gene integration or gene correction, DNA repair templates in the form of either dsDNA or ssDNA should be designed. After the delivery of Cas9, gRNA, and, if necessary, donor template, cells can be subcloned by fluorescence-activated cell sorting (FACS) or limiting dilution. An advantage of FACS is the possibility of enriching transfected cells, when applying fluorescent markers in CRISPR/Cas9 components, leading to an increase of genome editing events. Modifications introduced upon CRISPR/Cas9-mediated genome editing can be monitored by different assays. In the case of gene knockouts, mismatch-recognizing nucleases (Surveyor nuclease or T7 endonuclease I) or indel detection by amplicon analysis (IDAA) can be applied for the detection of indels [26]. Genomic changes can also be analyzed by DNA sequencing (Sanger sequencing or next-generation sequencing). These methods will confirm the indel length and position and will help to select clones with desired genomic modifications. For verification of gene knockout, it is also recommended to conduct functional confirmation of gene disruption, for example, by qPCR, western blot, and functional assays, to verify the loss of gene function.

8.4.1 Target Site Selection and Design of gRNAs

There are two main considerations in the selection of target site for genome editing: the presence of the PAM sequence in the vicinity of the targeted genomic site and the minimization of potential off-target activity. In the context of large mammalian genomes, SpCas9 may bind and cleave off-target sites with relatively high rates [27]. Therefore, methods to measure and enhance Cas9 specificity and improve computational tools for gRNA design are of great interest. The methods to improve Cas9 specificity include modification of gRNA (shortening of spacers to 17–18 nt or chemical modification of gRNA), discovery of Cas9 orthologs with higher specificity, Cas9 protein engineering, limitation of the time Cas9 is present in the nucleus, and development of better predictive models for the design of gRNA [28]. Several deep sequencing-based protocols, such as GUIDE-seq, Digenome-seq, Circle-seq, and Site-seq, have been applied for genome-wide measuring Cas9 off-target activity in human cell lines [29–32]. They confirmed that sequence homology alone is not fully predictive for Cas9 off-target sites. Thus, more advanced rules for gRNA design need further investigation.

Present computational tools for gRNA design can calculate gRNA specificity based on *in silico* prediction of off-target sites on the basis of sequence homology and *in vivo* and *in vitro* assessment of Cas9 specificity [28]. There are many online tools for the design of gRNA, which can be easily used to select the target sequence with the highest predicted editing specificity. These tools differ in the number of genomes and Cas9 orthologs supported, type

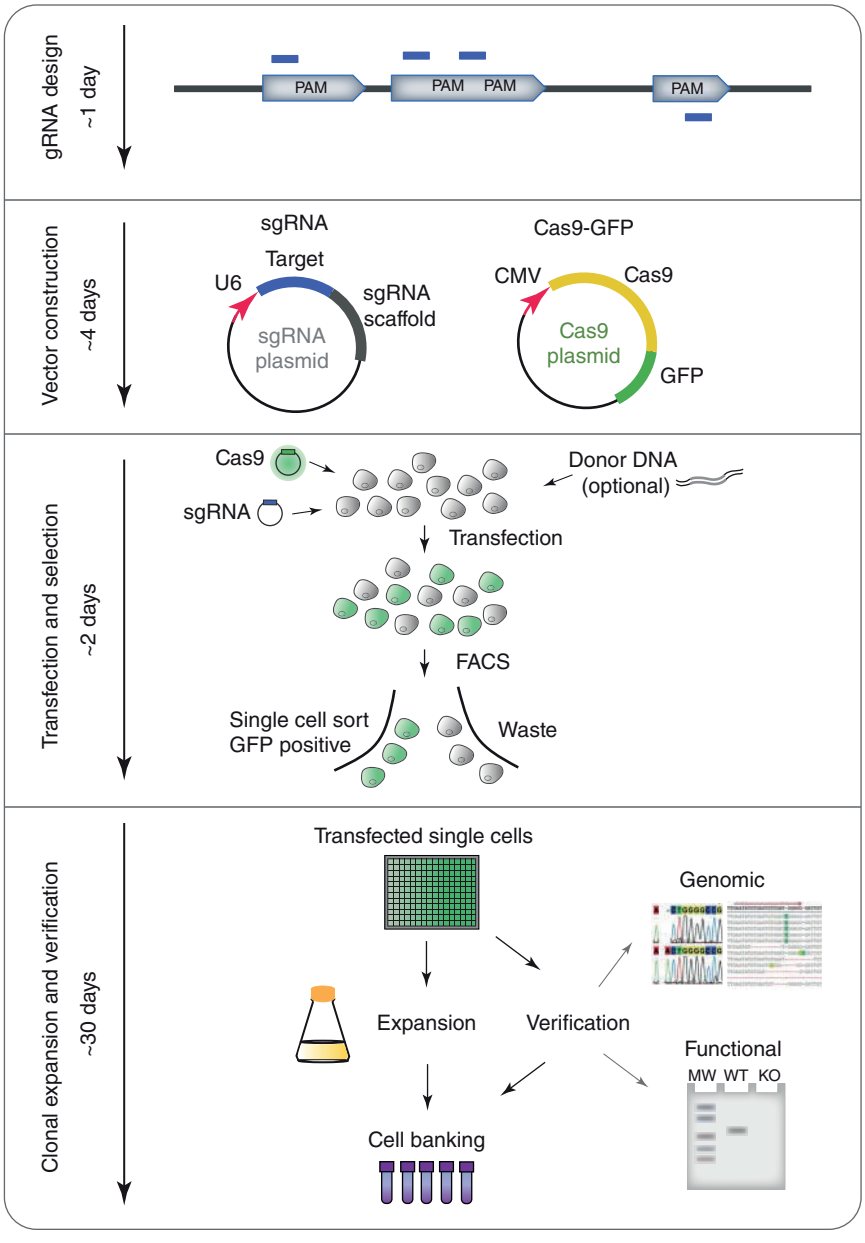


Figure 8.2 *Timeline of CRISPR/Cas9 genome editing experiment*. First, gRNA for a specific target is designed using *in silico* tools. Second, the target sequence is cloned into a sgRNA expression plasmid. Third, sgRNA and Cas9 plasmids and optional DNA repair template are transfected into cells. Then, transfected cells are enriched by FACS, clonally expanded, and verified to derive cell lines with desired modifications.

of gRNA design, and algorithms used to score gRNA specificity. Examples of online tools, which can be applied to the design of sgRNA for CRISPR editing in CHO cells, include CRISPy (http://staff.biosustain.dtu.dk/laeb/crispy/) [33] and Benchling (https://benchling.com/crispr/). As the efficiency of gRNA may vary for different genomic sites depending on many factors, including the chromatin state [34], it is recommended to design at least two gRNAs for each target locus and test their efficiency.

8.4.2 Delivery of CRISPR/Cas9 Components

Depending on the desired genome editing application, there are various ways in which CRISPR/Cas9 components can be introduced into the cell. gRNA and Cas9 can be delivered as DNA plasmids, RNA, or RNP complex. For short-term experiments in cell lines, the preferable method of knocking out or editing a gene by CRISPR/Cas9 is transient transfection, as a constitutive expression of CRISPR components is unwanted once the desired genome modification has occurred. For some applications such as genome-wide CRISPR screens, it is often desirable to make a stable cell line expressing Cas9 and gRNAs using viral transduction combined with selection to ensure genome editing in all cells and even dispersal of the many different gRNAs.

There are many mammalian expression vectors available that encode Cas9 variants and gRNA (see https://www.addgene.org). For gRNA expression, RNA polymerase III U6 promoter is typically used. The promoter driving the Cas9 expression in mammalian cells can be constitutive (e.g. cytomegalovirus [CMV] or elongation factor 1 alpha [EF1-α] promoters) or inducible. The plasmid may also contain a reporter gene (e.g. GFP) or selection marker (e.g. antibiotic resistance gene) to facilitate screening, enrichment, and selection of transfected cells as mentioned above.

The plasmid-based delivery of CRISPR/Cas9 has been widely used because of distinct benefits of this system, such as low-cost and ready-to-use transfection methods. However, there are numerous drawbacks to plasmid delivery, for example, potential risk of random integration of plasmid DNA into a host genome and high off-target cleavage. Although prolonged expression of CRISPR components increase on-target editing events over time, this also increases off-target editing. To improve the efficiency and specificity of CRISPR/Cas9, it was proposed to deliver preassembled gRNA:Cas9 RNPs directly into the cell [35]. To form RNP, purified Cas9 is assembled together with synthetic gRNA (sgRNA or crRNA:trRNA complex) and then RNP is delivered into the cell via electroporation or lipid-mediated transfection. RNP cleaves DNA within several hours after the delivery and is rapidly degraded in the cell. Such fast action and a short duration of the Cas9 presence in the nucleus increase the efficiency and reduce the off-target effects [36, 37].

Certain cell types, including primary cells and stem cells, can be difficult to transfect via the nonviral methods described above. Also, for CRISPR screening (described in Section 8.6), it is important to obtain an even dispersal of gRNA integration over a pool of cells. In these cases, it is advantageous to carry out viral delivery, in which CRISPR components are encapsulated by a viral vector

and enters the cells through transduction [38]. The most extensively studied viral vectors include retroviruses, adenoviruses, and adeno-associated viruses (AAVs). Some common drawbacks of using viral delivery include safety concerns, limitations of insert size, and less ease-of-use compared with plasmid-based transfection. Over the years, however, improved vector construction, viral particle production, transduction efficiency, and general safety have refined the system [39, 40].

8.5 Development of CRISPR/Cas9 Tools

Since the first publications of CRISPR/Cas9 application in mammalian cells, diverse CRISPR-based tools have been developed for gene editing, gene regulation, epigenetic modification, genome imaging, and CRISPR-based chromatin immunoprecipitation. Besides DNA targeting, CRISPR can be applied for RNA knockdown and modification (Figure 8.3). The following sections will cover recent advances in CRISPR/Cas tool development and their applications in mammalian cells.

8.5.1 CRISPR/Cas9-mediated Gene Editing

8.5.1.1 Gene Knockout

Gene disruption was the first application of CRISPR/Cas9 in mammalian cells and remains the most exploitable method for functional knockout of the target genes. The easiest way to create a knockout cell line is to introduce two components into the cell: gRNA specific to the target sequence and Cas9. DSBs induced by CRISPR/Cas9 are preferentially repaired by error-prone NHEJ, leading to indels. These mutations can cause a frame shift in the coding regions of genes that disrupts their proper translation and results in a functional knockout. When using Cas9 for gene knockout through the creation of indels, it is most common to target an early exon in the coding sequence or functional domain to disrupt as much of the protein as possible. Aside from NHEJ repair of CRISPR/Cas9-mediated DSBs, HDR can be applied to generate gene knockouts by introducing a premature stop codon. Codelivery of a DNA template bearing a stop codon together with CRISPR/Cas9 components can lead to increased knockout efficiency [43].

The introduction of multiple gRNAs along with a common Cas9 protein can lead to simultaneous modification of multiple target sequences located at the same or different genes, referred to as multiplexing. Multiplex genome engineering via CRISPR/Cas9 can be achieved by co-transfection of multiple gRNA plasmids, a single vector with gRNA arrays, or multiple RNP complexes [9, 10, 36]. Also, a multiplexing method using Cpf1 has been published, utilizing the special feature of Cpf1 to process its own crRNA array [17]. Simultaneous introduction of DSBs in different genomic loci can give rise to different types of genome modifications: multiallelic and multigene modifications, large deletions, or chromosomal rearrangements. Targeting multiple genes by a multiplexing strategy

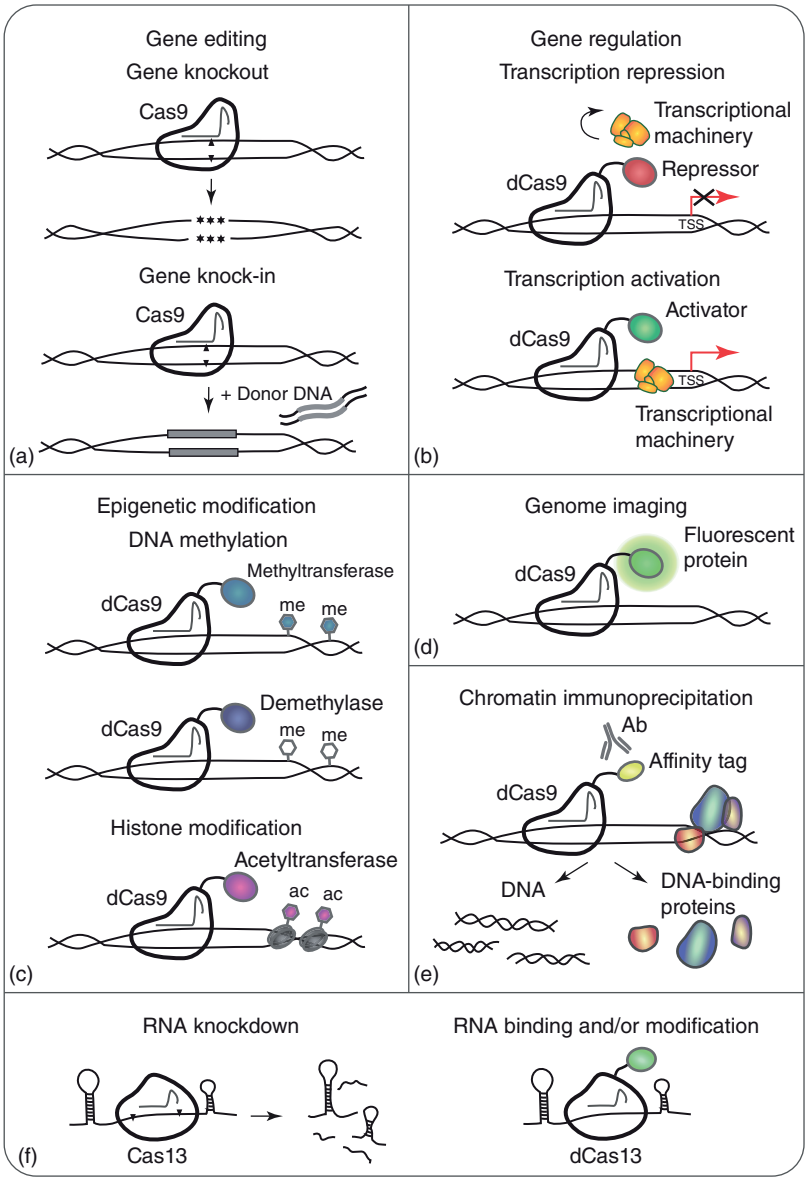


Figure 8.3 *CRISPR toolbox for mammalian cell engineering*. (a) CRISPR-mediated gene editing is based on Cas9 nuclease activity, where repair of Cas9-introduced DSBs can lead to gene knockout or knock-in. (b) Fusion of catalytically inactive dCas9 to transcriptional activation or repression domains can mediate transcriptional regulation by alteration of transcriptional machinery binding to transcription start site (TSS). (c) Fusion of dCas9 to DNA or chromatin modification domains enable epigenetic changes of target sites. (d) dCas9 can be fused to fluorescent domains for imaging of genomic loci [41]. (e) An affinity-tagged dCas9 can be used in chromatin immunoprecipitation assays to study protein interactions at specific genomic site [42]. (f) CRISPR/Cas13 can be used for RNA knockdown, binding, and modification in mammalian cells.

accelerates development of cell lines with complex genome edits, for example, multiple knockouts. Up to seven genomic loci were simultaneously modified in human cell line by using the all-in-one plasmid with gRNA expression array and Cas9 [44]. Besides multigene knockout, concurrent delivery of gRNAs designed to target two different genomic sites can result in targeted deletion of genomic segments ranging from several hundred base pairs to 1 Mbp [45]. This paired gRNAs strategy can be used for the purpose of gene knockout, when several exons in the gene are removed [46]. It was also shown that DNA cleavage by Cas9 at two genomic loci can result in chromosomal rearrangement and expression of fusion transcripts [47].

As was mentioned earlier, wild-type Cas9 is known to recognize and cleave many off-target sites. To minimize off-target mutagenesis, a double-nicking strategy can be used to introduce DSBs and disrupt the gene by nCas9 and a pair of offset gRNAs [48, 49]. Single-stranded nicks on the opposite strands of the target DNA lead to a production of composite DSBs, which are then repaired by NHEJ and result in indel formation or large genomic deletions. Because of the doubled number of bases that need to be specifically recognized by two nCas9s, this strategy increases the specificity of genome editing.

8.5.1.2 Site-Specific Gene Integration

CRISPR/Cas9 has facilitated site-specific integration of DNA by taking advantage of native DNA repair mechanisms [50]. HDR of CRISPR/Cas9-introduced DSBs near the targeted genomic loci can lead to the introduction of desired mutations or insertion of genes when the proper DNA template is presented. Single-nucleotide substitutions or modification of short sequences (e.g. insertion of tags or codon mutations) can be introduced using short oligonucleotide templates, called single-stranded oligodeoxynucleotides (ssODNs) [25]. The design of ssODNs requires addition of homology arms, whose length can vary from short (30–40 nt) to long (70–90 nt) [51]. Although ssODN-mediated integration was shown to be more efficient than double-stranded donor integration, ssODNs have a narrow insert length because of the limitation of single-stranded oligo synthesis. Alternatively, single-stranded DNA template can be provided in the form of recombinant AAV vectors, but their ssDNA length is also limited to 4.5 kb [52]. Therefore, for the integration of large DNA cassettes, dsDNA vectors with long homology arms (around 1 kb each) are used as donor template. Co-delivery of donor plasmids with CRISPR/Cas9 components was applied for the site-specific integration of transgenes encoding fluorescent markers, antibiotic resistance genes, recombinant proteins, and landing pads for recombinase-mediated cassette exchange [9, 53, 54]. The drawback of HDR-mediated strategy is that HDR appears at low frequency in many mammalian cell lines, leading to a low rate of integration events. The efficiency of HDR-mediated genome editing can be increased by chemical treatment, transient modulation of DNA repair proteins, or cell cycle synchronization [55, 56]. These approaches resulted in significant improvements in total integration efficiency but were shown to be cell-type specific (and context dependent) [57].

Other targeted knock-in strategies rely on homology-independent integration. An advantage of this strategy is the utilization of more frequent DNA repair

mechanisms such as NHEJ and simple construction of donor vectors because there is no requirement for long homology arms. Although NHEJ repair is error prone, it appears at a higher rate than HDR, which can be beneficial when increased targeted integration efficiency is needed. It was shown that CRISPR/Cas9-induced NHEJ can mediate site-specific knock-in more efficiently than HDR-based strategy [58]. The same trend was demonstrated by homology-independent targeted integration (HITI) method that relies on the simultaneous introduction of DSBs at genomic loci of interest and a donor vector by the gRNA [59]. MMEJ-mediated repair can also facilitate knock-in through an introduction of a donor with microhomology sequences, which was proved by PITCh method [60].

More sophisticated genome engineering can be done by multiplexed targeted integration of genes at multiple alleles and/or loci. As a proof of concept, multigenic homology-directed targeted integration of transgenes at different loci was performed in human hematopoietic stem and progenitor cells by using Cas9 RNP and AAVs [61]. In the future, this approach can be applied for the introduction of genes of multi-subunit protein complexes, building multigene pathways, and revealing functional gene networks in mammalian cells.

8.5.2 CRISPR/Cas9-mediated Genome Modification

With CRISPR gaining foothold as a powerful genome editing method, many have started repurposing the system for RNA-guided site-specific DNA modifications rather than cleavage. The catalytically inactive version of Cas9 (dCas9) combined with different effector molecules has been used for various applications. The following section will focus on the development of CRISPR/dCas9 technology for downregulation and upregulation of genes in mammalian cells via transcriptional and epigenetic modifications.

8.5.2.1 Transcriptional Regulation

The first mention of CRISPR being used as a transcriptional regulator was in a study by Qi, L. et al. in 2013. They engineered a catalytically inactive Cas9 and observed that, with coexpression of gRNA, it could sterically hinder the RNA polymerase from binding and elongating, leading to considerable repression of specific transcription in bacteria, a tool they coined CRISPR interference (CRISPRi) [62]. The effect in eukaryotic cells, however, was moderate, likely because of the complexity of transcriptional regulation. Follow-up studies have shown that fusing dCas9 with repressing regulatory domains, for example, Krüppel-associated box (KRAB), could yield increased transcriptional repression. RNA-sequencing analysis confirmed that CRISPRi knockdown is highly specific. Thus, dCas9-KRAB can be applied for efficient and targeted repression of multiple endogenous genes in mammalian cells [63].

In line with CRISPRi, it has been explored whether fusing dCas9 with activating regulatory elements could increase the expression of targeted genes, referred to as CRISPR activation (CRISPRa). Initial experiments have shown that fusing well-known transcription activators, such as VP16 and the p65 activation domain to dCas9 and coexpressing with gRNA, increases the expression of targeted genes

[63]. Several different combinations of gRNAs and activation effectors have since been tested across various cell types in various settings. In 2016 Chavez, A. et al. compared the effectiveness of the so-called second-generation activators including VP64, VPR (VP64-p65-Rta), SAM (synergetic activation mediator), and SunTag in human, mouse, and fly cell lines. The study was the first comprehensive side-by-side evaluation of their relative potential and revealed that SAM, SunTag, and VPR were consistently superior at increasing expression across cell lines [64]. Recently, dCpf1-based transcriptional activators were developed, which allows synergistic tuning of expression of endogenous genes, leveraging the multiplex capability of Cpf1 [65].

CRISPRi and CRISPRa have since been applied to and exerted regulatory effects in cells from a variety of organisms with varied success rates. For better reproducibility of CRISPR-based regulation of different genes, evidence suggests that careful design of the gRNA in relation to the transcription start site (TSS) as well as nucleosome occupancy of the target site is of great importance [34, 64, 66].

8.5.2.2 Epigenetic Modification

The role of epigenetic modifications in transcription, genome stability, and nuclear organization has solidified over the last decades. Even though numerous discoveries, such as involvement of histone modification, chromatin remodeling, and DNA methylation, have surfaced, technological limitations have obstructed insights into the precise mechanistic level [67–69]. Fusing dCas9 with epigenetic modifiers presents a promising tool for future epigenetic research. Addition of methylation modulators to dCas9 and targeting it to an unmethylated or methylated promoter region can result in repression or activation of the targeted gene, allowing for functional studies of epigenetic regulation [70]. In the same manner, fusing dCas9 with an acetyltransferase can catalyze acetylation of histones transforming chromatin to a more relaxed state and in response initiate transcriptional activation [71].

8.5.3 RNA Targeting

Although current CRISPR/Cas systems have been applied mostly for DNA editing, the recent discovery of RNA-targeting CRISPR/Cas13 systems provides a starting point for expanding the CRISPR toolbox for RNA manipulation. Upon binding to a complementary RNA target, Cas13 engages RNase activity both *in vitro* and *in vivo* [18]. This RNA cleavage activity of Cas13 was used for the development of *in vitro* methods for nucleic acid detection and specific RNA knockdown in bacterial and mammalian cells [72–74]. Thus, Cas13 provides the basis for improved tools for controlling cellular processes at the transcript level in mammalian cells.

CRISPR/Cas13 system opens new possibilities in the study of RNA function by targeted RNA binding and modification. Substitutions in the catalytic domain of Cas13 converted it into inactive programmable RNA-binding protein dCas13 [18]. dCas13 could be fused to effector domains with different functions to edit RNA sequence, enhance or inhibit translation, alter the splicing or visualize RNA trafficking, and localization [73–75].

8.6 Genome-Scale CRISPR Screening

In addition to the aforementioned applications of CRISPR/Cas9 for gene editing and regulation, the technique and said tools can be extended to large forward genetic screens that can be used not only to interrogate gene function but also to identify novel targets for treatment of disease or engineering industrial mammalian cells. In contrast to reverse genetic engineering, where known genes are modulated and the resulting phenotype is studied, the approach taken in forward genetic screening includes changing numerous genes at a pool level, applying a phenotypic selection and subsequently identifying the responsible genes. Traditionally, these screens have relied on random mutagenesis and isolating individuals with an interesting phenotype [76–78]. As the mutation is initially unknown, identification of causality is a very lengthy and difficult process and represents one of the main weaknesses of this system. With the advancement of RNA interference (RNAi) came a welcomed alternative approach to forward genetic screening. RNAi reagents degrading known sequences of mRNA replaced random mutagenesis, and the identification of causal mutations was hugely simplified [79]. However, the incomplete knockdown of targeted genes and substantial off-target effects have restrained the extent to which RNAi screens can be used. The advancement of CRISPR technologies presents a novel approach to attempt to solve these issues and increase the versatility for the next generation of forward screening methods. Consequently, the first records of CRISPR used as a screening tool in mammalian cells were published in 2014 by Shalem et al. [80] and Wang et al. [81]. CRISPR tools, such as CRISPRi and CRISPRa, have later been included in the screening setup, with activation/gain-of-function screens showing highest potential [5, 82, 83].

8.7 Applications of CRISPR/Cas9 for CHO Cell Engineering

The major objectives of mammalian cell engineering toward industrial production of therapeutic proteins are high productivity and product quality. To overcome cellular limitations in growth, productivity, and post-translational modifications (PTMs), various strategies have been applied in CHO cell engineering, such as overexpression, knock-in, knockout, or post-transcriptional silencing of genes [84]. CRISPR/Cas9 offers a new engineering tool for a broad field of facile-targeted genome engineering of CHO cells, described in detail in the following section.

In terms of programmable genome engineering, CRISPR/Cas9 was not the first tool to be applied. Before the era of CRISPR/Cas9, zinc finger nucleases (ZFNs) and transcription activator-like effector nucleases (TALENs) have been developed and used to generate DSBs and engineer mammalian cells. There were successful examples of CHO genome engineering by ZFN and TALENs for knockout of genes involved in glycosylation and metabolism as well as targeted gene integration [85–91]. In a short time, more simple and efficient CRISPR/Cas9

system has emerged as a fascinating alternative, which paved the way for accelerated genome editing for improved recombinant protein production together with advances in CHO genome sequencing and annotation [92].

The first demonstration of CRISPR/Cas9 activity in CHO cells was published in 2014 when two genes involved in glycosylation (C1GALT1-specific chaperone 1 [*COSMC*] and α-1,6-fucosyltransferase [*FUT8*]) were disrupted with the indels frequency up to 47% in a pool of cells [33]. Later, the enrichment of cells carrying GFP-Cas9 plasmid and multiplexing knockout strategies were developed [93]. Simultaneous introduction of three gRNA plasmids and fluorescent enrichment yielded 59% of clones harboring indels in *FUT8*, *BAX*, and *BAK* genes. Generated knockout cell lines showed improved resistance to apoptosis and disrupted fucosylation activity, relevant for prolonged cell cultivation and production of nonfucosylated therapeutic proteins.

More recently, a novel CRISPR/Cpf1 system was used for multiplexed gene knockout in CHO cells, where Cpf1 showed parallel and comparable efficiency to Cas9 efficiency of genome cleavage [94]. In this study, the multiplexed knockout strategy was based on a pair of gRNAs, expressed as a crRNA array. Introduction of two DSBs by paired gRNAs enables full deletion of genes and regulatory regions and has an advantage in predictable loss of function in contrast to indel formation strategy.

As the successful application of CRISPR-mediated gene editing, various CHO engineering strategies have been developed in an attempt to improve product quality and to expedite cell line development for high productivity. A characteristic example of such genome engineering efforts is glycoengineering. Desired glycosylation profiles of therapeutic proteins are the crucial property of product quality attributes for increased stability and efficacy of proteins. CRISPR-mediated knockout of *FUT8* allows production of monoclonal antibodies lacking the core fucose on *N*-glycans, which induce stronger antibody-dependent cell-mediated cytotoxicity (ADCC) [93, 95, 96]. To produce α-2,6-sialylated antibodies with superior effector functions, two α-2,3-sialyltransferases *ST3GAL4* and *ST3GAL6* were consecutively disrupted by CRISPR/Cas9 and α-2,6-sialyltransferase *ST6GAL1* was overexpressed in CHO cells that lack the expression of α-2,6-sialyltransferase but only express α-2,3-sialyltransferases [97]. As the glycosylation pathway and genes involved in this PTM are highly characterized, it makes it possible to predict optimal CRISPR/Cas9 engineering strategies to obtain a specific glycosylation profile with desired properties [98]. Further advanced CRISPR/Cas9 glycoengineering can provide cell lines with homogeneous glycosylation of therapeutic proteins, resolving the challenge of product heterogeneity during glycoprotein production.

Besides different glycan structures, heterogeneity of monoclonal antibodies is also caused by different C-terminal lysine levels. C-terminal lysine on antibody heavy chains is cleaved by carboxypeptidase D (CpD). To maintain consistent C-terminal lysine levels, CpD was knocked out by paired-gRNA-mediated deletion of CpD exons, leading to the production of homogeneous proteins [46].

Impurities, such as host cell proteins (HCPs), present another challenge in the production of biopharmaceuticals as they can cause an unpredictable immunogenic response and impair product quality and stability. Variable HCPs

secreted from viable cells and released from dead cells were identified, which represent appropriate targets to be removed from cell culture and purification process by CRISPR/Cas9 technology [99]. As an example, lipoprotein lipase (LPL) that has the ability to degrade antibody formulations was disrupted by CRISPR/Cas9 [100].

CRISPR/Cas9 engineering has expanded its potential for efficient cell line development. To engineer CHO cells for the rapid adaptation to a suspension culture, RNA sequencing was exploited to identify genes differentially expressed during the adaptation process. Then, two of identified downregulated genes, *Igfbp4* and *AqpI*, were disrupted by CRISPR/Cas9 RNPs, leading to reduced adaptation time [101]. CRISPR/Cas9 knockout of frequently used metabolic selection marker, glutamine synthetase (GS), was reported in CHO cells to facilitate CHO cell line development [102]. Moreover, CRISPRi was recently applied in CHO cells to enhance coamplification of another essential selection marker, dihydrofolate reductase (DHFR), and a gene of interest (GOI). Transcriptional repression of *dhfr* by dCas9-KRAB resulted in selecting clones with increased productivity [103].

In addition to such cell engineering efforts, targeted gene integration provides a new opportunity for cell line development. Conventional methods of CHO cell line development for production of biopharmaceuticals are based on a random integration of transgenes. These methods are time-consuming and labor-intensive, which yields cell lines exhibiting a wide range of expression, growth, and stability characteristics. Targeted integration of the transgene into high transcriptional active sites in the genome ("hot spots") would be an ideal solution for acceleration of cell line development for the production of recombinant proteins. The first application of CRISPR-mediated targeted integration of transgenes was demonstrated in CHO cells that employed HDR for insertion of large gene cassettes encoding recombinant proteins [104]. The efficiency of targeted integration after the drug or lectin enrichment varied between 7% and 28% depending on the target locus. Fluorescent enrichment of genome-edited cells was further applied to avoid usage of lectin and antibiotic selection [53]. It resulted in a threefold increase of cells with HDR-mediated integration relative to nonenriched samples, with ~7% frequency of successful integration. The same HDR-mediated approach was also used to integrate the Bxb1 and Flp/FRT recombinase target sites flanking fluorescent marker and thymidine kinase ("landing pads") into a defined locus in CHO cells with 27% efficiency [54]. These cell lines were subsequently used for recombinase-mediated cassette exchange with antibody-encoding donor plasmids for streamlining the antibody development process. The NHEJ-mediated targeted integration in CHO was also reported, although its efficiency was considerably low (0.45%) [105].

8.8 Conclusion

CRISPR has proven itself a highly valuable RNA-guided genome engineering technique. Despite the great potential already demonstrated, there is still room

for improvements in the current CRISPR system. The issue of off-target effects is one of the major concerns, especially in fields such as gene therapy, where random disruptions are non-negotiable. In line with this, the effect of CRISPR on chromosomal rearrangements and genome stability has yet to be uncovered. Furthermore, some aspects of the technique still have to be improved; for example, the knock-in efficiency as well as insert size in targeted integration. As discussed in this chapter, some CRISPR issues regarding off-target effects, endonuclease size, and type of target nucleic acid seems to have been alleviated with the development of next-generation RNA-guided endonucleases such as Cpf1 and Cas13. Further mining for novel CRISPR systems in bacterial and archaeal genomes will aid in establishing a catalog of nucleases and CRISPR systems that can potentially accommodate all applications.

Beyond technical issues of the CRISPR technology, the identification of target sites for genome engineering is another challenge, particularly in CHO cells. This issue can be resolved by improved annotation of CHO genome and revealing of epigenetic landscape, which will help to refine the prediction of gRNA target sites. Understanding of complex cellular interactions by analysis of omics data sets and their integration with the genome-scale model of CHO cell metabolism and the secretory network can provide novel targets for the rational engineering of CHO cells. Together with a better understanding of CHO biology, CRISPR/Cas9 technology has the promise to resolve current bottlenecks in biopharmaceutical production, such as protein folding, secretion, and PTMs, and create cell lines with superior capacity for therapeutic protein production. Overall, advanced cell line engineering using genome engineering tools will help to generate efficient mammalian cell factories, providing patients with new therapeutics of high quality.

Acknowledgment

The Novo Nordisk Foundation (NNF10CC1016517 and NNF16CC0020908) kindly supported this work. The authors declare no conflict of interest.

References

1 Wang, F. and Qi, L.S. (2016). Applications of CRISPR genome engineering in cell biology. *Trends Cell Biol.* 26: 875–888.
2 Xiong, X., Chen, M., Lim, W.A. et al. (2016). CRISPR/Cas9 for human genome engineering and disease research. *Annu. Rev. Genomics Hum. Genet.* 17: 131–154.
3 Jusiak, B., Cleto, S., Perez-Piñera, P., and Lu, T.K. (2016). Engineering synthetic gene circuits in living cells with CRISPR technology. *Trends Biotechnol.* 34: 535–547.
4 Porteus, M. (2016). Genome editing: a new approach to human therapeutics. *Annu. Rev. Pharmacol. Toxicol.* 56: 163–190.
5 Shalem, O., Sanjana, N.E., and Zhang, F. (2015). High-throughput functional genomics using CRISPR-Cas9. *Nat. Rev. Genet.* 16: 299–311.

6 Jakočiūnas, T., Jensen, M.K., and Keasling, J.D. (2016). CRISPR/Cas9 advances engineering of microbial cell factories. *Metab. Eng.* 34: 44–59.
7 Wright, A.V., Nuñez, J.K., and Doudna, J.A. (2016). Biology and applications of CRISPR systems: harnessing nature's toolbox for genome engineering. *Cell* 164: 29–44.
8 Wang, H., La Russa, M., and Qi, L.S. (2016). CRISPR/Cas9 in genome editing and beyond. *Annu. Rev. Biochem.* 85: 227–264.
9 Mali, P., Yang, L., Esvelt, K.M. et al. (2013). RNA-guided human genome engineering via Cas9. *Science* 339: 823–826.
10 Cong, L., Ran, F.A., Cox, D. et al. (2013). Multiplex genome engineering using CRISPR/Cas systems. *Science* 339: 819–823.
11 Ceccaldi, R., Rondinelli, B., and D'Andrea, A.D. (2016). Repair pathway choices and consequences at the double-strand break. *Trends Cell Biol.* 26: 52–64.
12 Bhargava, R., Onyango, D.O., and Stark, J.M. (2016). Regulation of single-strand annealing and its role in genome maintenance. *Trends Genet.* 32: 566–575.
13 Ran, F.A., Cong, L., Yan, W.X. et al. (2015). In vivo genome editing using *Staphylococcus aureus* Cas9. *Nature* 520: 186–191.
14 Zetsche, B., Gootenberg, J.S., Abudayyeh, O.O. et al. (2015). Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell* 163: 759–771.
15 Kim, D., Kim, J., Hur, J.K. et al. (2016). Genome-wide analysis reveals specificities of Cpf1 endonucleases in human cells. *Nat. Biotechnol.* 34: 863–868.
16 Kleinstiver, B.P., Tsai, S.Q., Prew, M.S. et al. (2016). Genome-wide specificities of CRISPR-Cas Cpf1 nucleases in human cells. *Nat. Biotechnol.* 34: 869–874.
17 Zetsche, B., Heidenreich, M., Mohanraju, P. et al. (2017). Multiplex gene editing by CRISPR-Cpf1 using a single crRNA array. *Nat. Biotechnol.* 35: 31–34.
18 Abudayyeh, O.O., Gootenberg, J.S., Konermann, S. et al. (2016). C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector. *Science* 353: aaf5573.
19 Jinek, M., Chylinski, K., Fonfara, I. et al. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 337: 816–821.
20 Kleinstiver, B.P., Pattanayak, V., Prew, M.S. et al. (2016). High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. *Nature* 529: 490–495.
21 Slaymaker, I.M., Gao, L., Zetsche, B. et al. (2015). Rationally engineered Cas9 nucleases with improved specificity. *Science* https://doi.org/10.1126/science.aad5227.
22 Kleinstiver, B.P., Prew, M.S., Tsai, S.Q. et al. (2015). Engineered CRISPR-Cas9 nucleases with altered PAM specificities. *Nature* 523: 481–485.
23 Kleinstiver, B.P., Prew, M.S., Tsai, S.Q. et al. (2015). Broadening the targeting range of *Staphylococcus aureus* CRISPR-Cas9 by modifying PAM recognition. *Nat. Biotechnol.* 33: 1293–1298.

24 Nuñez, J.K., Harrington, L.B., and Doudna, J.A. (2016). Chemical and biophysical modulation of Cas9 for tunable genome engineering. *ACS Chem. Biol.* 11: 681–688.

25 Ran, F.A., Hsu, P.D., Wright, J. et al. (2013). Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.* 8: 2281–2308.

26 Lonowski, L.A., Narimatsu, Y., Riaz, A. et al. (2017). Genome editing using FACS enrichment of nuclease-expressing cells and indel detection by amplicon analysis. *Nat. Protoc.* 12: 581–603.

27 Fu, Y., Foden, J.A., Khayter, C. et al. (2013). High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat. Biotechnol.* 31: 822–826.

28 Tycko, J., Myer, V.E., and Hsu, P.D. (2016). Methods for optimizing CRISPR-Cas9 genome editing specificity. *Mol. Cell* 63: 355–370.

29 Tsai, S.Q., Zheng, Z., Nguyen, N.T. et al. (2015). GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nat. Biotechnol.* 33: 187–197.

30 Kim, D., Bae, S., Park, J. et al. (2015). Digenome-seq: genome-wide profiling of CRISPR-Cas9 off-target effects in human cells. *Nat. Methods* 12: 237–243.

31 Tsai, S.Q., Nguyen, N.T., Malagon-Lopez, J. et al. (2017). CIRCLE-seq: a highly sensitive in vitro screen for genome-wide CRISPR-Cas9 nuclease off-targets. *Nat. Methods* 14: 607–614.

32 Cameron, P., Fuller, C.K., Donohoue, P.D. et al. (2017). Mapping the genomic landscape of CRISPR-Cas9 cleavage. *Nat. Methods* 14: 600–606.

33 Ronda, C., Pedersen, L.E., Hansen, H.G. et al. (2014). Accelerating genome editing in CHO cells using CRISPR Cas9 and CRISPy, a web-based target finding tool. *Biotechnol. Bioeng.* 111: 1604–1616.

34 Horlbeck, M.A., Witkowsky, L.B., Guglielmi, B. et al. (2016). Nucleosomes impede Cas9 access to DNA in vivo and in vitro. *eLife* 5: e12677.

35 Kim, S., Kim, D., Cho, S.W. et al. (2014). Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins. *Genome Res.* 24: 1012–1019.

36 Liang, X., Potter, J., Kumar, S. et al. (2015). Rapid and highly efficient mammalian cell engineering via Cas9 protein transfection. *J. Biotechnol.* 208: 44–53.

37 Liang, X., Potter, J., Kumar, S. et al. (2017). Enhanced CRISPR/Cas9-mediated precise genome editing by improved design and delivery of gRNA, Cas9 nuclease, and donor DNA. *J. Biotechnol.* 241: 136–146.

38 Cai, Y. and Mikkelsen, J.G. (2016). Lentiviral delivery of proteins for genome engineering. *Curr. Gene Ther.* 16: 194–206.

39 Nayerossadat, N., Maedeh, T., and Ali, P.A. (2012). Viral and nonviral delivery systems for gene delivery. *Adv. Biomed. Res.* 1: 27.

40 Yin, H., Kauffman, K.J., and Anderson, D.G. (2017). Delivery technologies for genome editing. *Nat. Rev. Drug Discovery* 16: 387–399.

41 Chen, B., Gilbert, L.A., Cimini, B.A. et al. (2013). Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. *Cell* 155: 1479–1491.

42 Fujita, T. and Fujii, H. (2013). Efficient isolation of specific genomic regions and identification of associated proteins by engineered DNA-binding molecule-mediated chromatin immunoprecipitation (enChIP) using CRISPR. *Biochem. Biophys. Res. Commun.* 439: 132–136.

43 Li, K., Wang, G., Andersen, T. et al. (2014). Optimization of genome engineering approaches with the CRISPR/Cas9 system. *PLoS One* 9: e105779.

44 Sakuma, T., Nishikawa, A., Kume, S. et al. (2014). Multiplex genome engineering in human cells using all-in-one CRISPR/Cas9 vector system. *Sci. Rep.* 4: https://doi.org/10.1038/srep05400.

45 He, Z., Proudfoot, C., Mileham, A.J. et al. (2015). Highly efficient targeted chromosome deletions using CRISPR/Cas9. *Biotechnol. Bioeng.* 112: 1060–1064.

46 Hu, Z., Zhang, H., Haley, B. et al. (2016). Carboxypeptidase D is the only enzyme responsible for antibody C-terminal lysine cleavage in Chinese hamster ovary (CHO) cells. *Biotechnol. Bioeng.* 113: 2100–2106.

47 Choi, P.S. and Meyerson, M. (2014). Targeted genomic rearrangements using CRISPR/Cas technology. *Nat. Commun.* 5: 3728.

48 Mali, P., Aach, J., Stranges, P.B. et al. (2013). CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nat. Biotechnol.* 31: 833–838.

49 Ran, F.A., Hsu, P.D., Lin, C.-Y. et al. (2013). Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell* 154: 1380–1389.

50 Liu, Z., Liang, Y., Ang, E.L., and Zhao, H. (2017). A new era of genome integration-simply cut and paste! *ACS Synth. Biol.* 6: 601–609.

51 Richardson, C.D., Ray, G.J., DeWitt, M.A. et al. (2016). Enhancing homology-directed genome editing by catalytically active and inactive CRISPR-Cas9 using asymmetric donor DNA. *Nat. Biotechnol.* 34: 339–344.

52 Kaulich, M. and Dowdy, S.F. (2015). Combining CRISPR/Cas9 and rAAV templates for efficient gene editing. *Nucleic Acid Ther.* 25: 287–296.

53 Lee, J.S., Grav, L.M., Pedersen, L.E. et al. (2016). Accelerated homology-directed targeted integration of transgenes in Chinese hamster ovary cells via CRISPR/Cas9 and fluorescent enrichment. *Biotechnol. Bioeng.* 113: 2518–2523.

54 Inniss, M.C., Bandara, K., Jusiak, B. et al. (2017). A novel Bxb1 integrase RMCE system for high fidelity site-specific integration of mAb expression cassette in CHO cells. *Biotechnol. Bioeng.* 114: 1837–1846.

55 Chu, V.T., Weber, T., Wefers, B. et al. (2015). Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells. *Nat. Biotechnol.* 33: 543–548.

56 Lin, S., Staahl, B.T., Alla, R.K., and Doudna, J.A. Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery. *eLife* 3: e04766.

57 Song, J., Yang, D., Xu, J. et al. (2016). RS-1 enhances CRISPR/Cas9- and TALEN-mediated knock-in efficiency. *Nat. Commun.* 7: https://doi.org/10.1038/ncomms10548.

58 He, X., Tan, C., Wang, F. et al. (2016). Knock-in of large reporter genes in human cells via CRISPR/Cas9-induced homology-dependent and independent DNA repair. *Nucleic Acids Res.* 44: e85.

59 Suzuki, K., Tsunekawa, Y., Hernandez-Benitez, R. et al. (2016). In vivo genome editing via CRISPR/Cas9 mediated homology-independent targeted integration. *Nature* 540: 144–149.

60 Sakuma, T., Nakade, S., Sakane, Y. et al. (2016). MMEJ-assisted gene knock-in using TALENs and CRISPR-Cas9 with the PITCh systems. *Nat. Protoc.* 11: 118–133.

61 Bak, R.O., Dever, D.P., Reinisch, A. et al. (2017). Multiplexed genetic engineering of human hematopoietic stem and progenitor cells using CRISPR/Cas9 and AAV6. *eLife* 6: e27873.

62 Qi, L.S., Larson, M.H., Gilbert, L.A. et al. (2013). Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* 152: 1173–1183.

63 Gilbert, L.A., Larson, M.H., Morsut, L. et al. (2013). CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell* 154: 442–451.

64 Chavez, A., Tuttle, M., Pruitt, B.W. et al. (2016). Comparison of Cas9 activators in multiple species. *Nat. Methods* 13: 563–567.

65 Tak, Y.E., Kleinstiver, B.P., Nuñez, J.K. et al. (2017). Inducible and multiplex gene regulation using CRISPR–Cpf1-based transcription factors. *Nat. Methods* 14: 1163–1166.

66 Horlbeck, M.A., Gilbert, L.A., Villalta, J.E. et al. (2016). Compact and highly active next-generation libraries for CRISPR-mediated gene repression and activation. *eLife* 5: e19760.

67 Weinhold, B. (2006). Epigenetics: the science of change. *Environ. Health Perspect.* 114: A160–A167.

68 Pfister, S.X. and Ashworth, A. (2017). Marked for death: targeting epigenetic changes in cancer. *Nat. Rev. Drug Discovery* 16: 241–263.

69 Enríquez, P. (2016). CRISPR-mediated epigenome editing. *Yale J. Biol. Med.* 89: 471–486.

70 Liu, X.S., Wu, H., Ji, X. et al. (2016). Editing DNA methylation in the mammalian genome. *Cell* 167: 233–47.e17.

71 Hilton, I.B., D'Ippolito, A.M., Vockley, C.M. et al. (2015). Epigenome editing by a CRISPR/Cas9-based acetyltransferase activates genes from promoters and enhancers. *Nat. Biotechnol.* 33: 510–517.

72 Gootenberg, J.S., Abudayyeh, O.O., Lee, J.W. et al. (2017). Nucleic acid detection with CRISPR-Cas13a/C2c2. *Science* 356: 438–442.

73 Abudayyeh, O.O., Gootenberg, J.S., Essletzbichler, P. et al. (2017). RNA targeting with CRISPR-Cas13. *Nature* 550: 280–284.

74 Cox, D.B.T., Gootenberg, J.S., Abudayyeh, O.O. et al. (2017). RNA editing with CRISPR-Cas13. *Science* 358: 1019–1027.

75 Mackay, J.P., Font, J., and Segal, D.J. (2011). The prospects for designer single-stranded RNA-binding proteins. *Nat. Struct. Mol. Biol.* 18: 256–261.

76 Patton, E.E. and Zon, L.I. (2001). The art and design of genetic screens: zebrafish. *Nat. Rev. Genet.* 2: 956–966.

77 St Johnston, D. (2002). The art and design of genetic screens: *Drosophila melanogaster*. *Nat. Rev. Genet.* 3: 176–188.

78 Jorgensen, E.M. and Mango, S.E. (2002). The art and design of genetic screens: *Caenorhabditis elegans*. *Nat. Rev. Genet.* 3: 356–369.

79 Boutros, M. and Ahringer, J. (2008). The art and design of genetic screens: RNA interference. *Nat. Rev. Genet.* 9: 554–566.

80 Shalem, O., Sanjana, N.E., Hartenian, E. et al. (2014). Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science* 343: 84–87.

81 Wang, T., Wei, J.J., Sabatini, D.M., and Lander, E.S. (2014). Genetic screens in human cells using the CRISPR-Cas9 system. *Science* 343: 80–84.

82 Gilbert, L.A., Horlbeck, M.A., Adamson, B. et al. (2014). Genome-scale CRISPR-mediated control of gene repression and activation. *Cell* 159: 647–661.

83 Joung, J., Konermann, S., Gootenberg, J.S. et al. (2017). Genome-scale CRISPR-Cas9 knockout and transcriptional activation screening. *Nat. Protoc.* 12: 828–863.

84 Fischer, S., Handrick, R., and Otte, K. (2015). The art of CHO cell engineering: a comprehensive retrospect and future perspectives. *Biotechnol. Adv.* 33: 1878–1896.

85 Yang, Z., Wang, S., Halim, A. et al. (2015). Engineered CHO cells for production of diverse, homogeneous glycoproteins. *Nat. Biotechnol.* 33: 842–844.

86 Fan, L., Kadura, I., Krebs, L.E. et al. (2012). Improving the efficiency of CHO cell line generation using glutamine synthetase gene knockout cells. *Biotechnol. Bioeng.* 109: 1007–1015.

87 Santiago, Y., Chan, E., Liu, P.-Q. et al. (2008). Targeted gene knockout in mammalian cells by using engineered zinc-finger nucleases. *Proc. Natl. Acad. Sci. U.S.A.* 105: 5809–5814.

88 Cost, G.J., Freyvert, Y., Vafiadis, A. et al. (2010). BAK and BAX deletion using zinc-finger nucleases yields apoptosis-resistant CHO cells. *Biotechnol. Bioeng.* 105: 330–340.

89 Orlando, S.J., Santiago, Y., DeKelver, R.C. et al. (2010). Zinc-finger nuclease-driven targeted integration into mammalian genomes using donors with limited chromosomal homology. *Nucleic Acids Res.* 38: e152.

90 Sakuma, T., Takenaga, M., Kawabe, Y. et al. (2015). Homologous recombination-independent large gene cassette knock-in in CHO cells using TALEN and MMEJ-directed donor plasmids. *Int. J. Mol. Sci.* 16: 23849–23866.

91 Cristea, S., Freyvert, Y., Santiago, Y. et al. (2013). In vivo cleavage of transgene donors promotes nuclease-mediated targeted integration. *Biotechnol. Bioeng.* 110: 871–880.

92 Lee, J.S., Grav, L.M., Lewis, N.E., and Faustrup Kildegaard, H. (2015). CRISPR/Cas9-mediated genome engineering of CHO cell factories: application and perspectives. *Biotechnol. J.* 10: 979–994.

93 Grav, L.M., Lee, J.S., Gerling, S. et al. (2015). One-step generation of triple knockout CHO cell lines using CRISPR/Cas9 and fluorescent enrichment. *Biotechnol. J.* 10: 1446–1456.

94 Schmieder, V., Bydlinski, N., Strasser, R. et al. (2017). Enhanced genome editing tools for multi-gene deletion knock-out approaches using paired CRISPR sgRNAs in CHO cells. *Biotechnol. J.* 13: 1700211.

95 Sun, T., Li, C., Han, L. et al. (2015). Functional knockout of FUT8 in Chinese hamster ovary cells using CRISPR/Cas9 to produce a defucosylated antibody. *Eng. Life Sci.* 15: 660–666.

96 Zong, H., Han, L., Ding, K. et al. (2017). Producing defucosylated antibodies with enhanced in vitro antibody-dependent cellular cytotoxicity via FUT8 knockout CHO-S cells. *Eng. Life Sci.* 17: 801–808.

97 Chung, C., Wang, Q., Yang, S. et al. (2017). Integrated genome and protein editing swaps α-2,6 sialylation for α-2,3 sialic acid on recombinant antibodies from CHO. *Biotechnol. J.* 12: 1600502.

98 Spahn, P.N., Hansen, A.H., Hansen, H.G. et al. (2016). A Markov chain model for N-linked protein glycosylation – towards a low-parameter tool for model-driven glycoengineering. *Metab. Eng.* 33: 52–66.

99 Park, J.H., Jin, J.H., Lim, M.S. et al. (2017). Proteomic analysis of host cell protein dynamics in the culture supernatants of antibody-producing CHO cells. *Sci. Rep.* 7: 44246.

100 Chiu, J., Valente, K.N., Levy, N.E. et al. (2017). Knockout of a difficult-to-remove CHO host cell protein, lipoprotein lipase, for improved polysorbate stability in monoclonal antibody formulations. *Biotechnol. Bioeng.* 114: 1006–1015.

101 Lee, N., Shin, J., Park, J.H. et al. (2016). Targeted gene deletion using DNA-free RNA-guided Cas9 nuclease accelerates adaptation of CHO cells to suspension culture. *ACS Synth. Biol.* 5: 1211–1219.

102 Grav, L.M., la Cour Karottki, K.J., Lee, J.S., and Kildegaard, H.F. (2017). Application of CRISPR/Cas9 genome editing to improve recombinant protein production in CHO cells. *Methods Mol. Biol. (Clifton, NJ)* 1603: 101–118.

103 Shen, C.-C., Sung, L.-Y., Lin, S.-Y. et al. (2017). Enhancing protein production yield from Chinese hamster ovary cells by CRISPR interference. *ACS Synth. Biol.* 6: 1509–1519.

104 Lee, J.S., Kallehauge, T.B., Pedersen, L.E., and Kildegaard, H.F. (2015). Site-specific integration in CHO cells mediated by CRISPR/Cas9 and homology-directed DNA repair pathway. *Sci. Rep.* 5: 8572.

105 Bachu, R., Bergareche, I., and Chasin, L.A. (2015). CRISPR-Cas targeted plasmid integration into mammalian cells via non-homologous end joining. *Biotechnol. Bioeng.* 112: 2154–2162.

Chapter 2 - CRISPR/Cas9 as a genome editing tool for targeted gene integration in CHO cells

This chapter describes a step-by-step protocol for CRISPR/Cas9-mediated targeted integration of genes into the genome of CHO cells. The method is based on the homology-mediated integration of a donor plasmid with a gene of interest and antibiotic selection marker into a specific genomic site. Fluorescence-activated cell sorting (FACS) is used for the enrichment of the cells that have correct integration. The method can be applied for the development of stable CHO cell lines for the production of recombinant proteins.

The chapter is reprinted from: 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". CRISPR Gene Editing: Methods and Protocols, Methods in Molecular Biology, vol. 1961 (eds Luo Y.), Humana Press. 2019

CRISPR/Cas9 as a genome editing tool for targeted gene integration in CHO cells

Daria Sergeeva[1], Jose Manuel Camacho-Zaragoza[1], Jae Seong Lee[2], Helene Faustrup Kildegaard[1]

[1] The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Kgs. Lyngby, Denmark
[2]Department of Molecular Science and Technology, Ajou University, Suwon, Republic of Korea

Abstract

The emergence of CRISPR/Cas9 system as a precise and affordable method for genome editing has prompted its rapid adoption for the targeted integration of transgenes in Chinese hamster ovary (CHO) cells. Targeted gene integration allows the generation of stable cell lines with a controlled and predictable behavior, which is an important feature for the rational design of cell factories aimed at the large-scale production of recombinant proteins. Here we present the protocol for CRISPR/Cas9-mediated integration of a gene expression cassette into a specific genomic locus in CHO cells using homology-directed DNA repair.

Keywords: Chinese hamster ovary cells, CRISPR/Cas9, genome editing, targeted integration

1. Introduction

CHO cells have a privileged position in biopharma industry as the preferred host for the production of recombinant therapeutic proteins. With a long record of regulatory approvals, therapeutic protein production in CHO cells now represents 60-70% of biopharmaceuticals market [1]. Development of new biopharmaceuticals and biosimilars has put more pressure in biopharma industry for shortening cell line generation time and lowering manufacturing costs. This has prompted a renewed interest in genome engineering of CHO cells to harness the whole biosynthetic capacity for recombinant protein production [1]. Achieving this goal calls for the generation of streamlined cell lines by means of genetic engineering approaches aimed at mitigating the metabolic and cellular bottlenecks that limit production.

Whereas small-scale production of recombinant proteins for screening purposes or pre-clinical studies can be easily achieved by transient gene expression systems, high productivity is largely dependent on the long-term retention of the plasmid encoding the gene of interest (GOI), which generally decreases over time due to cell division [2]. Consequently, the large-scale protein production is generally accomplished in stable cell lines where the GOI has been integrated into the genome.

Most methods for gene integration can be grouped into two classes: random integration and targeted integration approaches. Current CHO cell line development technologies are mostly based on random integration and amplification of a recombinant gene together with a selection marker, commonly, dihydrofolate reductase (DHFR) or glutamine synthetase (GS) [3]. Integration of transgene into random genomic sites yields cell lines with a wide range of expression, growth and stability characteristics and require extensive clone screening to identify stable high producers. In contrast to random integration, targeted integration allows the selection of a transcriptionally active genomic locus that promotes high and stable transgene expression with minimal perturbation of the genetic context, keeping out confounding factors such as chromosome position effects, copy number variability and unwanted mutations [4,5]. Thus, targeted integration approach accelerates development of cell lines with predictable performance and consistent behaviour.

Site-specific genome editing tools developed over the last decades rely on engineered nucleases and comprise zinc-finger nucleases (ZFN), transcription activator-like effector nucleases (TALENs) and, the most recent, clustered regularly interspaced short palindromic repeats/CRISPR-associated (CRISPR/Cas) system [6]. These tools can be customized to recognize and cleave a double-stranded DNA molecule at a specific site, which can be subsequently repaired by one of two endogenous cellular pathways mainly: non-homologous end-joining (NHEJ) or homology-directed repair (HDR). NHEJ generates indel mutations at the repair site, which can be harnessed for the knockout generation at the targeted locus. On the other hand, HDR can seamlessly repair DNA break in the presence of a DNA repair template with homologous regions spanning the cleavage site [6]. The DNA template used in HDR can be customized by placing a GOI in between two homology arms and then delivered exogenously in a plasmid vector. Even though HDR occurs at a lower frequency than NHEJ in mammalian cells (particularly in CHO cells) [7], this strategy has been successfully implemented for the targeted integration of a GOI in a genomic locus in several mammalian cells, including CHO cells [4,5,8–14].

Compared to ZFNs and TALENs, whose specific site recognition activity relies only on the protein structure of the nuclease by itself, site recognition in the CRISPR/Cas9 system is carried out within a nuclease/RNA complex. Base pairing between a small RNA fragment and the target locus guides the nuclease to the target site in a sequence-specific manner. The RNA molecule used by this system is called single guide RNA (sgRNA), which is a synthetic fusion of two RNA molecules: a targeting RNA molecule called CRISPR RNA (crRNA) and a scaffold RNA called trans-activating crRNA (tracrRNA) [15]. Site recognition occurs by base-pairing of the crRNA portion of the sgRNA and the genomic target sequence when it is immediately next to a small 3 nt sequence called the protospacer adjacent motif (PAM). Upon target recognition, two endonuclease domains within Cas9 (RuvC and HNH) become activated, generating a blunt double-stranded break (DSB) between the 3rd and the 4th nucleotide upstream from the PAM site (Figure 1A) [15]. The ease, versatility and speed of creating sgRNAs have fostered the use and evolution of CRISPR/Cas9 as a tool over previous methods, making this system the preferred genome engineering tool currently available.

In this chapter we describe the protocol for CRISPR/Cas9-mediated targeted integration of a gene expression cassette in a predefined genomic site in CHO cells. This method is based on the delivery of three plasmids: Cas9, sgRNA and a donor plasmid (repair template). The donor plasmid harbors the GOI, a selection marker and homology arms (HAs) to promote homology-directed integration. Upon drug selection and fluorescence-activated cell sorting (FACS), cell lines with homogeneous and stable transgene expression can be generated. This method is easily adaptable to site-specific knock-in of genes in a predefined genomic site in various mammalian cell lines.

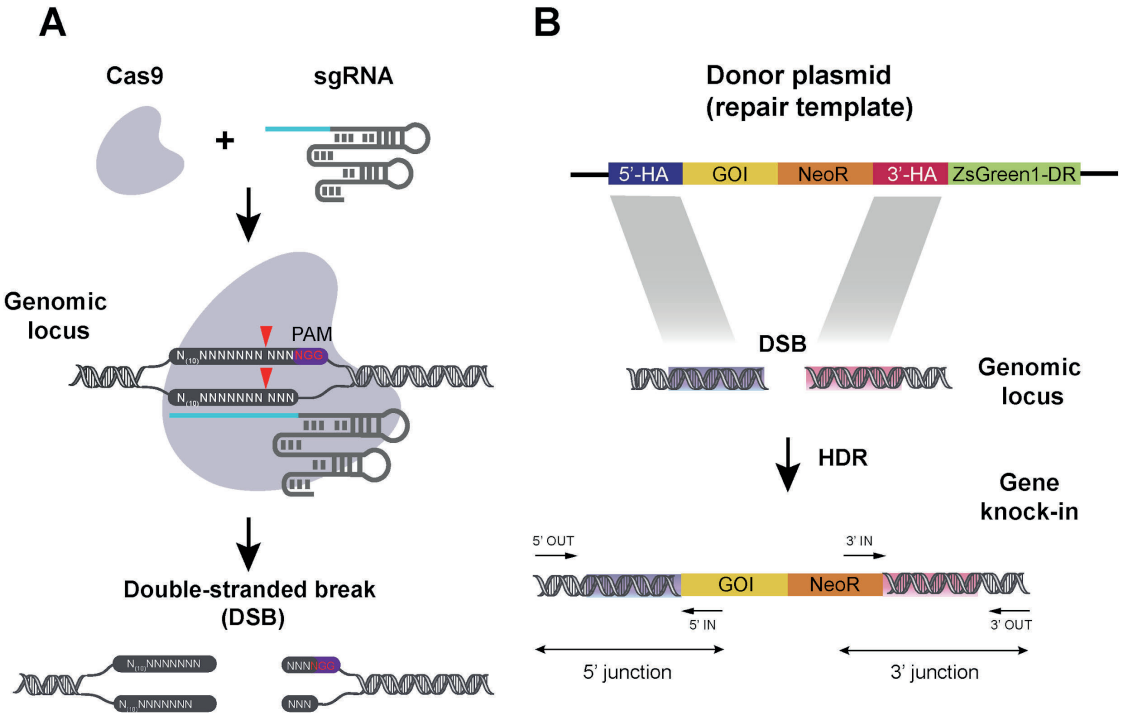


Figure 1. The use of CRISPR/Cas9 for targeted gene integration. (A) Complex of Cas9 and sgRNA recognizes genomic target sequence in the proximity of PAM site and generates double-stranded break (DSB). (B) DSB can be repaired by homology-directed repair (HDR) when a repair template with homology arms (HAs) is provided. It results in the integration of the GOI in the predefined locus.

2. Materials

2.1 Construction of sgRNA and donor plasmid

1. Glycerol stocks of *E. coli* transformed with sgRNA expression plasmid (from Ronda et al. [16]) and donor plasmid (from Lee et al. [5]).
2. 2x YT medium.
3. Kanamycin.
4. Ampicillin.
5. 250-500 mL baffled Erlenmeyer shake flask.
6. Sterile pipette tips.
7. Incubator with shaker, 37°C, 250 rpm.
8. Plasmid mini- and midiprep kit.
9. Nuclease-free water.
10. NanoDrop 2000.
11. Oligonucleotides containing sgRNA sequence (for design instructions see section 3.1 and 3.2).
12. PCR primers for amplification of sgRNA backbone and donor plasmid backbone (see sections 3.2 and 3.4).
13. PCR primers for amplification of homology arms, drug resistance cassette and gene expression cassette (see section 3.3 and 3.4).
14. 2x Phusion U Hot Start PCR Master Mix (Thermo Scientific).
15. PCR tubes.
16. Thermocycler.
17. FastDigest DpnI enzyme (Thermo Scientific).
18. 10x FastDigest Green Buffer (Thermo Scientific).
19. 1 kb DNA ladder.
20. 1% agarose gel: 1 g agarose powder dissolved in 100 mL 1x TAE buffer.
21. Gel chamber and power source.
22. PCR and gel purification kit.
23. Heat block.
24. USER enzyme (New England Biolabs).
25. CutSmart buffer (New England Biolabs).

26. 1.5 mL Eppendorf tubes.
27. Mach1 competent *E. coli* cells (Thermo Scientific).
28. Heat block, 37°C, 300 rpm.
29. Table top centrifuge.
30. Sterile spatula.
31. LB-ampicillin and LB-kanamycin agar plates: 15 g/L Agar, 10 g/L Tryptone, 10 g/L NaCl, 5 g/L Yeast Extract, 60 µg/mL ampicillin or 50 µg/mL kanamycin.

2.2 Preparation of Cas9 expression plasmid

1. Glycerol stock with *E. coli* transformed with the codon-optimized Cas9 expression plasmid (from Ronda et al. [16]).
2. Ampicillin.
3. 2x YT medium.
4. 250-500 mL baffled Erlenmeyer shake flask.
5. Sterile pipette tips.
6. Incubator with shaker, 37°C, 250 rpm.
7. Plasmid midiprep kit.
8. Nuclease-free water.
9. NanoDrop 2000.

2.3 Transfection of CHO-S cells

1. CHO-S cells (Life Technologies).
2. Cell counter.
3. Growth medium: CD CHO medium (Life Technologies), 8 mM L-glutamine.
4. 15 or 50 mL centrifuge tubes.
5. 6-well plate, flat bottom, not-treated (Corning).
6. Humidified incubator, 37°C, 5% CO_2, 120 rpm.
7. OptiPro SFM (Life Technologies).
8. FreeStyle MAX reagent (Life Technologies).
9. sgRNA expression plasmid generated in section 3.5.
10. Donor plasmid generated in section 3.5.

11. Cas9 expression plasmid prepared in section 3.6.

2.4 Antibiotic selection

1. Growth medium: CD CHO medium, 8 mM L-glutamine.
2. Selection drug (hygromycin B or G418).
3. 6-well plate, advanced TC (Greiner bio-one).
4. Humidified incubator, 37°C, 5% CO_2
5. TrypLE select (Life Technologies).
6. Phosphate buffered saline (PBS).
7. 6-well plate, flat bottom, not-treated (Corning).

2.5 Fluorescence Activated Cell Sorting (FACS)

1. Fluorescence-activated cell sorter.
2. 384-well plates, flat bottom, TC-treated (Corning).
3. FACS sorting medium: CD CHO medium, 8 mM L-glutamine, 1x Antibiotic-Antimycotic (Gibco), and 1.5% HEPES (Life Technologies).
4. FACS tubes.
5. 30 µm cell strainers.
6. Image cytometer or microscope.
7. Humidified incubator, 37°C, 5% CO_2, no shake.
8. 96-well plates, flat bottom, not-treated (Corning)..
9. Clone expansion medium: CD CHO medium, 8 mM L-glutamine, 1x Antibiotic-Antimycotic, and 1 µL/mL Anti-clumping agent (Life Technologies)
10. 96-well plates, V-Shaped (Greiner bio-one).
11. Breathable plastic bag.

2.6 PCR verification of clones

1. QuickExtract DNA extraction solution (Epicentre).
2. Primers (for design instructions see section 3.10).
3. 2x Phusion PCR Master Mix (Thermo Scientific).
4. Nuclease-free water.

5. PCR tubes or plates.
6. Thermocycler.
7. 1 kb DNA ladder.
8. 1% agarose gel.
9. Gel chamber and power source.

2.7 Expansion of clones

1. 12-well plates, flat bottom.
2. 6-well plates, flat bottom.
3. 125 mL shake flask.
4. Clone expansion medium: CD CHO medium, 8 mM L-glutamine, 1x Antibiotic-Antimycotic, and 1 µL/mL Anti-clumping agent.
5. DMSO.
6. Cryotubes.
7. Humidified incubator, 37°C, 5% CO_2, 120 rpm.

2.8 Copy number analysis

1. GeneJet Genomic DNA purification kit (Thermo Scientific)
2. NanoDrop 2000.
3. 20x TaqMan Gene Expression Assays for GOI and COSMC (see section 3.11 for instructions).
4. 2x TaqMan Gene Expression Master Mix (Applied Biosystems).
5. Nuclease-free water.
6. Real-time PCR instrument.
7. Reaction tubes or plates suitable for qRT-PCR.

3. Methods

The following section is a general protocol that we use for site-specific integration in CHO-S cells in our lab (Figure 2). The protocol includes generation of sgRNA plasmid, Cas9 plasmid and donor plasmid with GOI by USER cloning [16–18]. After co-delivery of these plasmids and antibiotic selection, cells can be single-cell sorted using FACS. Clones

with targeted integration can be verified using PCR analysis of junction sequences and copy number analysis by quantitative real time PCR (qRT-PCR).

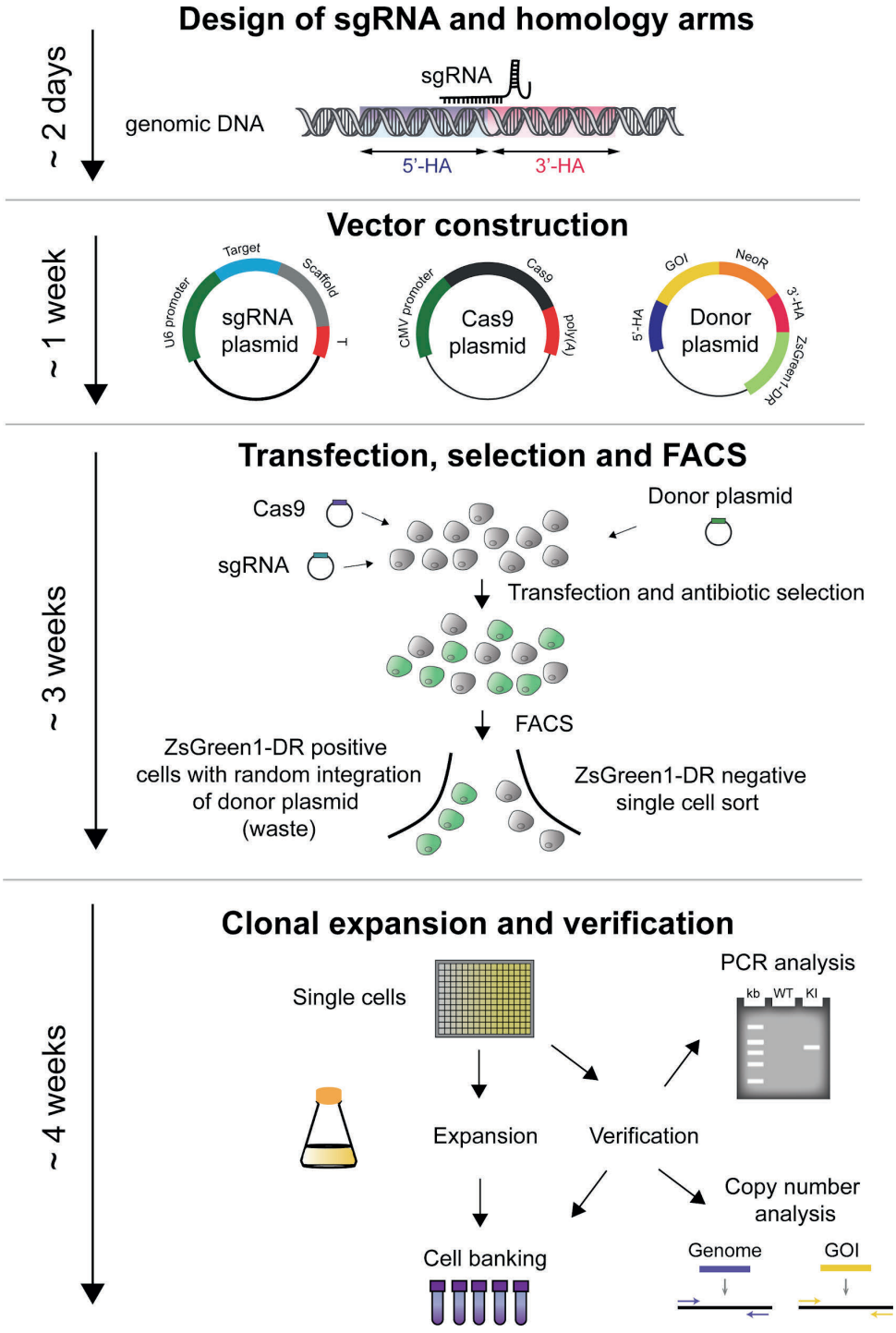


Figure 2. An overview of the protocol for the generation of CHO cell lines with CRISPR/Cas9-mediated targeted gene integration. The first step of the protocol is to design a sgRNA and homology arms for the specific genomic locus. Second, sgRNA, Cas9 and donor plasmids are constructed. Next, CHO cells are transfected with constructed plasmids and subjected to antibiotic selection and FACS sorting. Clones with targeted integration should be

ZsGreen1-DR negative since this fluorescent marker should not be integrated into the genome. Then cells are expanded and verified using PCR and copy number analysis to isolate cell lines with site-specific gene integration.

3.1 Design of sgRNA

1. Find a sequence of the genomic locus you want to use for integration in CHO genome (download genomic sequence from NCBI).
2. Copy ~1500 - 2000 bp of the genomic region and paste it to a CRISPR design tool (e.g. CRISPOR http://crispor.tefor.net/).
3. Select appropriate genome and PAM and search for target sequences.
4. In the list of predicted target sequences, select one with the highest specificity score and the lowest off-targets (see Note 1).
5. Proceed with sgRNA plasmid construction.

3.2 sgRNA plasmid construction by USER cloning

The sgRNA plasmid consists of two DNA bricks, which can be assembled by USER cloning (Figure 3): (1) sgRNA backbone, which contains elements for replication in the bacterial host, U6 promoter, sgRNA scaffold and termination signal to generate sgRNA expression cassette. (2) sgRNA annealed oligonucleotides, containing target sequence.

1. Design and order a pair of sgRNA oligos using the following sequences by replacing the N's with your 20 nt target sequence. The first position of the target sequence is a GC pair, shown in bold (if your target sequence starts with another nucleotide, replace it with G for proper U6-driven transcription).
   sgRNA_FW_oligo:
   5'-GGAAAGGACGAAACACCGNNNNNNNNNNNNNNNNNNNGTTTTAGAGCTAGAAAT-3'.
   sgRNA_RV_ oligo:
   5'-CTAAAACNNNNNNNNNNNNNNNNNNNCGGTGTTTCGTCCTTTCCACAAGATAT-3'.
2. Anneal sgRNA oligos. Mix the following in an Eppendorf tube: 10 µL 10x NEBuffer 4, 10 µL sgRNA forward oligo (100 µM), 10 µL sgRNA reverse oligo (100 µM), 70 µL nuclease-free water.

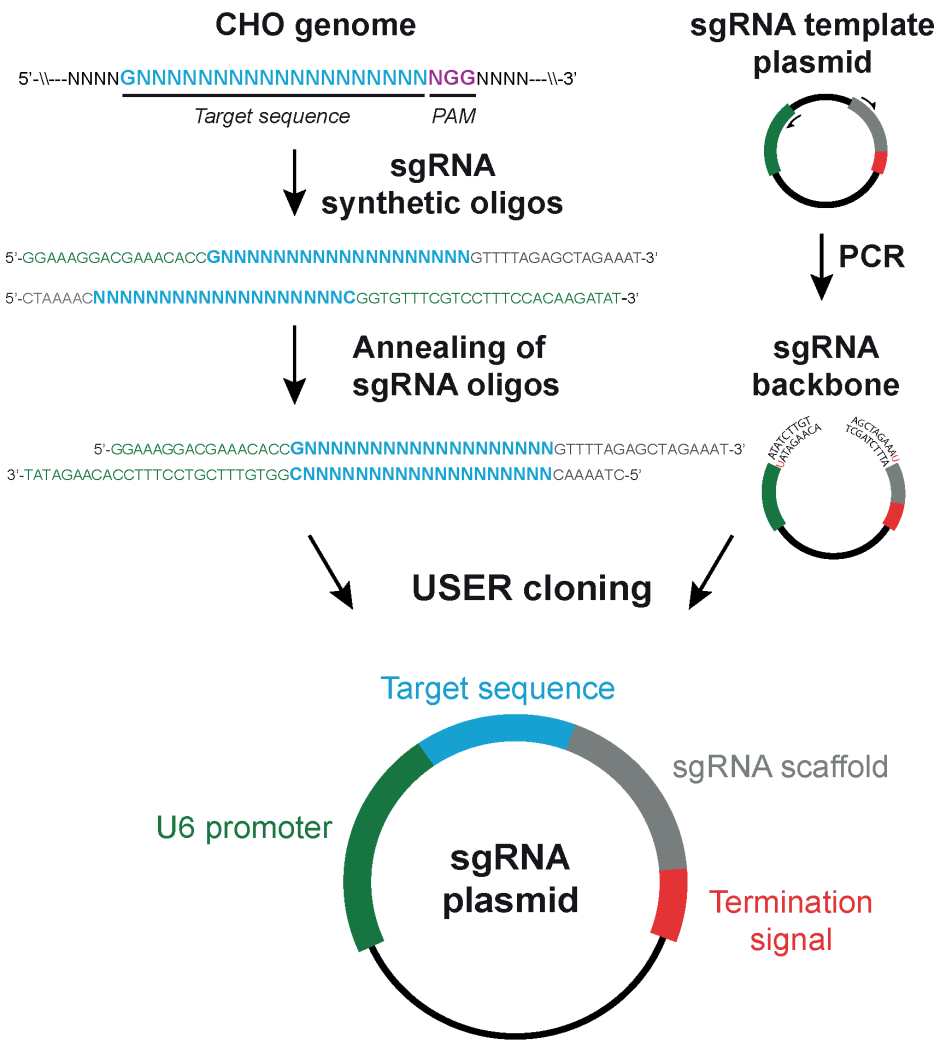


Figure 3. An overview of the sgRNA plasmid construction. The target sequence with extensions is ordered as two short synthetic oligonucleotides. After annealing these oligos form 3' single-stranded overhangs, which will be used in USER cloning. The universal sgRNA plasmid backbone is amplified using PCR primers containing a single deoxyuracil residue. USER enzyme treatment creates 3'-overhangs on sgRNA plasmid backbone, which allows assembly of the target sequence and backbone, generating sgRNA plasmid.

3. Incubate at 95°C for 5 minutes on a heat block, then turn off the heat block and leave overnight for oligo annealing.
4. Request the sgRNA backbone plasmid from Ronda et al. [16] and make a bacterial glycerol stock.
5. Inoculate bacterial stock in 4 mL of 2x YT medium in a 14 mL bacterial culture tube with 50 µg/mL kanamycin. Incubate overnight at 37°C with shaking at 250 rpm.

6. Extract plasmid using a miniprep kit. Resuspend in nuclease-free water and quantify using NanoDrop 2000.
7. Order the following uracil-containing primers for sgRNA plasmid backbone amplification:
   sgRNA_BB_FW: 5'-AGCTAGAAAUAGCAAGTTAAAATAAGGC -3'.
   sgRNA_BB_RV: 5'-ACAAGATAUATAAAGCCAAGAAATCGA -3'.
8. Amplify the sgRNA plasmid backbone. Mix the following in a PCR tube: 2.5 µL primer sgRNA_BB_FW (10 µM), 2.5 µL primer sgRNA_BB_RV (10 µM), 25 µL 2x Phusion U Hot Start PCR Master Mix, 1 µL sgRNA backbone plasmid (2.5pg - 25ng), 19 µL nuclease-free water.
9. Place the PCR tube in the thermocycler and run the following PCR program: 98 °C for 30 s; 35 cycles of 98 °C for 10 s, 57 °C for 30 s, 72 °C for 1 min 15 s; 72 °C for 10 min.
10. Treat the PCR product with DpnI enzyme to degrade the methylated plasmid template by mixing the following components: 44 µL sgRNA backbone PCR product, 5 µL 10x FastDigest Green buffer, 1 µL FastDigest DpnI enzyme.
11. Incubate mixture in the thermocycler at 37°C for 15 min.
12. Run the whole reaction mixture on a preparative 1% agarose gel at 100V for 30 min along with 1 kb DNA ladder.
13. Cut the band close to 4.2 kb and purify using a gel purification kit. Measure concentration of backbone DNA brick using NanoDrop 2000.
14. Assemble sgRNA plasmid by USER cloning. Mix the following components in a PCR tube for sgRNA reaction: 7 µL annealed sgRNA oligos, 1 µL backbone DNA brick, 1 µL CutSmart buffer, 1 µL USER enzyme. For the negative control mix the following: 7 µL nuclease-free water, 1 µL backbone DNA brick, 1 µL CutSmart buffer, 1 µL USER enzyme.
15. Incubate the mix at 37°C for 40 min and 25°C for 30 min.
16. Add 1.5 µL of USER reaction to 15 µL of *E. coli* Mach1 competent cells in an Eppendorf tube and incubate on ice for 30 min.
17. Heat shock at 42°C for 30 s.
18. Place the cells on ice for 1 min and then add 950 µL of 2x YT medium.
19. Incubate at 37°C for 1 h at 300 rpm.
20. Spin down the cells at 2000 g for 5 min.
21. Discard the supernatant and resuspend pellet in 100 µL of 2x YT medium.

22. Plate all the sample (100 µL) in an agar plate with kanamycin and incubate at 37°C overnight.

3.3 Design of homology arms

1. Open genomic locus sequence from section 3.1 in your preferred sequence viewer software.
2. Identify CRISPR/Cas9 cleavage site as the site between the 3rd and the 4th nucleotide upstream from the PAM sequence within your target site. Your 5′ and 3′ homology arms (5′-HA and 3′-HA) will correspond to the sequence extending ~750 bp to the left and to the right of the cleavage site, respectively (Figure 4).
3. Design primers for amplification of homology arms so that they bind approximately 750 bp upstream and downstream of the cleavage site. Using these primers, amplify ~1500 bp of genomic region and use this PCR amplicon as a template for the generation of homology arms.

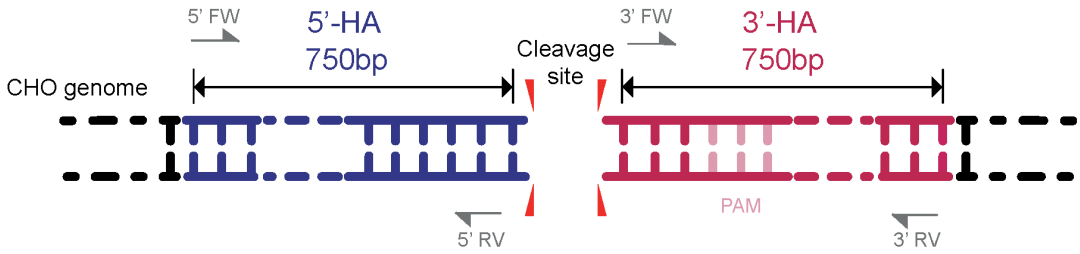


Figure 4. Design of homology arms. Homology arms are ~750bp sequences upstream and downstream of CRISPR/Cas9 cleavage site in the genome. The Cas9 cleavage site is located 3 nt upstream of the PAM site.

3.4 Donor plasmid construction by USER cloning

The donor plasmid consists of the following fragments, which can be assembled by USER cloning (Figure 5): (1) Donor plasmid backbone (amplified from plasmid from Lee et al. [5]). This backbone includes elements for replication in the bacterial host and ZsGreen1-DR expression cassette as a fluorescent reporter to rule out cells with random integration of the donor plasmid using FACS. (2) 5′ and 3′ homology arms (~750 bp each) to mediate homology-directed targeted integration (amplified from genomic DNA or synthetic template). (3) Drug resistance cassette, harboring neomycin or hygromycin

resistance gene together with promoter and terminator sequences (amplified from another plasmid or synthetic template). (4) Gene expression cassette, harboring gene of interest (GOI) with the desired promoter and terminator sequence (amplified from another plasmid or synthetic template).

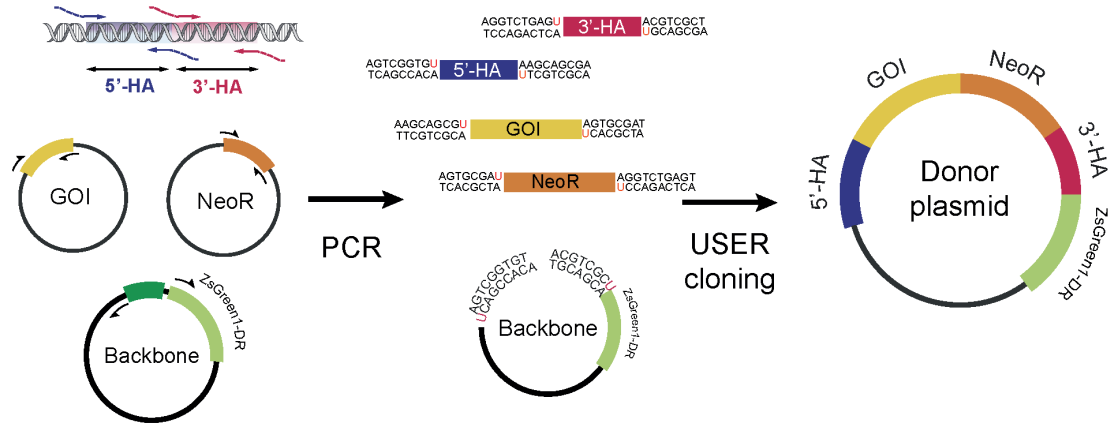


Figure 5. An overview of the donor plasmid construction. Five DNA bricks (5′ and 3′-homology arms (HAs), gene expression cassette (with GOI), drug resistance cassette (e.g. NeoR) and plasmid backbone) are amplified using primers containing USER-linkers. After USER treatment PCR fragments are assembled in the desired order and form the donor plasmid.

1. Request the donor backbone plasmid from Lee et al. [5] and make a bacterial glycerol stock.
2. Inoculate bacterial stock in 4 mL of 2x YT medium in a 14 mL bacterial culture tube with 60 µg/mL ampicillin. Incubate overnight at 37°C with shaking at 250 rpm.
3. Extract plasmid using a miniprep kit. Resuspend in nuclease-free water and quantify using NanoDrop 2000.
4. Design primers for amplification of homology arms (HAs), gene expression cassette (GOI) and drug resistance (DR) cassette. Primers for HAs should bind to 5′ and 3′ ends of genomic region and at the cleavage site (Figure 4). Add following uracil-containing overhangs on 5′-end of your primers for USER cloning (see Note 2):

   5′-HA_FW primer tail (Linker A): 5′-AGTCGGTGU-3′.

   5′-HA_RV primer tail (Linker B): 5′-ACGCTGCTU-3′.

   GOI_FW primer tail (Linker B): 5′-AAGCAGCGU-3′.

   GOI_RV primer tail (Linker O2): 5′-ATCGCACU-3′.

DR_FW primer tail (Linker O2): 5'-AGTGCGAU-3'.

DR_RV primer tail (Linker D): 5'-ACTCAGACCU-3'.

3'-HA_FW primer tail (Linker D): 5'-AGGTCTGAGU-3'.

3'-HA_RV primer tail (Linker O1): 5'-AGCGACGU-3'.

5. Order following uracil-containing primers for amplification of donor plasmid backbone:
   Donor_BB_FW primer (Linker O1): 5'-ACGTCGCUGTTGACATTGATTATTGACT-3'.
   Donor_BB_RV primer (Linker A): 5'-ACACCGACUGAGTCGAATAAGGGCGACACCCCA-3'.
6. Amplify DNA bricks for USER cloning. For homology arms amplification use the DNA of genomic region (25 - 250 ng) amplified in the section 3.3. For amplification of donor plasmid backbone use plasmid (1 - 10 ng) prepared in step 3. For amplification of DR and GOI use your source of sequence (another plasmid or synthetic template).
7. Mix the following in a PCR tube: 2.5 µL primer forward (10 µM), 2.5 µL primer reverse (10 µM), 25 µL 2x Phusion U Hot Start PCR Master Mix, 1 µL template, 19 µL nuclease-free water
8. Place the PCR tubes in the thermocycler and run the following program: 98 °C for 30 s; 35 cycles of 98 °C for 10 s, Tm for 30 s, 72 °C for 15-30 s/kb; 72 °C for 10 min.
9. If plasmid was used as a template for PCR reaction, treat the PCR product with DpnI enzyme as was described in the section 3.2. Purify DNA bricks and measure concentration using NanoDrop 2000.
10. Perform USER cloning reaction. Mix the components with an equimolar ratio in a PCR tube as shown in Table 1.
11. Incubate the mix at 37°C for 40 min and 25°C for 30 min.
12. Add 1.5 µL of USER reaction to 15 µL of *E. coli* Mach1 competent cells in an Eppendorf tube and incubate on ice for 30 min.
13. Heat shock at 42°C for 30 s.
14. Place the cells on ice for 1 min and then add 950 µL of 2x YT medium.
15. Incubate at 37°C for 1 h at 300 rpm.
16. Spin down the cells at 2000 g for 5 min.
17. Discard the supernatant and resuspend pellet in 100 µL of 2x YT medium.
18. Plate the sample (100 µL) in an agar plate with ampicillin and incubate at 37°C overnight.

Table 1. Components for donor plasmid assembly

Component	USER reaction (µL)	Negative control (µL)
Donor plasmid backbone	1	1
5'-HA	1	-
3'-HA	1	-
Drug resistance cassette	1	-
Gene expression cassette	1	-
CutSmart buffer	1	1
USER enzyme	1	1
Nuclease-free water	3	7

3.5 Analysis and preparation of sgRNA and donor plasmids

1. Pick colonies from each plate of sgRNA and donor plasmid transformation using a sterile pipette tip and inoculate 4 mL of 2x YT medium with the corresponding antibiotic in a 10 mL bacterial culture tube.
2. Incubate at 37°C and 250 rpm.
3. Extract plasmid using a miniprep kit and use this preparation for Sanger sequencing. Use sequencing primers that bind before the U6 promoter of the sgRNA plasmid and sequencing primers covering both 5' and 3' homology arms and sequence in between these elements of the donor plasmid.
4. Analyze sequencing results using a sequencing analysis software to make sure that your plasmids do not contain any mutations.
5. Grow a bacterial culture of correct transformants for plasmid purification. Inoculate 100-200 mL of 2x YT medium in baffled shake flasks, supplemented with the corresponding antibiotics, and incubate at 37°C and 250 rpm overnight.
6. Extract plasmid using a midiprep kit and resuspend in endotoxin-free water. Quantify plasmid yield using NanoDrop 2000 and dilute plasmids to 500-1000 ng/µL. These preparations will be used for the CHO-S transfection.

3.6 Preparation of Cas9 expression plasmid

1. Request Cas9 plasmid from Ronda et al. [16].
2. Inoculate 100-200 mL of 2x YT medium with 60 μg/mL ampicillin in shake flasks with the bacterial stock using a sterile pipette tip and incubate overnight at 37°C and 250 rpm.
3. Extract plasmid using a midiprep kit and resuspend in endotoxin-free water. Quantify plasmid DNA using NanoDrop 2000, dilute it to 500-1000 ng/μL and save this preparation for the transfection.

3.7 Transfection of CHO-S cells

1. Cultivate CHO-S cells in the shake flask using CD CHO medium supplemented with 8 mM L-glutamine. For transfection use CHO-S cells at low passage and viability above 95%.
2. One day before transfection measure viable cell density (VCD) and calculate the volume needed to seed at 7 x 10^5 cells/mL in 30 mL of fresh medium.
3. Spin down the required volume 200 g for 5 min and discard the supernatant.
4. Resuspend cells in 5 mL of prewarmed medium (CD CHO + 8mM L-glutamine) and transfer to 25 mL of media in a 125 mL shake flask.
5. Incubate cells at 37°C, 5% CO_2, and shake 120 rpm for 16-24 h.
6. On the day of transfection, measure VCD and dilute cells in prewarmed media (CD CHO + 8mM L-glutamine) to a final VCD of 1 x 10^6 cells/mL.
7. Add 3 mL of diluted cells to a 6-well plate and place the plate in the incubator at 37°C, 5% CO_2, and shake 120 rpm.
8. Dilute plasmids in OptiPRO SFM to a final volume 60 μL. Mix sgRNA, Cas9 and donor plasmids using a total amount of 3.75 μg of DNA in a 1:1:1 (w/w) ratio (Cas9:sgRNA:donor).
9. Invert transfection reagent tube (FreeStyle MAX). Dilute 3.75 μL of transfection reagent with OptiPRO SFM to a total volume 60 μL. Mix gently. Incubate for 5 minutes.
10. Add diluted transfection reagent to diluted plasmid mix.
11. Incubate lipid-DNA mix for 8 minutes.
12. Immediately after incubation, transfer the transfection mix (120 μL) to cells in the 6-well plate.
13. Incubate cells at 37°C, 5% CO_2, and shake 120 rpm for 3 days (see Note 3).

3.8 Antibiotic selection

1. Measure VCD of transfected cells and calculate the volume needed to seed at 3 x 10^5 cells/mL in 3 mL of medium (CD CHO + 8mM L-glutamine).
2. Spin down the calculated volume 200 g for 5 min and discard the supernatant.
3. Resuspend the pellet in 3 mL of fresh growth medium with added selection drug (see Note 4) and seed in a 6-well advanced TC plate.
4. Incubate at 37°C, 5% CO_2, no shake for 3-4 days. Live cells will attach to the plate.
5. After 3-4 days carefully exchange spent medium for fresh medium with selection drug, removing dead cells present in the spent media.
6. Repeat step 5 every 3-4 days.
7. Detach cells after around 2 weeks of selection. Remove old media, briefly wash cells with 3 mL PBS, add 0.3 mL of TrypLE reagent and incubate 3-5 min at 37°C. Then add 3 mL of pre-warmed selection media to the cells.
8. Measure VCD. Cell concentration and viability of selected cells should reach 5-10 x 10^5 cells/mL and 60-80 %, respectively.
9. Transfer detached cells to non-treated 6-well plate and cultivate in suspension at 37°C, 5% CO_2, 120 rpm.
10. When cells have recovered to >90% of viability, transfer stable cell pool to the 125 mL shake flask in 15-20 mL of media at a seeding density of 3 x 10^5 viable cells/mL. Keep passaging cells in selection media until FACS.

3.9 Fluorescence activated cell sorting (FACS)

1. Prepare 384-well flat bottom plates with 30 µL of sorting media for single cell sorting.
2. Filter cells through a 30 µm cell strainer into a FACS tube to eliminate clumps and debris.
3. Use wild-type CHO-S and CHO-S transiently transfected with ZsGreen1-DR as gating control for FACS.
4. Single cell sort stable cell pool, selecting for ZsGreen1-DR negative cells (ZsGreen1-DR positive corresponds to cells with a random integration of the donor plasmid), into one or more pre-warmed (37°C) 384-well plates. If FACS is not available, limited dilution can be used instead (see Note 5).
5. Spin plates at 200 g for 5 minutes to make sure cells reach the medium.

6. Place cells in a breathable plastic bag to limit evaporation and incubate cells at 37°C, 5% CO_2, no shake.
7. After 10-14 days check for surviving cells using a microscope or image cytometer. Cell count should preferably be > 1000 cells in a well or confluency > 50%.
8. Transfer selected colonies from 384-well plates to 96-well plates. Carefully pipet cells up and down and move 30 µL to a 96-well flat bottom plate with 180 µL of clone expansion medium.
9. Leave cells for maximum 4 days. When the clones have a confluency > 50%, carefully pipet up and down and transfer 50 µL cell suspension to a 96-well V-shaped plate. Refill the flat-shaped plate with 50 µL of fresh media.
10. Spin down the V-shaped 96-well plate at 1000 x g for 5 minutes. Remove the supernatant and add 20 µL of QuickExtract DNA extraction solution to cell pellets.
11. Resuspend the pellets and move them to PCR tubes or plates. Incubate at 65°C for 15 minutes and 95°C for 5 minutes. Store at -20°C. This DNA will be used for PCR verification of clones.
12. Within 4 following days verify clones using PCR analysis.

3.10 PCR verification of clones

1. Design primers that flank 5′ and 3′ donor:genome junctions (Figure 1B). You should have primers binding outside of homology arms in the genomic locus (5′ and 3′ OUT primers), a primer specific to 5′ end of gene expression cassette and a primer recognizing 3′ end of drug resistance cassette (5′ and 3′ IN primers). Use 5′ OUT primer with 5′ IN primer and 3′ IN primer with 3′ OUT primer to amplify 5′ and 3′ junction sequences, respectively.
2. Mix the following components in a PCR tube for 5′ junction PCR (for one reaction): 10 µL 2x Phusion Master Mix, 1 µL primer 5′ forward OUT (10 µM), 1 µL primer 5′ reverse IN (10 µM), 1 µL DNA template (genomic DNA extract from section 3.9), 7 µL nuclease-free water. Use genomic DNA extracted from wild-type CHO-S as a template in the negative control.
3. Place the PCR tubes in a thermocycler and run the program for touchdown PCR as shown in Table 2.
4. Run PCR products on a 1% agarose gel and select clones with expected amplicon size. No PCR amplicons with expected size should be present in the negative control.

5. If desired, sequence respective junction PCR amplicons by Sanger sequencing to ensure that the junction sequences are correct.
6. Repeat steps 2-5 for 3′ junction PCR on selected 5′ junction-positive clones using 3′ forward IN primer with 3′ reverse OUT primer.

Table 2. PCR program for PCR verification of clones

Temperature (°C)	Time (min)	Number of cycles
98	00:30	1
98	00:10	10
Tm+10 -1 per cycle	00:30	
72	15-30 s/kb	
98	00:10	30
Tm	00:30	
72	15-30 s/kb	
72	10:00	1
4	∞	1

3.11 Expansion of clones

1. Select 5′ junction- and 3′ junction-positive clones. It indicates that your GOI was inserted in the selected genomic locus.
2. Move the selected clones from the 96-well plate when > 90% confluent to a 12-well flat bottom plate.
3. Maintain clones in the 12-well plate until confluent, then move 1 mL of cells to a 6-well flat bottom plate with 2 mL media
4. When cells are confluent in 6-well plate, harvest 1 x 10^6 cells for copy number analysis (see section 3.12) and replenish the harvested volume with fresh media.
5. When confluent, move cells from 6-well plate to a 125 mL shake flask and seed at 3 x 10^5 cells/mL in 15-20 mL.

6. Freeze the clones that have one-copy integration (see section 3.12). Spin 1 x 10^7 cells 200 g 5 min, discard media, resuspend pellet in 1 mL medium (CD CHO + 8mM L-glutamine) with 5-10% DMSO and transfer cells to cryotube. Freeze in a Styrofoam box at -80°C the first 24 h before moving to permanent storage at -180°C.

3.12 Copy number analysis by qRT-PCR

1. Extract DNA from 1 x 10^6 cells using Genomic DNA purification kit. Measure the concentration using NanoDrop 2000. Dilute genomic DNA to 10 ng/µL using nuclease-free water.
2. Design TaqMan assay for your GOI using e.g. PrimerQuest software (https://www.idtdna.com/PrimerQuest/). Order the primers and assess their specificity and efficiency (see Note 6). Select specific primers with efficiency between 90 and 105%. Order respective FAM dye-labeled TaqMan probe and test efficiency of your TaqMan assay.
3. Order the following TaqMan assay with VIC dye-labeled MGB probe for endogenous one-copy gene COSMC and test its efficiency:
   COSMC_FW primer: 5'-ACCCGAACCAGGTAGTAGAA-3'.
   COSMC_RV primer: 5'-ACATGTCCAAAGGCCCTAAG-3'.
   COSMC probe: 5'-AGTGACAGCCATATTGGAACAGCATCC-3'.
4. Calculate the number of reactions that you need (including no template control) to perform copy number analysis of your clones by qRT-PCR. Have at least three replicates of each reaction.
5. Prepare the reaction mix and pipet in PCR tubes or plates. Mix components (for one reaction): 10 µL 2x TaqMan Gene Expression Master Mix, 1 µL 20x TaqMan Gene Expression Assay (FAM), 1 µL 20x TaqMan Gene Expression Assay (VIC), 2 µL DNA template, 6 µL nuclease-free water.
6. Run samples on a qRT-PCR instrument with following conditions: 50 °C for 2 min; 95 °C for 1 min; 40 cycles of 95 °C for 15 s, 60 °C for 1 min.
7. Calculate GOI copy number for each clone using the formula:

$$CN = \frac{(1 + Ef(COSMC))^{Ctmean(COSMC)}}{(1 + Ef(GOI))^{Ctmean(GOI)}}$$

where: CN is the number of GOI copies in the genome, *Ct* is the threshold cycle, *Ef(COSMC)* is the efficiency of COSMC TaqMan assay, *Ef(GOI)* is the efficiency of GOI TaqMan assay.

8. Select clones that have one copy of your GOI (see Note 7).

4. Notes

1. We recommend to design and test at least two different sgRNA. Select the sgRNA with the highest DNA cleavage efficiency. To validate sgRNA efficiency you can use T7 endonuclease assay (see protocol from [18]).
2. USER cloning allows direct assembly of multiple DNA bricks by using unique 3' single-stranded DNA overhangs (USER linkers). Here we provide universal USER linkers used in our laboratory, although other linkers can be designed and used as well (see [17]).
3. Other transfection methods can be used for plasmid delivery, e.g. electroporation. Follow instructions provided by the supplier of electroporation system.
4. The drug concentration recommended for antibiotic selection depends on the selection marker used and must be adjusted for the specific cell line by performing a killing curve experiment with different concentrations of the selection drug. For CHO-S cells harboring neomycin resistance cassette, we recommend using 500 μg/mL G418, for hygromycin resistance cassette we recommend using 600 μg/mL hygromycin. Prepare media containing selection drug fresh each time (drug can be added to each well individually).
5. Limiting dilution is an alternative way to isolate single cells. In this case, you will need to screen more clones, as you cannot rule out cells with random integration of the donor plasmid by FACS enrichment.
6. It is important to validate your primers for qRT-PCR and assess amplification efficiency of a qPCR reaction for accurate measurement of GOI copy number. Design several sets of primers and screen them using the standard curve and melt curve analysis.
7. Digital PCR can be used instead of qRT-PCR for copy number analysis using the same TaqMan assays. Digital PCR is less dependent on primer efficiency and provides a linear response to the number of copies present, which can result in more accurate copy number estimation.

References

1. Templeton, N.; Young, J. D. Biochemical and Metabolic Engineering Approaches to Enhance Production of Therapeutic Proteins in Animal Cell Cultures. Biochem. Eng. J. 2018
2. Lee, J.-H.; Park, J.-H.; Park, S.-H.; Kim, S.-H.; Kim, J. Y.; Min, J.-K.; Lee, G. M.; Kim, Y.-G. Co-amplification of EBNA-1 and PyLT through dhfr-mediated gene amplification for improving foreign protein production in transient gene expression in CHO cells. Appl. Microbiol. Biotechnol. 2018
3. Kim, J. Y.; Kim, Y.-G.; Lee, G. M. CHO cells in biotechnology for production of recombinant proteins: current state and further potential. Appl. Microbiol. Biotechnol. 2012, 93, 917–930.
4. Lombardo, A.; Cesana, D.; Genovese, P.; Di Stefano, B.; Provasi, E.; Colombo, D. F.; Neri, M.; Magnani, Z.; Cantore, A.; Lo Riso, P.; Damo, M.; Pello, O. M.; Holmes, M. C.; Gregory, P. D.; Gritti, A.; Broccoli, V.; Bonini, C.; Naldini, L. Site-specific integration and tailoring of cassette design for sustainable gene transfer. Nat. Methods 2011, 8, 861–869.
5. Lee, J. S.; Kallehauge, T. B.; Pedersen, L. E.; Kildegaard, H. F. Site-specific integration in CHO cells mediated by CRISPR/Cas9 and homology-directed DNA repair pathway. Sci. Rep. 2015, 5
6. Carroll, D. Genome engineering with targetable nucleases. Annu. Rev. Biochem. 2014, 83, 409–439.
7. Cristea, S.; Freyvert, Y.; Santiago, Y.; Holmes, M. C.; Urnov, F. D.; Gregory, P. D.; Cost, G. J. In vivo cleavage of transgene donors promotes nuclease-mediated targeted integration. Biotechnol. Bioeng. 2013, 110, 871–880.
8. Hockemeyer, D.; Soldner, F.; Beard, C.; Gao, Q.; Mitalipova, M.; DeKelver, R. C.; Katibah, G. E.; Amora, R.; Boydston, E. A.; Zeitler, B.; Meng, X.; Miller, J. C.; Zhang, L.; Rebar, E. J.; Gregory, P. D.; Urnov, F. D.; Jaenisch, R. Efficient targeting of expressed and silent genes in human ESCs and iPSCs using zinc-finger nucleases. Nat. Biotechnol. 2009, 27, 851–857.
9. Cong, L.; Ran, F. A.; Cox, D.; Lin, S.; Barretto, R.; Habib, N.; Hsu, P. D.; Wu, X.; Jiang, W.; Marraffini, L. A.; Zhang, F. Multiplex genome engineering using CRISPR/Cas systems. Science 2013, 339, 819–823.
10. Mali, P.; Yang, L.; Esvelt, K. M.; Aach, J.; Guell, M.; DiCarlo, J. E.; Norville, J. E.; Church, G. M. RNA-guided human genome engineering via Cas9. Science 2013, 339, 823–826.
11. Miller, J. C.; Tan, S.; Qiao, G.; Barlow, K. A.; Wang, J.; Xia, D. F.; Meng, X.; Paschon, D. E.; Leung, E.; Hinkley, S. J.; Dulay, G. P.; Hua, K. L.; Ankoudinova, I.; Cost, G. J.; Urnov, F. D.; Zhang, H. S.; Holmes, M. C.; Zhang, L.; Gregory, P. D.; Rebar, E. J. A TALE nuclease architecture for efficient genome editing. Nat. Biotechnol. 2011, 29, 143–148.
12. Orlando, S. J.; Santiago, Y.; DeKelver, R. C.; Freyvert, Y.; Boydston, E. A.; Moehle, E. A.; Choi, V. M.; Gopalan, S. M.; Lou, J. F.; Li, J.; Miller, J. C.; Holmes, M. C.; Gregory, P. D.; Urnov, F. D.; Cost, G. J. Zinc-finger nuclease-driven targeted integration into mammalian genomes using donors with limited chromosomal homology. Nucleic Acids Res. 2010, 38, e152.
13. Lee, J. S.; Grav, L. M.; Pedersen, L. E.; Lee, G. M.; Kildegaard, H. F. Accelerated homology-directed targeted integration of transgenes in Chinese hamster ovary cells

via CRISPR/Cas9 and fluorescent enrichment. Biotechnol. Bioeng. 2016, 113, 2518–2523.
14. Gaidukov, L.; Wroblewska, L.; Teague, B.; Nelson, T.; Zhang, X.; Liu, Y.; Jagtap, K.; Mamo, S.; Tseng, W. A.; Lowe, A.; Das, J.; Bandara, K.; Baijuraj, S.; Summers, N. M.; Lu, T. K.; Zhang, L.; Weiss, R. A multi-landing pad DNA integration platform for mammalian cell engineering. Nucleic Acids Res. 2018, 46, 4072–4086.
15. Jinek, M.; Chylinski, K.; Fonfara, I.; Hauer, M.; Doudna, J. A.; Charpentier, E. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science 2012, 337, 816–821.
16. Ronda, C.; Pedersen, L. E.; Hansen, H. G.; Kallehauge, T. B.; Betenbaugh, M. J.; Nielsen, A. T.; Kildegaard, H. F. Accelerating genome editing in CHO cells using CRISPR Cas9 and CRISPy, a web-based target finding tool. Biotechnol. Bioeng. 2014, 111, 1604–1616.
17. Lund, A. M.; Kildegaard, H. F.; Petersen, M. B. K.; Rank, J.; Hansen, B. G.; Andersen, M. R.; Mortensen, U. H. A versatile system for USER cloning-based assembly of expression vectors for mammalian cell engineering. PLoS One 2014, 9, e96693.
18. Grav, L. M.; la Cour Karottki, K. J.; Lee, J. S.; Kildegaard, H. F. Application of CRISPR/Cas9 Genome Editing to Improve Recombinant Protein Production in CHO Cells. In Methods in Molecular Biology; 2017; pp. 101–118.

Chapter 3 - Minimizing clonal variation during mammalian cell line engineering for improved systems biology data generation

This chapter presents a CHO cell line development platform based on targeted gene integration, which combines the use of CRISPR/Cas9-mediated integration and recombinase-mediated cassette exchange (RMCE). The system was used for the creation of isogenic CHO cell lines producing a set of recombinant proteins that were analyzed by RNA-seq. The results show that cell lines generated by this method have minimal clonal variation, facilitating robust comparative studies of CHO cells.

Reprinted with permission from: Grav, L. M., Sergeeva, D., Lee, J. S., Marin de Mas, I., Lewis, N. E., Andersen, M. R., Nielsen, L. K., Lee, G. M., and Kildegaard, H. F. (2018) "Minimizing Clonal Variation during Mammalian Cell Line Engineering for Improved Systems Biology Data Generation". ACS Synth. Biol. 7, 2148–2159.

ACS Synthetic Biology
Cite This: ACS Synth. Biol. 2018, 7, 2148−2159
Research Article
pubs.acs.org/synthbio

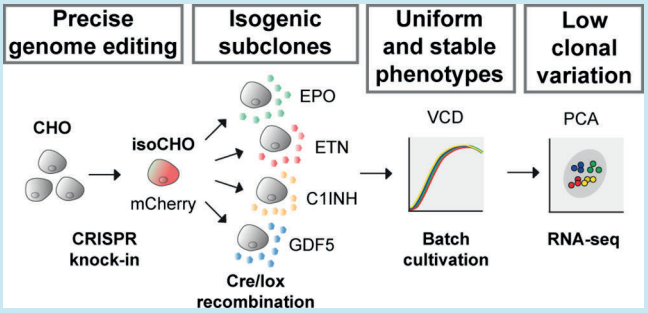
Minimizing Clonal Variation during Mammalian Cell Line Engineering for Improved Systems Biology Data Generation

Lise Marie Grav,[†] Daria Sergeeva,[†] Jae Seong Lee,[†,‡] Igor Marin de Mas,[†] Nathan E. Lewis,[§,∥] Mikael Rørdam Andersen,[⊥] Lars Keld Nielsen,[†,#] Gyun Min Lee,[†,∇] and Helene Faustrup Kildegaard*[,†]

[†]The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark
[‡]Department of Molecular Science and Technology, Ajou University, Suwon 16499, Republic of Korea
[§]Department of Pediatrics, University of California, San Diego, La Jolla, California 92093, United States
[∥]The Novo Nordisk Foundation Center for Biosustatainability, University of California, San Diego, La Jolla, California 92093, United States
[⊥]Department of Biotechnology and Biomedicine, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark
[#]Australian Institute for Bioengineering and Nanotechnology, University of Queensland, St Lucia, QLD 4072, Australia
[∇]Department of Biological Sciences, KAIST, Daejeon 34141, Republic of Korea

Supporting Information

ABSTRACT: Mammalian cells are widely used to express genes for basic biology studies and biopharmaceuticals. Current methods for generation of engineered cell lines introduce high genomic and phenotypic diversity, which hamper studies of gene functions and discovery of novel cellular mechanisms. Here, we minimized clonal variation by integrating a landing pad for recombinase-mediated cassette exchange site-specifically into the genome of CHO cells using CRISPR and generated subclones expressing four different recombinant proteins. The subclones showed low clonal variation with high consistency in growth, transgene transcript levels and global transcriptional response to recombinant protein expression, enabling improved studies of the impact of transgenes on the host transcriptome. Little variation over time in subclone phenotypes and transcriptomes was observed when controlling environmental culture conditions. The platform enables robust comparative studies of genome engineered CHO cell lines and can be applied to other mammalian cells for diverse biological, biomedical and biotechnological applications.



KEYWORDS: *mammalian cells, CRISPR/Cas9, targeted integration, recombinase-mediated cassette exchange, transcriptome, clonal variation*

Stable expression of transgenes in mammalian cells is crucial for basic biology studies of gene functions and for producing proteins for diverse applications including biopharmaceuticals in Chinese hamster ovary (CHO) cell lines.[1] Current mammalian cell line generation platforms are based on random integration of transgenes into the genome, yielding clones exhibiting a wide range of expression, growth and stability characteristics, referred to as clonal variation. This clonal heterogeneity requires time-consuming and labor-intensive screening to find cell lines with the desired performance.[2]

The phenomenon of clonal variation can originate from different sources. It can be partly explained by the plasticity of CHO genomes, which is reflected in frequent chromosomal rearrangements, high mutation rates and genome instability.[3,4] Genomic variation also occurs due to random integration of vectors with transgenes, which can be inserted in multiple copies in different genomic loci. This variation is often explained as the "position effect" and emphasizes the importance of genomic environment surrounding the transgene.[5] Moreover, upon random integration, the transgene cassette can be rearranged. This affects the original vector elements, which may confer unpredictable expression of the transgene.[6] Another source of heterogeneity is nongenetic variation due to epigenetics, stochastic gene expression and changing environmental conditions.[7] Currently, clonal variation is a barrier that hinders our ability to understand the biology of mammalian cell lines, and to engineer more productive CHO cells.

In the development of biopharmaceutical-producing CHO cell lines, it remains unclear how specific transgene impact the host cell. Elucidation of the cell's response to a transgene is crucial for the discovery of potential engineering targets to

Received: March 30, 2018
Published: July 30, 2018

DOI: 10.1021/acssynbio.8b00140
ACS Synth. Biol. 2018, 7, 2148−2159

improve CHO cell protein production performance. Cell line engineering approaches based on transient transgene expression and random chromosomal integration, introduce experimental biases and are not reproducible.[8] Thus, there is a need for a robust cell line generation platform to study the cell's response to specific transgene expression.

Recent advances in mammalian genome engineering technologies provide new opportunities for the development of platforms to assess the impact of transgenes on the cell. Methods for site-specific integration of transgenes allow the generation of cell lines with low genomic variation and predictable and stable transgene expression.[9] The first generation of targeted integration approaches was based on site-specific recombination and recombinase-mediated cassette exchange (RMCE) and has been used for genetic modification in mammalian cells.[10] These methods employ recombinases that recognize specific recombinase-attachment sites and can mediate targeted excision or integration of large transgenic cassettes into mammalian cell lines (up to 100 kb[11]). The second generation of targeted integration tools—programmable endonucleases, such as zinc finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs) and clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated protein 9 (Cas9)—have been successfully used for site-specific integration of transgenes in mammalian cells,[12] particularly in CHO.[9,13] However, when using CRISPR/Cas9 there is a risk of aberrant recombination events around the cut site and off-target effects, while recombinase-based genome editing is highly specific.[14]

Several site-specific recombinases have been described and used for engineering of mammalian genomes. One of the most commonly used recombinases is the P1 bacteriophage-derived Cre that mediates site-specific recombination between direction-sensitive DNA sequences named *loxP*. Alternative systems employ Flp-recombinase with *FRT* sites or Bxb1-recombinase with *att* sites.[10,15] By using a pair of heterotypic sites for a particular recombinase, site-directional RMCE can occur. In RMCE platforms, the master cell line (MCL) contains a marker gene flanked by recombination sites commonly referred to as a "landing pad". The MCL can then be used to insert any genes of interest into the landing pad by exchanging it with the marker gene.

Several studies have applied recombinase-mediated targeted integration in CHO cell lines. The Flp/*FRT* system has been used to generate cell lines expressing either monoclonal antibodies (mAbs)[16−18] and antibody fragments,[19] or other recombinant proteins, such as erythropoietin (EPO),[20] tissue plasminogen activator,[21] secreted alkaline phosphatase,[21] bone morphogenic protein 2[14] and G-protein coupled receptors.[22] Analogous to Flp/*FRT* system, Cre/*lox* recombinase system has been used to generate mAb-expressing CHO cell lines.[23] Initially, RMCE platforms were developed by random integration of the landing pad into the genome. In some cases, primary screening was performed for the selection of parental clones with the highest expression level of the marker (fluorescent protein,[20] surface marker[16] or antibody[17]), in order to preselect transcriptionally active integration site. Site-specific integration of an RMCE landing pad into predefined loci is an evident improvement of the RMCE platform. Targeted integration ensures that the landing pad will not disrupt coding genes or essential genetic elements, which otherwise could have phenotypic consequences. This has previously been done using CRISPR/Cas9-mediated homology-directed targeted integration to create MCLs with Flp/FRT- and Bxb1/*att*-landing pads and used for efficient generation of mAb-expressing cell lines.[11,24]

Use of the RMCE-based site-specific integration platform allows faster cell line generation, compared to traditional random integration approach, and this has already been adopted by industry. It has been employed by Pfizer to speed up the process of mAbs development, allowing early assessment of mAb candidates.[17,25] Interestingly, RMCE-derived clones expressing similar mAbs or antibody fragments can have differences in productivity.[19,25] However, only one study (to our knowledge) went beyond a descriptive analysis of clone performance and tried to reveal the cause of different productivity between clones expressing two similar antibody fragments by proteomic analysis.[26]

In this study, we developed a two-staged approach to generate isogenic CHO-S cell lines (isoCHO) with minimal clonal variation to evaluate potential changes in the transcriptome caused by transgene expression. As a first stage, we created MCLs with stable and homogeneous expression of reporter fluorescent protein using CRISPR/Cas9 integration of a landing pad into a specific genomic site. By directly flanking the reporter gene with lox sites, we constructed an efficient promoter/poly(A) trap to only permit transgene expression from the targeted genomic site. This allowed us to avoid the subsequent use of antibiotic selection pressure, keeping the culture conditions unchanged. As a second stage, we used RMCE and fluorescence-activated cell sorting (FACS) to generate isogenic stable producer cell lines, expressing four recombinant proteins: etanercept (ETN), EPO, growth/differentiation factor 5 (GDF5) and C1 esterase inhibitor (C1INH). The sets of isogenic producer cell lines, which are genetically identical aside from the expressed gene of interest (GOI), showed highly comparable phenotypes and transcriptomes, even during long-term cultivation. However, large transcriptional variation was observed between isogenic cell lines derived from two different MCLs (with the landing pad integrated in the same genomic site), underlining the importance of minimizing clonal variation when conducting comparative studies.

RESULTS

Stable MCLs Generated Using CRISPR/Cas9 and a Preselected Target Locus. To facilitate the construction of stable recombinant CHO cell lines in a more controlled manner, we established a two-stage approach using CRISPR/Cas9-mediated targeted integration followed by RMCE. For the first stage of targeted integration, we selected a noncoding genomic region located between an essential gene and a highly expressed gene as our target locus, to avoid disruption of gene-encoding regions, loss of transgenes and transcriptional silencing. Based on our previously published method for CRISPR/Cas9-mediated homology-directed targeted integration,[9] we inserted the RMCE-"landing pad" into the preselected target locus, creating a MCL for the second stage of cell line generation (RMCE with GOIs). To ensure exchange of the GOI and expression exclusively from our target locus upon RMCE, we introduced *loxP/lox2272* sites[27] flanking the mCherry reporter gene (Figure 1a). The EF-1α promoter (EP) driving the mCherry expression and the poly(A) signal are consequently located outside *lox* sites, creating a promoter/poly(A) trap for the incoming promoterless GOI upon cotransfection with Cre recombinase (Figure 1b). This ensures that expression of the GOI only occurs upon exchange with mCherry, while avoiding the use of antibiotic selection in order to keep the culture

DOI: 10.1021/acssynbio.8b00140
ACS Synth. Biol. 2018, 7, 2148−2159

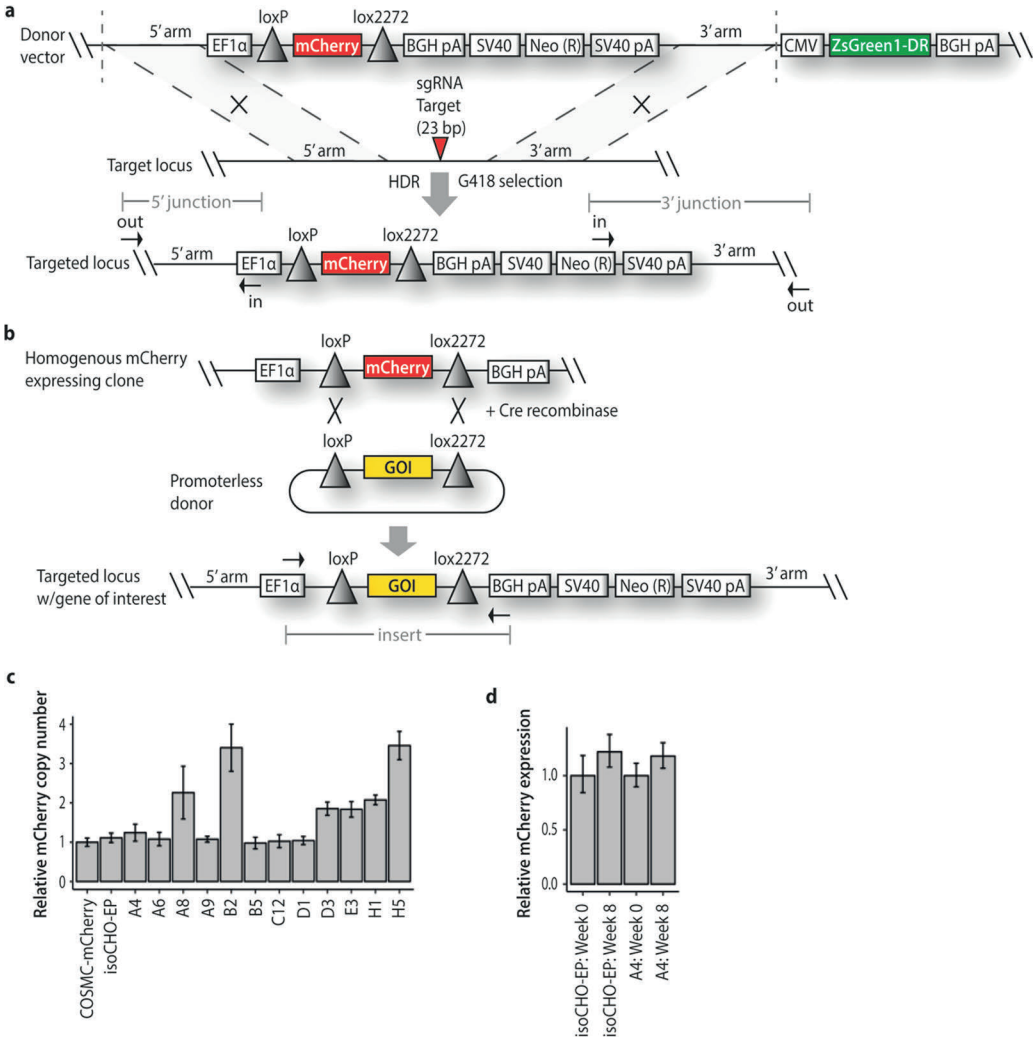


Figure 1. Targeted integration of the landing pad into the preselected locus using CRISPR/Cas9 and RMCE strategy. (a) Schematic representation of the targeting strategy of the landing pad into the preselected target locus. The donor plasmid contains the landing pad with *loxP* (34bp) and *lox2272* (34bp) sequences flanking mCherry gene. EF-1α promoter and BGHpA are located outside the *lox* sequences allowing for both promoter and poly(A) trapping upon recombinase-mediated cassette exchange (RMCE). A gene encoding the short-lived green fluorescent protein ZsGreen1-DR was placed outside the homology arms on the donor plasmid to exclude random integration events. Primer positions for 5′/3′ junction PCR are denoted. (b) Schematic of RMCE strategy: cotransfection of promoterless RMCE donor + Cre recombinase ensures the incoming gene of interest (GOI) will only be expressed when exchanged with mCherry in the genomic landing pad. Primer positions for insert PCR are denoted. (c) Relative copy number of mCherry regions in clonal cells. The plot shows the relative copy number of mCherry in all targeted cell lines verified by junction PCR, in comparison to reference sample COSMC-mCherry from Lee *et al.*[9] The error bars represent the standard deviations of technical replicates (n = 3). (d) Stability of mCherry gene expression tested over an eight-week period measured by qRT-PCR, for two selected one-copy clonal cell lines. The error bars represent the standard deviations of technical replicates (n = 3).

conditions unchanged. After CRISPR/Cas9-mediated targeted integration of the landing pad, 13 targeted integrants were selected upon FACS with enrichment for mCherry-positive (red fluorescent)/ZsGreen1-DR-negative (green fluorescent) cells followed by junction PCR screening. The junction PCR primer binding sites are annotated in Figure 1a. Among the targeted integrants, seven clones contained one copy of the mCherry expression cassette (Figure 1c). Out of two further characterized clones, showing highly stable mCherry expression after an eight-week cultivation period (Figure 1d), one clone (isoCHO-EP) was selected as the MCL based on homogeneous and high levels of mCherry expression (Supporting Figure S1 and Table S1).

Isogenic Subclones Expressing Recombinant Proteins Show Similar Growth and Low Transcriptional Variation. To generate a set of isogenic subclones from the MCL, we cloned four promoterless vectors for RMCE with GOIs encoding recombinant proteins ETN, EPO, GDF5, or C1INH. After cotransfections of RMCE donor vector and Cre-recombinase into isoCHO-EP, we single-cell sorted mCherry-negative cells to enrich for cells that have exchanged mCherry with the incoming GOI, using isoCHO-EP as gating control. The percentage of mCherry negative clones using this setup was in the range of 2−6% 7 days after the cotransfection. For each GOI exchange, we verified the GOI exchange in the single-cell sorted subclones using insert PCR with primers aligning outside the *lox* sequences and sequenced the products (Figure 1b). Around 85−95% of the clones were positively verified upon insert PCR depending on the GOI; we selected 12 PCR positive clones (three clones for each GOI) for further analysis. To study the growth behavior of these subclones, batch cultures in shake flasks were performed. All 12 subclones

DOI: 10.1021/acssynbio.8b00140
ACS Synth. Biol. 2018, 7, 2148−2159

derived from isoCHO-EP displayed highly similar growth phenotypes (Figure 2a).

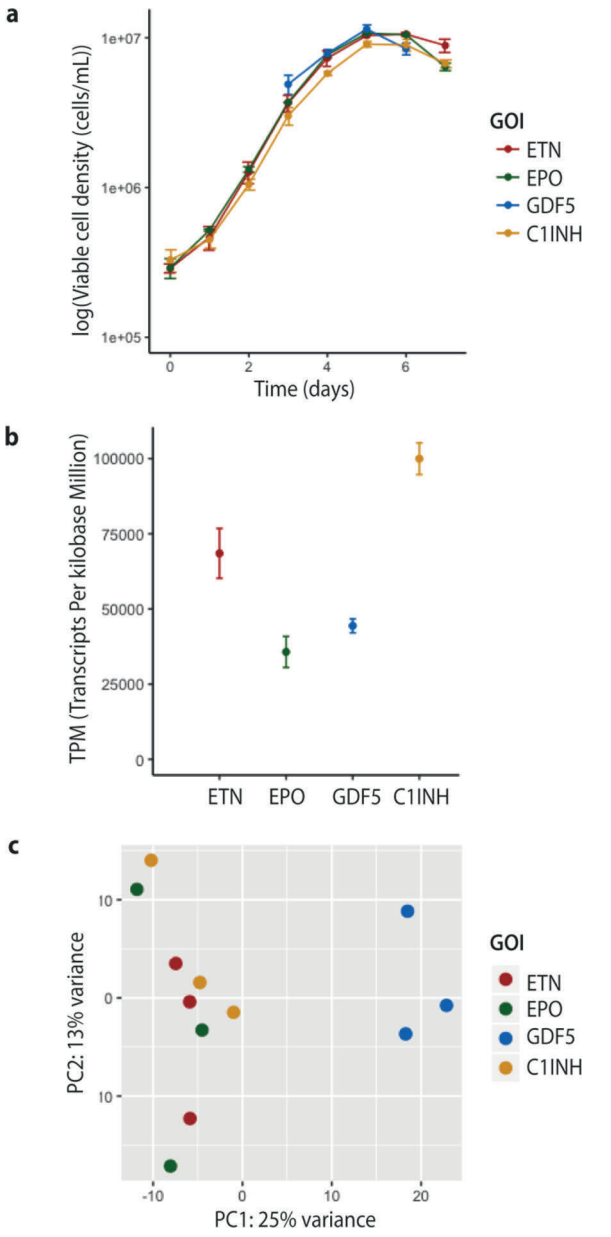


Figure 2. Isogenic subclones generated by RMCE show low variation in phenotypes and transcriptomes. (a) Viable cell densities (VCD) of isoCHO-EP subclones expressing etanercept (ETN), erythropoietin (EPO), growth differentiation factor 5 (GDF5) or C1 esterase inhibitor (C1INH). The error bars of each line represent the standard deviations of three isogenic subclones expressing the same gene of interest (GOI) (n = 3). (b) Relative levels of transgene expression, as measured in transcripts per kilobase million (TPM) of ETN, EPO, GDF5 or C1INH producing subclones sampled at day 4 in late exponential phase. The error bars represent the standard deviations of three isogenic clones expressing the same GOI (n = 3). (c) Principal component analysis (PCA) of whole transcriptome data from isoCHO-EP subclones expressing ETN, EPO, GDF5 or C1INH sampled at day 4 in late exponential phase.

On day four in late exponential phase, cells were harvested for RNA-seq to capture the transcriptome for further analysis of the variations between all subclones expressing different recombinant proteins (data set 1). On the basis of the transcriptomics data, highly comparable relative transgene expression levels, in terms of transcript per kilobase million (TPM) values, were observed within biological replicates (three clones expressing the same GOI) (Figure 2b). TPM is a value for relative transcript abundances in RNA-seq data, and the transgenes were consistently among the top ten most expressed genes within the isogenic cell lines (alongside endogenous genes such as Eef1a1, Actb, Hsp90b1 and Gapdh). The highly comparable relative transgene expression levels were further confirmed by qRT-PCR (Supporting Figure S2). The reproducibility among the biological replicates was evaluated following guidelines for technical replicates.[28] There was high Spearman correlation R^2 > 0.90 between the clones (Supporting Figure S3). To assess the variation in global gene expression between subclone sets expressing different GOIs, principal component analysis (PCA) of the transcriptomics data was performed. Transgene expression was excluded from the PCA to avoid potential biased subclone separation due to their different transgene expression. Little variation in the gene expression profiles of the isoCHO-EP subclones was observed: the first principal component (PC) accounts for only 25% of the variation and the second PC accounts for 13% of the variation. The variation in PC1 is clearly explained by GDF5 expressing subclones, as they provide the separation according to PC1 (Figure 2c). To further assess the variation in all genes expressed in the subclones, differential gene expression (DGE) analysis was performed (data set 1, Supporting Figure S4a). Only 160 out of 12,647 tested genes (1.27%) were significantly differentially expressed (DE) (Supporting Data S1), indicating highly similar gene expression profiles of all 12 clones. From these data, we can conclude that our two-stage isogenic cell line generation platform facilitated efficient generation of highly reproducible subclones with very low variation in phenotypes and transcriptomes.

Isogenic Subclones Show High Stability in Growth and Transcriptome during Three Months of Cultivation. Generation of clones by the traditional approach of random integration of the GOI into the genome of cells often results in unstable clones that show loss of expression over time. Therefore, we were interested in studying how the generated clones behaved in long-term culture. For the analysis, six isoCHO-EP subclones expressing ETN and C1INH were cultivated up to three months without antibiotic selection pressure by passaging three times a week, with a cryopreservation at each passage. Upon completion, the cryopreserved cultures from the beginning of the maintenance culture (0 months), midculture (1.5 months) and end of culture (3 months) were thawed and cultivated in batch cultures simultaneously. One biological replicate of both ETN and C1INH was lost after 1.5 months and were therefore not included in the batch. The viable cell densities (VCDs) were analyzed revealing almost no variation in growth irrespective of product or time of harvest indicating very stable and similar growth of the subclones (Figure 3a). From the batch cultures, cells were harvested in late exponential phase for RNA-seq analysis (data set 2). TPM values from the RNA-seq analysis revealed low variation in relative transgene expression levels over the time in culture, indicating stable expression of GOIs from the integration site (Figure 3b). PCA of the transcriptomics data (not including transgenes) revealed little variation in subclone populations over time with PC1 explaining only 23% of the variation and PC2 16% of the variation. Clustering was observed for each biological replicate, indicating that the largest variation is between individual

DOI: 10.1021/acssynbio.8b00140
ACS Synth. Biol. 2018, 7, 2148−2159

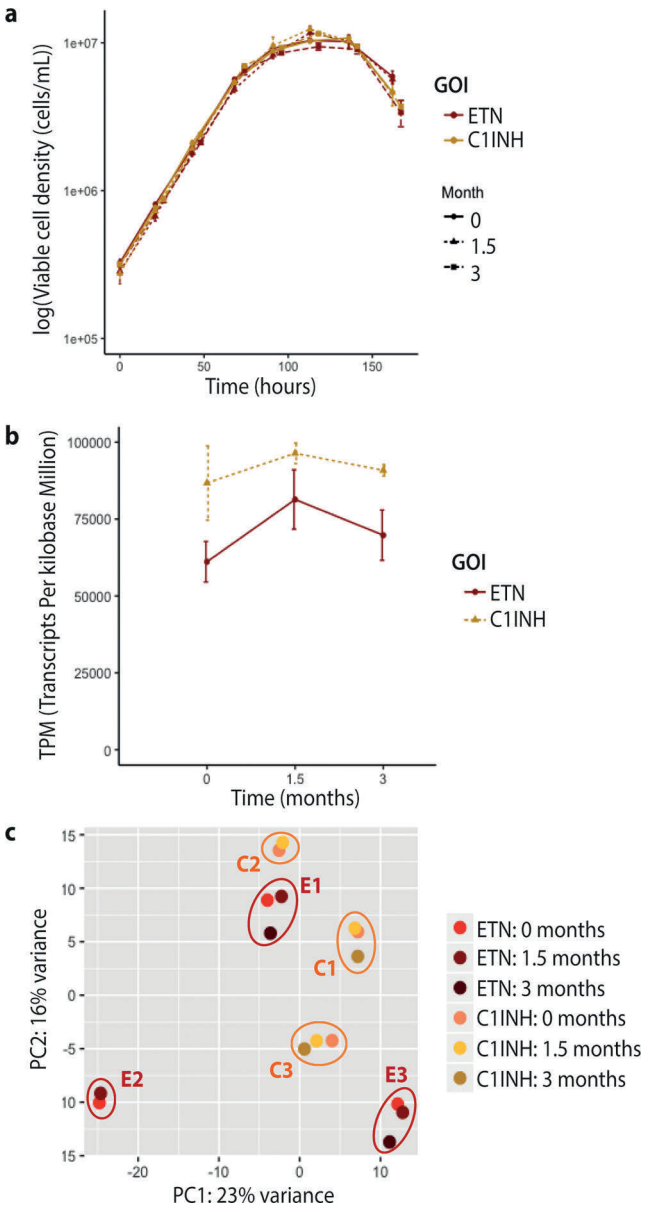


Figure 3. Long-term stability of gene expression in isogenic subclones. (a) Viable cell densities (VCD) of isoCHO-EP subclones expressing etanercept (ETN) and C1 esterase inhibitor (C1INH) cultivated upon thawing of cells harvested after 0, 1.5, and 3 months in culture. The error bars represent the standard deviations of three (0 and 1.5 months) or two (3 months) subclones expressing the same gene of interest (GOI) (n = 2−3). (b) Levels of transgene expression over time of isoCHO-EP ETN and C1INH expressing clones cultivated upon thawing of cells harvested after 0, 1.5, and 3 months in culture as measured in transcripts per kilobase million (TPM). The error bars represent the standard deviations of three (0 and 1.5 months) or two (3 months) subclones expressing the same GOI (n = 2−3). (c) Principal component analysis (PCA) of whole transcriptome data from isoCHO-EP subclones expressing ETN or C1INH cultivated upon thawing of cells harvested after 0, 1.5, and 3 months in culture. Clustering is observed within each biological replicate of ETN producing subclones (E1, E2, E3) and C1INH producing subclones (C1, C2, C3), and no clustering observed due to time in culture.

subclones (Figure 3c). The gene expression of biological replicates was deemed as reproducible, with Spearman correlation R^2 > 0.90 (Supporting Figure S5). High Spearman correlation (R^2 > 0.90) was also observed between biological replicates measured at different time points (Supporting Figure S6). From the DGE analysis of subclones producing the same recombinant protein (Supporting Figure S4b), only 10 out of 12 937 tested genes (0.08%) were statistically significant DE (adj$P$-value <0.05) over the three months period of passaging (Supporting Data S1). This indicates very low changes in gene expression levels over long-term cultivation. Moreover, only 4 out of 12,937 tested genes (0.03%) were detected as DE (Supporting Data S1) when comparing the ETN cultures at 0, 1.5, and 3 months *versus* the C1INH cultures at the same time points (data set 2, Supporting Figure S4c), indicating a low or similar effect of ETN and C1INH expression on the transcriptome in these cells. Together, the data indicate that the generated clones behave very similarly and show nearly no transcriptomic changes over time.

EPO Expression Levels Were Doubled When Expressed from a Composite Promoter. To facilitate analysis of the direct impact of different transgenes on the transcriptome, we aimed to generate a MCL with higher basal expression levels of mCherry/GOIs, potentially imposing a larger burden on the cells by increased expression of recombinant proteins. To achieve this, we replaced the EF-1α promoter (EP) with a composite promoter that shows the highest GFP expression in CHO cells upon transient transfection out of a panel of ten composite promoters.[29] We created a new donor vector for targeted integration of the landing pad, where the EP was replaced with a composite promoter (CP, consisting of mouse CMV enhancer, minimal human EF-1α promoter and human T cell leukemia virus 1 untranslated region). After CRISPR/Cas9-mediated targeted integration of the landing pad, 11 targeted integrants were selected upon FACS and junction PCR screening. Among the verified targeted integrants, nine clones contained one copy of the mCherry expression cassette (Figure 4a). Of two selected clones, both showed stable mCherry expression after an eight-week long cultivation period (Figure 4b). One clone (isoCHO-CP) was selected as a new MCL based on homogeneous mCherry expression (Supporting Table S2), and it showed about 1.9 times higher mCherry expression level than isoCHO-EP (Supporting Figure S7).

Similar to isoCHO-EP, isoCHO-CP was used to generate subclones expressing four recombinant proteins by RMCE. To facilitate comparisons with producing subclones, nonproducing subclones of both isoCHO-EP and isoCHO-CP were generated by low-frequency recombineering of *loxP* and *lox2272* sites in the absence of a RMCE donor vector. For each recombination event, subclones were verified by insert PCR and sequencing of inserted GOI. Fifteen clones (three clones expressing each GOIs and three nonproducer clones) derived from isoCHO-CP were selected for further analysis. To study the growth behavior of these subclones, batch cultures were performed. Of the isoCHO-CP subclones, the three EPO and three GDF5 subclones grow comparably to nonproducing isoCHO-CP subclones. Interestingly, C1INH and ETN expressing clones grow to slightly lower maximum VCDs (Supporting Figure S8a). On day four in late exponential phase, cells were harvested for RNA-seq to analyze the variation in transcriptome of all 15 subclones (data set 3). From the RNA-seq data, highly comparable relative expression levels were observed within ETN and GDF5 subclones, while larger error bars were observed within the EPO and C1INH expression subclones (Supporting Figure S8b). However,

DOI: 10.1021/acssynbio.8b00140
ACS Synth. Biol. 2018, 7, 2148−2159

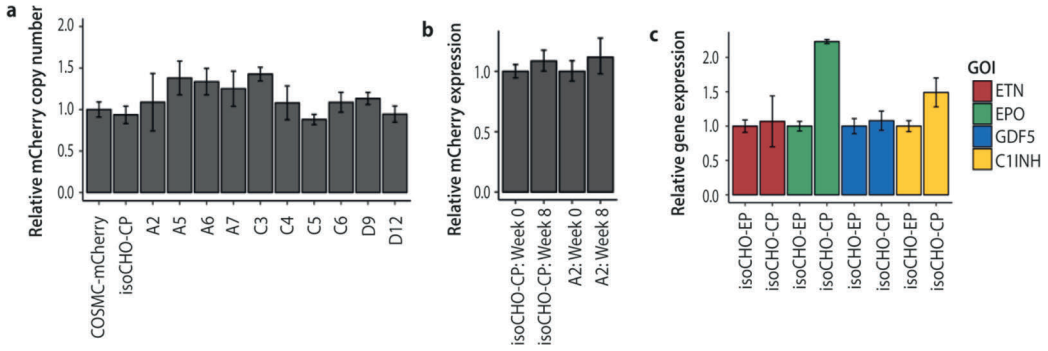


Figure 4. Generation of clones with a composite promoter. (a) Relative copy number of mCherry transgene in clonal cells. The plot shows the relative copy number of mCherry in of all targeted cell lines verified by junction PCR, in comparison to reference sample COSMC-mCherry from Lee *et al.*[9] The error bars represent the standard deviations of technical replicates (n = 3). (b) Stability of mCherry gene expression tested over an eight-week period measured by qRT-PCR. The error bars represent the standard deviations of technical replicates (n = 3). (c) Relative levels of transgene expression in isoCHO-EP and isoCHO-CP-derived subclones, measured by qRT-PCR, normalized to the mean value of isoCHO-EP subclones. The error bars represent the standard deviations of three isogenic clones expressing the same GOI (n = 3).

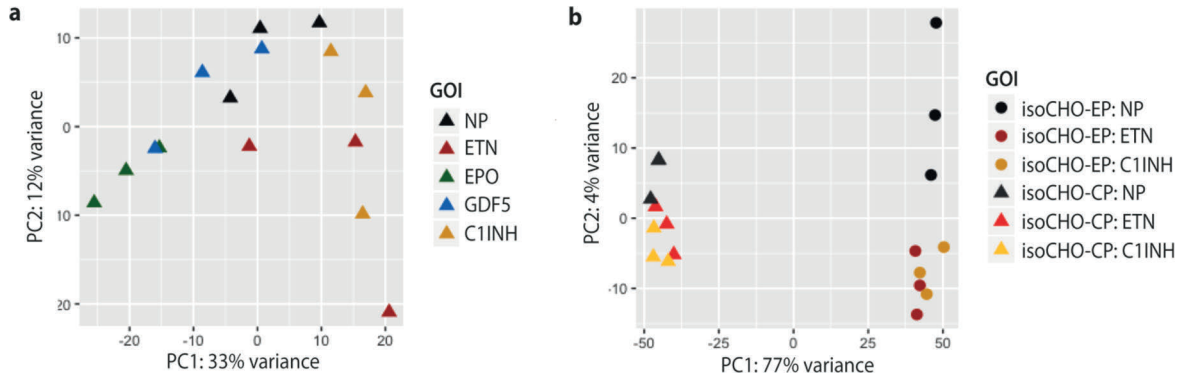


Figure 5. Variation within subclone populations. (a) Principal component analysis (PCA) of whole transcriptome data from isoCHO-CP subclones expressing etanercept (ETN), erythropoietin (EPO), growth differentiation factor 5 (GDF5) or C1 esterase inhibitor (C1INH) and nonproducer (NP) subclones sampled at day 4 in late exponential phase. (b) PCA of whole transcriptome data from isoCHO-EP and isoCHO-CP subclones expressing ETN or C1INH and nonproducer subclones sampled at day 4 in late exponential phase.

qRT-PCR showed lower variation in gene expression among the EPO and C1INH clones (Supporting Figure S9). Unexpectedly, only EPO expression levels were increased by 2.2-fold in isoCHO-CP clones compared to isoCHO-EP clones (Figure 4c). The expression level of ETN, GDF5 and C1INH were similar in both isoCHO-EP and isoCHO-CP clones.

Increase in EPO Expression Leads to Large Transcriptional Changes in Isogenic Cell Lines. To study the variation in global gene expression between isoCHO-CP subclones expressing ETN, EPO, GDF5 or C1INH, PCA of the transcriptomics data from the 15 isoCHO-CP subclones was performed (data set 3). PCA of gene expression data from the isoCHO-CP subclones, not including transgenes, displayed slightly higher variation between subclones compared to isoCHO-EP subclones, resulting in PC1 explaining 33% of the variation and PC2 explaining 12% of the variation (Figure 5a). The gene expression of biological replicates was compared and deemed to be reproducible with Spearman $R^2 > 0.90$ (Supporting Figure S10). A large number of statistically significant (adj$P$-value < 0.05) DE genes were observed when comparing all five groups of subclones (Supporting Figure S4d), with 4,622 DE genes out of 12,967 tested genes (35.6%) (Supporting Data S1). With a log fold change cutoff of 1.5, the number of statistically significant DE genes is reduced to 229 (5%) (Supporting Data S1). To check if the increase in DE genes between isoCHO-CP subclones could be due to the increase in EPO expression, we compared the gene expression of all subclones excluding EPO producers (Supporting Figure S4e). This yielded a much lower number of statistically significant (adj$P$-value < 0.05) DE genes: 137 out of a total of 12,967 tested genes (1.1%) (Supporting Data S1). To analyze if the DE genes identified when including EPO producers in the comparison were related to the increased transgene expression, the DE genes with human homologues were subjected to enrichment analysis of canonical pathways. The results showed that DE genes are most enriched in mitochondrial dysfunction, oxidative phosphorylation, and EIF2 signaling canonical pathways (Supporting Figure S11). Analysis of nonredundant gene interaction networks showed that the most represented cellular functions in our set of DE genes are molecular transport, post-translational modification, protein synthesis, lipid metabolism and cell signaling.

Large Transcriptional Variation Was Observed between Subclones Derived from Different MCLs. To assess clonal variation between the two MCLs, we compared transcriptomic profiles of the subclones expressing ETN and C1INH derived from both isoCHO-EP and isoCHO-CP together with their respective nonproducing clones (data set 4). The PCA showed a clear separation between subclones based on which MCL they originate from. They are separated by the PC1 that accounts for 77% of the variation (Figure 5b). PC2 only accounts for 4% of the distance between subclones

DOI: 10.1021/acssynbio.8b00140
ACS Synth. Biol. 2018, 7, 2148−2159

derived from the same MCL. When comparing subclones from the different MCLs (Supporting Figure S4f), 6,830 out of a total of 13,031 genes (52.4%), were identified as DE (Supporting Data S1). With a log fold change cutoff of 1.5, the DE genes are reduced to 954 (7.3%) (Supporting Data S1). The DGE analysis indicates a high level of clonal variation between the two MCLs, even when the same locus in the CHO-S genome was used for targeting when generating both MCLs.

DISCUSSION

Traditionally, recombinant protein producing mammalian cell lines are generated by random integration of transgenes, believed to give rise to large uncontrollable variations in growth and transgene expression levels. Upon random integration a high diversity of clones is created, which can be beneficial in the search for high producing clones.[4] However, clonal variation complicates transcriptomics data analyses due to difficulties in deciphering if transcriptional changes originate from transgene expression or clonal variation.[30] Here we present a platform for generating isogenic cell lines using CRISPR/Cas9-mediated targeted integration combined with RMCE for improved transcriptomic analyses by minimizing clonal variation interference.

With CRISPR/Cas9-mediated targeted integration of a landing pad for subsequent recombineering of transgenes, we have established a two-stage approach for generating isogenic transgene expressing clones. CRISPR/Cas9 allows flexible selection of target sites for landing pad integration in the first step of the approach. Our selected target site close to essential genes exhibited both stable mCherry expression and stable expression of GOIs during a 3-month cultivation period. A similar strategy for the selection of target sites close to essential genetic elements has been applied in yeast.[31] Our preliminary finding might suggest that this strategy will be a successful criterion for selection of target sites in other mammalian cells as well. The second step of our approach allows fast and reproducible generation of cell lines originating from the same MCL that only differ in the transgene coding sequence, by enriching for non-mCherry expressing clones with FACS upon RMCE.

Combination of CRISPR/Cas9-mediated targeted integration with RMCE has also been proposed by Inniss *et al.*, which integrated a Bxb1/*att* and Flp/*FRT*-landing pad into the Fer1L4 site of CHO-K1 genome.[24] They recombined GFP and Thymidine kinase genes from the landing pad with incoming promoterless blasticidin resistance gene and genes encoding IgG heavy chain (HC) and light chain (LC) driven by CMV promoters. In their system, a selection with blasticidin and ganciclovir was required upon RMCE. Selection pressure applied after RMCE exchange can lead to some perturbations in the cell,[4] which could explain the variable levels of expression in RMCE-derived clones reported in their study. The design of the RMCE promoter/poly(A) trap without resistance gene in our study hinders the expression of GOI from possible random integration, allowing the generation of isogenic clones with highly comparable phenotypes while avoiding the use of additional antibiotic selection step. Due to the efficient generation of recombineered mCherry negative clones and selection with FACS and high stability of expression in the established system, we propose that removal of neoR gene during RMCE could be feasible by moving the *lox2272* site downstream of the neoR gene before its poly(A) tail. Removal of the NeoR marker during RMCE could potentially release more transcriptional and translational capacity for growth and protein production in the host cell line as observed previously.[32] In addition, this could facilitate reuse of neoR selection marker for efficient future CRISPR/Cas9-mediated targeted integration events.

With our two-staged approach, the generated subclones expressing ETN, EPO, GDF5 and C1INH showed very similar growth profiles between biological replicates as well as between different transgene expressing clones (Figure 2a and Supporting Figure S8a). Although corresponding recombinant proteins differ in size and post-translational modifications, their expression in the MCLs did not seem to affect the growth, even when the expression level of EPO was increased. It is possible that the respective level of transgene expression was not high enough to repress the growth as growth repression is typically regarded as a trait of high-producing CHO cell lines.[33] The similar growth performance might also be due to a mild impact of RMCE on the MCLs phenotype, the subcloning of CHO-S cells during the first CRISPR-mediated targeted integration step or the selection of confluent clones upon RMCE. CRISPR-mediated targeted integration of mCherry into COSMC site in the same CHO-S cells as used in this study generated clones with different growth profiles while showing similar mCherry expression levels.[9] This could indicate that the CHO-S cells used as a starting point in both studies contain a high variation of growth among the cells or that the CRISPR-mediated genome editing in combination with antibiotic selection introduced the growth variation observed in Lee *et al.*, 2015.

Major challenges in transcriptomics studies arise from the genome plasticity of immortalized mammalian cells, and the relatively easy adaptation of the cells to environmental changes. It was therefore surprising to us that the isoCHO-EP clones expressing ETN and C1INH showed such high stability in their transcriptome over three months in culture. This could be explained by the constant culture conditions and stability of transgene expression. There was no indication of drift in the transcriptome, as would be expected due to the large variation observed in the genome over long-term culture,[34] which additionally makes the platform highly attractive for general stable expression of transgenes in CHO.

The recent advances in obtaining large-scale data sets for CHO cells including transcriptomics allow researchers to improve their understanding of the basic biology of CHO cells underlying recombinant protein production. To reveal the impact of expressed transgenes on the global transcriptomic levels, DE gene analysis was performed. The transcriptomic data showed a low variation of the transcriptome among the subclones originating from the same MCL (isoCHO-EP or isoCHO-CP) that display similar transgene expression levels. Especially the isoCHO-EP-derived clones showed few DE genes (about 1%). We expect that the low variation obtained is due to the two-stage approach applied to generate the isogenic clones, which resulted in similar transgene expression levels among clones and similar growth profiles. Especially the low variation in growth facilitate easier transcriptomics comparisons as subclones might be experiencing the same nutrient depletion and thereby the same environmental impact on their transcriptome over time in culture. For future studies, it would be interesting to analyze if the clones indeed show similar uptake of nutrients and excretion of byproducts.

The transcriptomic analysis of C1INH- and ETN-producing clones together with nonproducing clones originating from

DOI: 10.1021/acssynbio.8b00140
ACS Synth. Biol. 2018, 7, 2148−2159

isoCHO-EP and isoCHO-CP showed low variation among the subclones originating from both isoCHO-EP and isoCHO-CP, indicating a low or similar effect of ETN and C1INH expression on the transcriptome in these cells. On the contrary, large variation was observed between the isogenic cell lines originating from different MCLs, even though they had similar growth profiles and transgene expression level (or no transgene expression) (Figure 5b). In these MCLs the landing pad is integrated into the exact same genomic site, and they both have EF-1α based promoters that only differ slightly in composition. The large transcriptomic variation observed challenges the view that mammalian cell lines are isogenic as long as transgenes are integrated in the same genomic site.[35] It further indicates that even two MCLs (only slightly differing in promoter composition) with similar growth profiles can have quite different transcriptomic profiles, as subclones derived from the different MCLs display a large amount of DE genes (52.8%). A similar observation was reported in Orellana *et al.*[30] in cell lines with a randomly integrated transgene, where two cell lines originating from the same cell pool had 58% of DE genes despite only modest differences in growth rate. With this high degree of variation, it is challenging to reliably use transcriptomics data to identify potential effect of transgenes without minimizing clonal variation, as performed in the present study. Most likely, the transcriptional variation observed between the two different MCLs in our study can be attributed to the inherent heterogeneity of the parental CHO cell population, caused by genetic and epigenetic changes.[4] In this case, any clones derived from the CHO parental population will have quite different transcriptomic profiles, even without any form of gene editing. The divergence of MCLs might also be affected by CRISPR/Cas9 editing and the antibiotic selection process itself, which can potentially introduce evolutionary pressure, forcing cells to perturb biological processes in order to adapt to the changed conditions. Thus, any MCL, regardless of target site and genetic elements used, will have its unique transcriptomic signature, corresponding to its individual adaptation pathway. In our study, we show that the global transcriptome can be isogenic within subpopulation (subclones of MCL) when we keep the conditions (including passage number) and clone generation timeline unchanged. Hence, we conclude that it is possible to reduce the "intrinsic" clonal variation of CHO cells by controlling the cell line generation process.

The higher level of EPO expression in isoCHO-CP-derived clones allowed us to further investigate the transcriptional impact of EPO expression in the subclones. Increase in EPO expression was coupled to large transcriptional changes, especially in mitochondrial dysfunction, oxidative phosphorylation, and EIF2 signaling pathways. Mitochondrial dysfunction and oxidative phosphorylation are highly interconnected and are both shown to be up-regulated in EPO producing clones, indicating a transcriptional change to meet the higher energy demands for recombinant protein synthesis—consistent with a previous study by Yusufi *et al.*[6] EIF2 signaling is known to be involved in global translational control and has been shown to regulate translation and protein production in CHO cells.[36] The most represented cellular functions in our set of DE genes were molecular transport, post-translational modification, protein synthesis, lipid metabolism and cell signaling. These are all cellular functions expected to change when recombinant protein production is increased, indicating that the increase in DE genes could be explained by the increased expression of EPO. Extended analysis of the gene networks and their major regulators could lead to the discovery of interesting engineering targets for improving protein expression levels.

In summary, the presented cell line generation approach minimizes clonal variation in CHO cells, making it optimal for comparative studies such as DGE analysis. Specifically, the combination of the stable integration site, ensuring transgene expression from the landing pad only, and avoiding additional selection pressure gives advantages over previous cell line generation platforms. The developed approach can be applied in numerous studies ranging from multiomics studies, investigating the effect of media supplementations or the impact of specific genes or genetic elements. The platform can be further advanced by inserting the landing pad into multiple sites,[11] combining it with inducible promoters or promoters of different strengths, in order to study multiple genes in coherence to each other. Furthermore, the platform can be applied for studying the effects of transgene expression in other mammalian systems including human and mouse cells where system biology data analyses are also being challenged by clonal variation.[37]

METHODS

Plasmids and Plasmid Construction. GFP_2A_Cas9 and sgRNA vectors were constructed as described previously.[38] sgRNA target sequence was designed by manual bioinformatics analysis after selection of target loci, the sequence is listed in Supporting Table S5. Donor plasmids with RMCE landing pad and promoterless RMCE vectors were constructed *via* uracil-specific excision reagent (USER) cloning method. The donor plasmids harboring the RMCE landing pad was made of seven PCR amplified DNA parts, including 5′ and 3′ homology arms, the promoter (EF-1α (EP) or composite promoter (CP)), mCherry coding sequence with *loxP* sequence at 5′ end and *lox2272* sequence at 3′ end and BGHpA, NeoR expression cassette (pSV40-NeoR-SV40pA), ZsGreen1-DR expression cassette (pCMV-NeoR-BGHpA) and backbone. CHO-S genomic DNA was used as PCR template for the amplification of the 750 bp long homology arm sequences matching the noncoding target locus, the sequences are listed in Supporting Table S4. Plasmids used as PCR templates for EF-1α sequence, backbone, NeoR and ZsGreen1-DR expression cassettes have been described previously,[9] *lox*-sequences were synthesized as a part of primers. Plasmid lox-mCherryOri with loxP-mCherry-lox2272-BGHpA DNA part was constructed in-house by USER cloning, the template used for mCherry amplification have previously been described.[9] CP was amplified from pDRIVE5-GFP-2 plasmid included in Invivogen's PromTest kit (version # 13F06-MM). Promoterless RMCE vectors were made of two parts, the backbone and GOI (ETN, EPO, GDF5, C1INH) with *loxP* sequence at 5′ end and *lox2272* sequence at 3′ end. Plasmids used as PCR templates for ETN, EPO and C1INH have previously been described.[39−41] The plasmid with the coding sequence for human GDF5 was ordered *via* GeneArt and codon optimized for CHO, coding sequence is listed in Supporting Table S3. Primers are listed in Supporting Table S5. Assembled PCR fragments were transformed into *E. coli* Mach1 competent cells (Life Technologies). All constructs were verified by sequencing and purified using NucleoBond Xtra Midi EF (Macherey-Nagel) according to manufacturer's instructions. For Cre recombinase expression, Sigma-Aldrich PSF-CMV-CRE recombinase expression vector was used directly (OGS591).

DOI: 10.1021/acssynbio.8b00140
ACS Synth. Biol. 2018, 7, 2148−2159

Generation of RMCE MCLs Using CRISPR/Cas9. CHO-S cells (Thermo Fisher Scientific) were maintained in CD CHO medium supplemented with 8 mM L-Glutamine (Thermo Fisher Scientific) and cultivated in 125 mL Erlenmeyer shake flasks (Corning Inc., Acton, MA), incubated at 37 °C, 5% CO_2 at 120 rpm and passaged every 2−3 days. Cell growth and viability were monitored using the NucleoCounter NC-200 Cell Counter (ChemoMetec). Cells at a concentration of 1 × 10^6 cells/mL were transfected with donor plasmid and vectors encoding GFP_2A_Cas9 and sgRNA targeting the integration site at a ratio of 1:1:1 (w:w:w) in 6-well plates (BD Biosciences) using FreeStyle MAX transfection reagent (Thermo Fisher Scientific) according to the manufacturer's recommendations. Stable cell pools were generated by seeding cells in CELLSTAR 6 well Advanced TC plates (Greiner Bio-one) on day 3 after transfection in medium containing G418 (500 μg/mL; Sigma-Aldrich). During selection with G418, the medium was exchanged every 3−4 days. After 2 weeks of selection, cells were detached with TrypLE (Life Technologies) and adapted to grow in suspension in untreated plates or Erlenmeyer flask depending on cell concentrations. For clonal selection, cell pools were subjected to single cell sorting using a BD FACSJazz cell sorter (BD Biosciences), and GFP negative/mCherry positive clones were isolated. Single cells were seeded in flat-bottom Corning 384-well plates (Sigma-Aldrich) in 30 μL of CD CHO medium, supplemented with 8 mM L-Glutamine, 1.5% HEPES (Gibco) and 1× Antibiotic-Antimycotic (Gibco). Fourteen days after single-cell sorting, the entire volume of subconfluent clones was transferred to 180 μL of CD CHO medium supplemented with 1× Antibiotic-Antimycotic (Gibco) in flat-bottom 96-well plates using an epMotion 5070 liquid handling workstation (Eppendorf). Subsequently, cells were expanded in suspension and verified by fluorescent level analysis, junction PCR and qRT-PCR.

Fluorescent Level Analysis. The generated mCherry expressing colonies were analyzed by a Celigo Imaging Cell Cytometer (Nexcelom Bioscience) applying the mask (blue fluorescent channel, Hoechst-based staining of live cells) + target1 (red fluorescence channel, mCherry signal). Master mix (200 μL), containing CD CHO + 8 mM L-glutamine + 5 μg/mL Hoechst-33342 (Life Technologies), and cell suspension (3 μL) were mixed in a 96-well optical-bottom microplate (Greiner Bio-One), and cells were incubated for 40 min at room temperature. Colonies, homogeneous in mCherry expression, were selected as having ≥90% of mCherry positive cells.

PCR Amplification of Target Regions. For junction and insert PCR genomic DNA was extracted from the cell pellets using QuickExtract DNA extraction solution (Epicenter, Illumina) according to the manufacturer's instructions. 1−2 μL of genomic DNA mixture was used as PCR template. 5′/3′ junction PCR and insert PCR was carried out using 2× Phusion Master Mix (Thermo Fisher Scientific) in touchdown PCR (98 °C for 30 s; 98 °C for 10 s; 66−56 °C (insert PCR) or 68−58 °C (junction PCR) [−1 °C/cycle] for 30 s; 72 °C for 2 min; 30×: 98 °C for 10 s; 56 °C (insert PCR) or 58 °C (junction PCR) for 30 s; 72 °C for 2 min; 72 °C for 10 min). PCR primers for junction and insert PCR are listed in Supporting Table S5. PCR products were visualized on a 1% agarose gel and verified by sequencing.

Quantitative Real-Time PCR (qRT-PCR) for Copy Number Analysis. qRT-PCR was carried out on genomic DNA samples to determine the relative copy number of mCherry gene. GeneJET Genomic DNA Purification Kit (Thermo Fisher Scientific) was used for genomic DNA preparation for copy number analysis according to the manufacturer's instructions. The qRT-PCR was run in the QuantStudio 5 Real-Time PCR System (Agilent Technologies). Amplification was executed with the following conditions: 95 °C for 10 min; 40×: 95 °C for 20 s, 60 °C for 30 s. Copy number of mCherry gene was determined using Brilliant III Ultra-Fast SYBR Green QPCR Master Mix (Agilent Technologies). A delta−delta threshold cycle ($\Delta\Delta C_T$) method was applied to calculate copy number of mCherry transgene compared to COSMC-mCherry clone (calibrator),[9] using GAPDH as an internal control gene for normalization. Clones were considered having one copy when the calculated value was in the range 0.5−1.5 (including standard deviations). Primers are listed in Supporting Table S6 and were validated by melting curve analysis and primer efficiency test. Each experiment included no template controls in every PCR run and had 3 replicates with 2 times repetition.

Quantitative Real-Time PCR (qRT-PCR) for Relative RNA Expression Levels Analysis. qRT-PCR was carried out on RNA samples to measure relative RNA expression level of transgenes. RNA was extracted from a minimum of 1.0 × 10^6 cells using TRIzol Reagent (Thermo Fisher Scientific), followed by DNase treatment to remove contaminating DNA (TURBO DNA-free DNase Treatment and Removal Reagents, Thermo Fisher Scientific). cDNAs were synthesized from 1 μg of total RNAs using Maxima First Strand cDNA Synthesis Kit for RT-qPCR (Thermo Fisher Scientific). The qRT-PCR was run in the QuantStudio 5 Real-Time PCR System (Agilent Technologies). Relative mCherry and recombinant protein expression levels were determined using TaqMan Gene Expression Master Mix and custom-made TaqMan probes (Thermo Fisher Scientific). Primers and probes are listed in Supporting Table S6 and were validated by melting curve analysis and primer efficiency test. Amplification was executed with the following conditions: 50 °C for 2 min, 95 °C for 10 min; 40×: 95 °C for 15 s, 60 °C for 1 min. Using $\Delta\Delta C_T$ method, the relative expression level was calculated by normalization to expression levels of two reference genes (Fkbp1 and Gnb1).[42] Each experiment included no template controls in every PCR run and had 3 replicates with 2 times repetition.

RMCE Subclone Generation. MCLs at concentration 1 × 10^6 cells/mL were transfected with promoterless expression vector and Cre-recombinase vector in 3:1 ratio (w:w) in 6-well plates using FreeStyle MAX transfection reagent. For the generation of nonproducer clones only Cre-recombinase vector was transfected. Cells were passaged two times after transfection. After 7 days the cell pools were single cell sorted as described above, using the respective MCLs as gating control for mCherry negative cells. Clones were expanded and verified by insert PCR and sequencing of the inserted GOI or empty region (for nonproducer clones).

Batch Cultivation. Cells were seeded at 3 × 10^5 cells/mL in 60 mL CD CHO medium, supplemented with 8 mM L-Glutamine and 1 μL/mL anticlumping agent, in 250 mL Erlenmeyer shake flasks. Cells were incubated in a humidified incubator at 37 °C, 5% CO_2 at 120 rpm. VCD and viability were monitored daily using the NucleoCounter NC-200 Cell Counter (ChemoMetec). Cultures were discontinued after 7 days in culture.

Long-Term Cultivation. 1 × 10^7 cells were thawed in 30 mL CD CHO medium, supplemented with 8 mM L-Glutamine

DOI: 10.1021/acssynbio.8b00140
ACS Synth. Biol. 2018, 7, 2148−2159

and 1 μL/mL anticlumping agent, in 125 mL Erlenmeyer shake flasks. They were passaged to 3×10^5 cells/mL three times a week for three months. Cells were cryopreserved at every passage: 1×10^7 cells were harvested, centrifuged at 250 g and resuspended in spent media with 5% DMSO. Cells from passage 2 (viability >95%), passage 17 and passage 35, was selected as our 0 month, 1.5 months and 3 months samples, respectively. The selected samples were run in batch cultivation as described above.

RNA Extraction for RNA-Seq. 5×10^6 cells were harvested on day four in late-exponential phase during batch cultivation. Cells were centrifuged at 1000 *g* for 4 min and the supernatant was discarded. The pellet was completely resuspended in 1 mL Invitrogen Trizol Reagent and stored at −80 °C. RNA was extracted following the Trizol manufacturer instructions. RNA concentrations were measured with Qubit fluorometric analysis (Life Technologies) and the quality was assessed with Agilent 2100 Bioanalyzer and Fragment analyzer automated CE system (Advanced Analytical Technologies, Inc.). Only samples of good quality were used for RNA sequencing (RIN or RIQ > 9).

Library Preparation and RNA-Seq. The total RNA samples were processed by the NGS lab at the Novo Nordisk Foundation Center for Biosustainability (Technical University of Denmark), and prepared and depleted for rRNA with Illumina's TruSeq Stranded mRNA sample preparation kit, according to manufacturer's instructions. The samples were pooled and sequenced on Illumina's NextSeq 500 using reagents from the NextSeq 500/550 Mid Output v2 kit and 2 × 150 bp paired-end reads for data set 1 with about 10 million reads per sample, and NextSeq 500/550 High Output v2 kit and 300 bp paired-end reads for the remaining subclone data sets analyzed (data sets 2, 3 and 4) with about 20 million reads per sample (Supporting Table S7).

RNA-Seq Analysis and Differential Gene Expression Analysis. The transcript levels in all the experimental data sets were quantified following the pipeline depicted in Supporting Figure S12. Raw sequence reads from multiple lanes were first merged into one for each biological replicate of each sample. The merged sequence files were subjected to Trim Galore v0.4.4 (Babraham Bioinformatics) to remove adaptor sequences, and FastQC (Babraham Bioinformatics) to examine data quality. Gene-level counts were produced by estimating the transcript abundances with Salmon v0.8.2[43] in quasi-mapping-based mode (with default parameters) using the transcriptome from the Cricetulus griseus (Chinese hamster) representative genome assembly CriGri_1.0 (GCA_000223135.1) as a reference. The transgenes (ETN, EPO, GDF5, C1INH, and NeoR) coding sequences used in this study were all added to the transcriptome reference and included in the DGE analysis. The salmon output was imported with tximport[44] into R[45] and analyzed. PCA was performed by applying the prcomp R function to the complete data set (excluding transgenes) and visualized in conjunction with ggplot2.[46] DGE was determined with limma-voom[47] using ANOVA-style F-test on the comparisons of interest. Genes were considered DE when the adjusted *p* value of the F-test was <0.05.

Canonical Pathway and Network Analysis. Functional enrichment analysis of canonical pathways and network analysis was carried out using QIAGEN's Ingenuity pathway analysis (IPA) tool. All statistically significant DE genes with a human homologue were ranked based on their log fold change, the 1000 most down-regulated and the 1000 most up-regulated genes were selected, which provided the same information as using the whole list of DE genes while reducing the noise. These genes and their corresponding log fold change (between EPO and nonproducer subclones) were integrated into a core analysis using the IPA tool.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acssynbio.8b00140.

mCherry expression levels; transgene expression levels; expression level comparisons of transcriptomics replicates; overview of DE analysis data sets; phenotypes of isoCHO-CP subclones; canonical pathway enrichment analysis; overview of transcriptomics analysis pipeline; fluorescence level of mCherry clones; GDF5 nucleotide sequence; homology arms sequences; primers for plasmids construction, junction and insert PCR; qRT-PCR primers and probes (PDF)

Lists of differentially expressed genes (XLSX)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: hef@biosustain.dtu.dk.

ORCID

Helene Faustrup Kildegaard: 0000-0003-3727-5721

Author Contributions

L.M.G, J.S.L and D.S. designed and performed the experiments and analyzed the results. L.M.G. analyzed the RNA-seq data. I.M.M. conducted the Ingenuity pathway analysis. H.F.K. and J.S.L. contributed with the main ideas and supervised the project. L.M.G., D.S. and H.F.K. wrote the manuscript and Supporting Information. N.E.L., M.R.A., L.K.N. and G.M.L. provided valuable scientific inputs. All authors discussed the results and implications of the research and commented on the manuscript.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors thank Nachon Charanyanonda Petersen and Karen Kathrine Brøndum for assistance with the FACS. We greatly acknowledge support from the Novo Nordisk Foundation (NNF10CC1016517 and NNF16CC0020908).

■ REFERENCES

(1) Walsh, G. (2014) Biopharmaceutical Benchmarks 2014. *Nat. Biotechnol. 32* (10), 992−1000.

(2) Noh, S. M., Sathyamurthy, M., and Lee, G. M. (2013) Development of Recombinant Chinese Hamster Ovary Cell Lines for Therapeutic Protein Production. *Curr. Opin. Chem. Eng. 2* (4), 391−397.

(3) Vcelar, S., Melcher, M., Auer, N., Hrdina, A., Puklowski, A., Leisch, F., Jadhav, V., Wenger, T., Baumann, M., and Borth, N. (2018) Changes in Chromosome Counts and Patterns in CHO Cell Lines upon Generation of Recombinant Cell Lines and Subcloning. *Biotechnol. J. 13*, 1700495.

(4) Wurm, F., and Wurm, M. (2017) Cloning of CHO Cells, Productivity and Genetic Stability—A Discussion. *Processes 5* (2), 20.

(5) Wilson, C., Bellen, H. J., and Gehring, W. J. (1990) Position Effects on Eukaryotic Gene Expression. *Annu. Rev. Cell Biol. 6* (1), 679−714.

DOI: 10.1021/acssynbio.8b00140
ACS Synth. Biol. 2018, 7, 2148−2159

(6) Yusufi, F. N. K., Lakshmanan, M., Ho, Y. S., Loo, B. L. W., Ariyaratne, P., Yang, Y., Ng, S. K., Tan, T. R. M., Yeo, H. C., Lim, H. L., et al. (2017) Mammalian Systems Biotechnology Reveals Global Cellular Adaptations in a Recombinant CHO Cell Line. *Cell Syst 4* (5), 530−542.e6.
(7) Pilbrough, W., Munro, T. P., and Gray, P. (2009) Intraclonal Protein Expression Heterogeneity in Recombinant CHO Cells. *PLoS One 4* (12), e8432.
(8) Hansen, H. G., Pristovšek, N., Kildegaard, H. F., and Lee, G. M. (2017) Improving the Secretory Capacity of Chinese Hamster Ovary Cells by Ectopic Expression of Effector Genes: Lessons Learned and Future Directions. *Biotechnol. Adv. 35* (1), 64−76.
(9) Lee, J. S., Kallehauge, T. B., Pedersen, L. E., and Kildegaard, H. F. (2015) Site-Specific Integration in CHO Cells Mediated by CRISPR/Cas9 and Homology-Directed DNA Repair Pathway. *Sci. Rep. 5*, 8572.
(10) Wirth, D., Gama-Norton, L., Riemer, P., Sandhu, U., Schucht, R., and Hauser, H. (2007) Road to Precision: Recombinase-Based Targeting Technologies for Genome Engineering. *Curr. Opin. Biotechnol. 18* (5), 411−419.
(11) Gaidukov, L., Wroblewska, L., Teague, B., Nelson, T., Zhang, X., Liu, Y., Jagtap, K., Mamo, S., Tseng, W. A., Lowe, A., et al. (2018) A Multi-Landing Pad DNA Integration Platform for Mammalian Cell Engineering. *Nucleic Acids Res. 46* (8), 4072−4086.
(12) Kim, H., and Kim, J.-S. (2014) A Guide to Genome Engineering with Programmable Nucleases. *Nat. Rev. Genet. 15* (5), 321−334.
(13) Lee, J. S., Grav, L. M., Pedersen, L. E., Lee, G. M., and Kildegaard, H. F. (2016) Accelerated Homology-Directed Targeted Integration of Transgenes in Chinese Hamster Ovary Cells via CRISPR/Cas9 and Fluorescent Enrichment. *Biotechnol. Bioeng. 113* (11), 2518−2523.
(14) Phan, Q. V., Contzen, J., Seemann, P., and Gossen, M. (2017) Site-Specific Chromosomal Gene Insertion: Flp Recombinase versus Cas9 Nuclease. *Sci. Rep. 7* (1), 17771.
(15) Merrick, C., Zhao, J., and Rosser, S. (2018) Serine Integrases: Advancing Synthetic Biology. *ACS Synth. Biol. 7*, 299.
(16) Baumann, M., Gludovacz, E., Sealover, N., Bahr, S., George, H., Lin, N., Kayser, K., and Borth, N. (2017) 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* (11), 2616−2627.
(17) Zhang, L., Inniss, M. C., Han, S., Moffat, M., Jones, H., Zhang, B., Cox, W. L., Rance, J. R., and Young, R. J. (2015) Recombinase-Mediated Cassette Exchange (RMCE) for Monoclonal Antibody Expression in the Commercially Relevant CHOK1SV Cell Line. *Biotechnol. Prog. 31* (6), 1645−1656.
(18) Nehlsen, K., Schucht, R., da Gama-Norton, L., Krömer, W., Baer, A., Cayli, A., Hauser, H., and Wirth, D. (2009) Recombinant Protein Expression by Targeting Pre-Selected Chromosomal Loci. *BMC Biotechnol. 9*, 100.
(19) Mayrhofer, P., Kratzer, B., Sommeregger, W., Steinfellner, W., Reinhart, D., Mader, A., Turan, S., Qiao, J., Bode, J., and Kunert, R. (2014) Accurate Comparison of Antibody Expression Levels by Reproducible Transgene Targeting in Engineered Recombination-Competent CHO Cells. *Appl. Microbiol. Biotechnol. 98* (23), 9723−9733.
(20) Kim, M. S., and Lee, G. M. (2008) Use of Flp-Mediated Cassette Exchange in the Development of a CHO Cell Line Stably Producing Erythropoietin. *J. Microbiol. Biotechnol 18* (7), 1342−1351.
(21) Zhou, H., Liu, Z.-G., Sun, Z.-W., Huang, Y., and Yu, W.-Y. (2010) Generation of Stable Cell Lines by Site-Specific Integration of Transgenes into Engineered Chinese Hamster Ovary Strains Using an FLP-FRT System. *J. Biotechnol. 147* (2), 122−129.
(22) Schucht, R., Lydford, S., Andzinski, L., Zauers, J., Cooper, J., Hauser, H., Wirth, D., and May, T. (2011) Rapid Establishment of G-Protein-Coupled Receptor-Expressing Cell Lines by Site-Specific Integration. *J. Biomol. Screening 16* (3), 323−331.
(23) Kito, M., Itami, S., Fukano, Y., Yamana, K., and Shibui, T. (2002) Construction of Engineered CHO Strains for High-Level Production of Recombinant Proteins. *Appl. Microbiol. Biotechnol. 60* (4), 442−448.
(24) Inniss, M. C., Bandara, K., Jusiak, B., Lu, T. K., Weiss, R., Wroblewska, L., and Zhang, L. (2017) A Novel Bxb1 Integrase RMCE System for High Fidelity Site-Specific Integration of mAb Expression Cassette in CHO Cells. *Biotechnol. Bioeng. 114* (8), 1837−1846.
(25) Scarcelli, J. J., Shang, T. Q., Iskra, T., Allen, M. J., and Zhang, L. (2017) Strategic Deployment of CHO Expression Platforms to Deliver Pfizer's Monoclonal Antibody Portfolio. *Biotechnol. Prog. 33* (6), 1463−1467.
(26) Sommeregger, W., Mayrhofer, P., Steinfellner, W., Reinhart, D., Henry, M., Clynes, M., Meleady, P., and Kunert, R. (2016) Proteomic Differences in Recombinant CHO Cells Producing Two Similar Antibody Fragments. *Biotechnol. Bioeng. 113* (9), 1902−1912.
(27) Araki, K., Araki, M., and Yamamura, K.-I. (2002) Site-Directed Integration of the Cre Gene Mediated by Cre Recombinase Using a Combination of Mutant Lox Sites. *Nucleic Acids Res. 30* (19), e103.
(28) Conesa, A., Madrigal, P., Tarazona, S., Gomez-Cabrero, D., Cervera, A., McPherson, A., Szcześniak, M. W., Gaffney, D. J., Elo, L. L., Zhang, X., et al. (2016) A Survey of Best Practices for RNA-Seq Data Analysis. *Genome Biol. 17*, 13.
(29) InvivoGen 2011 Catalog, https://www.invivogen.com/sites/default/files/invivogen/old/docs/InvivoGen2011Catalog.pdf (accessed Jul 5, 2018).
(30) Orellana, C. A., Marcellin, E., Palfreyman, R. W., Munro, T. P., Gray, P. P., and Nielsen, L. K. (2018) RNA-Seq Highlights High Clonal Variation in Monoclonal Antibody Producing CHO Cells. *Biotechnol. J. 13*, 1700231.
(31) Mikkelsen, M. D., Buron, L. D., Salomonsen, B., Olsen, C. E., Hansen, B. G., Mortensen, U. H., and Halkier, B. A. (2012) Microbial Production of Indolylglucosinolate through Engineering of a Multi-Gene Pathway in a Versatile Yeast Expression Platform. *Metab. Eng. 14* (2), 104−111.
(32) Kallehauge, T. B., Li, S., Pedersen, L. E., Ha, T. K., Ley, D., Andersen, M. R., Kildegaard, H. F., Lee, G. M., and Lewis, N. E. (2017) Ribosome Profiling-Guided Depletion of an mRNA Increases Cell Growth Rate and Protein Secretion. *Sci. Rep. 7*, 40388.
(33) Hefzi, H., Ang, K. S., Hanscho, M., Bordbar, A., Ruckerbauer, D., Lakshmanan, M., Orellana, C. A., Baycin-Hizal, D., Huang, Y., Ley, D., et al. (2016) A Consensus Genome-Scale Reconstruction of Chinese Hamster Ovary Cell Metabolism. *Cell Syst. 3* (5), 434−443.e8.
(34) Vcelar, S., Jadhav, V., Melcher, M., Auer, N., Hrdina, A., Sagmeister, R., Heffner, K., Puklowski, A., Betenbaugh, M., Wenger, T., et al. (2018) Karyotype Variation of CHO Host Cell Lines over Time in Culture Characterized by Chromosome Counting and Chromosome Painting. *Biotechnol. Bioeng. 115* (1), 165−173.
(35) Gaj, T., Epstein, B. E., and Schaffer, D. V. (2016) Genome Engineering Using Adeno-Associated Virus: Basic and Clinical Research Applications. *Mol. Ther. 24* (3), 458−464.
(36) Underhill, M. F., Birch, J. R., Smales, C. M., and Naylor, L. H. (2005) eIF2alpha Phosphorylation, Stress Perception, and the Shutdown of Global Protein Synthesis in Cultured CHO Cells. *Biotechnol. Bioeng. 89* (7), 805−814.
(37) Hansen, K. D., Wu, Z., Irizarry, R. A., and Leek, J. T. (2011) Sequencing Technology Does Not Eliminate Biological Variability. *Nat. Biotechnol. 29* (7), 572−573.
(38) Ronda, C., Pedersen, L. E., Hansen, H. G., Kallehauge, T. B., Betenbaugh, M. J., Nielsen, A. T., and Kildegaard, H. F. (2014) Accelerating Genome Editing in CHO Cells Using CRISPR Cas9 and CRISPy, a Web-Based Target Finding Tool. *Biotechnol. Bioeng. 111* (8), 1604−1616.
(39) Pristovšek, N., Hansen, H. G., Sergeeva, D., Borth, N., Lee, G. M., Andersen, M. R., and Kildegaard, H. F. (2018) Using Titer and Titer Normalized to Confluence Are Complementary Strategies for

DOI: 10.1021/acssynbio.8b00140
ACS Synth. Biol. 2018, 7, 2148−2159

Obtaining Chinese Hamster Ovary Cell Lines with High Volumetric Productivity of Etanercept. *Biotechnol. J. 13*, 1700216.
(40) Kol, S., Kallehauge, T. B., Adema, S., and Hermans, P. (2015) Development of a VHH-Based Erythropoietin Quantification Assay. *Mol. Biotechnol. 57* (8), 692−700.
(41) Hansen, H. G., Nilsson, C. N., Lund, A. M., Kol, S., Grav, L. M., Lundqvist, M., Rockberg, J., Lee, G. M., Andersen, M. R., and Kildegaard, H. F. (2016) Versatile Microscale Screening Platform for Improving Recombinant Protein Productivity in Chinese Hamster Ovary Cells. *Sci. Rep. 5*, 18016.
(42) Brown, A. J., Gibson, S., Hatton, D., and James, D. C. (2018) Transcriptome-Based Identification of the Optimal Reference CHO Genes for Normalisation of qPCR Data. *Biotechnol. J. 13* (1), 1700259.
(43) Patro, R., Duggal, G., Love, M. I., Irizarry, R. A., and Kingsford, C. (2017) Salmon Provides Fast and Bias-Aware Quantification of Transcript Expression. *Nat. Methods 14* (4), 417−419.
(44) Soneson, C., Love, M. I., and Robinson, M. D. (2015) Differential Analyses for RNA-Seq: Transcript-Level Estimates Improve Gene-Level Inferences. *F1000Research 4*, 1521.
(45) R Core Team (2014) *R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing*.
(46) Wickham, H. (2009) *ggplot2: Elegant Graphics for Data Analysis*, Springer-Verlag, New York.
(47) Law, C. W., Chen, Y., Shi, W., and Smyth, G. K. (2014) Voom: Precision Weights Unlock Linear Model Analysis Tools for RNA-Seq Read Counts. *Genome Biol. 15* (2), R29.

DOI: 10.1021/acssynbio.8b00140
ACS Synth. Biol. 2018, 7, 2148−2159

Chapter 4 - Multi-copy targeted integration for accelerated development of high-producing CHO cells

This chapter focuses on increasing the productivity of CHO cell lines generated by targeted integration. It describes the development and optimization of a multi-copy targeted integration platform for the production of therapeutic proteins, providing a detailed characterization of the method and making it available for the scientific community. We show that upon optimization of genetic elements and increase in gene copy number, cell lines created by targeted integration can reach commercially-relevant productivities and titers. Moreover, the study sheds the light on the relationship between gene copy number and protein expression, revealing the limitations of high-level protein production in CHO.

Accepted in ACS Synthetic Biology:
Sergeeva, D., Lee, G. M., Nielsen, L. K., Grav, L. M. "Multi-copy targeted integration for accelerated development of high-producing CHO cells", 2020
https://doi.org/10.1021/acssynbio.0c00322

Multi-copy targeted integration for accelerated development of high-producing CHO cells

Daria Sergeeva[1], Gyun Min Lee[1,2], Lars Keld Nielsen[1,3], Lise Marie Grav[1*]

[1]The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark

[2]Department of Biological Sciences, KAIST, Daejeon 34141, Republic of Korea

[3]Australian Institute for Bioengineering and Nanotechnology, University of Queensland, Brisbane, 4072, Australia

*Corresponding author: Lise Marie Grav, lgrav@biosustain.dtu.dk

Keywords: Chinese hamster ovary cells, cell line development, recombinase-mediated cassette exchange, targeted integration, specific productivity

Abstract

The ever-growing biopharmaceutical industry relies on the production of recombinant therapeutic proteins in Chinese hamster ovary (CHO) cells. The traditional timelines of CHO cell line development can be significantly shortened by the use of targeted gene integration (TI). However, broad use of TI has been limited due to the low specific productivity (q_P) of TI-generated clones. Here, we show a 10-fold increase in the q_P of therapeutic glycoproteins in CHO cells through the development and optimization of a multi-copy TI method. We used a recombinase-mediated cassette exchange (RMCE) platform to investigate the effect of gene copy number, 5' and 3' gene regulatory elements and landing pad features on q_P. We evaluated the limitations of multi-copy expression from a single genomic site as well as multiple genomic sites and found that a transcriptional bottleneck can appear with an increase in gene dosage. We created a dual-RMCE system for simultaneous multi-copy TI in two genomic sites and generated isogenic high-producing clones with q_P of 12-14 pg/cell/day and product titer close to 1 g/L in fed-batch. Our study provides an extensive characterization of the multi-copy TI method and elucidates the relationship between gene copy number and protein expression in mammalian cells. Moreover, it demonstrates that TI-generated CHO cells are capable of producing therapeutic proteins at levels that can support their industrial manufacture.

The foundation of the $200 billion biopharmaceutical industry continues to be the production of complex therapeutic proteins such as monoclonal antibodies (mAbs), fusion proteins, hormones, enzymes and coagulation factors in mammalian cells, with Chinese hamster ovary (CHO) cells being the leading production host[1]. For each therapeutic product, a stable cell line with a high product titer and a desirable protein quality must be developed[2]. The conventional cell line development process is long and unpredictable, spanning several months and involving the screening of several hundred cell clones for high productivity[3]. To accelerate cell line development, more efficient and predictable methods for generating high-producing CHO cell lines are required.

For the past 30 years, random gene integration has been the method of choice for developing protein-producing CHO cell lines. Using this method, the gene of interest (GOI) is integrated into randomly occurring double-stranded breaks in the genomic DNA, resulting in GOI integration into variable sites with an uncontrolled number of copies. This approach generates a set of clones with diverse and unpredictable phenotypes attributed to clonal variation[4], which necessitates tedious screening of clones and bioprocess re-optimization for the chosen clone.

Targeted gene integration (TI) has the potential to significantly reduce clonal variation, thus accelerating CHO cell line development[4,5]. Integration of the transgene into a pre-selected genomic site enables predictable generation of isogenic cell lines with uniform phenotypes and consistent expression[6,7]. Since the first reports of TI in mammalian cells in late 1980s[8,9] and early 1990s[10–12], several methods for targeted gene integration have been developed, which utilize site-specific recombinases such as Cre, Flp and Bxb1 for single-site recombination or recombinase-mediated cassette exchange (RMCE), and programmable nucleases such as CRISPR/Cas9 (clustered regularly interspaced short palindromic repeat/CRISPR-associated protein 9)[13–15]. However, while TI has been used in special applications in CHO[16], it has not been used widely in industry, largely due to low specific productivity (q_P) of TI-generated clones, caused by single-copy GOI integration, and the challenge to identify highly active transcriptional sites (hotspots) in the genome.

In recent years, the interest in TI has been rising in the biopharmaceutical industry, with several companies publishing reports on mAb production using RMCE systems. In 2013, Genentech established a Cre-based RMCE platform for mAb expression by the

screening of transcriptionally active sites in the CHO genome[17]. Insertion of a single expression cassette containing heavy and light chain genes led to the generation of stable cell lines for five different mAbs with reproducible q_P of 3-4 pg/cell/day (pcd). Insertion of two mAb cassettes into the same genomic site doubled specific productivity with a maximum of 10 pcd achieved. Similarly, Pfizer reported the development of Flp-based RMCE[18] and Bxb1-based RMCE[19] platforms for mAb expression, reaching q_P of 3 pcd for single-copy mAb insertion. In 2020, Genentech showed that mAb expression can reach q_P of 20-50 pcd in the RMCE system by the integration of multiple copies of heavy and light chains into a single genomic site[20]. However, the commercial nature means that these studies fail to provide details on the design of the expression vectors and do not report any limitations of the methods, making them difficult to reproduce. Moreover, while these and other studies[21] indicate a positive correlation between gene copy number and mAb production in TI-generated clones, this relationship can be obscured by the ratio of expression between light and heavy chains.

In this study, we focused on increasing q_P upon TI for the production of therapeutic proteins not limited to mAbs. We developed and extensively characterized a multi-copy TI method for the production of biopharmaceutical proteins and used it to study the relationship between gene copy number and protein production to reveal if any expression bottlenecks appear with an increase in gene dosage. We selected two industrially-relevant single-chain glycoproteins: the hormone erythropoietin (EPO) and the Fc-fusion protein etanercept (ETN), which have different biochemical properties and complex post-translational modifications, with ETN generally considered as a difficult-to-express protein[22]. Compared to mAb-producing CHO cells generated by random integration that can reach q_P of 20-40 pcd[23], EPO- and ETN-producing CHO cells generated by random integration usually have a lower q_P in the range 1-10 pcd[24,25].

In our previous works, we established a TI method for CHO cell line development based on the Cre-mediated RMCE, where EPO specific productivity of 1-3 pcd was achieved from a single-copy integration[7,26]. We used this RMCE system as a starting point to develop and optimize the multi-copy TI method described in the current paper. Here, we investigated the effect of different gene regulatory elements in the RMCE donor vector, landing pad features and gene copy number in order to increase q_P. We evaluated the limitations of multi-copy expression from a single genomic site as well as

multiple genomic sites and showed that a transcriptional bottleneck can appear when the copy number is increased. We developed a dual-RMCE master cell line with two landing pads for simultaneous multi-copy TI in two genomic sites. With this system, we were able to site-specifically introduce up to four copies of GOI into the genome simultaneously, leading to the generation of high-producing EPO and ETN clones with q_P of 12-14 pcd and titer of nearly 1 g/L in shake flask cultivation, reaching and excelling q_P of high-producing clones generated by random integration. To support further use of multi-copy TI by the scientific community, we have made our vectors for the multi-copy RMCE system available from Addgene.

Results

Optimal 5' and 3' gene regulatory elements for increased protein production upon multi-copy targeted integration.

To increase q_P upon targeted integration in CHO cells, we first focused on finding strong gene regulatory elements that can support high protein expression in the RMCE system. As a platform for screening, we used the previously described CHO-S derived AD1 MCL with a landing pad for poly(A) trap-based RMCE integrated into the genomic site A, which is a proven safe harbor site for TI supporting high and stable protein expression[26]. The landing pad in AD1 MCL has a mCherry fluorescent marker expressed from an EF1α promoter surrounded by loxP and lox2272 recombination sites and a BGH poly(A) tail outside the lox sites, thus allowing a RMCE of EF1α-mCherry to GOI with different 5' regulatory sequences (Figure 1a). To find optimal 5' genetic elements, we designed RMCE donor plasmids D1-D3 with two different promoters (CP and 100RPU.2) and two Kozak sequences (k1 and k2) driving expression of the EPO gene (Figure 1b). CP is a composite promoter consisting of mouse CMV enhancer, minimal human EF1α promoter and human T cell leukemia virus 1 untranslated region. CP was previously found to be stronger than CMV, SV40, and EF1α promoters when integrated into site A in the RMCE system[26]. The 100RPU.2 promoter is a *de novo*-designed synthetic promoter composed of CHO-specific transcription factor regulatory elements and it showed a 2-fold higher expression level compared to CMV promoter in transient transfection and stable CHO-K1 pools generated by random integration[27]. Kozak sequence k1 is a

consensus mammalian translation initiation site (TIS) CGCCACC[28]. Kozak sequence k2 is a TCGGTC TIS found to be the strongest in yeast and showed almost 50% higher expression of eGFP in CHO cells when compared to the consensus Kozak[29]. The combination of CP and k1 showed the best performance in our poly(A)-trap RMCE system previously[26]. We co-transfected Cre recombinase together with each D1-D3 donor plasmid into AD1 MCL and bulk sorted mCherry-negative cells six days after transfection. The cell pools were verified to have the desired one-copy insert by genomic PCR of the target region and copy number analysis by digital PCR (dPCR) (Supporting Figure 1a,b). Generated cell pools were grown in batch culture, and EPO q_P, titer and mRNA were evaluated (Figure 1c,d, Supporting Figure 1e). EPO q_P and titer of D2 cells with 100RPU.2 promoter were doubled (2.8 pcd, 61mg/L) compared to D1 cells with CP promoter (1.4 pcd, 28 mg/L) (multiple testing, adj. p = 0.001). Kozak sequence k2 had about 30% lower EPO expression in D3 cells (2.1 pcd, 43 mg/L) compared to Kozak k1 in D2 cells (multiple testing, adj. p = 0.01), which was attributed to lower mRNA levels. Hence we chose a combination of 100RPU.2 and k1 as optimal 5' gene regulatory elements for GOI expression in the RMCE system.

To evaluate 3' gene regulatory elements for multi-copy RMCE integration, we designed RMCE donor plasmids R1-R6 (Figure 1e), which had one copy (R1, R2) or two copies (R3-R6) of the expression cassette 100RPU.2_k1_EPO. We assessed the effects of Woodchuck hepatitis virus post-transcriptional regulation element (WPRE) and chicken β-globin insulator HS4. WPRE has been known to affect the transport of mRNA and increase transgene expression in viral vectors[30], and enhance mAb production in HEK 293E cells[31]. The chicken HS4 insulator can act as a barrier to block promoter interference[32], can prevent DNA methylation and maintain open chromatin state[33], and has been used for the construction of multi-genic plasmids for expression in mammalian cells[34]. WPRE was inserted in the 3' untranslated region (UTR) of the EPO gene between the stop codon and the BGH poly(A) tail in plasmids R2, R5, and R6. Two 250-bp core HS4 insulators were placed between poly(A) tail and 100RPU.2 promoter in plasmids R4 and R6 to separate two EPO expression cassettes. Each RMCE donor plasmid R1-R6 was co-transfected with Cre recombinase into AD1 MCL, mCherry-negative cells were single-cell sorted by FACS and clones were verified by genomic PCR of the target region and copy number analysis (Supporting Figure 1c,d). Two clones for each cell design R1-R6 were

cultivated in batch mode, and EPO q_P, titer and mRNA were evaluated (Figure 1f,g, Supporting Figure 1f). We observed that the introduction of WPRE in each case (i.e., R2 vs R1, R5 vs R3, and R6 vs R4) had a significant negative effect on specific productivity (multiple testing, adj. $p < 10^{-10}$). The HS4 insulator significantly enhanced specific productivity in two copy expression vectors (i.e, R4 vs R3 and R6 vs R5) (multiple testing, adj. $p < 3 \times 10^{-9}$). The effect of HS4 and WPRE was also reflected in changes of EPO mRNA levels (Supporting Figure 1f). The best two copy plasmid R4 (with insulator and no WPRE) enabled cells to almost double (1.9-fold increase) EPO expression compared to then best single copy plasmid R1. Thus, the use of an insulator was crucial to block the interference between expression cassettes, and plasmid design R4 was the optimal for building multi-copy plasmids.

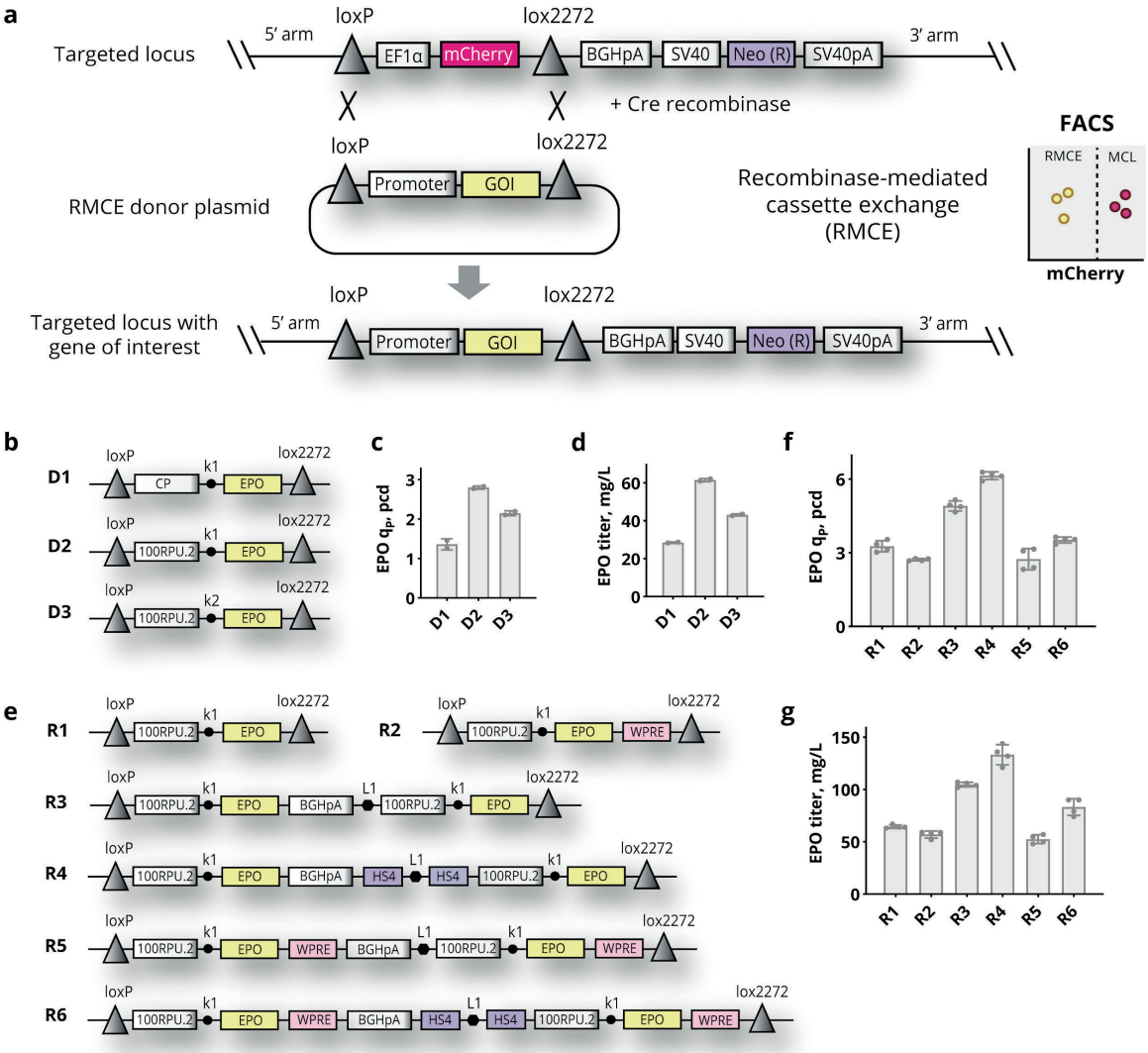


Figure 1. RMCE platform and screening of gene regulatory elements for increased protein production upon multi-copy targeted integration. **a**. Overview of the RMCE strategy. First, a landing pad with poly(A) trap is integrated into the specific genomic locus using CRISPR/Cas9-mediated targeted integration, resulting in the creation of a master cell line (MCL). Then the MCL is transfected with a RMCE donor plasmid together with Cre recombinase, leading to the site-specific recombination between lox sites and exchange of EF1α-mCherry cassette to promoter-GOI cassette in the targeted locus. mCherry-negative cells with recombinase-mediated integration are isolated using FACS. **b**. Design of RMCE donor plasmids D1-D3 used for screening of 5' regulatory elements (promoters (CP, 100RPU.2) and Kozak sequences (k1, k2)). Cell pools D1-D3 were grown in batch culture (2 technical replicates), EPO specific productivity (**c**) and EPO titer on day 5 (**d**) are shown. **e**. Design of RMCE donor plasmids R1-R6 used for screening of 3' regulatory elements (WPRE and HS4 insulator). Two clones for each cell design were grown in batch culture (2 technical replicates per clone), EPO specific productivity (**f**) and EPO titer on day 5 (**g**) are shown.

Targeted integration of up to 6 gene copies in a single genomic site.

To further increase recombinant protein expression upon TI, we constructed multi-copy plasmids expressing four and six copies of GOI for RMCE-based integration into a single genomic site. To assemble such long and repetitive constructs, we used USER cloning, a flexible method for the plasmid assembly of up to seven DNA fragments in one reaction enabled by single-stranded DNA overhangs (linkers)[35]. USER linkers flanked by unique restriction sites (L1-L6) were placed between each GOI expression cassette to ease the plasmid and clone verification (Figure 2a). The four- and six-copy plasmids were confirmed by sequencing and restriction analysis (Supporting Figure 2), no undesirable recombination events between repetitive elements were observed. We integrated the multi-copy plasmids expressing one (EPO1), two (EPO2), four (EPO4) and six (EPO6) copies of erythropoietin into the AD1 MCL using the described RMCE method. The RMCE efficiency decreased from 1.8% (EPO1) to 0.06% (EPO6) as the size of the RMCE insert increased (Supporting Table 1), but it was high enough to select desired clones. The single-cell sorted cells were expanded and verified by genomic PCR (Supporting Figure 3a-f,j) and copy number analysis (Figure 2b). While most of EPO1 and EPO2 mCherry-negative clones had correctly inserted expression cassettes in site A, we faced some difficulties verifying the integration of EPO4 and EPO6 plasmids in AD1 MCL. Although

genomic PCR of the 5' and 3' junctions of EPO4 and EPO6 cassettes confirmed that plasmids were integrated into site A (Supporting Figure 3e,f), PCR amplification of the whole four- and six-copy expression cassettes in the landing pad (insert PCR) consistently resulted in amplicons with a smaller than expected size (2.1 kb instead of 7.4 kb (EPO4) and 10.9 kb (EPO6)) (Supporting Figure 3j). The observed amplicons were sequence-specific to the landing pad with EPO cassette in site A (Supporting Sequence Data 1). These PCR artifacts can be explained by difficulties of PCR amplification of repetitive DNA sequences resulting from polymerase disengagements, which have been reported previously[36]. To analyze GOI copy number of multi-copy clones we used dPCR as a reportedly accurate method for copy number analysis[37]. While EPO1, EPO2 and EPO4 clones had the desired EPO copy number (Figure 2b), we noticed that dPCR measurements of six-copies EPO6 clones consistently showed lower copy number than expected. This inaccuracy can be attributed to a high partitioning error[38] or to incomplete separation of tandem gene copies by restriction digestion prior to dPCR analysis. We also cannot exclude possible recombination of repetitive genetic elements happening inside the CHO cell that could lead to a decrease in copy number in EPO6 clones.

We selected three EPO1 and EPO2 clones and four EPO4 and EPO6 clones, which had integration of RMCE cassettes confirmed by genomic insert PCR (EPO1) or junction PCR (EPO2, EPO4, EPO6) and copy number analysis (EPO1, EPO2, EPO4), and cultivated these clones in batch mode to evaluate EPO expression. The growth of the clones was uniform irrespective of inserted EPO copies (Figure 2c). A two-fold increase in titer and q_P was observed for EPO2 clones compared to EPO1, but no further increase was observed for EPO4 and EPO6 clones (Figure 2d,e). EPO mRNA expression levels followed the same trend as q_P (Figure 2f).

To confirm that the observed effects were not specific to EPO, we next constructed multi-copy plasmids ETN1, ETN2, ETN4, and ETN6 encoding etanercept and co-transfected them into AD1 MCL together with Cre recombinase to generate RMCE clones. The efficiency of RMCE dropped with the size of the plasmid similarly to EPO plasmids (Supporting Table 1), but no ETN6 clones could be FACS sorted, possibly due to the big insert size (15 kb). We verified ETN1, ETN2 and ETN4 clones by genomic junction PCR (Supporting Figure 3g,h,i) and copy number analysis (Figure 2g), although ETN4 copy

number was lower than expected. Moreover, we observed similar PCR artifacts for insert PCR of ETN4 as discussed above (Supporting Figure 3j). We cultivated three clones for each cell design ETN1, ETN2 and ETN4 in batch mode to assess ETN expression. The growth, ETN titer, q_P and mRNA levels of ETN clones followed the same trend as was noticed for EPO clones (Figure 2h-k), with no increase in ETN productivity observed for ETN4 clones.

The single-site multi-copy targeted integration of plasmids with more than two GOI copies proved challenging due to difficulties in clone verification, possible gene recombination and transcriptional limitation. Thus, we looked for other ways to increase recombinant protein expression in the RMCE system.

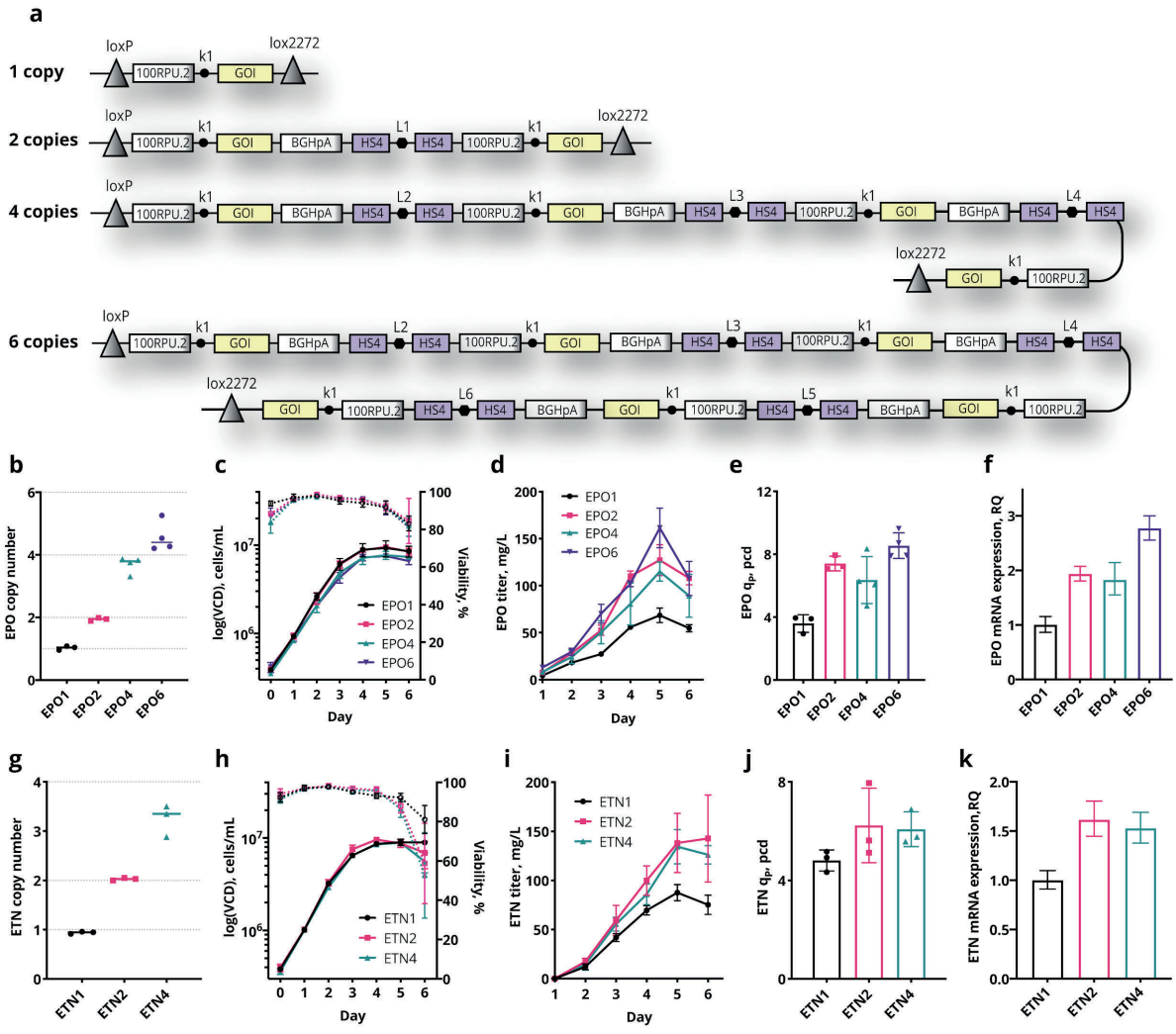


Figure 2. Targeted integration of one, two, four and six copies of EPO and ETN into a single genomic site. **a**. Schematic representation of RMCE donor plasmids with up to six GOI expression units. Each expression unit was flanked by a specific USER linker (L1-L6). **b-f**. Characterization of EPO1, EPO2, EPO4 and EPO6 clones created by targeted integration of RMCE donor plasmids into

AD1 master cell line (MCL). Three (EPO1, EPO2) or four (EPO4, EPO6) clones for each cell design were grown in batch culture. **b**. Copy number analysis of EPO1-EPO6 clones, showing median value for biological replicates. **c**. Viable cell densities (VCD) and viability of EPO1-EPO6 clones, showing mean with SD for biological replicates. **d**. EPO titer, showing mean with SD for biological replicates. **e**. EPO specific productivity, showing mean with SD for biological replicates. **f**. EPO mRNA expression, showing mean and SD for biological replicates. **g-k.** Characterization of ETN1, ETN2 and ETN4 clones created by targeted integration of RMCE donor plasmids into AD1 MCL. Three clones for each cell design were grown in batch culture. **g**. Copy number analysis of ETN1-ETN4 clones, showing median value for biological replicates. **h**. Viable cell densities (VCD) and viability of ETN1-ETN4 clones, showing mean and SD for biological replicates. **i**. ETN titer, showing mean and SD for biological replicates. **j**. ETN specific productivity, showing mean and SD for biological replicates. **k**. ETN mRNA expression, showing mean and SD for biological replicates.

Optimal landing pad features.

To increase protein expression in our RMCE system, we optimized features of the landing pad, evaluating the effect of a poly(A) tail outside recombination sites and the position of the antibiotic selection marker. Previous investigation of the global translation in the recombinant CHO cells have observed that non-essential and highly abundant mRNA encoding an antibiotic selection marker could occupy about 5% of the translatome and its elimination could improve the production of recombinant proteins[39]. Since the cultivation of RMCE-derived subclones does not require the presence of an antibiotic in the medium, we designed new landing pads to be able to remove the NeoR antibiotic selection marker after RMCE. In AD1 MCL used in the initial experiments NeoR expression cassette was placed outside of lox sites in the landing pad, hence NeoR was stably expressed in all RMCE-derived subclones. The new landing pads MP3 and LC1 (Figure 3a) had the NeoR expression cassette moved inside the lox sites, allowing it to be exchanged together with EF1α-mCherry to GOI expression cassettes after RMCE, thus eliminating NeoR expression. We placed BGH poly(A) tail outside of lox sites in the landing pad MP3, and SV40 poly(A) tail in the landing pad LC1 to assess the effect of these poly(A) tails on recombinant protein expression. Using CRISPR/Cas9-mediated targeted integration[40] we inserted a single copy of the landing pads MP3 or LC1 into the genomic site A of CHO-S cells, creating corresponding MCLs. The integration of the landing pads was confirmed by junction PCR and copy number analysis

(Supporting Figure 4a,b). We used MCLs MP3 and LC1 to generate RMCE subclones expressing one copy of EPO by co-transfecting EPO1 plasmid and Cre recombinase (Figure 3a). Resulting clones were verified by genomic PCR (data not shown) and copy number analysis (Supporting Figure 4c). We cultivated EPO-expressing subclones of MCLs AD1, MP3 and LC1 in batch mode and measured EPO expression (Figure 3b,c). To investigate the effect of NeoR cassette removal after RMCE, we compared subclones AD1_EPO1 and MP3_EPO1, which both had BGH poly(A) tail outside of lox sites, but have kept or removed NeoR after RMCE, accordingly. EPO mRNA expression and q_P both increased by ~20% in MP3-derived clones compared to AD1-derived clones (multiple comparison, p. adj. = 10^{-5}), showing that depletion of the NeoR marker can increase recombinant protein expression in the RMCE system. To study the effect of poly(A) tail on EPO expression, we compared MP3_EPO1 and LC1_EPO1 subclones, which had BGH poly(A) or SV40 poly(A) outside of lox sites, accordingly. The use of SV40 poly(A) tail resulted in a 2-fold decrease in EPO mRNA expression and a 3-fold decrease in q_P when comparing MP3- and LC1-derived subclones (multiple comparison, p. adj. = 2×10^{-10}). Thus, we conclude that the design of the landing pad in the MP3 MCL supports the highest recombinant protein expression upon RMCE due to removal of NeoR marker and usage of BGH poly(A) tail.

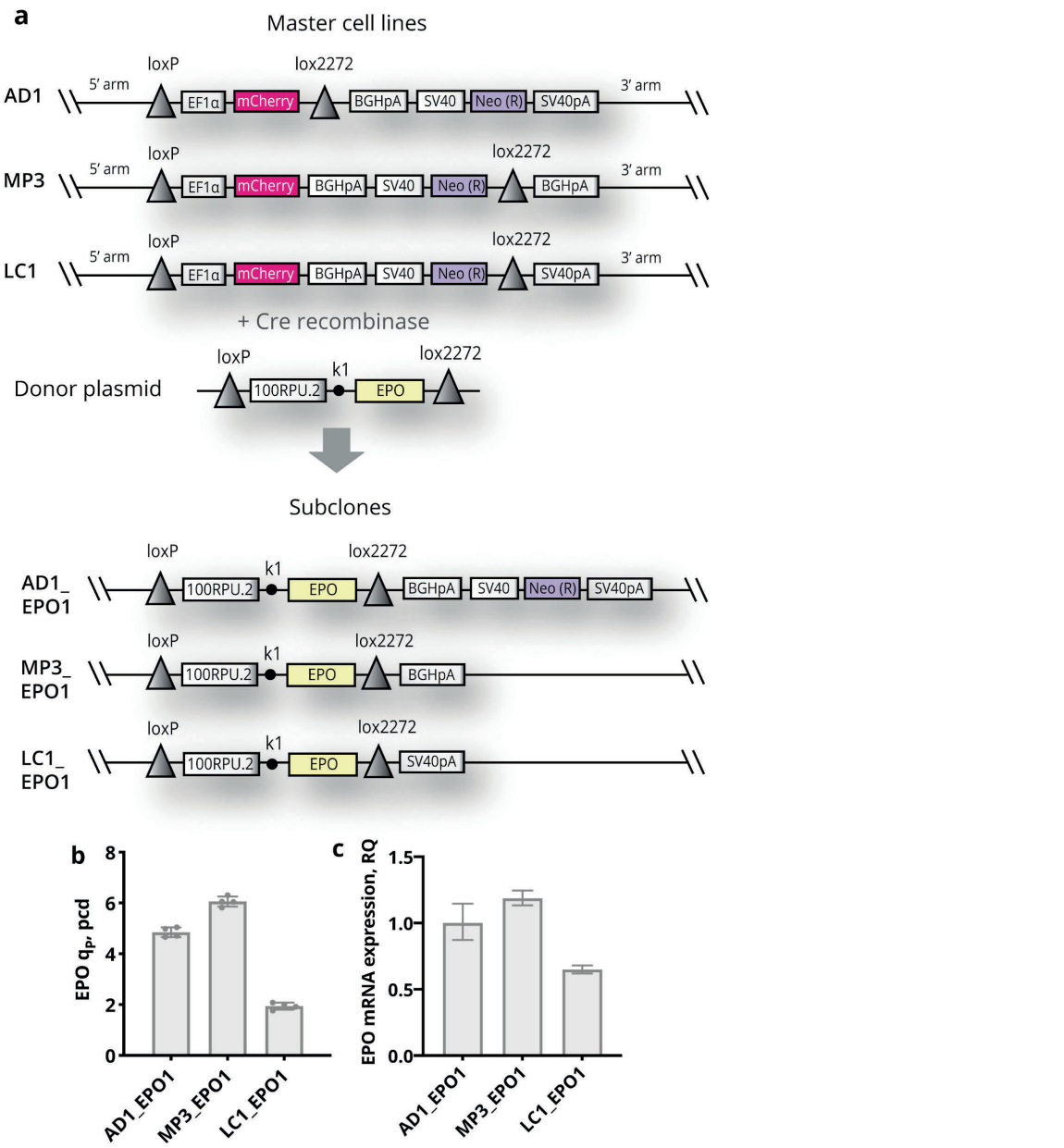


Figure 3. Optimization of landing pad features: poly(A) tail and antibiotic selection marker NeoR. **a**. Three master cell lines (AD1, MP3 and LC1) were created by CRISPR/Cas9-mediated targeted integration of landing pads into genomic site A. AD1 MCL had NeoR cassette outside the loxP and lox2272 sites, with BGH poly(A) outside lox sites. MP3 MCL had NeoR expression cassette moved inside loxP and lox2272 sites, with BGH poly(A) outside lox sites. LC1 MCL had NeoR expression cassette moved inside loxP and lox2272 sites, with SV40 poly(A) outside lox sites. The MCLs were used to generate subclones using a single copy (EPO1) RMCE donor plasmid. Two EPO-producing subclones per MCL were grown in batch culture (2 technical replicates). EPO specific productivity (**b**) and EPO mRNA expression (**c**) are shown with mean and SD for biological replicates.

Simultaneous integration of multi-copy plasmids in two genomic sites using the dual-RMCE system.

As demonstrated earlier, single-site targeted integration of >2-copy RMCE plasmids is challenging due to difficulties in verification and transcriptional limitation. As a solution to this, we designed a dual-RMCE system for simultaneous plasmid integration in two genomic sites to create CHO cell lines expressing multiple copies of GOI from independent locations. Our dual-RMCE system has two landing pads integrated into two predefined chromosomal loci enabling simultaneous RMCE in both genomic sites within a single transfection (Figure 4a). To build the dual-RMCE system, we used MP3 MCL having loxP-EF1α-mCherry-NeoR-lox2272 landing pad integrated into site A as a parental cell line. Using CRISPR/Cas9 we integrated a second landing pad loxP-EF1α-tagBFP-HygR-lox2272 into site T9 of MP3 MCL, creating the dual-RMCE master cell line BP5 (Supporting Figure 5a,b). Genomic site T9 was previously identified as a safe harbor site for TI and showed high and stable expression of recombinant proteins in CHO cells[26]. The use of landing pads encoding different fluorescent markers (mCherry and tagBFP) enables selection of clones with RMCE integration in either or both target sites using FACS. The transfection with Cre recombinase and a single RMCE donor plasmid into the dual-RMCE master cell line will result in a mixture of cells, where mCherry-negative/tagBFP-positive cells will have the donor plasmid integrated into site A, mCherry-positive/tagBFP-negative cells will have integration in site T9, double-negative cells will have integration in both site A and site T9. Thus, performing a single transfection and selecting the desired population using FACS, it is possible to generate cell lines with GOI integrated into single or dual sites, controlling the GOI copy number.

To test our dual-RMCE system for expression of therapeutic proteins, we generated dual-RMCE subclones expressing one, two and four copies of EPO and ETN as well as non-producing (NP) clones serving as a control. NP clones were created by co-transfection of Cre recombinase together with RMCE donor plasmid having three stop codons and short synthetic poly(A) tail, mCherry-negative/tagBFP-negative cells were single-cell sorted by FACS and clones were verified to have integration in both site A and site T9 using genomic PCR (Supporting Figure 5c). To create EPO-producing cell lines, we transfected one-copy EPO1 plasmid together with Cre recombinase into dual-RMCE MCL to generate one-copy clones in site A (EPO1_A), one-copy clones in site T9 (EPO1_T9) and

two-copy clones with the one-copy plasmid integrated into both site A and T9 (EPO1_EPO1). Similarly, we transfected two-copy EPO2 plasmid into dual-RMCE MCL to generate two-copy clones in site A (EPO2_A), two-copy clones in site T9 (EPO2_T9) and four-copy clones with two-copy plasmid integrated into both site A and site T9 (EPO2_EPO2). RMCE efficiency measured by FACS was 2-3% for one-copy clones EPO1_A and EPO1_T9, 0.6% for two copy clones EPO1_EPO1, EPO2_A, EPO2_T9, and 0.01% for four-copy clones EPO2_EPO2 (Supporting Table 2). We single-cell sorted and verified the clones using genomic junction PCR (Supporting Figure 6) and copy number analysis (Figure 4b), without any issues with the verification of four-copy clones. Using the same set-up and transfecting ETN1 or ETN2 donor plasmids, we generated ETN-producing cell lines: one-copy clones ETN1_A and ETN1_T9, two-copy clones ETN1_ETN1, ETN2_A, ETN2_T9, and four-copy clones ETN2_ETN2. RMCE efficiency was similar to the efficiency of EPO integration (Supporting Table 2), clones were confirmed to have desired integration using genomic junction PCR (Supporting Figure 7) and copy number analysis (Figure 4e). Thus, we showed that the dual-RMCE system enables simultaneous integration of up to four copies of therapeutic proteins site-specifically into the CHO genome.

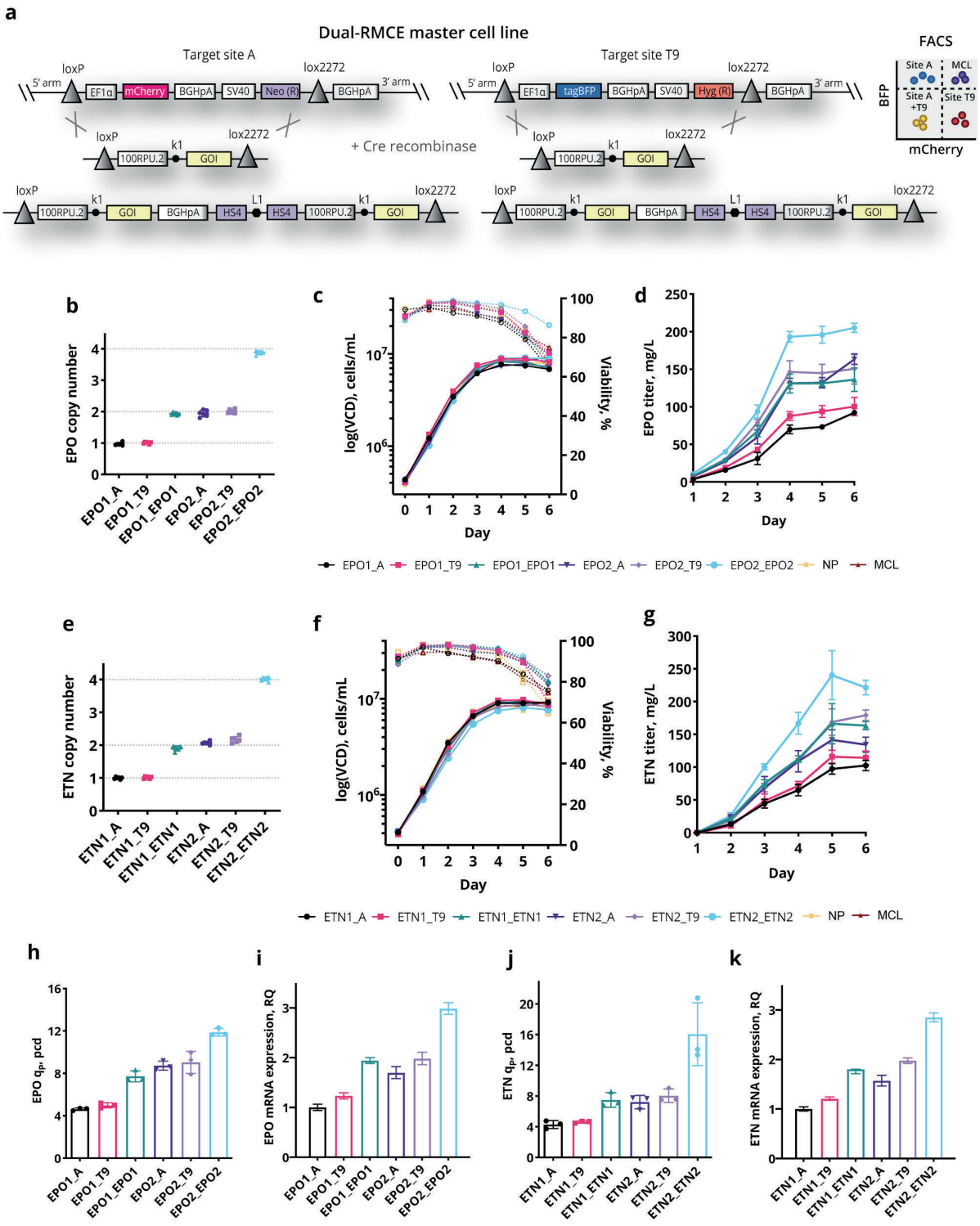


Figure 4. Simultaneous multi-copy integration of GOI in a dual-RMCE system. **a**. Overview of the dual-RMCE system. The dual-RMCE master cell line has two landing pads integrated in two different genomic sites: loxP-EF1α-mCherry-NeoR-lox2272 landing pad in site A and loxP-EF1α-tagBFP-HygR-lox2272 landing pad in site T9. When Cre recombinase is transfected together with a single RMCE donor plasmid (with one GOI copy or two GOI copies), RMCE can occur in site A only, in site T9 only or both sites simultaneously. To select the cells with integration in one specific site or both sites, FACS is used. **b-k**. Characterization of EPO and ETN clones created by targeted

integration of RMCE donor plasmids EPO1, EPO2, ETN1 and ETN2 into the dual-RMCE MCL. Three clones for each cell design were grown in batch culture, together with the MCL and three non-producer (NP) clones. **b, e.** Copy number analysis, showing median value for three biological replicates. **c, f**. Viable cell densities (VCD) and viability, showing mean and SD for three biological replicates. **d, g**. Titer, showing mean and SD for three biological replicates. **h, j**. Specific productivity, showing mean and SD for three biological replicates. **i, k**. mRNA expression, showing mean and SD for three biological replicates.

Dual-RMCE system supports high expression of EPO and ETN

To assess the expression capability of the dual-RMCE system, we selected three dual-RMCE subclones for each EPO-expressing cell design (EPO1_A, EPO1_T9, EPO1_EPO1, EPO2_A, EPO2_T9, EPO2_EPO2) and ETN-expressing cell design (ETN1_A, ETN1_T9, ETN1_ETN1, ETN2_A, ETN2_T9, ETN2_ETN2) and cultivated them in batch mode together with three NP clones and the dual-RMCE MCL. The growth of EPO and ETN clones was uniform and similar to the growth of NP clones and MCL, only the ETN2_ETN2 clones had slightly lower VCD (Figure 4c,f). Titer (Figure 4d,g), q_P (Figure 4h,j) and mRNA levels (Figure 4i,k) of both EPO and ETN were continuously increasing with the number of GOI copies integrated. Expression of EPO and ETN in site A and site T9 was similar, with about 7% higher q_P in site T9 for both proteins (insignificant, F-test, $p = 0.578$). Two-copies EPO and ETN clones showed an expected ~2-fold increase in titer, q_P, and mRNA compared to one-copy clones: q_P of one-copy clones EPO1_A, EPO1_T9 and ETN1_A, ETN1_T9 was around 4.6 pcd, q_P of two-copies clones EPO1_EPO1, EPO2_A, EPO2_T9 and ETN1_ETN1, ETN2_A, ETN2_T9 was around 8 pcd (Figure 4h,j). Instead of an ideal 4-fold change in expression in four-copies clones, the increase in q_P of EPO2_EPO2 clones was only 2.6-fold (12 pcd), while the q_P of ETN2_ETN2 clones was increased only to 3.2-fold (14 pcd, excluding one outlier clone with 21 pcd, which showed q_P of 14 pcd in the follow-up experiments). It can indicate the possible appearance of transcription, translation or secretion bottlenecks with the increased copy number load. mRNA expression of EPO and ETN suggest that the protein production in EPO2_EPO2 and ETN2_ETN2 clones was limited on the level of transcription (Figure 4i,k). Overall, four-copy clones of both EPO and ETN showed the same high specific productivity 12-14 pcd and both EPO2_EPO2 and ETN2_ETN2 reached over 200 mg/L in batch culture with no process optimization.

The increase in protein production puts an additional load on the secretory pathway that may influence folding and post-translational modifications of the secreted proteins. To evaluate if the increase in q_P affected the protein quality, we analyzed ETN in the supernatants of four-copy clones (ETN2_ETN2) versus one-copy clones (ETN1_A). ETN is a homodimer with 13 intra-chain and 3 inter-chain disulfide bonds, two N-glycosylation sites in TNFR domain, one N-glycosylation site in Fc-domain and 13 potential O-glycosylation sites on each monomer[41]. We analyzed ETN N-glycans by LC-MS and found the same N-glycan structures in ETN1_A and ETN2_ETN2 samples, with ETN2_ETN2 having only a minor decrease in N-glycan maturation (Supporting Figure 8a). It is known that when ETN is expressed in CHO cells, it can be secreted as a dimer held together by disulfide bonds, but there is a high probability of disulfide bond scrambling, leading to the formation of high-molecular-weight multimers[41]. Western blot analysis of ETN1_A and ETN2_ETN2 supernatants during batch culture showed no difference between one-copy and four-copies clones in the distribution of dimers and multimers (Supplementary Figure 8b). We conclude that increased protein expression in high-producing four-copies clones generated by the dual-RMCE system did not affect the protein quality.

Dual-RMCE subclones show long-term stability and high protein production in fed-batch culture

Industrial production of biopharmaceuticals requires an assessment of growth and titer stability of the cell lines and optimization of cultivation conditions to ensure consistent and high product yield upon bioprocess. To evaluate the long-term stability of recombinant protein expression and growth of dual-RMCE subclones, we cultivated two cell lines for each EPO- and ETN-expressing cell designs (EPO1_A, EPO1_T9, EPO1_EPO1, EPO2_A, EPO2_T9, EPO2_EPO2, ETN1_A, ETN1_T9, ETN1_ETN1, ETN2_A, ETN2_T9, ETN2_ETN2) for two months by passaging the cells three times a week, with titer measurements in the beginning (week 1), middle (week 4 and 6) and end (week 8) of the cultivation. None of the clones decreased in growth rate or lost specific productivity (Figure 5a,b). The growth rate of ETN clones slightly increased by week 8 leading to a small decrease in q_P, which may be an adaptation to culture conditions in

repeated passages. Overall, all dual-RMCE subclones retained their growth and productivity over 80 generations, proving their stability.

We aimed to increase the volumetric productivity of four-copies clones ETN2_ETN2 and EPO2_EPO2 by bioprocess optimization to show their suitability in an industrial set-up. We cultivated two ETN2_ETN2 clones in the fed-batch mode in shake flasks for 11 days with three different basal media: OptiCHO, FortiCHO, and ActiPro. In all media, cell lines reached high viable cell densities above 10 million cells/mL (Figure 5c), with OptiPro supporting higher VCD at the end of the culture compared to ActiPro and FortiCHO. It resulted in the final ETN titer of ~900 mg/L in OptiCHO, ~800 mg/L in ActiPro and ~620 mg/L in FortiCHO media (Figure 5d). ETN q_P in all three media was in the range of 12-14 pcd (Figure 5e), the same as previously noted in batch culture with CD CHO media. Similarly, EPO2_EPO2 clones cultivated in the fed-batch mode with ActiPro media showed high cell densities, high final titer ~980 mg/L and median q_P of 14.6 pcd (Supporting Figure 9). Thus, high ETN and EPO titers in fed-batch cultivation were achieved by the increase in the duration of the culture and final cell concentrations and were not attributed to the changes in the specific productivity of the cell lines. With further bioprocess development aimed at the maximization of cell mass of these clones, it would be possible to reach the scale of >1g/L protein production in the fed-batch culture.

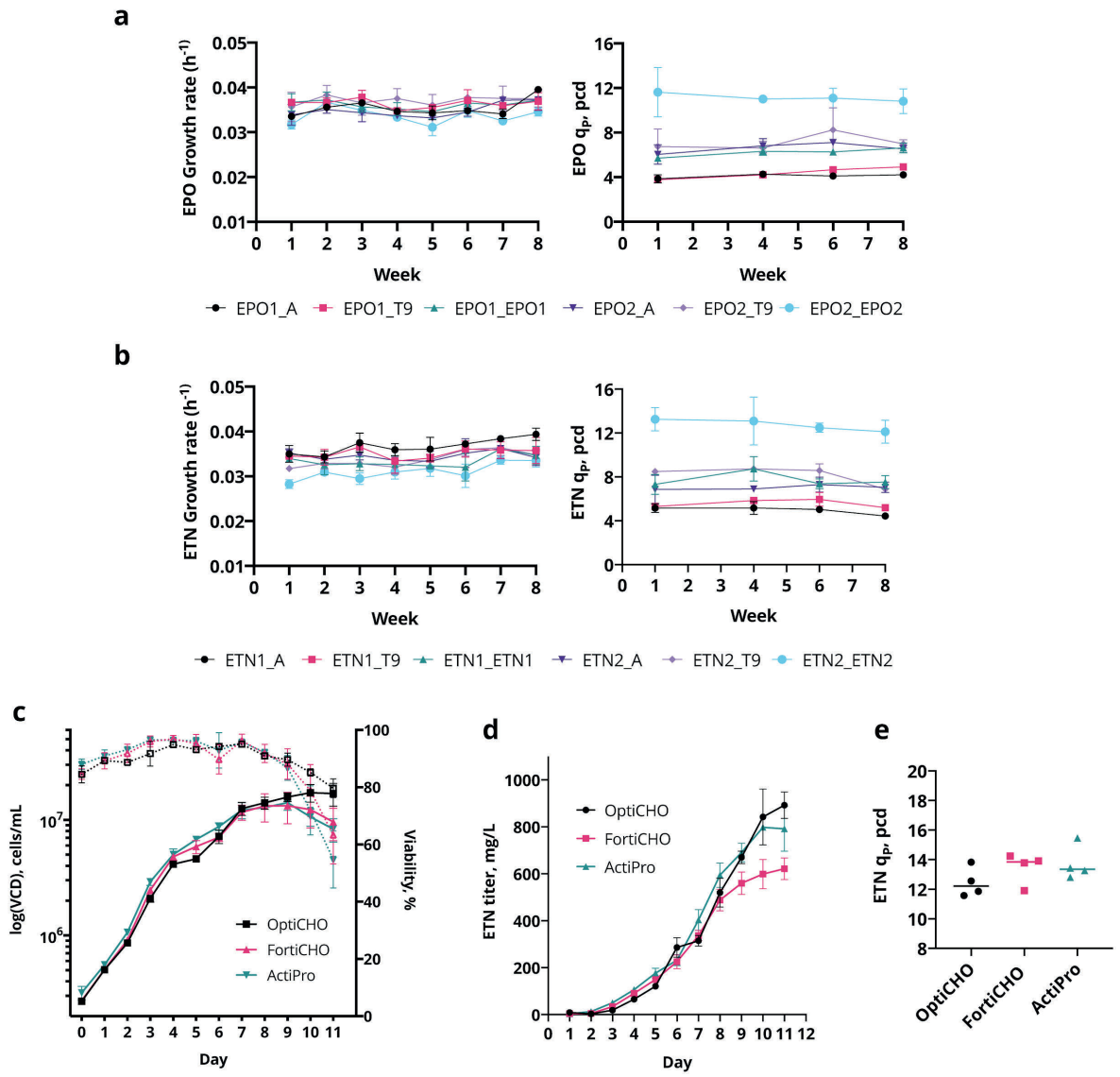


Figure 5. Long-term stability and fed-batch culture performance of multi-copy clones generated by the dual-RMCE system. **a**. Stability of the dual-RMCE EPO clones, represented by growth rate and specific productivity over eight weeks, showing mean and SD for two biological replicates for each cell design. **b.** Stability of the dual-RMCE ETN clones, represented by growth rate and specific productivity over eight weeks, showing mean and SD for two biological replicates for each cell design. **c-e**. Fed-batch cultivation of two ETN2_ETN2 clones in different production media (OptiCHO, FortiCHO, ActiPro), 2 technical replicates per clone. **c**. Viable cell densities (VCD) and viability, showing mean and SD for biological replicates. **d.** ETN titer, showing mean and SD for biological replicates. **e**. ETN specific productivity, showing median value for two clones with two technical replicates.

Discussion

The overarching aim of this study was to increase q_P of therapeutic proteins using the TI system for accelerated and predictable generation of high-producing CHO cell lines. There are four main engineering strategies that can be used to increase q_P upon TI: (i) selecting hotspots in the genome, (ii) using strong gene regulatory elements in the expression vector, (iii) increasing gene copy number and (iv) improving the secretory capacity of the cell. In this study, we combined these strategies to improve the expression of complex therapeutic proteins EPO and ETN in a RMCE system, demonstrating a 10-fold increase in q_P compared to our initial RMCE set-up. In the process, we have explored the importance of different elements in the expression cassette as well as the capacity for introducing multiple gene copies in a single RMCE or multiple RMCE sites. By developing and optimizing the dual-RMCE system for multi-copy integration, we created cell lines that reached q_P of 12-14 pcd for both EPO and ETN, exceeding q_P of the best high-producing clones generated by random integration in our laboratory. Interestingly, both EPO and ETN were expressed at similar q_P, although their protein size and post-translational modifications are different.

To increase q_P upon RMCE, we first explored the effect of gene regulatory elements in the RMCE expression vectors on protein production. We found that the strong synthetic promoter 100RPU.2 supports high expression of recombinant proteins upon TI, resulting in a two-fold increase in q_P compared to the best promoter (CP) from our previous study (Figure 1c). The 100RPU.2 promoter was also protein- and site-independent: the expression of 100RPU.2-EPO (EPO1) and 100RPU.2-ETN (ETN1) constructs in target sites A and T9 was equally high (Figure 4h,j). This illustrates the promise of synthetic biology approaches for designing gene regulatory elements to improve biopharmaceuticals production. In this study we tested just a few regulatory elements to maximize expression from multi-copy vectors due to a small list of mammalian gene regulatory elements characterized to date, selecting only the elements with reported ability to increase protein expression in mammalian cells. Some genetic elements that may positively affect gene expression and stability such as epigenetic regulatory elements S/MAR and UCOE[42] were not included in our study due to long-term

stability of genomic sites A and T9 shown previously[26]. Additionally, the large size of these epigenetic elements (over 1 kb) would negatively affect the integration efficiency of RMCE plasmids. Two regulatory elements tested in our study with previously reported ability to increase protein expression - WPRE that was able to improve protein expression in viral vectors[30] and Kozak sequence k2 that showed increased expression of eGFP in CHO cells[29] - did not perform well, suggesting a significant degree of a construct, cell- or gene-specificity. Construct, cell and gene-specificity of WPRE have been previously noted in other studies[43–45]. Interestingly, both post-transcriptional elements WPRE and Kozak sequence k2 affected transgene mRNA levels (Supporting Figure 1e,f), correlating with specific productivity. A similar effect of 5'UTR sequence on mRNA levels has been observed in yeasts, showing a high correlation between protein and mRNA levels for different 5'UTR sequences in position -10 to -1[46]. Since both WPRE and Kozak sequences are located in UTR, we hypothesise that their presence can affect the overall structure of mRNA, and so may change its stability. The molecular mechanisms of such effects are yet to be uncovered. Another tested element, the chicken HS4 insulator, had a positive effect on the expression from two-copies vectors (Figure 1f,g) presumably by blocking transcriptional interference, thus proving the importance of insulator usage in multi-copy expression cassettes. Vector engineering remains one of the main strategies to boost therapeutic protein expression upon TI, with advances in mammalian synthetic biology and machine learning being crucial for the discovery of novel elements[47]. However, the reported crosstalk between the integration sites and vector regulatory elements might contribute to additional levels of transgene regulation upon TI[26,48], so evaluation of newly discovered vector elements in each target site is required.

The hypothesis that q_P strongly correlates with GOI copy number, although intuitive, was not supported by studies in CHO cell lines generated by random integration. Transgene vector rearrangements and genomic positions appear to have a greater influence on the level of expression than the number of integrated gene copies[23]. With the development of TI methods for mammalian cell line generation, it has become possible to test this hypothesis under controlled conditions with minimal clonal variation. Recent reports of multi-copy mAb expression in TI systems showed that mAb production can be linearly increased with the copy number[21] or increased in a non-linear

fashion[20], and affected by the ratio and position of heavy and light chains in RMCE expression cassette[20]. To assess the effect of gene copy number on protein expression independently from the effect of chain ratio, we expressed single-chain proteins EPO and ETN in our RMCE system. It allowed us to study the correlation between copy number and productivity in a more defined manner and reveal bottlenecks in protein production that occur with an increase in gene copy number.

There are two ways of controlling GOI copy number upon TI: one is multi-copy GOI integration into a single genomic site, another is GOI integration into multiple genomic sites. Here, we evaluated both strategies in the RMCE system in order to improve q_P and identified their limitations. First, we demonstrated that the introduction of optimized two-copy expression vectors in a single site leads to a two-fold increase in q_P and is feasible for both EPO and ETN (Figure 2). However, the insertion of larger four- and six-copy plasmids in a single site was more difficult, limited by a capacity to validate the clones and the low integration efficiency. Overall, the challenge of insert verification makes TI of large multi-copy plasmids less preferable for routine CHO cell line development until more robust methods of clone verification are developed. The observed low integration efficiency can be attributed to two factors: the transfection efficiency of large multi-copy plasmids and RMCE efficiency. The transfection efficiency of large plasmids in mammalian cells is known to be low[49], so improved methods for plasmid delivery such as minicircle DNA may increase TI efficiency[50]. The principle of RMCE places a limitation on TI efficiency due to the requirement of two recombination events happening simultaneously in two recombination sites. The use of a single recombination site for TI can increase integration efficiency of multi-copy plasmids, which was demonstrated for mAbs production in CHO cells[21]. However, the reported mAb expression levels were low (maximum 30 mg/L) and no q_P estimates were shown in the mentioned study, making it difficult to assess the applicability of such TI system for industrial production of biopharmaceuticals. Moreover, TI using a single recombination site has a disadvantage in the integration of bacterial plasmid DNA together with GOI that can affect cell growth and cause transcriptional silencing of transgenes[50].

Importantly, when integrating four and six copies of GOI into a single RMCE site, we saw no increase in protein expression in EPO4, EPO6 and ETN4 clones, which was accompanied by low transgene mRNA level (Figure 2). There are several potential causes of the observed transcriptional limitation, including recombination or silencing of repetitive elements in the genome[51] or transcriptional interference between tandemly assembled gene copies[52,53], which was not prevented by the use of insulators. Similar to our study, accumulative TI of scFv-Fc antibody into a single genomic site, which led to the generation of four- and six-copy clones, showed the appearance of promoter interference with an increase in the number of expression cassettes[54]. Follow-up experiments are required to reveal the cause of low transgene expression in the EPO4, EPO6 and ETN4 clones.

To remove possible promoter interference between multiple expression cassettes in our RMCE system, we created a dual-RMCE MCL for multi-copy integration into separate genomic sites. First, we showed that the integration of two copies in either of our two target sites or one copy in both sites can generate cell lines with almost double the productivity (Figure 4). Then, we created four-copy cell lines EPO2_EPO2 and ETN2_ETN2 with two-copy plasmids being integrated into two genomic sites simultaneously. We observed that the protein expression of both EPO and ETN did not increase proportionally to the gene copy number, as the q_P increase from one to four copies was only 2.6 and 3.2-fold, respectively (Figure 4h,j). The decrease in GOI expression efficiency in four-copy clones was reflected in transgene mRNA levels (Figure 4i,k), pointing to a transcriptional bottleneck. The transcriptional bottleneck was not caused by a transcriptional interference of integrated cassettes in their target sites, because two-copy clones showed a 2-fold increase in q_P individually. Moreover, the verification of clones confirmed that no recombination occurred between the gene copies. Thus, opposed to the common notion that high protein expression in CHO cells is hindered by secretion bottlenecks[55], we observed that a transcriptional bottleneck can appear with an increase in gene dosage.

What can limit transgene mRNA expression when its copy number is increased? A possible reason for ineffective transcription could be a depletion of required transcription factors[56]. Such depletion of the transcription machinery would have caused

changes in global transcription and, consequently, cell growth. However, the growth of the clones was not hampered with an increase in protein production (Figure 4c,f; Figure 5a,b), which suggests that the transcription machinery was not in limited supply. It could be that dual-RMCE clones showing uniform growth adapted to avoid utilizing all available resources towards recombinant protein production, and the selective repression of transgene expression occurred as a feedback regulation to keep the cell growth unchanged. Various post-transcriptional mechanisms can lead to a negative regulation of transgene mRNA expression and turnover, altering its transport and stability. The selectivity of such regulation would be controlled by specific RNA-binding proteins and/or non-coding RNAs. Curiously, experiments in yeasts have shown that mRNA of genes with an increased copy number expressed from endogenous loci can be selectively targeted and degraded by RNA interference[57]. A similar mechanism might be responsible for a decrease in mRNA levels of multi-copy transgenes in CHO cells. To understand the cause of transcriptional bottlenecks observed upon multi-copy expression in our system, further systematic studies of the global cellular response are required.

In summary, our study is the first comprehensive characterization of multi-copy targeted integration for the production of complex therapeutic proteins in CHO cells. The use of the dual-RMCE system enabled accelerated and predictable generation of stable CHO cell lines with high productivity and high titer, thus demonstrating the suitability of TI for industrial production of therapeutic proteins. Due to minimal clonal variation, it was possible to elucidate the relationship between gene copy number and protein expression and reveal a transcriptional bottleneck appearing with an increase in gene dosage. We make our multi-copy RMCE system available for the scientific community and believe that it can aid a better understanding of the response to recombinant protein production in mammalian cells and help develop new strategies for mammalian cell engineering towards the efficient production of biopharmaceuticals.

Methods

Plasmid construction. Plasmids created and used in this study are listed in Supporting Table 3 (with links to Addgene for selected plasmids). All plasmids were constructed by uracil-specific excision reagent (USER) cloning using flexible assembly sequence tags[35]. Fragments for USER assembly (DNA bricks) were generated by PCR amplification with Phusion U Hot Start DNA polymerase (Thermo Fisher Scientific) and uracil-containing primers (IDT). Landing pad plasmids used for the generation of master cell lines MP3, LC1 and dual-RMCE BP5 included 5' homology arm for targeted site (site A or site T9), loxP (ATAACTTCGTATAGCATACATTATACGAAGTTAT), EF1α promoter, mCherry or tagBFP gene, BGH poly(A) tail, SV40 promoter, NeoR or HygR gene, lox2272 (ATAACTTCGTATAGGATACTTTATACGAAGTTAT), BGH poly(A) or SV40 poly(A) tail, 3' homology arm and ZsGreen1-DR expression cassette (CMV_ZsGreen1-DR_BGHpA). RMCE donor plasmids were assembled by USER cloning by combining up to 7 DNA bricks in one reaction. Features of donor plasmids are described in Supporting table 3 and depicted in Figure 1 and Figure 2. They contained one or more copies of GOI (EPO or ETN) and other genetic elements (promoter, poly(A) tail, WPRE, insulator) flanked by loxP site at the 5' end and lox2272 at the 3' end. For cloning of multi-copy plasmids EPO4, EPO6, ETN4 and ETN6 smaller plasmids HS4_100RPU.2_GOI_BGHpA_HS4 were created first, and these were used as a template for PCR amplification of expression units for the USER assembly of full multi-copy plasmids. Each expression unit was flanked by specific restriction sites that were introduced on 5' and 3' ends of USER linkers (listed in Supporting Table 4) to ease the following plasmid and clone verification. RMCE donor plasmid for generation of NP clones had 3 stop codons and short synthetic poly(A) tail[58], flanked by loxP and lox2272 sites. Plasmids and sequences used for the amplification of genetic elements have been described previously: site A and site T9 homology arms, mCherry, NeoR, EF1α, SV40 promoters and poly(A) tails, ZsGreen1-DR expression cassette, *Cricetulus griseus* codon-optimized EPO were amplified from plasmids described in Pristovšek et al.[26]; ETN sequence was described in Pristovšek et al.[24] Sequence of 100RPU.2 promoter (aka 100 RPU construct 2) was described in Brown et al.[27] and was synthesized as gBlock (IDT). tagBFP was amplified from plasmid Addgene#60903, WPRE was amplified from plasmid Addgene#17492, chicken HS4

insulator was amplified from plasmid Addgene#36880. Plasmid backbone was cloned from pJ204 plasmid (DNA 2.0). Assembled DNA bricks were transformed into *E. coli* Mach1 competent cells (Thermo Fisher Scientific) for cloning of landing pad plasmids and one and two-copies RMCE donor plasmids. For the cloning of multi-copy RMCE donor plasmids (EPO4, EPO6, ETN4 and ETN6) Stbl3 competent cells were used (Thermo Fisher Scientific). All constructs were verified by restriction digestion and Sanger sequencing. Multi-copy plasmids EPO4, EPO6, ETN4 and ETN6 were digested prior to sequencing with linker-specific restriction enzymes, generating fragments with separate expression units. These units were then isolated from agarose gel and sequenced. For the transfection in CHO cells, plasmids were purified using NucleoBond Xtra Midi EF kit (Macherey-Nagel) according to manufacturer's instructions.

Cell cultivation. CHO-S cells (Thermo Fisher Scientific) were maintained in CD CHO medium supplemented with 8 mM L-Glutamine (Thermo Fisher Scientific) and cultivated in 125 mL Erlenmeyer shake flasks (Corning), incubated at 37°C, 5% CO_2 at 120 rpm in humidified incubator and passaged every 2−3 days. Viable cell density and viability were monitored using the NucleoCounter NC-200 Cell Counter (ChemoMetec), using Via1-Cassettes and "Viability and Cell Count Method 2" assay.

Master cell lines generation using CRISPR/Cas9. 1×10^6 cells/mL CHO-S cells were transfected with the sgRNA plasmid and landing pad plasmid for the corresponding integration sites, together with a GFP_2A_Cas9 plasmid at a ratio of 1:1:1 (w:w:w) in 6-well plates using FreeStyle MAX transfection reagent (Thermo Fisher Scientific) according to the manufacturer's recommendations. Three days after transfection cells were seeded at 0.3×10^6 cells/mL in 6-well plates in medium containing G418 (600 µg/mL; Sigma-Aldrich) for MP3 and LC1 MCLs generation or Hygromycin (1000 µg/mL; Sigma-Aldrich) for the dual-RMCE BP5 MCL generation. During antibiotic selection, the antibiotic-containing medium was exchanged every 2-3 days. After 9-12 days of selection in suspension, cell pools were subjected to single cell sorting by FACS, where the mCherry-positive/ZsGreen1-DR-negative cells were isolated for the generation of MCLs MP3 and LC1. ZsGreen1-DR was used as a marker to exclude cells that had random integration of the landing pad plasmid. For the generation of dual-RMCE BP5 MCL, the loxP-EF1α-mCherry-NeoR-lox2272-BGHpA landing pad (PL0627) was first integrated in

the site A resulting in creation of MP3 MCL, and then the loxP-EF1α-tagBFP-HygR-lox2272-BGHpA landing pad (PL0626) was integrated into the site T9 of MP3 MCL. Generation of AD1 master cell line was described previously[26]. Detailed protocol for CRISPR/Cas9-mediated targeted integration in CHO cells is described in[40].

Fluorescence activated cell sorting (FACS). For clonal isolation, cell pools were single cell sorted using a FACSJazz cell sorter (BD Biosciences) equipped with 405 nm, 488 nm and 561 nm lasers. Single cells were seeded in flat-bottom Corning 384-well plates (Sigma-Aldrich) in 30 µL of CD CHO medium, supplemented with 8 mM L-Glutamine, 1.5% HEPES (Gibco) and 1×Antibiotic-Antimycotic (Gibco). Ten-fourteen days after single cell sorting, plates were visualized using Celigo image cytometer (Nexcelom Bioscience), and the entire volume of subconfluent clones was transferred to 180 µL of CD CHO medium supplemented with 1×Antibiotic-Antimycotic (Gibco) in flat-bottom 96-well plates using an epMotion 5070 liquid handling workstation (Eppendorf). Subsequently, verified cells were expanded in suspension to 12-well plates, then to 6-well plates and shake flasks.

RMCE subclones generation. MCLs at concentration 1 × 10^6 cells/mL were transfected with RMCE donor plasmid and Cre-recombinase plasmid in 3:1 ratio (w:w) in 6-well plates using FreeStyle MAX transfection reagent (Thermo Fisher Scientific). Cells were passed two times after transfection and then were FACS sorted, using the respective MCLs as a gating control. To generate D1-D3 cell pools, mCherry-negative cells were bulk sorted as described before[26]. For the generation of R1-R6, EPO1-EPO6, ETN1-ETN4 cell lines and MP3 and LC1 RMCE subclones, mCherry-negative cells were single cell sorted as described above. For the generation of dual-RMCE sublones the following sorting scheme was applied: mCherry-negative/tagBFP-positive cells were sorted for selecting clones with recombinase-mediated integration in site A (EPO1_A, EPO2_A, ETN1_A, ETN2_A), mCherry-positive/tagBFP-negative cells were sorted for selecting clones with recombinase-mediated integration in site T9 (EPO1_T9, EPO2_T9, ETN1_T9, ETN2_T9), mCherry-negative/tagBFP-negative cells were sorted for selecting clones with simultaneous recombinase-mediated integration in sites A and T9 (EPO1_EPO1, ETN1_ETN1, EPO2_EPO2, ETN1_ETN1, NP). Subsequently, clones were expanded and verified by PCR of the targeted regions and copy number analysis.

Genomic PCR amplification of targeted regions. Targeted integration was verified by PCR amplification of genomic DNA of targeted locus using specific primers binding outside of the integration site in the genomic locus (5′ and 3′ OUT primers) and primers specific to the 5′ and 3′ ends of the integrated cassette (5′ and 3′ IN primers). Insert PCR was performed by combining forward 5′ OUT primer with reverse 3′ OUT primer. 5′ and 3′ junction PCR was performed by pairing forward 5′ OUT primer with reverse 5′ IN primer and forward 3′ IN primer with reverse 3′ OUT primer, respectively. Schematic representation of PCR set up is shown on Supporting Figure 3a. PCR primers for junction and insert PCR are listed in Supporting Table 5. For junction and insert PCR genomic DNA was extracted from the cell pellets harvested from 96 well plates using QuickExtract DNA extraction solution (Lucigen) according to the manufacturer's instructions. PCR was carried out using 2×Phusion Master Mix (Thermo Fisher Scientific) in touchdown PCR: 98 °C for 30 s; 10×: 98 °C for 10 s, (Tm+10)−Tm°C [−1°/cycle] for 30 s, 72 °C for 30 s or more (6 min for the insert PCR for EPO4, EPO6 and ETN4 clones); 25×: 98 °C for 10 s, Tm°C for 30 s, 72 °C for 30 s or more (6 min for the insert PCR for EPO4, EPO6 and ETN4 clones); 72 °C for 5 min. 1 µL of genomic DNA was used as a PCR template in 10-20 µL of total PCR reaction volume. PCR products were visualized on a 1% agarose gel, and only clones showing correct amplicons were selected.

Gene copy number analysis by qPCR. Copy number analysis of master cell lines and RMCE subclones designed to have one copy of GOI was performed by qPCR on QuantStudio 5 Real-Time PCR System (Applied Biosystems). Genomic DNA was purified using GeneJet Genomic DNA extraction kit (Thermo Fisher Scientific) according to manufacturer's instructions and diluted to 10 ng/µL. TaqMan Gene Expression Master Mix (Thermo Fisher Scientific), custom-made TaqMan assays for GOI (mCherry, tagBFP, EPO, ETN) and one-copy endogenous reference gene C1GALT1C1 (*Cosmc*)[59] were used in duplex assay (one GOI and one reference gene). Sequences of TaqMan probes and primers are listed in Supporting Table 6. qPCR amplification was performed under the following conditions: 50°C for 2 min, 95°C for 10 min; 40×: 95°C for 15 s, 60°C for 1 min. Each experiment included no template control and three technical replicates for each sample. Results were analysed in QuantStudio Design and Analysis software. A comparative C_t method was applied to calculate GOI copy number by comparing $\Delta\Delta C_t$

values of analysed clones to $\Delta\Delta C_t$ values of one-copy GOI clones generated previously[7]. Only clones with a single copy of GOI (RQ 0.7-1.3) were expanded and banked.

Gene copy number analysis by digital PCR. RMCE subclones designed to have two, four and six copies of GOI as well as cell pools D1-D3 were analysed by digital PCR (dPCR) on QuantStudio 3D Digital PCR System (Applied Biosystems). Genomic DNA was purified using GeneJet Genomic DNA extraction kit (Thermo Fisher Scientific) according to manufacturer's instructions. Prior dPCR DNA of multi-copy subclones was digested using restriction enzyme Eco32I (EcoRV) by incubating 1000 ng of genomic DNA with 2 µL FastDigest Eco32I (Thermo Fisher Scientific) at 37°C for 20 hours, then diluted to 6 ng/uL with TE buffer. DNA was mixed with QuantStudio 3D Digital PCR Master Mix v2 and two TaqMan assays (endogenous one-copy gene *Cosmc* and GOI, Supporting Table 6) and loaded on QuantStudio 3D Digital PCR 20K Chips v2 (Thermo Fisher Scientific). dPCR amplification was performed under the following conditions: 96°C for 10 min; 39x: 60°C for 2 min, 98°C for 30 s; 60°C for 2 min. Results were evaluated using Applied Biosystems AnalysisSuite. GOI copy number was calculated as a ratio of GOI copies/µL to *Cosmc* copies/µL.

Batch cultivation. For batch cultivation cells were seeded at 4×10^5 cells/mL in 40 mL CD CHO medium, supplemented with 8 mM L-Glutamine and 2 µL/mL anti-clumping agent in 125 mL Erlenmeyer shake flasks. Cells were incubated in a humidified incubator at 37°C, 5% CO_2 at 120 rpm. VCD and viability were monitored daily using the NucleoCounter NC-200 Cell Counter, samples for titer measurements were harvested daily, samples for RNA expression analysis were harvested on day 3. Cultures were discontinued on day 6.

Titer measurements and q_P calculation. Supernatants during a batch, fed-batch, and stability test were collected by harvesting cell culture and centrifugation at 200g 5 min and stored at -80°C. EPO titers were measured with an Octet RED96 system (ForteBio) using Streptavidin biosensors (ForteBio), functionalized with the anti-EPO VHH biotin conjugate (Life Technologies, cat. no.:7103372100), as previously described[60]. Absolute EPO titers were calculated using a 7-point calibration curve generated from a dilution series of commercially available EPO (Genscript, cat. no.:Z02975-50) in range

1.95-125 mg/L. ETN titers were measured by fluorescent polarization assay with ValitaTiter plates (Valitacell) following manufacturer protocol using Spark plate reader (Tecan). Absolute ETN titers were calculated using a 10-point calibration curve generated from a dilution series of Enbrel (Pfizer) in range 3.125-150 mg/L. EPO and ETN standards and samples, when necessary, were diluted with spent CD CHO media. Specific productivity (q_P, pg/cell/day) was calculated using titer and integral of viable cell (IVC) values for exponential phase of cultivation (from day 1 to day 4 for batch cultures, from day 1 to day 5 for fed-batch). q_P was calculated as a slope of the linear regression line through data points of daily IVC (x-axis) and titer (y-axis) measurements. IVC value for each day was calculated as shown in the equation:

$$IVC = \frac{VCD + VCD_{-1}}{2} + IVC_{-1}$$

where VCD is the viable cell density measured on a specific day, VCD_{-1} is the viable cell density recorded on the previous day, IVC_{-1} is the IVC calculated for the previous day.

Relative mRNA expression analysis by RT-qPCR. Total RNA was extracted using RNeasy Mini Plus Kit (QIAGEN) following the manufacturer's protocol. Samples were treated with DNase and cDNA was synthesized using Maxima First Strand cDNA Synthesis Kit (Thermo Fisher Scientific). RT-qPCR was performed on the QuantStudio 5 Real-Time PCR System using TaqMan Multiplex Master Mix (Thermo Fisher Scientific) in a triplex assay (GOI and two reference genes) using the following amplification conditions: 50°C for 2 min, 95°C for 10 min; 40×: 95°C for 15 s, 60°C for 1 min. Sequences of custom-made Taqman probes for EPO and ETN, reference genes *Gnb1* and *Fkbp1a*[61] as well as primers are listed in Supporting Table 6. Relative mRNA expression levels of GOIs were calculated using Applied Biosystems Analysis Software by comparative C_t method. Each experiment included a "no template" control and three technical replicates for each sample.

Analysis of ETN protein quality by Western blot. SDS-PAGE of culture supernatants was performed in non-reduced conditions using Novex 4-12% Bis-Tris protein gels. Around 100 ng of ETN were loaded per lane according to ETN titer measurements. Proteins were transferred to iBlot2 nitrocellulose membranes using iBlot2 dry blotting system (Thermo Fisher Scientific). After blocking with 5% milk in PBS

with 0.1% Tween buffer, the membranes were incubated with IgG Fc Rabbit anti-human HRP-conjugated antibodies (Thermo Fisher Scientific, cat. no.:10475535) and were visualized using Amersham ECL Prime Western Blotting Detection Reagent on Amersham Imager 600 (GE Healthcare).

N-glycan analysis. For N-glycan analysis, ETN was purified from supernatant samples harvested on day 6 in batch culture. Supernatants were centrifuged at 6000 g for 5 min and purified by protein A chromatography using HiTrap MabSelect Xtra columns (GE Healthcare) on ÄKTA pure protein purification system (GE Healthcare). N-glycans were released from purified ETN and derivatized using GlycoWorks RapiFluor-MS N-Glycan kit (Waters) following the manufacturer's instructions. Labeled N-glycans were analysed by LC-MS using Thermo Ultimate 3000 HPLC with fluorescence detector coupled to Orbitrap Fusion Tribrid mass spectrometer (Thermo Fisher Scientific). Separation was performed on ACQUITY UPLC Glycan BEH Amide 130Å column (Waters). The amount of N-glycans was measured by integrating the areas under the normalized fluorescence spectrum peaks with Thermo Xcalibur software, giving a normalized relative amount of glycans.

Stability test. Prior to the stability test, clones were recovered from cryopreservation for one week by passaging every 2-3 days in 125 mL shake flasks in CD CHO medium supplemented with 8mM L-glutamine, 1% Antibiotic-antimycotic and 2 µL/mL anti-clumping agent. For stability test cells were seeded at 3×10^5 cells/mL in 6-well plates (Corning) and passed every 2-3 days at the same concentration for 8 weeks (80 generations) without antibiotic selection pressure. VCD and viability were monitored using the NucleoCounter NC-200 Cell Counter. Growth rate (µ, h^{-1}) was determined using the following equation:

$$\mu = \frac{Ln(VCD) - Ln(VCD_s)}{\Delta t}$$

where VCD is the viable cell density measured on a specific day, VCD_s is the seeding density, Δt is time in hours.

Fed-batch cultivation. Prior fed-batch cultivation cells were adapted to corresponding media 3 passages before inoculation. Fed-batch cultures were inoculated

at 3×10^5 cells/mL in 30 mL OptiCHO (Thermo Fisher Scientific), FortiCHO (Thermo Fisher Scientific) or ActiPro (GE Healthcare) medium, supplemented with 8 mM L-Glutamine, 1% Antibiotic-antimycotic and 1 µL/mL anti-clumping agent in 125 mL Erlenmeyer shake flasks. Cells were incubated in a humidified incubator at 37°C, 5% CO_2 at 120 rpm, on day 3 temperature was shifted to 33°C. Accelerated feeding was started on day 3 and continued for 4 days with Cell Boost 7a (GE Healthcare) at 1.5%, 3%, 4.5% and 6% of working volume and Cell Boost 7b (GE Healthcare) at 0.15%, 0.3%, 0.45% and 0.6% of working volume. Cultures were maintained at a minimum of 24mM of glucose by supplementing with a 400mM glucose concentrate and at a minimum of 4mM of glutamine by supplementing with a 200mM glutamine concentrate. Metabolite concentrations were determined daily using Bioprofile 400 (Nova Biomedical). VCD and viability were monitored daily using the NucleoCounter NC-250 Cell Counter (ChemoMetec). Cell cultures were terminated when viability was below 80%.

Statistics. All statistics was performed in R. Data were fitted using linear models and the multcomp package used to perform the contrasts described. Adjusted p values corrected for multiple testing is reported.

Supporting Information

Verification of D1-D3 cell pools and R1-R6 clones; Multi-copy plasmids verification (EPO4, EPO6); Verification of EPO1-EPO6 and ETN1-ETN4 clones; Verification of MP3 and LC1 master cell lines and clones AD1_EPO1(R1), MP3_EPO1 and LC1_EPO1; Verification of dual-RMCE BP5 master cell line and non-producer (NP) clones; Verification of dual-RMCE EPO clones; Verification of dual-RMCE ETN clones; Analysis of ETN protein quality by N-glycan analysis and Western blot; Fed-batch cultivation of two EPO2_EPO2 clones; RMCE integration efficiency for cell designs EPO1-EPO6 and ETN1-ETN6; RMCE integration efficiency for dual-RMCE EPO and ETN clones; Plasmids used and created in this study; List of extended USER linkers with restriction sites; Primers used for verification of targeted integration; TaqMan assays; Supporting sequence data.

Acknowledgments

The authors thank Helene Faustrup Kildegaard for the initiation of the project, Saranya Nallapareddy for assistance with landing pad plasmid cloning, Nachon Charanyanonda Petersen and Karen Kathrine Brøndum for assistance with the FACS, Elham Maria Javidi and Marianne Decker for performing fed-batch cultures, Helle Munck Petersen for assistance with the protein purification, Anders Holmgaard Hansen for assistance with N-glycan analysis and Sara Petersen Bjørn for valuable discussions. We greatly acknowledge funding support from the Novo Nordisk Foundation (NNF10CC1016517). D.S. is receiving the funding from NNF Copenhagen Bioscience PhD programme (NNF16CC0020908).

Notes

The authors declare no competing financial interests. A patent application has been filed by third parties on the promoter 100RPU.2 used in this paper, permission for deposition of plasmids containing the promoter sequence has been provided by AstraZeneca and The University of Sheffield.

Author contributions

D.S., L.M.G., G.M.L., L.K.N. designed the study. D.S. performed experiments and analyzed the data. L.K.N. performed statistical analysis. L.M.G., G.M.L., L.K.N. supervised the project. D.S. drafted the manuscript. L.M.G., G.M.L. and L.K.N. reviewed and corrected the manuscript.

References

(1) Walsh, G. (2018) Biopharmaceutical benchmarks 2018. *Nat. Biotechnol. 36*, 1136–1145.
(2) Noh, S. M., Shin, S., and Lee, G. M. (2019) Cell Line Development for Therapeutic Protein Production. In *Cell Culture Engineering* (Lee, G. M., Kildegaard, H. F., Ed.), pp 23-47, Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.
(3) Seth, G., Charaniya, S., Wlaschin, K. F., and Hu, W.-S. (2007) In pursuit of a super producer—alternative paths to high producing recombinant mammalian cells. *Curr. Opin. Biotechnol. 18,* 557-64.
(4) 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.
(5) Kelley, B. (2020) Developing therapeutic monoclonal antibodies at pandemic pace. *Nat. Biotechnol. 38*, 540–545.
(6) Fukushige, S., and Sauer, B. (1992) Genomic targeting with a positive-selection lox integration vector allows highly reproducible gene expression in mammalian cells. *Proc. Natl. Acad. Sci. U. S. A. 89*, 7905–7909.
(7) Grav, L. M., Sergeeva, D., Lee, J. S., Marin de Mas, I., Lewis, N. E., Andersen, M. R., Nielsen, L. K., Lee, G. M., and Kildegaard, H. F. (2018) Minimizing Clonal Variation during Mammalian Cell Line Engineering for Improved Systems Biology Data Generation. *ACS Synth. Biol. 7*, 2148–2159.
(8) Smithies, O., Gregg, R. G., Boggs, S. S., Koralewski, M. A., and Kucherlapati, R. S. (1985) Insertion of DNA sequences into the human chromosomal β-globin locus by homologous recombination. *Nature 317*, 230-234.
(9) Thomas, K. R., and Capecchi, M. R. (1987) Site-directed mutagenesis by gene targeting in mouse embryo-derived stem cells. *Cell 51*, 503–512.
(10) Sauer, B., and Henderson, N. (1990) Targeted insertion of exogenous DNA into the eukaryotic genome by the Cre recombinase. *New Biol. 2*, 441–449.
(11) O'Gorman, S., Fox, D. T., and Wahl, G. M. (1991) Recombinase-mediated gene activation and site-specific integration in mammalian cells. *Science 251*, 1351–1355.
(12) Schlake, T., and Bode, J. (1994) Use of mutated FLP recognition target (FRT) sites for the exchange of expression cassettes at defined chromosomal loci. *Biochemistry 33*, 12746–12751.

(13) 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.
(14) Wirth, D., Gama-Norton, L., Riemer, P., Sandhu, U., Schucht, R., and Hauser, H. (2007) Road to precision: recombinase-based targeting technologies for genome engineering. *Curr. Opin. Biotechnol. 18*, 411–419.
(15) Turan, S., Zehe, C., Kuehle, J., Qiao, J., and Bode, J. (2013) Recombinase-mediated cassette exchange (RMCE) — A rapidly-expanding toolbox for targeted genomic modifications. *Gene 515*, 1-27.
(16) Wiberg, F. C., Rasmussen, S. K., Frandsen, T. P., Rasmussen, L. K., Tengbjerg, K., Coljee, V. W., Sharon, J., Yang, C.-Y., Bregenholt, S., Nielsen, L. S., Haurum, J. S., and Tolstrup, A. B. (2006) Production of target-specific recombinant human polyclonal antibodies in mammalian cells. *Biotechnol. Bioeng. 94*, 396–405.
(17) Crawford, Y., Zhou, M., Hu, Z., Joly, J., Snedecor, B., Shen, A., and Gao, A. (2013) Fast identification of reliable hosts for targeted cell line development from a limited-genome screening using combined φC31 integrase and CRE-Lox technologies. *Biotechnol. Prog. 29*, 1307-1315.
(18) Zhang, L., Inniss, M. C., Han, S., Moffat, M., Jones, H., Zhang, B., Cox, W. L., Rance, J. R., and Young, R. J. (2015) Recombinase-mediated cassette exchange (RMCE) for monoclonal antibody expression in the commercially relevant CHOK1SV cell line. *Biotechnol. Prog. 31*, 1645–1656.
(19) Inniss, M. C., Bandara, K., Jusiak, B., Lu, T. K., Weiss, R., Wroblewska, L., and Zhang, L. (2017) A novel Bxb1 integrase RMCE system for high fidelity site-specific integration of mAb expression cassette in CHO Cells. *Biotechnol. Bioeng. 114*, 1837–1846.
(20) Carver, J., Ng, D., Zhou, M., Ko, P., Zhan, D., Yim, M., Shaw, D., Snedecor, B., Laird, M. W., Lang, S., Shen, A., and Hu, Z. (2020) Maximizing antibody production in a targeted integration host by optimization of subunit gene dosage and position. *Biotechnol. Prog. 36*, e2967.
(21) Gaidukov, L., Wroblewska, L., Teague, B., Nelson, T., Zhang, X., Liu, Y., Jagtap, K., Mamo, S., Tseng, W. A., Lowe, A., Das, J., Bandara, K., Baijuraj, S., Summers, N. M., Lu, T. K., Zhang, L., and Weiss, R. (2018) A multi-landing pad DNA integration platform for mammalian cell engineering. *Nucleic Acids Res. 46*, 4072–4086.
(22) O'Callaghan, P. M., Berthelot, M. E., Young, R. J., Graham, J. W. A., Racher, A. J., and Aldana, D. (2015) Diversity in host clone performance within a Chinese hamster ovary cell line. *Biotechnol. Prog. 31*, 1187–1200.
(23) Noh, S. M., Shin, S., and Lee, G. M. (2018) Comprehensive characterization of glutamine synthetase-mediated selection for the establishment of recombinant CHO cells producing monoclonal antibodies. *Sci. Rep. 8*, 5361.
(24) Pristovšek, N., Hansen, H. G., Sergeeva, D., Borth, N., Lee, G. M., Andersen, M. R., and Kildegaard, H. F. (2018) Using Titer and Titer Normalized to Confluence Are Complementary Strategies for Obtaining Chinese Hamster Ovary Cell Lines with High Volumetric Productivity of Etanercept. *Biotechnol. J. 13*, e1700216.

(25) Ley, D., Seresht, A. K., Engmark, M., Magdenoska, O., Nielsen, K. F., Kildegaard, H. F., and Andersen, M. R. (2015) Multi-omic profiling -of EPO-producing Chinese hamster ovary cell panel reveals metabolic adaptation to heterologous protein production. *Biotechnol. Bioeng. 112*, 2373–2387.
(26) Pristovšek, N., Nallapareddy, S., Grav, L. M., Hefzi, H., Lewis, N. E., Rugbjerg, P., Hansen, H. G., Lee, G. M., Andersen, M. R., and Kildegaard, H. F. (2019) Systematic Evaluation of Site-Specific Recombinant Gene Expression for Programmable Mammalian Cell Engineering. *ACS Synth. Biol. 8*, 758–774.
(27) Brown, A. J., Gibson, S. J., Hatton, D., and James, D. C. (2017) In silico design of context-responsive mammalian promoters with user-defined functionality. *Nucleic Acids Res. 45*, 10906–10919.
(28) Kozak, M. (1987) At least six nucleotides preceding the AUG initiator codon enhance translation in mammalian cells. *J. Mol. Biol. 196*, 947–950.
(29) Petersen, S. D., Zhang, J., Lee, J. S., Jakočiūnas, T., Grav, L. M., Kildegaard, H. F., Keasling, J. D., and Jensen, M. K. (2018) Modular 5'-UTR hexamers for context-independent tuning of protein expression in eukaryotes. *Nucleic Acids Res*. *46*, e127.
(30) Xu, Z.-L., Mizuguchi, H., Mayumi, T., and Hayakawa, T. (2003) Woodchuck hepatitis virus post-transcriptional regulation element enhances transgene expression from adenovirus vectors. *Biochim. Biophys. Acta 1621*, 266–271.
(31) Kim, K.-S., Kim, M. S., Moon, J. H., Jeong, M. S., Kim, J., Lee, G. M., Myung, P.-K., and Hong, H. J. (2009) Enhancement of recombinant antibody production in HEK 293E cells by WPRE. *Biotechnol. Bioprocess Eng*. *14*, 633.
(32) Yahata, K., Maeshima, K., Sone, T., Ando, T., Okabe, M., Imamoto, N., and Imamoto, F. (2007) cHS4 insulator-mediated alleviation of promoter interference during cell-based expression of tandemly associated transgenes. *J. Mol. Biol. 374*, 580–590.
(33) Mutskov, V. J., Farrell, C. M., Wade, P. A., Wolffe, A. P., and Felsenfeld, G. (2002) The barrier function of an insulator couples high histone acetylation levels with specific protection of promoter DNA from methylation. *Genes Dev. 16*, 1540–1554.
(34) Duportet, X., Wroblewska, L., Guye, P., Li, Y., Eyquem, J., Rieders, J., Rimchala, T., Batt, G., and Weiss, R. (2014) A platform for rapid prototyping of synthetic gene networks in mammalian cells. *Nucleic Acids Res. 42*, 13440–13451.
(35) Lund, A. M., Kildegaard, H. F., Petersen, M. B. K., Rank, J., Hansen, B. G., Andersen, M. R., and Mortensen, U. H. (2014) A versatile system for USER cloning-based assembly of expression vectors for mammalian cell engineering. *PLoS One 9*, e96693.
(36) Hommelsheim, C. M., Frantzeskakis, L., Huang, M., and Ülker, B. (2014) PCR amplification of repetitive DNA: a limitation to genome editing technologies and many other applications. *Sci. Rep. 4*, 5052.
(37) Hindson, B. J., Ness, K. D., Masquelier, D. A., Belgrader, P., Heredia, N. J., Makarewicz, A. J., Bright, I. J., Lucero, M. Y., Hiddessen, A. L., Legler, T. C., Kitano, T. K., Hodel, M. R., Petersen, J. F., Wyatt, P. W., Steenblock, E. R., Shah, P. H., Bousse, L. J., Troup, C. B., Mellen, J. C., Wittmann, D. K., Erndt, N. G., Cauley, T. H., Koehler, R. T., So,

A. P., Dube, S., Rose, K. A., Montesclaros, L., Wang, S., Stumbo, D. P., Hodges, S. P., Romine, S., Milanovich, F. P., White, H. E., Regan, J. F., Karlin-Neumann, G. A., Hindson, C. M., Saxonov, S., and Colston, B. W. (2011) High-throughput droplet digital PCR system for absolute quantitation of DNA copy number. *Anal. Chem. 83*, 8604–8610.
(38) Basu, A. S. (2017) Digital Assays Part I: Partitioning Statistics and Digital PCR. *SLAS Technol. 22*, 369–386.
(39) Kallehauge, T. B., Li, S., Pedersen, L. E., Ha, T. K., Ley, D., Andersen, M. R., Kildegaard, H. F., Lee, G. M., and Lewis, N. E. (2017) Ribosome profiling-guided depletion of an mRNA increases cell growth rate and protein secretion. *Sci. Rep. 7*, 40388.
(40) 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. In *CRISPR Gene Editing. Methods in Molecular Biology vol 1961* (Luo Ed.), pp 213-232, Humana Press, New York.
(41) Cho, I. H., Lee, N., Song, D., Jung, S. Y., Bou-Assaf, G., Sosic, Z., Zhang, W., and Lyubarskaya, Y. (2016) Evaluation of the structural, physicochemical, and biological characteristics of SB4, a biosimilar of etanercept. *MAbs 8*, 1136–1155.
(42) Harraghy, N., Calabrese, D., Fisch, I., Girod, P.-A., LeFourn, V., Regamey, A., and Mermod, N. (2015) Epigenetic regulatory elements: Recent advances in understanding their mode of action and use for recombinant protein production in mammalian cells. *Biotechnol. J. 10*, 967–978.
(43) Klein, R., Ruttkowski, B., Knapp, E., Salmons, B., Günzburg, W. H., and Hohenadl, C. (2006) WPRE-mediated enhancement of gene expression is promoter and cell line specific. *Gene 372*, 153–161.
(44) Schambach, A., Wodrich, H., Hildinger, M., Bohne, J., Kräusslich, H. G., and Baum, C. (2000) Context dependence of different modules for posttranscriptional enhancement of gene expression from retroviral vectors. *Mol. Ther. 2*, 435–445.
(45) Mariati, Mariati, Ho, S. C. L., Yap, M. G. S., and Yang, Y. (2010) Evaluating post-transcriptional regulatory elements for enhancing transient gene expression levels in CHO K1 and HEK293 cells. *Protein Expression Purif*. *69,* 9-15.
(46) Dvir, S., Velten, L., Sharon, E., Zeevi, D., Carey, L. B., Weinberger, A., and Segal, E. (2013) Deciphering the rules by which 5'-UTR sequences affect protein expression in yeast. *Proc. Natl. Acad. Sci. U. S. A. 110*, E2792–801.
(47) de Jongh, R. P. H., van Dijk, A. D. J., Julsing, M. K., Schaap, P. J., and de Ridder, D. (2020) Designing Eukaryotic Gene Expression Regulation Using Machine Learning. *Trends Biotechnol. 38*, 191–201.
(48) Maricque, B. B., Chaudhari, H. G., and Cohen, B. A. (2019) A massively parallel reporter assay dissects the influence of chromatin structure on cis-regulatory activity. *Nat. Biotechnol*. doi:10.1038/nbt.4285
(49) Kreiss, P. (1999) Plasmid DNA size does not affect the physicochemical properties of lipoplexes but modulates gene transfer efficiency. *Nucleic Acids Res*. *27*, 3792-3798.

(50) Wang, X., Kawabe, Y., Hada, T., Ito, A., and Kamihira, M. (2018) Cre-Mediated Transgene Integration in Chinese Hamster Ovary Cells Using Minicircle DNA Vectors. *Biotechnol. J. 13*, e1800063.
(51) Garrick, D., Fiering, S., Martin, D. I. K., and Whitelaw, E. (1998) Repeat-induced gene silencing in mammals. *Nat. Genet*. *18*, 56-59.
(52) Shearwin, K., Callen, B., and Egan, J. (2005) Transcriptional interference – a crash course. *Trends Genet*. *21*, 339-345.
(53) Eszterhas, S. K., Bouhassira, E. E., Martin, D. I. K., and Fiering, S. (2002) Transcriptional interference by independently regulated genes occurs in any relative arrangement of the genes and is influenced by chromosomal integration position. *Mol. Cell. Biol. 22*, 469–479.
(54) Wang, X., Kawabe, Y., Kato, R., Hada, T., Ito, A., Yamana, Y., Kondo, M., and Kamihira, M. (2017) Accumulative scFv-Fc antibody gene integration into the hprt chromosomal locus of Chinese hamster ovary cells. *J. Biosci. Bioeng. 124*, 583–590.
(55) Hansen, H. G., Pristovšek, N., Kildegaard, H. F., and Lee, G. M. (2017) Improving the secretory capacity of Chinese hamster ovary cells by ectopic expression of effector genes: Lessons learned and future directions. *Biotechnol. Adv. 35*, 64–76.
(56) Brewster, R. C., Weinert, F. M., Garcia, H. G., Song, D., Rydenfelt, M., and Phillips, R. (2014) The transcription factor titration effect dictates level of gene expression. *Cell 156*, 1312–1323.
(57) Cruz, C., and Houseley, J. (2014) Endogenous RNA interference is driven by copy number. *Elife 3*, e01581.
(58) Levitt, N., Briggs, D., Gil, A., and Proudfoot, N. J. (1989) Definition of an efficient synthetic poly(A) site. *Genes Dev. 3*, 1019–1025.
(59) Yang, Z., Halim, A., Narimatsu, Y., Jitendra Joshi, H., Steentoft, C., Schjoldager, K. T.-B. G., Alder Schulz, M., Sealover, N. R., Kayser, K. J., Paul Bennett, E., Levery, S. B., Vakhrushev, S. Y., and Clausen, H. (2014) The GalNAc-type O-Glycoproteome of CHO cells characterized by the SimpleCell strategy. *Mol. Cell. Proteomics 13*, 3224–3235.
(60) Kol, S., Kallehauge, T. B., Adema, S., and Hermans, P. (2015) Development of a VHH-Based Erythropoietin Quantification Assay. *Mol. Biotechnol. 57*, 692–700.
(61) Brown, A. J., Gibson, S., Hatton, D., and James, D. C. (2018) Transcriptome-Based Identification of the Optimal Reference CHO Genes for Normalisation of qPCR Data. *Biotechnol. J. 13*. doi:10.1002/biot.201700259

Chapter 5 - Transcriptional response to recombinant protein production in isogenic multi-copy CHO cells

This chapter studies how CHO cells respond to increased recombinant protein production. We analyzed isogenic multi-copy cell lines created in the previous study using RNA-seq and examined the effect of different recombinant proteins and increased protein load on the global transcriptome. The use of isogenic cell lines enabled to minimize the clonal variation and reveal common and protein-specific patterns of differential gene expression. Our analysis also pointed to possible mechanisms underlying the limitations of protein expression that appear with the copy number increase.

Manuscript in preparation

Transcriptional response to recombinant protein production in isogenic multi-copy CHO cells

Daria Sergeeva[1], Lise Marie Grav[1], Gyun Min Lee[1,2], Lars Keld Nielsen[1,3]

[1]The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Kgs. Lyngby, Denmark

[2]Department of Biological Sciences, KAIST, Daejeon, Republic of Korea

[3]Australian Institute for Bioengineering and Nanotechnology, University of Queensland, Brisbane, Australia

Abstract

Recombinant therapeutic proteins are crucial medicines for the treatment of human diseases. The production of these proteins relies on mammalian cell factories - Chinese hamster ovary (CHO) cells. Yet, our understanding of how CHO cells react to the burden of recombinant protein expression is limited and greatly hampered by clonal variation. Targeted gene integration mitigates clonal variation during CHO cell line development, offering a robust platform for systems biology studies. Here we surveyed isogenic CHO cell lines generated by multi-copy targeted integration with varying productivity of two biopharmaceuticals (erythropoietin and etanercept) to reveal the adaptive response to increased recombinant protein production. Using RNA-seq analysis, we found common and protein-specific responses that appear with increasing recombinant protein loads. The majority of the differential expression involved relatively small yet significant changes in common sets of genes functionally related to protein processing, lysosomes and cell cycle. We saw that an increase in recombinant protein expression activates predominantly ATF6 and IRE1a branches of unfolded protein response without inducing PERK signaling, therefore increasing protein folding capacity of the cell without inducing ER stress-mediated apoptosis. Moreover, we studied how the choice of genomic sites for targeted integration affects the cell and found that the integration site has a minor influence on the global transcriptome, altering only the expression of adjacent genes. In summary, this study provides insights into the response of CHO cells to recombinant protein production and can help to rationally design CHO cell lines with improved production of biopharmaceuticals.

Introduction

Chinese hamster ovary (CHO) cells are the preferred host for the industrial production of recombinant glycoproteins, which are the foundation of the multi-billion biopharmaceutical industry[1]. Over the past three decades, advances in CHO bioprocessing have led to significant improvements in protein titer; however, the gains through cell line engineering have been modest[2]. The emergence of CRISPR/Cas genome editing tools and the availability of genome sequencing data brought new capabilities for rational CHO cell engineering[3]. Nevertheless, due to the complexity of protein synthesis, folding and secretion in mammalian cells[4–6], the targets for rational engineering remain elusive. To develop strategies for the engineering of high-producing CHO cell lines, it is crucial to understand the cellular mechanisms involved in protein production and unravel how cells respond to recombinant protein production.

There has been a longstanding aspiration of finding the traits that distinguish mammalian cells with low and high production of recombinant proteins in order to engineer more productive cell lines. Since protein productivity is a complex trait[7], analysis of the global cellular response by omics technologies can facilitate our understanding of protein synthesis and help to identify engineering targets. However, previous studies comparing transcriptome, proteome, and metabolome of low- and high-producing mammalian cell lines (CHO, HEK293, NS0) display limited overlap with respect to which pathways and molecules are changing in response to increased recombinant protein production[8,9]. The major issue of these studies was a high biological variability of analyzed cell lines, known as clonal variation. Clonal variation can originate from various factors, but a significant contribution belongs to random transgene integration used for the generation of stable cell lines[10]. When transgene encoding recombinant protein is randomly integrated into the genome, genomic rearrangements might happen and affect multiple processes in the cell. In omics studies, clonal variation hinders the unraveling of productivity traits, masking the genes directly involved in protein production[11,12]. With a high degree of clonal variation, big datasets are required to eliminate effects not related to protein production *per se*.

To minimize clonal variation in protein-producing cell lines, targeted integration can be used instead of random integration. By removing the variability in genome integration sites and copy number of the transgene, targeted integration leads to the generation of isogenic cell lines and gives an advantage for omics studies where smaller datasets become sufficient[13]. Targeted integration has already been exploited in the biopharmaceutical industry for CHO cell line development for therapeutic protein production[14–16]. Therefore, CHO cell lines generated by targeted integration are a perfect system to study the response to high recombinant protein production by omics technologies.

In our previous study[17], we generated isogenic CHO cell lines using multi-copy targeted integration where genes encoding recombinant glycoproteins (erythropoietin (EPO) or etanercept (ETN)) were integrated precisely into pre-selected genomic sites that support high and stable heterologous gene expression (site A and site T9). These multi-copy CHO cells, with a controlled copy number of transgenes ranging from 1 to 4, were expressing increasingly high amounts of recombinant proteins, reaching industrially relevant levels of protein production. Surprisingly, we found that protein expression was limited on the level of transgene mRNA when copy number was increased from 2 to 4 copies.

We used this set of isogenic multi-copy clones to understand how cells react to increased protein production and reveal possible mechanisms of the transcriptional bottleneck appearing at high copy numbers. We evaluated a global transcriptional response to recombinant protein production using RNA sequencing (RNA-seq) and examined (a) the effect of genomic sites of integration, (b) the effect of recombinant proteins, and (c) the effect of increased protein production on the coding transcriptome. The use of isogenic cell lines allowed us to minimize the variation between clones and uncover patterns of differential gene expression caused by the burden of recombinant protein production.

Results

Low transcriptional variation in isogenic multi-copy CHO cell lines

Using multi-copy targeted gene integration, we previously generated isogenic CHO cell lines, where EPO or ETN was integrated into genomic sites A, T9 or both sites simultaneously at increased copy number (see Methods). Cell designs with one (EPO1, ETN1), two (EPO2, ETN2) and four (EPO4, ETN4) copies of GOI were generated, which expressed increasingly high levels of recombinant proteins (Figure 1a). We chose three biological replicates (clones) per cell design and compared them to their common progenitor - master cell line (MCL - BP5C9), and three non-producing subclones (NP). These cell lines were grown in batch culture and were characterized in terms of growth and protein expression (Figure 1b,c,d). We observed that the growth of all cell lines was uniform, and the specific productivity (q_P) increased with GOI copy number. However, a transcriptional limitation was observed when GOI copy number was increased from 1 to 4: a fold change increase in specific productivity and transgene mRNA levels was only 2.5-3-fold instead of the ideal 4-fold change in 4 copy clones EPO4 and ETN4 (Figure 1c,d).

To reveal the global response to recombinant protein production, we collected samples from all clones on day three in the exponential phase of batch cultivation for RNA-seq of the coding transcriptome. We performed two RNA-seq runs: run 1 with EPO-expressing clones (P1-P18) and run 2 with ETN-expressing clones (T1-T18); samples from NP and MCL were included in both runs as controls for batch effects. 15-20 million single-end reads were generated per sample and mapped onto the Chinese hamster transcriptome. After filtering to remove non-expressed genes, 12374 CHO transcripts plus 6 transgenes (EPO, ETN, mCherry, tagBFP, NeoR and HygR) (total 12380 genes) were retained for further analysis. The Pearson's correlation coefficient showed a strong reproducibility of biological replicates and overall high similarity between clones expressing EPO or ETN at different copy number, MCL and NP clones (Pearson's $r>0.95$) (Supporting Figure 1). To explore the data, we performed principal component analysis (PCA). Transgenes (EPO, ETN, mCherry, tagBFP, NeoR, HygR) were excluded from the PCA to avoid biased clone separation due to their differential expression. The PCA based on top 500 variable Chinese hamster genes showed distorted clustering due to top variable genes being repetitive sequences, which are homologous to endogenous retroviruses

(Supporting Figure 2 and 3; Supporting Table 1). To remove the effect of the cross-mapping repetitive sequences, we performed PCA based on the top 500 variable genes that had human homologs (Figure 2). Little variation in the gene expression profiles was observed: the first principal component (PC1) accounted for only 13% of the variation and the PC2 accounted for 9% of the variation. PC1 showed some separation of the samples according to the recombinant protein expressed in the clones. Overall, these data show low transcriptional variation between isogenic CHO cells expressing different recombinant proteins at increasing protein load.

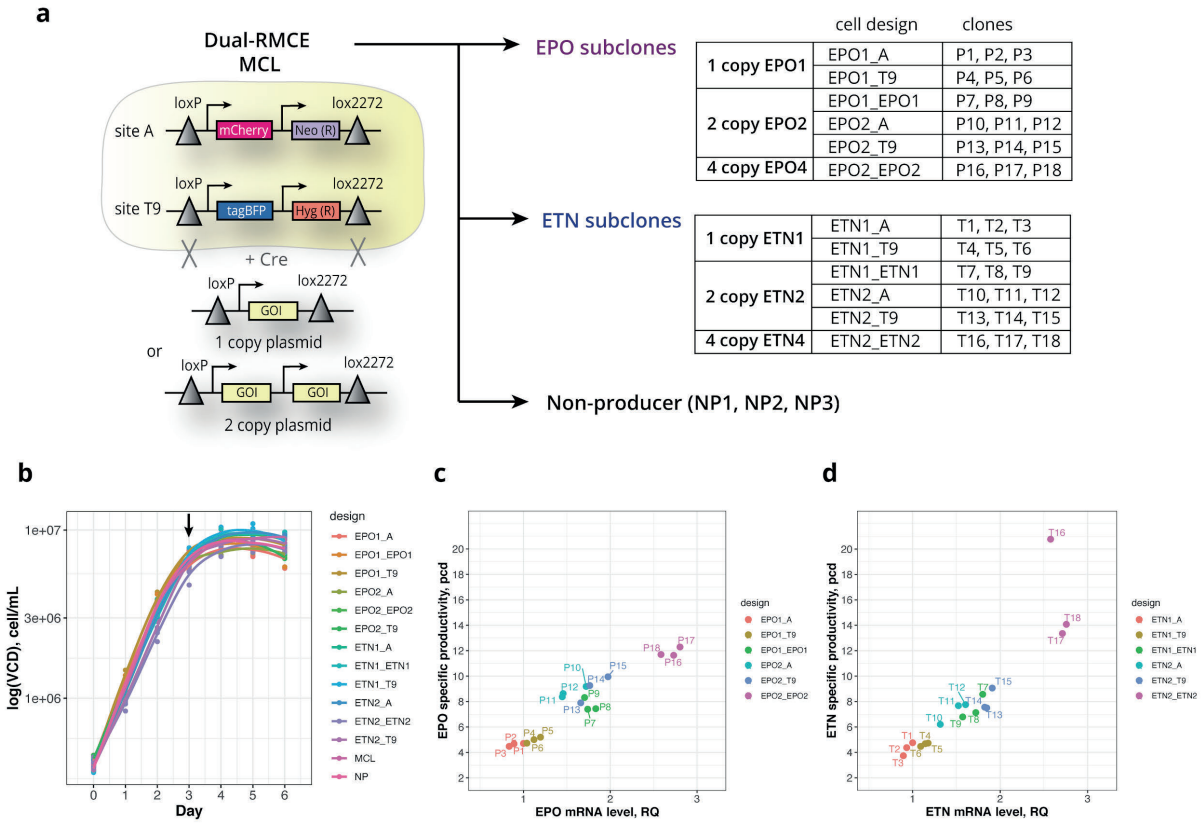


Figure 1. Expression of EPO and ETN in isogenic multi-copy CHO cells. **a.** Overview of CHO cell lines used in this study. Dual-RMCE MCL is the progenitor of protein-producing EPO and ETN subclones and non-producing (NP) subclones. EPO or ETN were integrated into genomic site A, T9, or both, generating cell lines with 1, 2 and 4 copies of GOI (cell designs). Three biological replicates (clones) were chosen for each cell design. **b-d**. Characterization of multi-copy clones in batch cultivation. Three clones for each cell design were grown in batch culture, together with the MCL and three NP clones. **b**. Viable cell densities (VCD), showing a mean for three biological replicates. **c.** EPO specific productivity (pg/cell/day (pcd)) and relative mRNA levels (RQ) measured by RT-qPCR, shown for each clone. **d.** ETN specific productivity and relative mRNA levels measured by RT-qPCR, shown for each clone. RQ values were normalized to cell design EPO1_A or ETN1_A.

We analysed the level of transgene expression and found that EPO and ETN were the top expressed transcripts in all protein-producing clones. EPO expression level (transcripts per million (TPM)) was ~3.3 fold higher than the level of the top expressed endogenous gene GAPDH in clones EPO4; ETN expression level was 4.8-fold higher than GAPDH in clones ETN4. Transcripts of recombinant proteins accounted for a significant share of the whole transcriptome: EPO1 clones - 0.96%, EPO2 - 1.72%, EPO4 - 2.66%, ETN1 - 1.48%, ETN2 - 2.42%, ETN4 - 3.63% of total TPM. Relative TPM of EPO and ETN corresponded to q_P and mRNA levels measured previously by RT-qPCR. A transcriptional limitation in transgene TPM was observed in 4 copy clones EPO4 and ETN4 (2.6-3.6-fold change in transcripts from 1 copy to 4 copy instead of the ideal 4-fold) (Supporting Figure 4). Thus, RNA-seq confirms the appearance of a transcriptional bottleneck of recombinant proteins in the 4 copy clones EPO4 and ETN4.

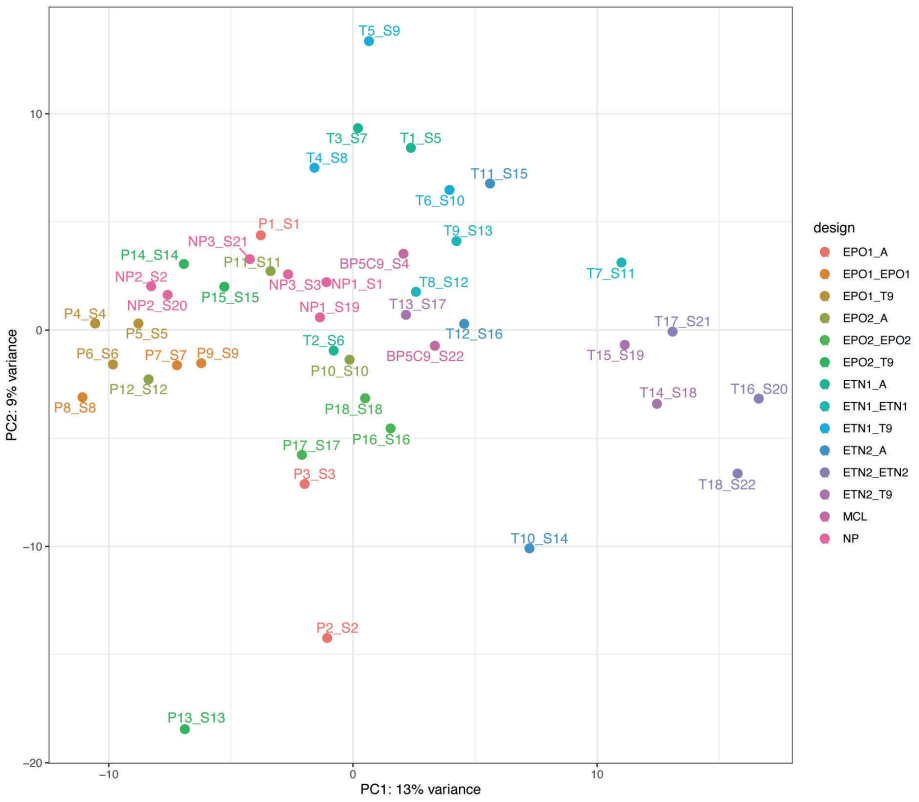


Figure 2. Principal component analysis (PCA) of all samples based on top 500 variable transcripts excluding transgenes (only genes with human homologs were included in top 500).

Effect of integration sites and landing pad genes on global expression

We performed differential expression analysis between MCL and NP clones to reveal the effects of landing pad genes (mCherry, tagBFP, NeoR, HygR) on the global transcriptome. Only 7 genes were found as significantly differentially expressed (no log fold change (LFC) cut-off, false discovery rate (FDR) < 0.05) (Figure 3a). Five genes (PRNP, BAG3, CYLD, HSPH1, and DNAJB1) were upregulated in MCL, three of which are related to the unfolded protein response (HSPH1, DNAJB1, and BAG3) and one is prion protein (PRNP). It indicates that even expression of reporter (mCherry, tagBFP) and/or selection marker genes (NeoR, HygR) can lead to unfolded protein response.

Next, we examined if the choice of the integration site had an effect on global transcription. We compared the following cell lines with the same GOI copy number, but different integration sites (site A or site T9): EPO1_A vs EPO1_T9, EPO2_A vs EPO2_T9, ETN1_A vs ETN1_T9, ETN2_A vs ETN2_T9, EPO1_EPO1 vs EPO2_A, EPO1_EPO1 vs EPO2_T9, ETN1_ETN1 vs ETN2_A, ETN1_ETN1 vs ETN2_T9. We found only 3 significantly differentially expressed genes (RRM1, POLK and COL4A3BP) across all 6 comparisons (no LFC cut-off, FDR < 0.05) (Figure 3b). RRM1 was upregulated in clones with GOI inserted into site T9, POLK and COL4A3BP were upregulated in clones with GOI integrated into site A. All three genes are located in the proximity of the integration sites[18]. RRM1 (a subunit of ribonucleotide reductase involved in the DNA synthesis) is a downstream gene (+ strand) of the intergenic site T9. Site A locates in the intron of POLK, which is a DNA polymerase involved in DNA repair during DNA replication. COL4A3BP is an upstream gene (- strand) of target site A and encodes a ceramide transfer protein CERT. Overall, the effect of integration sites on the global transcriptome was small, therefore in the further analysis we grouped together cell designs with different integration sites but the same GOI copy number into groups EPO1, EPO2, EPO4 and ETN1, ETN2, ETN4 (Figure 1a).

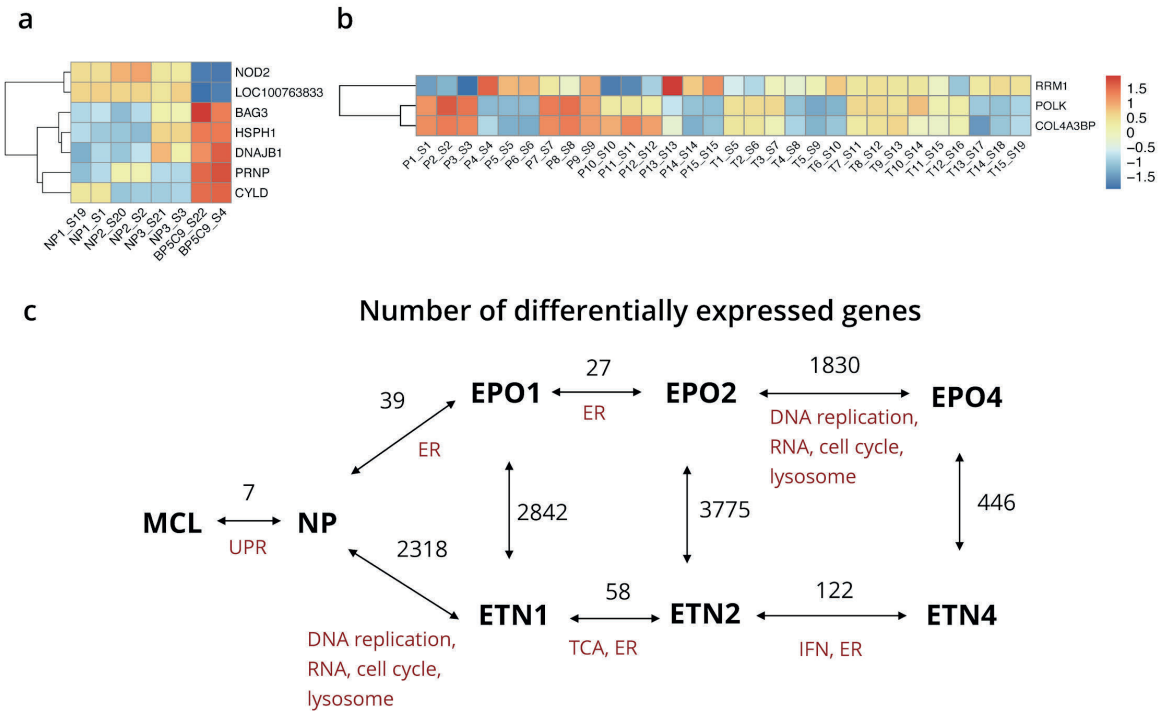


Figure 3. Differential expression analysis of multi-copy CHO cells. **a.** Heatmap for genes that are differentially expressed between NP and MCL. **b.** Heatmap for genes that are site-specifically differentially expressed. **c.** Overview of differential expression analysis. Number of differentially expressed genes (no LFC cut-off, FDR < 0.05) and top enriched pathways are indicated.

Differential expression analysis - EPO

To evaluate the response of CHO cells to increased EPO production, we did a pairwise differential expression analysis between groups with increased EPO expression NP-EPO1-EPO2-EPO4 (Figure 3c). We revealed that the largest transcriptional changes were between groups EPO2 vs EPO4 and the smallest changes were between EPO1 vs EPO2 and EPO1 vs NP (Figure 3c). Pairwise comparison of groups EPO1 and NP showed 39 differentially expressed genes (no LFC cut-off, FDR < 0.05) (Figure 4a). Eight genes were upregulated in EPO1 cells, 3 genes of them were site-specific (RRM1, POLK, and COL4A3BP), and 3 genes were related to protein transport (SEC63, LOC100752724 (SEC16B), and DNAJC13). Gene enrichment analysis showed that differentially expressed genes were represented in KEGG pathways “protein processing in endoplasmic reticulum (ER)” (including genes RPN1, SSR2, ATF6B, LMAN2, and SEC63), “lysosome”

(FUCA1, GUSB, DNASE2, and CTLB) and “other glycan degradation” (FUCA1 and MAN2B2) (Figure 4b). Comparison between the groups EPO1 and EPO2 showed that 27 genes were significantly differentially expressed: 17 genes were downregulated and 10 genes were upregulated in EPO2 (no LFC cut-off, FDR < 0.05) (Figure 4c). The genes were significantly enriched in the KEGG pathways “protein processing in ER” (ATF6, ERO1A, GANAB, PRKCSH, RPN2, SSR1, and SYVN1) and “N-glycan synthesis” (GANAB and RPN2) and “lysosome” (AGA and GUSB) (Figure 4d). Among the significantly upregulated genes in EPO2 were TMEM50B (transmembrane protein associated with ER), ISYNA1 (inositol-3-phosphate synthase 1), SDF2L1 (ER-localized protein), NANS (sialic acid synthase), and MANF (protect cells from ER stress-induced cell death). LOC100752363 (nodal modulator 1) and GUSB (beta-glucuronidase) were top significantly downregulated genes (Figure 4c).

We observed a significant shift in global transcription when EPO copy number was increased from 2 to 4 (EPO2 vs EPO4), 1830 genes were differentially expressed out of 12374 (14.8%). 844 genes were downregulated and 986 genes were upregulated in EPO4 samples (Figure 5a). Upregulated genes were enriched in KEGG pathways “DNA replication”, “spliceosome”, “cell cycle”, “protein processing in ER”, “RNA transport” and pathways related to DNA repair (Figure 5b). Downregulated genes were enriched in KEGG pathways “lysosome”, “oxidative phosphorylation” and “other glycan degradation” (Figure 5c). Gene ontology analysis of these genes showed a significant upregulation of genes in mRNA processing, response to stress, DNA metabolic process, biosynthetic process and cell cycle-related ontologies, and downregulation in lipid metabolic process, autophagy, catabolic process and transport-related ontologies (Figure 5d). Top significantly upregulated genes were ATP2A2 (calcium ATPase SERCA2), GMPPB (catalyzes synthesis of GDP-mannose), DNAJC3 (chaperone), ERO1B (oxidoreductase), UNG (uracil-DNA glycosylase), TMEM50B, SDF2L1 and ISYNA1. CTSF (cathepsin F) was the top significantly downregulated gene.

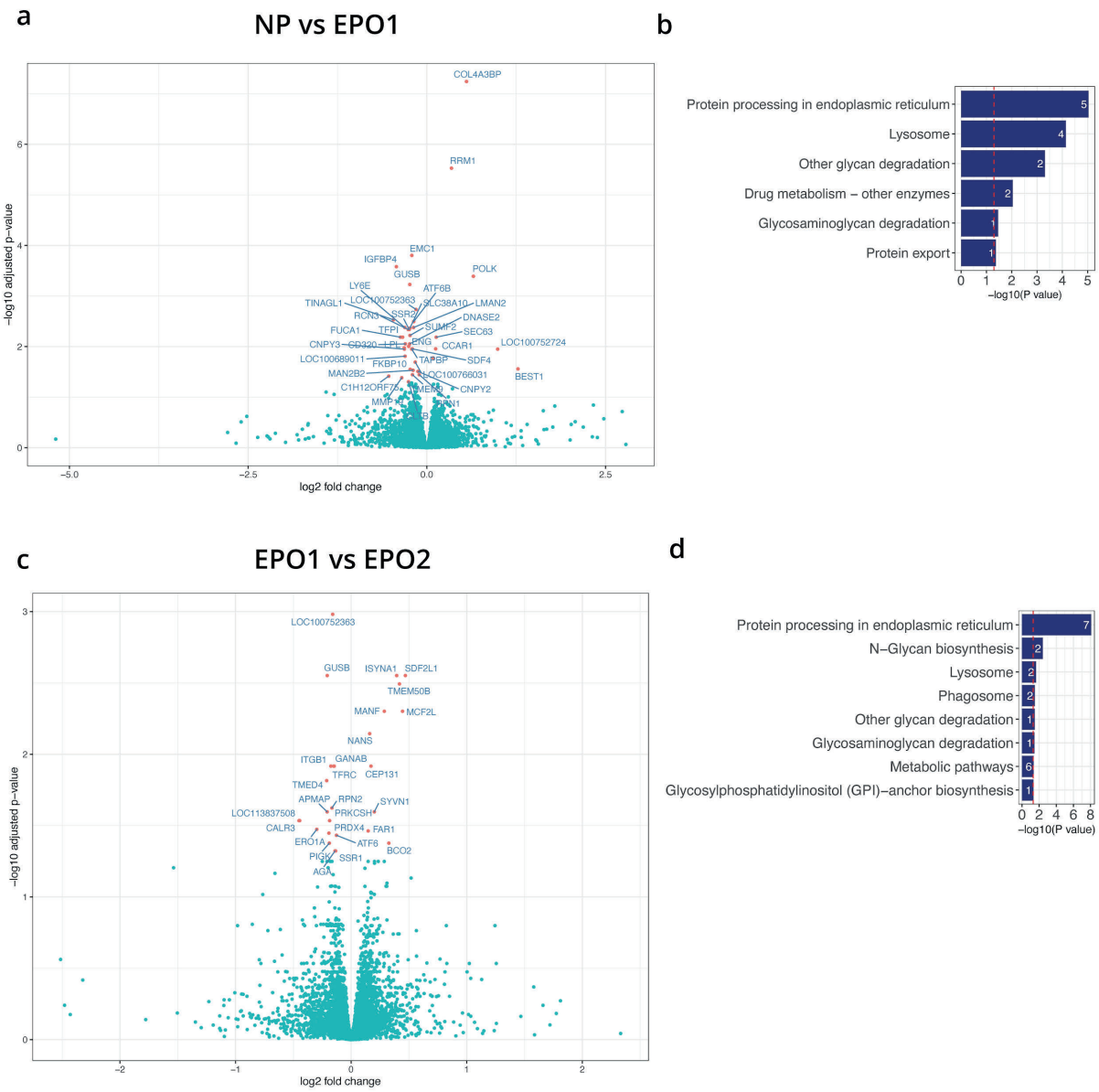


Figure 4. Analysis of groups NP vs EPO1 and EPO1 vs EPO2. **a.** Volcano plot showing fold-change and p-value for the comparison NP vs EPO1. Differentially expressed genes with p-value <0.05 and no LFC cut-off are depicted in red and are labeled. **b.** Top KEGG pathways enriched for significantly differentially expressed genes between EPO1 and NP. **c.** Volcano plot showing fold-change and p-value for the comparison EPO1 vs EPO2. Differentially expressed genes with p-value <0.05 and no LFC cut-off are depicted in red and are labeled. **d.** Top KEGG pathways enriched for significantly differentially expressed genes between EPO1 and EPO2. In KEGG barplots, the red line shows p-value cut-off (0.05), the white number is the number of genes in the pathway that were among differentially expressed genes.

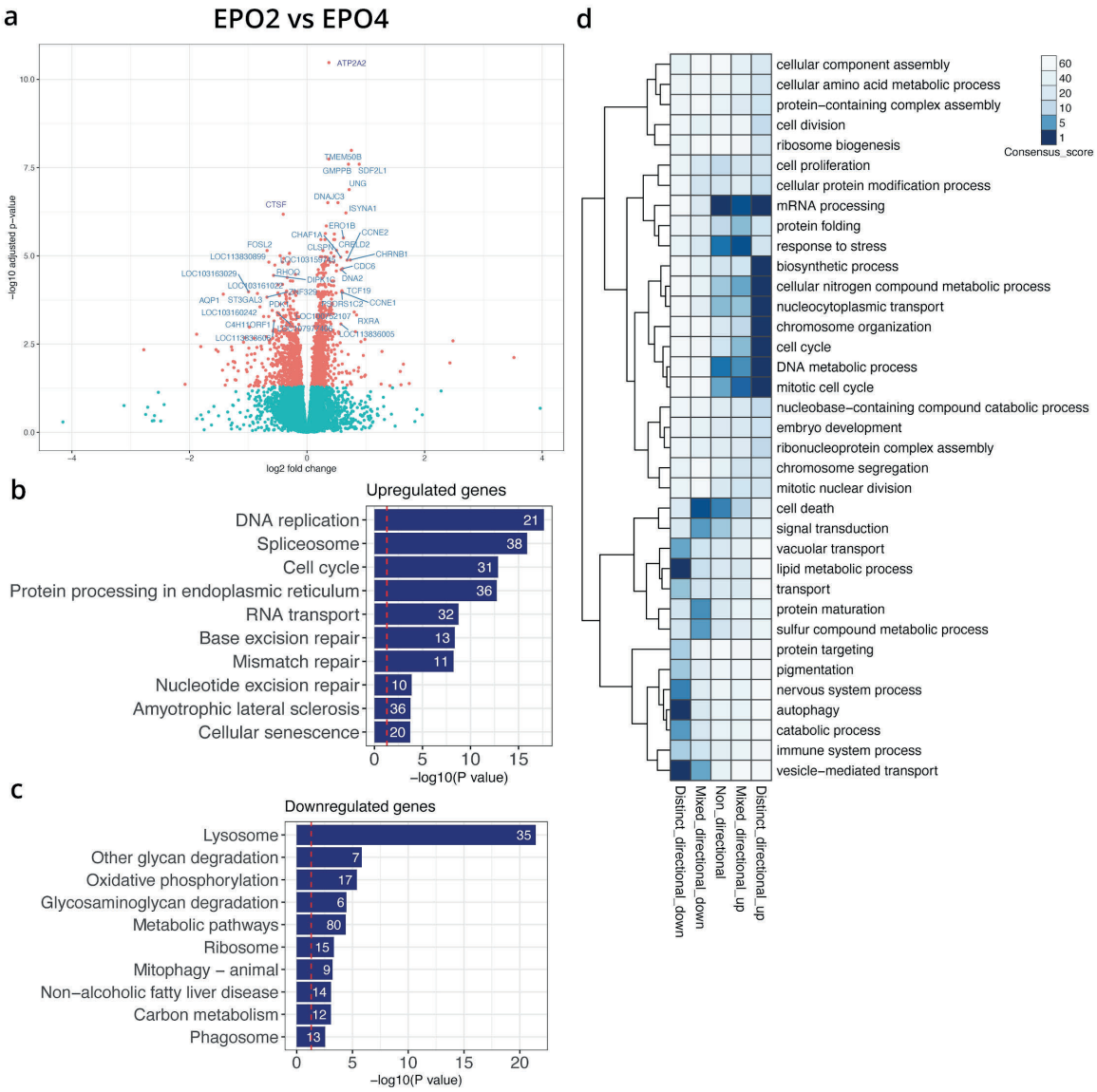


Figure 5. Analysis of groups EPO2 vs EPO4. **a.** Volcano plot showing fold-change and p-value for the comparison EPO2 vs EPO4. Differentially expressed genes with p-value <0.05 and no LFC cut-off are depicted in red. Genes with LFC > 0.5, p-value < 0.001 are labeled, including ATP2A2 and CTSF with LFC < 0.5. **b.** Top 10 KEGG pathways enriched for significantly upregulated genes between EPO2 and EPO4. **c.** Top 10 KEGG pathways enriched for significantly downregulated genes between EPO2 and EPO4. **d.** Consensus heatmap for gene ontology categories “biological process” for differentially expressed genes between EPO2 and EPO4. Gene sets that are significant in at least one directionality class (p-value < 0.05) are shown.

Next, we performed unsupervised clustering of all differentially expressed genes between conditions NP, EPO1, EPO2, and EPO4 to find synexpression gene clusters - common patterns in gene expression. We identified 4 synexpression clusters that displayed correlated expression during the increase in EPO copy number: cluster 1 (616 genes, linearly upregulated), cluster 2 (567 genes, linearly downregulated), cluster 3 (699 genes, non-linearly upregulated), and cluster 4 (708 genes, non-linearly downregulated) (Figure 6a). For each cluster, we extracted the core contributing genes (Supporting Table 2) and performed KEGG pathway enrichment analysis on genes that had human homologs. Cluster 1 was represented by core genes that linearly increased with the increase in recombinant protein expression. It included 65 core genes that were enriched in pathways related to protein processing in ER and protein export (HERPUD1, HSP90B1, SEC61B, SEL1L, and SRPRA) (Figure 6b). Cluster 2 included 90 core genes that linearly decreased with the increase in copy number and were significantly enriched in pathways related to the lysosome (ASAH1, CD164, CTSO, GALNS, GLB1, GUSB, LAMP1, HEXB, HEXB, NAGA, SMPD1, and SORT) (Figure 6c). Cluster 3 included 97 core genes that had a non-linear increase with the increased copy number, which were upregulated in transition from EPO2 to EPO4. They were enriched in pathways related to DNA repair and replication (EXO1, POLD1, POLD3, RFC2, RFC5, and MCM2), spliceosome and RNA transport (EIF4A3, PPIL1, SF3A2, SNRPA, EIF2B1, PRMT5, and RAN) (Figure 6d). Cluster 4 included 135 core genes that were non-linearly downregulated in the transition from EPO2 to EPO4, which were enriched in mitochondria-related pathways (ATF4 (transcription factor involved in unfolded protein response), BNIP3L, CALCOCO2, USP8, BCL2L11, NDUFB4, NDUFS4, and NDUFA10), ribosome (RPL36A, RPL23, RPL5, and RPS15A) and RNA polymerase III subunits (POLR3E and POLR3GL) (Figure 6e).

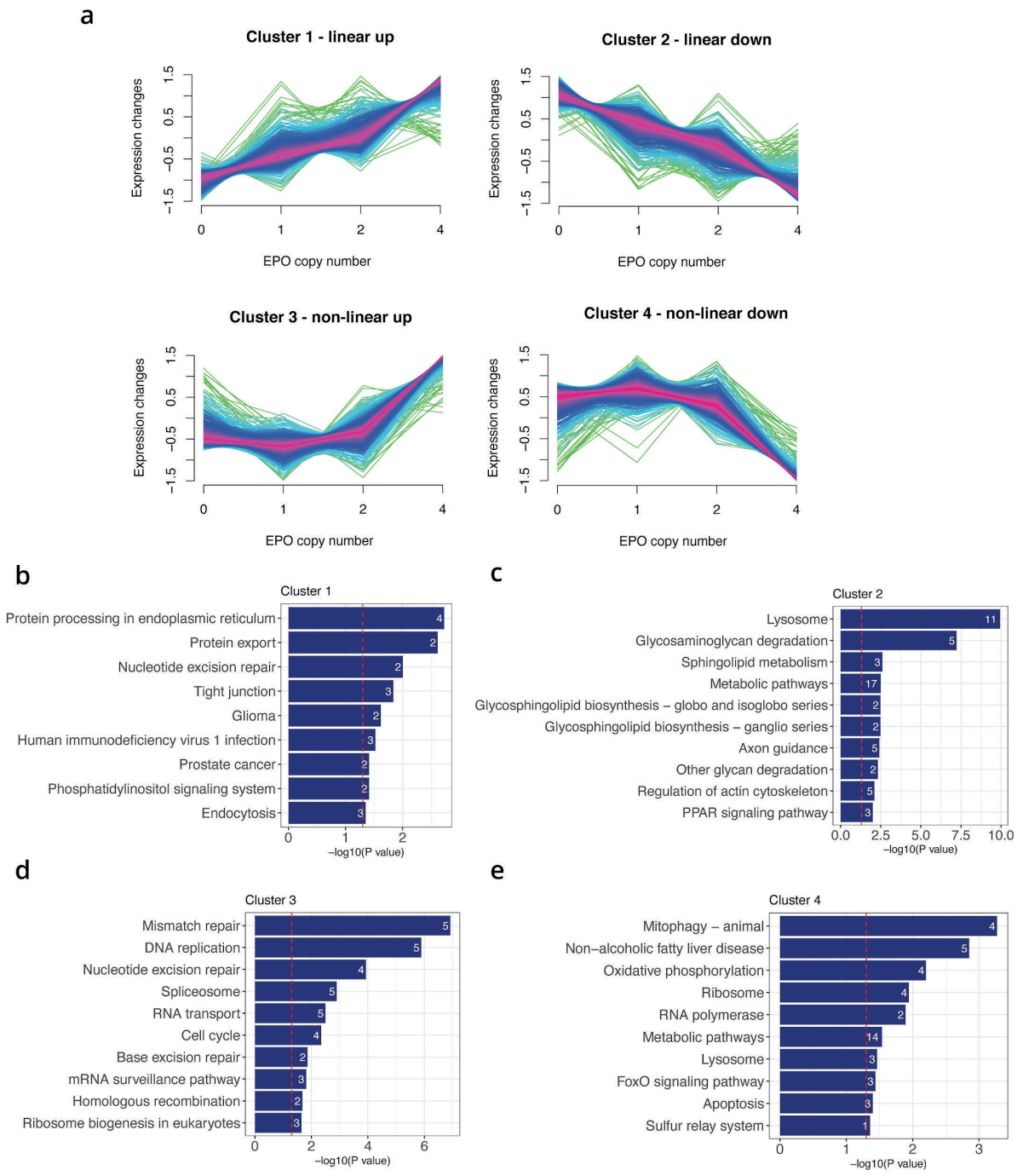


Figure 6. Patterns of gene expression correlated with EPO copy number increase. **a.** Four gene synexpression clusters identified by fuzzy clustering. **b-e.** Top KEGG pathways enriched for core genes in EPO clusters 1-4.

Differential expression analysis - ETN

Analogous to EPO cell lines, we performed differential expression analysis between groups with increased ETN expression NP-ETN1-ETN2-ETN4. The largest transcriptional changes were between NP vs ETN1 while the smallest changes were between ETN1 vs ETN2 and ETN2 vs ETN4 (Figure 3c). Pairwise comparison ETN1 vs NP revealed 2318 differentially expressed genes: 1163 genes were downregulated and 1155 genes were upregulated in ETN1 samples. CTSF was the top significantly downregulated gene (Figure 7a). Gene enrichment analysis showed that upregulated differentially expressed genes were represented in KEGG pathways related to spliceosome, DNA replication and DNA repair, RNA transport and ribosome, protein processing in ER (Figure 7b). Downregulated genes were represented in KEGG pathways related to lysosome and glycan degradation (Figure 7c). Gene ontology analysis of these genes showed a significant upregulation of genes related to translation, biosynthetic process, chromosome organization, mRNA processing, DNA metabolic process, and cell cycle-related ontologies, and downregulation in extracellular matrix organization, cell adhesion and vesicle-mediated transport (Figure 7d).

Differential expression analysis between groups ETN1 and ETN2 showed that 58 genes were significantly differentially expressed: 25 genes were downregulated and 33 genes were upregulated in ETN2 cells (Figure 8a). KEGG pathway analysis showed that differentially expressed genes were enriched in TCA cycle (OGDH and SDHA) and protein processing in ER (ATF6, CRYAB, and ERO1B) (Figure 8b). NANS, MANF, TMEM50B, ERO1B, and SDF2L1 were among the top significantly upregulated genes. Curiously, interferon regulatory factor 7 (IRF7) and a group of interferon-induced proteins were among significantly upregulated genes in ETN2 cell lines.

The increase in ETN copy number from 2 to 4 (ETN2 vs ETN4) resulted in differential expression of 122 genes: 30 genes were downregulated and 92 genes were upregulated in ETN4. The majority of significantly differentially expressed genes with LFC ≥ 1 were related to interferon response (Figure 8c). KEGG pathway analysis on genes with human homologs showed significant enrichment in virus infection-related pathways (CXCL10, IRF7, IRF9, OAS1 MX1, SLC25A5, STAT1, STAT2, and ISG15), NOD-like receptor signaling (IRF7, IRF9, GBP2, GBP3, OAS1, STAT1, and STAT2) and

protein processing in ER (DNAJB11, DNAJC3, GANAB, PDIA3, RAD23A, SEC31B, and SEC63) (Figure 8d).

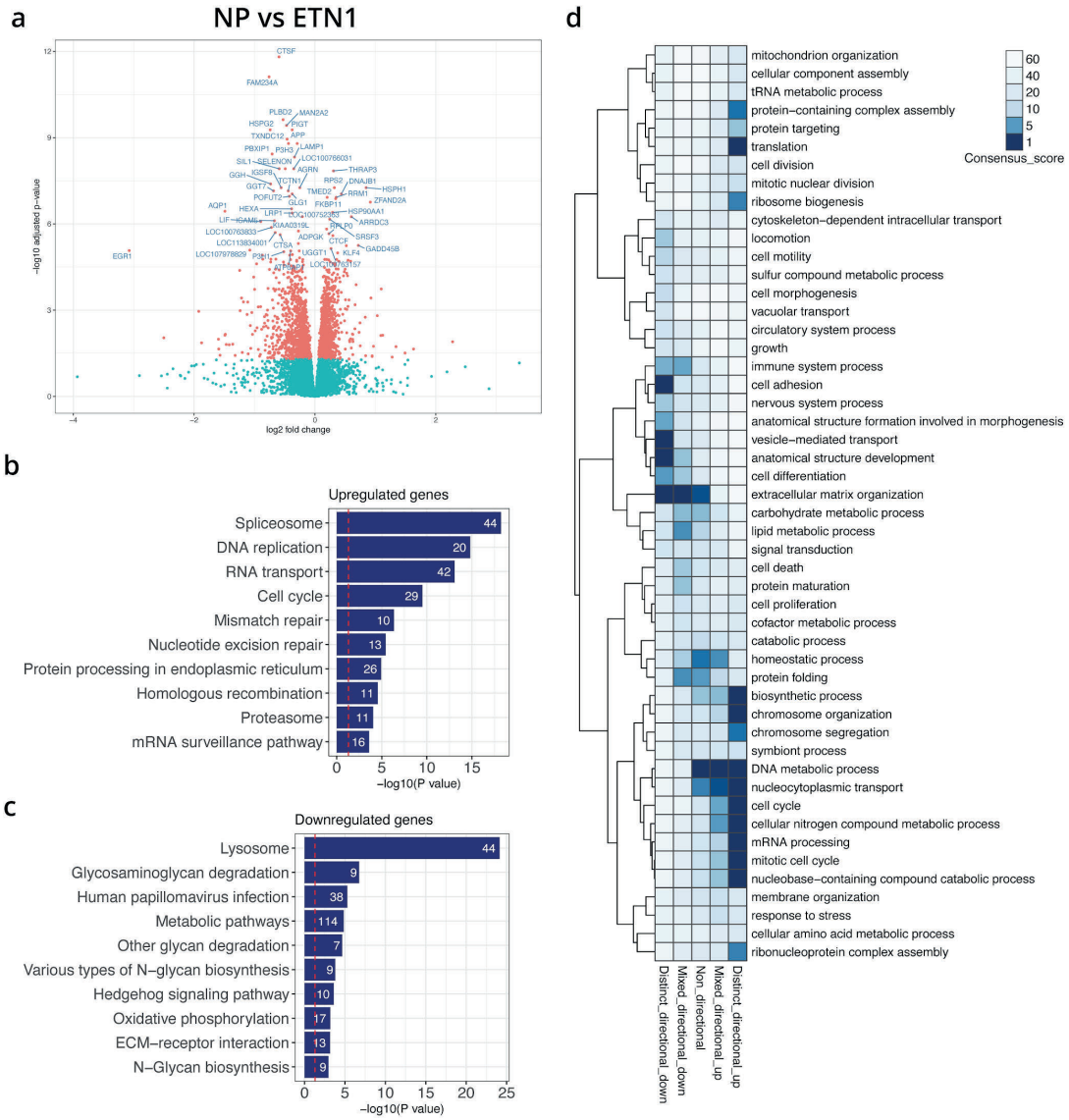


Figure 7. Analysis of groups NP vs ETN1. **a.** Volcano plot showing fold-change and p-value for the comparison NP vs ETN1. Differentially expressed genes with p-value <0.05 and no LFC cut-off are depicted in red. Genes with LFC > 0.2, p-value < 0.00001 are labeled. **b.** Top 10 KEGG pathways enriched for significantly upregulated genes between NP and ETN1. **c.** Top 10 KEGG pathways enriched for significantly downregulated genes between NP and ETN1. **d.** Consensus heatmap for gene ontology categories "biological process" for differentially expressed genes between NP and ETN1. Gene sets that are significant in at least one directionality class (p<0.05) are shown.

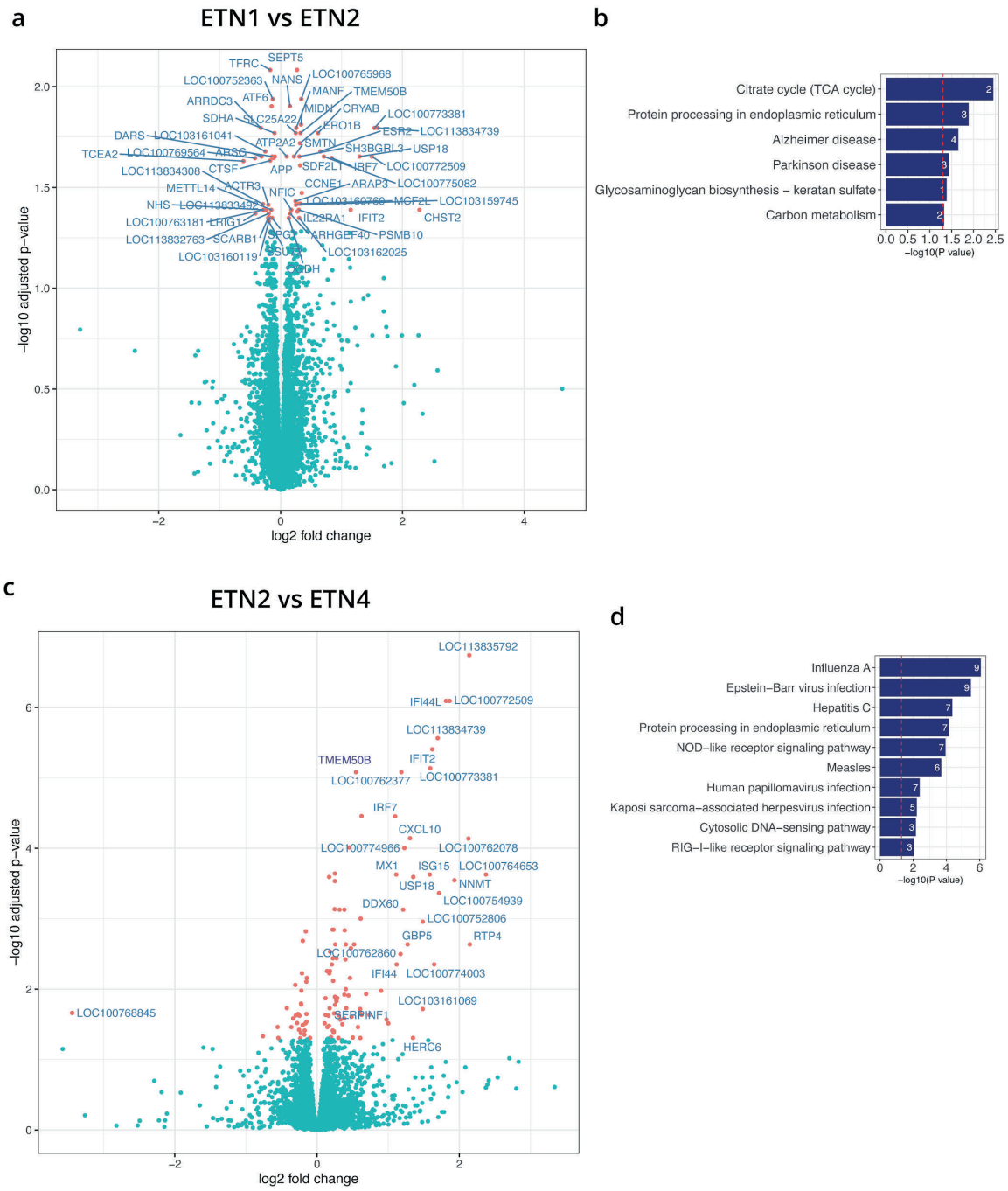


Figure 8. Analysis of groups ETN1 vs ETN2 and ETN2 vs ETN4. **a.** Volcano plot showing fold-change and p-value for the comparison ETN1 vs ETN2. Differentially expressed genes with p-value <0.05 and no LFC cut-off are depicted in red and are labeled. **b.** Top 10 KEGG pathways enriched for significantly upregulated genes between ETN1 and ETN2. **c.** Volcano plot showing fold-change and p-value for the comparison ETN2 vs ETN4. Differentially expressed genes with p-value <0.05 and no LFC cut-off are depicted in red. Genes with LFC ≥ 1, p-value < 0.05 are labeled including TMEM50B which has LFC < 1. **d.** Top 10 KEGG pathways enriched for significantly upregulated genes between ETN2 and ETN4.

Next, we performed unsupervised clustering of all differentially expressed genes between conditions NP, ETN1, ETN2, and ETN4 to find synexpression gene clusters - common patterns in gene expression. We identified 4 synexpression clusters that displayed correlated expression during the increase in ETN copy number: cluster 1 (912 genes, linearly upregulated), cluster 2 (1026 genes, linearly downregulated), cluster 3 (1056 genes, non-linearly upregulated), cluster 4 (928 genes, non-linearly downregulated) (Figure 9a). For each cluster, we extracted the core contributing genes (Supporting table 3) and performed KEGG pathway enrichment analysis on genes that had human homologs. Cluster 1 included 99 core genes that linearly increased with the increase in copy number. Genes were enriched in pathways protein processing in ER and protein export (DNAJB11, HSPA5, PDIA4, PDIA6, SELENOS, SYVN1, VCP, XBP1, SEC11C, and SPCS2) (Figure 9b). Cluster 2 included 158 core genes with linearly decreased expression when the copy number was increased. These genes were enriched in KEGG pathways related to NADH metabolism, TCA cycle and oxidative phosphorylation (ACO2, SDHA, SDHC, UQCR11, NDUFB5, and NDUFS1) (Figure 9c). Cluster 3 included 161 core genes that were non-linearly changed, increasing in the transition from NP to ETN1. Genes were enriched in pathways related to DNA replication (MCM4, MCM6, MCM7, POLE, PRIM1, and RNASEH2A), RNA transport (AAAS, KPNB1, NUP107, NUP160, NUP214, NUP54, PRMT5, RAN, and XPO1), spliceosome (DDX23, HNRNPM, PRPF19, RBMX, RBMXL1, SF3A3, SRSF7, and TRA2B), RNA polymerase II (POLR2B, POLR2D, and POLR2E) (Figure 9d). Cluster 4 had 92 core genes that were non-linearly changed, decreasing in the transition from NP to ETN1. Genes were enriched in pathways related to lysosome (ARSA, ATP6V0B, CTSO, DNASE2, GALNS, and GLA) and glutathione metabolism (GGT7, GSS, and TXNDC12) (Figure 9e).

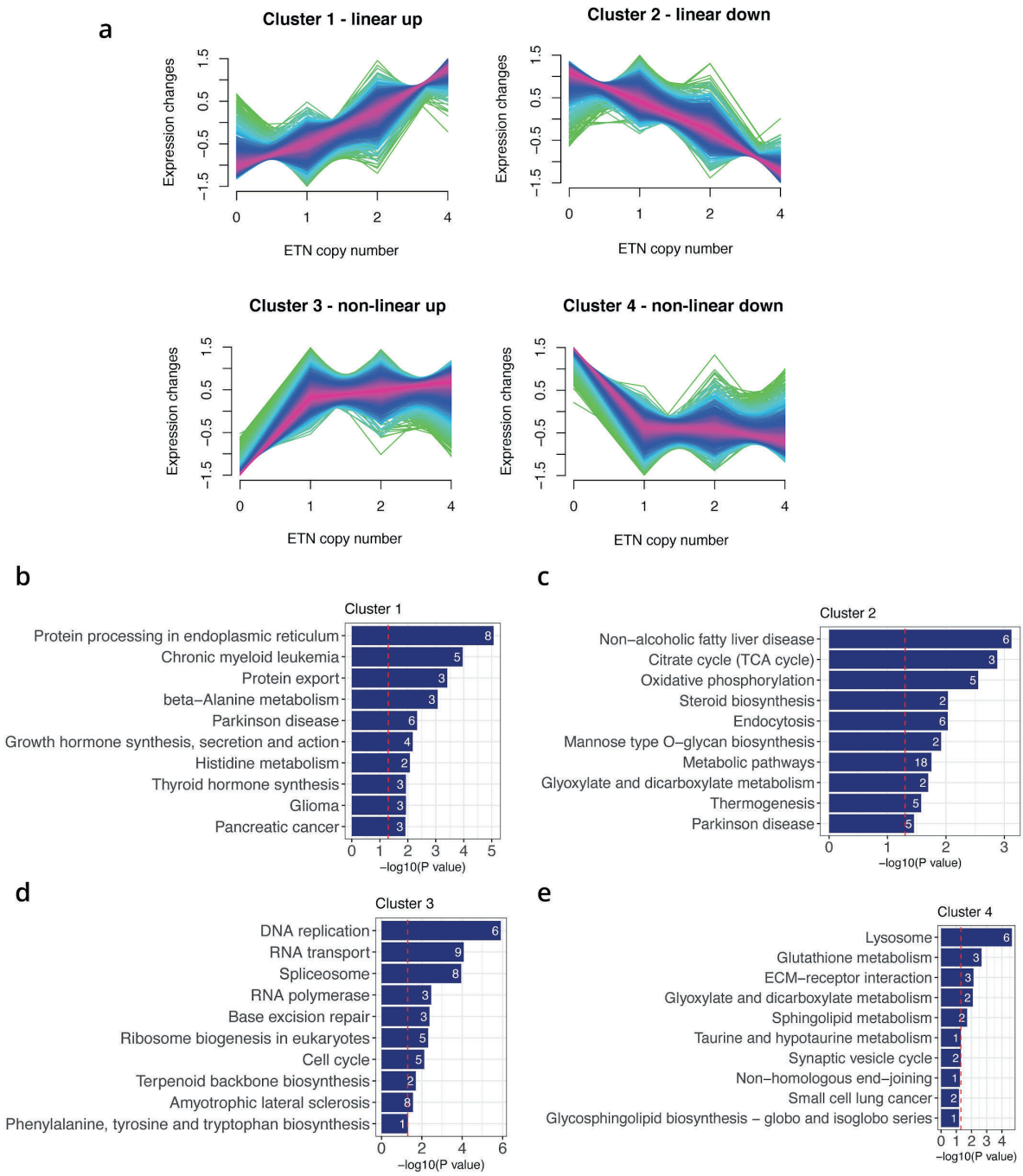


Figure 9. Patterns of gene expression correlated to ETN copy number increase. **a.** Four gene synexpression clusters identified by fuzzy clustering. **b-e.** Top KEGG pathways enriched for core genes in ETN clusters 1-4.

Recombinant protein-specific response in high-producing cells

Differential expression analysis showed that EPO4 and ETN4 cell lines were highly similar, having only 446 differentially expressed genes (Figure 3c). These genes reflect a protein-specific response to high recombinant protein production. By performing KEGG enrichment analysis of up- or down-regulated genes in this set, we found EPO-specific pathways and genes: protein processing in ER (HSPA5, HSPA8, MAN1C1, MAP2K7, SEL1L, SIL1, SYVN1, UBE2G1, UGGT1, and UGGT2), GPI-anchor biosynthesis (PGAP1, PIGA, and PIGT), ubiquitin-mediated proteolysis (HERC2, SMURF2, SYVN1, UBB, UBC, and UBE2G1), lysine degradation (ASH1L, EHMT1, EZH1, and KMT2C) (Supporting Figure 5). Among significantly differentially expressed genes in EPO cell lines were EGR1, EGR2, EGR3, and JUNB. ETN-specific genes were enriched in the pathways: protein processing in ER (DAD1, DNAJB1, ERLEC1, HSP90AA1, HSPH1, PRKCSH, RPN1, RPN2, SEC31B, SEC61B, SSR1, SSR4, SSR3, and STT3A), N-glycan synthesis (DAD1, HEXB, RPN1, RPN2, STT3A, and ALG5), NOD-like signalling (GSDMD, HSP90AA1, IRF7, GBP2, GBP3, IFNAR2, NAIP, OAS1, STAT1, and TXNIP) and interferon-stimulated genes (IFIT2, IFI44L, MX1, LOC100772509 (IFIT1), GBP5, LOC100754939 (IRGM), LOC100762078 (IRGM), LOC100762377 (IRGM), LOC100762860 (GBP1), LOC100764653 (MX1-like), LOC100774966 (OAS2-like), ISG15, and USP18) (Supporting Figure 5).

Common response to recombinant protein production

To reveal common responses to recombinant protein production in the transcriptome, we first looked at 25 common differentially expressed genes between comparisons EPO1 vs NP and ETN1 vs NP (Supporting Figure 5). 2 genes upregulated in protein-producing cells were site-specific (RRM1 and COL4A3BP), 3 other genes were components of ER (SEC63, RPN1, and LOC100752724 (SEC16B)). Among downregulated genes, 6 were extracellular (LOC100766031 (NID1), LPL, MAN2B2, MMP19, TFPI, and TINAGL1), 4 of which (NID1, TINAGL1, MMP19, and LPL) are among the top secreted host cell proteins found in CHO-S supernatant, whose knockout shown to improve the growth and productivity of CHO cells[19]. This suggests that protein-producing cell lines reduce the expression of unnecessary secreted proteins, which can release resources for recombinant protein synthesis.

To find common genes that changed due to the increase in the transgene copy number, we looked at the genes that overlap between all differentially expressed in groups NP-EPO1-EPO2-EPO4 and NP-ETN1-ETN2-ETN4. 2590 genes were significantly differentially expressed in the group NP-EPO1-EPO2-EPO4, 3922 genes were significantly differentially expressed in the group NP-ETN1-ETN2-ETN4, 1721 of them were common between the groups. To reveal which genes have common patterns of expression when the GOI copy number is increased, we found the overlap between synexpression clusters described earlier for EPO and ETN datasets separately (Figure 10). We defined four common trends in differential gene expression in response to copy number: linear up-regulation (cluster 1), linear down-regulation (cluster 2), non-linear upregulation (cluster 3) and non-linear downregulation (cluster 4). After assigning each differentially expressed gene in groups NP-EPO1-EPO2-EPO4 and NP-ETN1-ETN2-ETN4 to one of four clusters, we found common genes that belong to the same cluster (without extracting core genes) (Supporting table 4). We found 181 common genes in cluster 1, which were linearly increased with the increase in copy number. KEGG pathway analysis showed that genes were enriched in pathways related to protein processing in ER (DNAJB11, DNAJC3, EDEM1, EIF2AK3 (PERK), HERPUD1, HSPA5 (BIP), HYOU1, PDIA3, PDIA4, PDIA6, SEC24D, SEC61B, SELENOS, SYVN1, and XBP1), protein export (HSPA5, SEC61B, and SRPRA) and ubiquitin-mediated proteolysis (ANAPC2, BRCA, NEDD4, SKP2, SOCS1, and SYVN1). Cluster 2 included 153 common genes, whose expression was linearly decreased when copy number was increased. They were enriched in pathways related to lysosome and glycan degradation (AGA, ASAH1, CTSZ, FUCA1, GLB1, GNPTG, GUSB, HEXB, M6PR, MAN2B2, and SORT1) and protein processing in ER (ATF6B, SCARB1, ERLEC1, ERN1 (IRE1), ERO1A, GANAB, MBTPS1, and SSR2). Cluster 3 included 243 common genes that were non-linearly upregulated in response to copy number, which were related to DNA replication and DNA repair (FEN1, MCM2-MCM7, POLD2, POLD3, POLE, PRIM1, RFC2, RFC5, RNASEH2A, and RPA2), cell cycle (ANAPC5, CDC7, E2F1, E2F2, MCM2-MCM7, ORC1, RBL1, and YWHAB), spliceosome and RNA transport. Cluster 4 included 130 common genes that were non-linearly downregulated, which belong to diverse pathways and were enriched in pathways related to lysosome (CTSA, LAMP2, and PSAP) and oxidative phosphorylation (ATP6V1D (lysosomal ATPase), UQCRH, and NDUFC1).

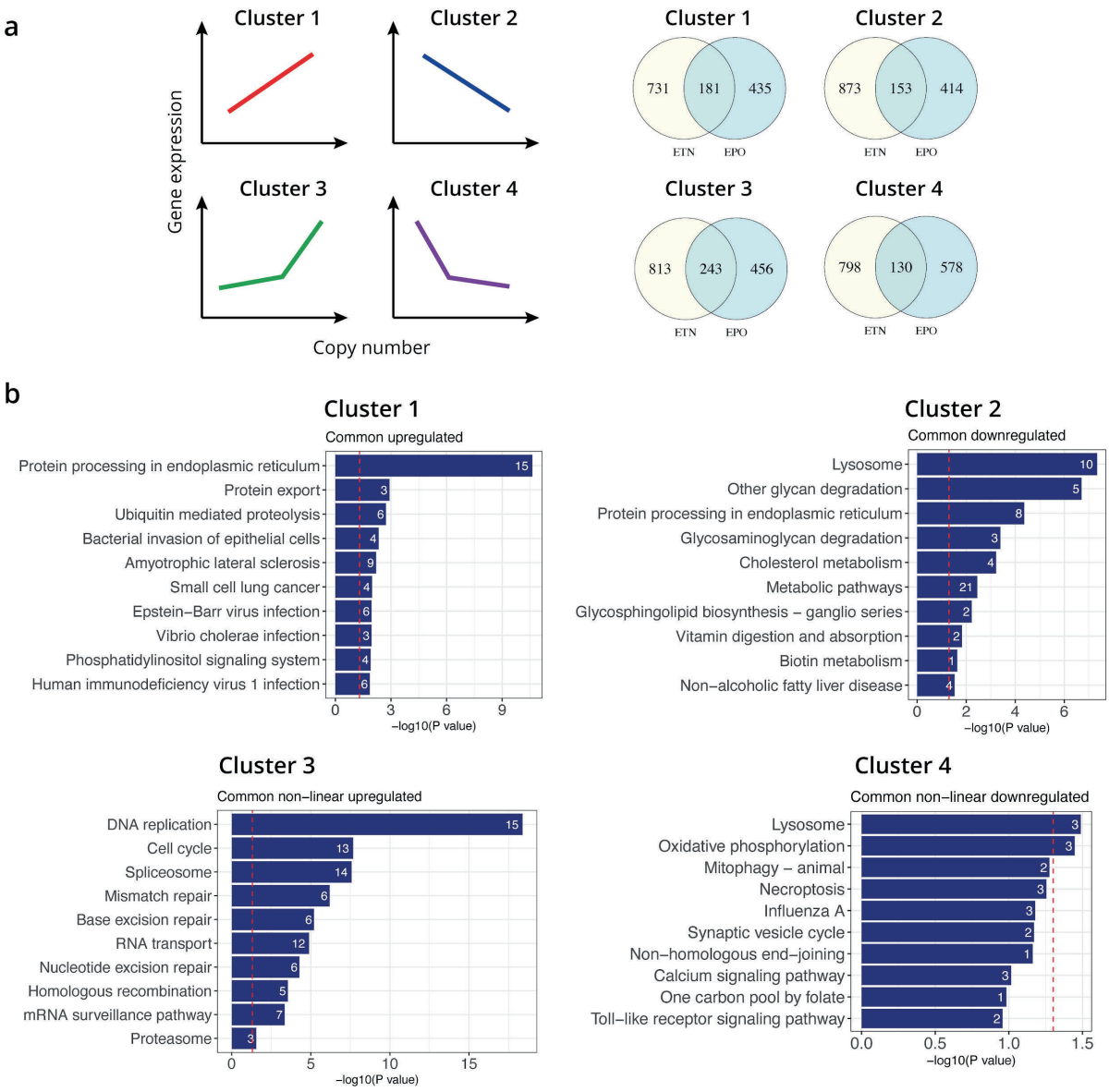


Figure 10. Common response to increased recombinant protein production. **a.** Four clusters of synexpressed genes and the overlap of significantly differentially expressed genes between EPO and ETN clusters. **b**. Top KEGG pathways enriched for common genes between EPO and ETN clusters 1-4 are shown.

Discussion

The overall aim of this study was to investigate the adaptive response of CHO cell lines to the increased production of recombinant proteins using transcriptome analysis. We used a set of multi-copy isogenic CHO cell lines producing EPO and ETN to answer the following questions: i) how the choice of the integration site affects global transcription, ii) what is the common response to increased protein production, iii) how the recombinant protein (EPO or ETN) affects the transcriptome, iv) what are potential causes of the transcriptional bottleneck in 4 copy clones EPO4 and ETN4. Using isogenic cell lines with minimal clonal variation enabled us to capture small changes in differentially expressed genes and identify shared and specific transcriptional responses to recombinant protein production.

First, we found that the site of transgene integration does not change transcription globally, only in the proximity of the integration site. In the analysed cell lines, the integration sites A and T9 are located close to or within the genes RRM1, POLK and COL4A3BP, whose expression was increased. It is possible that the upregulation of these genes affected DNA replication or ceramide transport in CHO cells. Previous studies have shown that the overexpression of COL4A3BP is able to increase production of recombinant proteins in CHO[20], so its upregulation might have had a positive effect on protein production in multi-copy CHO cell lines.

Second, we revealed a common transcriptional response to increased recombinant protein production in CHO cells. A controlled increase in recombinant protein expression using multi-copy targeted integration allowed us to study step changes in the transcriptome in adaptation to increased recombinant protein load. The clustering of synexpression genes common between EPO and ETN cell lines revealed four common patterns of differential expression (Figure 10). The majority of commonly linearly upregulated genes (cluster 1) were related to ER and protein export, which is a reasonable response to recombinant protein production. Upregulated genes included chaperones (BIP, DNAJB11, DNAJC3, and HYOU1), ER-associated degradation genes (EDEM1, HERPUD1, SELENOS, and SYVN1), protein disulfide-isomerases (PDIA3, PDIA4, and PDIA6), translocon components (SEC61B and SRPRA), component of COPII transport (SEC24D), and unfolded protein response (UPR) factors (PERK and XBP1). Some of the

ER-related genes were linearly downregulated (cluster 2), including UPR factors ATF6B (negative regulator of ATF6[21,22]), IRE1 and oxidoreductase ERO1A. This data indicates that an increase in protein production activates specific branches of UPR in CHO cells, augmenting the protein folding capacity without inducing ER stress-mediated apoptosis. UPR is coordinated by three signaling branches via ER transmembrane protein sensors IRE1a, ATF6 and PERK, whose activation upon ER stress leads to complex transcriptional and translational responses helping to restore proteostasis[6]. Transcriptional signatures of distinct UPR branches have been recently delineated by Perturb-seq (single-cell RNA-seq of pooled CRISPR screens)[23]. According to these transcriptional signatures, significantly upregulated genes in our CHO cell line dataset were mainly related to the ATF6 branch of UPR, seen in the expression of genes HSPA5, HERPUD1, SDF2L1, DNAJB11, MANF, CRELD2, TMEM50B, DNAJC3, HYOU1, PDIA4 and PDIA6. In spite of the fact that the IRE1 gene was downregulated, we saw the upregulation of genes in the IRE1a branch of UPR, specifically the expression of SEC61B, OSTC, XBP1 and NANS. UPR signaling through IRE1a is initiated by homodimerization and autophosphorylation of IRE1a, thereby changes in mRNA levels of IRE1a may not correspond to the changes in activation status of this UPR branch. Furthermore, although the expression of UPR sensor PERK was increased, we observed no activation of expression of downstream genes including DDIT3 (encodes pro-apoptotic transcription factor CHOP), and no signs of ER stress-mediated apoptosis. UPR signaling through PERK initiates by the phosphorylation of translation initiation factor 2 (eIF2a) leading to attenuation of protein translation. To confirm the activation status of distinct UPR branches and get a more complete picture of ER response to recombinant protein production, a further investigation of UPR on protein level is required. Still, based on transcriptional signatures, it seems that the common ER response to increased recombinant protein production in CHO cells mostly involves the activation of ATF6 and IRE1 branches of UPR without induction of PERK signaling. This response is similar to UPR in antibody-producing B cells, which do not activate PERK (B cells lacking PERK develop normally), but induce XBP1 and ATF6 signaling[24]. This differs from the cell response to chemically-induced ER stress e.g. by the addition of tunicamycin or thapsigargin, which triggers activation of all three UPR signaling pathways[23]. Thus, ER stress upon recombinant

protein (EPO and ETN) production in the CHO cell lines is mild and does not induce pro-apoptotic signals.

Additional common responses to increased recombinant protein production in the CHO cell lines were related to lysosomal and cell cycle pathways. We observed a significant linear downregulation of lysosomal acid hydrolases and mannose-6-phosphate receptor M6PR when the production of recombinant proteins was increased (cluster 2). This indicates that cells adapt to increased recombinant protein production by decreasing lysosomal enzyme synthesis and lysosome biogenesis, which could possibly reduce protein trafficking through endosomes and increase trafficking through secretory vesicles. Cluster 3 showed an interesting common pattern of non-linear upregulation of genes related to cell cycle and DNA replication, which was acquired at a different load of recombinant protein expression (early in ETN cell lines - in the transition from NP to ETN1, later in EPO cell lines - in the transition from EPO2 to EPO4). These genes play a role in G1/S transition in cell cycle[25,26] and are regulated by E2F transcription factors (two of which - E2F1 and E2F2 were found as significantly upregulated in the cluster). Although cell growth and proliferation were not affected by the increase in protein production in multi-copy CHO cells, the upregulation of cell cycle genes might point to some interplay between cell cycle machinery and ER, which might increase the overall biosynthetic capacity of the cell.

Third, we investigated distinct transcriptional responses in CHO cells producing two different recombinant proteins - EPO and ETN. These proteins differ in their physicochemical properties and post-translational modifications as well as their nature: EPO is natural cytokine while ETN is an artificial fusion protein. At first, different changes in the transcriptome were appearing at different protein loads: the largest transcriptional changes were observed between EPO2 and EPO4, while for ETN large significant changes appeared between NP and ETN1 (Figure 3c). These transcriptional changes turned out to involve a group of shared genes, seen in the cluster 3 (DNA replication and cell cycle response). So, when the copy number was increased, EPO4 and ETN4 cell lines showed high similarity in their transcriptomes and had only 446 differentially expressed genes. These genes are specific to the produced recombinant protein. In EPO cell lines, the gene with the highest fold change was EGR1 (early growth response 1), which is stimulated by EPO in human cells[27]. It shows that recombinant

human EPO may activate EPO-specific signaling in CHO cells, and thus can have a mitogenic activity. Specific response to EPO production was linked to upregulation of specific ER chaperons (HSPA5, HSPA8, and SIL1), ERAD genes (SEL1L, SYVN1, and UBE2G1), ubiquitin-mediated proteolysis factors (HERC2, SMURF2, SYVN1, UBB, UBC, and UBE2G1), N-glycan synthesis genes (MAN1C1, UGGT1, and UGGT2), GPI-anchor biosynthesis genes (PGAP1, PIGA, and PIGT) and histone-lysine methyltransferases (ASH1L, EHMT1, EZH1, and KMT2C). Protein-specific response to ETN, in turn, was seen in the specific upregulation of ER chaperones and folding cofactors (DNAJB1, HSP90AA1, and HSPH1), N-glycan synthesis genes (DAD1, PRKCSH, RPN1, RPN2, STT3A, and ALG5), ERAD gene (ERLEC1), a component of COPII (SEC31B), and components of translocon (SEC61B, SSR1, SSR3, and SSR4). An unexpected response to ETN was the activation of type I interferon-stimulated pathways and NOD-like signaling. It included upregulation of transcription factor IRF7, IFNa receptor and many IFN-stimulated genes (OAS1, OAS2-like, MX1, IFIT1, IFIT2, STAT1, ISG15, and USP18). Thus, recombinant proteins can induce specific responses in ER and glycosylation pathways, revealing their need for special protein processing machinery, as well as promote specific effects related to biological function of recombinant proteins.

Lastly, we sought to unravel possible mechanisms of the transcriptional bottleneck observed upon the transgene copy number increase from 2 to 4 copies. Differential expression analysis of the coding transcriptome between EPO2 and EPO4 cell lines did not point to any apparent mechanisms that can explain a bottleneck in EPO mRNA levels. The large transcriptional changes related to cell cycle and DNA replication observed between EPO2 and EPO4 cells might have masked the genes that could have caused limitation in EPO mRNA levels. It is also possible that the bottleneck appeared post-transcriptionally, e.g. by endogenous RNA interference via small non-coding RNA[28], which were not captured by RNA-seq of the coding transcriptome. A transcriptional response between ETN2 and ETN4 revealed upregulation of IFN-stimulated genes, which might be related to the transcriptional bottleneck in ETN expression. Normally, IFN-mediated innate immune response provides defense against pathogens, especially viruses[29]. The upregulation of OAS genes in ETN cell lines indicates that the response was stimulated by some pathogen-like RNA (double-stranded or single-stranded). Probably, the ETN sequence, which is unnatural to the cells and accumulated at very high

levels, was recognised as a viral RNA by the innate immune system of CHO cells. OAS recognition can lead to cleavage of pathogenic RNA by RNaseL[29], which may explain the transcriptional limitation of ETN transcripts in ETN4 cells. Interestingly, IFN signaling induction was reported previously for plasmid-based transfections in mammalian cells[30] and was shown to suppress gene expression from plasmids[31]. A downregulation of IFN pathway components might serve as a strategy to alleviate the transcriptional bottleneck in ETN cell lines and increase its expression.

In this study, we captured only one level of gene expression regulation - the coding transcriptome. To fully unravel the response to recombinant protein production, we plan next to study post-transcriptional, translational and post-translational levels of gene expression regulation. While mRNA expression to some extent correlates with protein levels[32], RNA-seq analysis may not reveal some functional changes that are regulated by post-transcriptional or post-translational mechanisms. Follow-up studies of small RNAs, proteome and phosphoproteome of multi-copy CHO cells can help to uncover the complete picture of the adaptive response to recombinant protein production. In summary, this study advances our understanding of adaptation to high protein production in mammalian cells and can help to design new CHO cell engineering strategies to improve the production of biopharmaceuticals in CHO cells.

Methods

Cell lines and batch cultivation. Generation of CHO-S-derived cell lines analysed in this study was previously described in[17]. Briefly, a dual-RMCE master cell line (MCL - BP5C9) was created by CRISPR/Cas9-mediated targeted integration of landing pad loxP-EF1α-mCherry-NeoR-lox2272-BGHpA into genomic site A and landing pad loxP-EF1α-tagBFP-HygR-lox2272-BGHpA into genomic site T9. This MCL was used for recombinase-mediated cassette exchange (RMCE) by co-transfection of Cre recombinase and RMCE donor plasmids encoding one or two copies of gene of interest (GOI) - EPO or ETN. RMCE led to the exchange of fluorescent markers (mCherry, tagBFP) together with selection markers (NeoR, HygR) to GOI in either or both of the target sites simultaneously. Upon FACS sorting, EPO- and ETN-producing subclones as well as non-producing (NP) clones were generated, where one or two copies of GOI were integrated into genomic site A (EPO1_A, EPO2_A, ETN1_A, ETN2_A), genomic site T9 (EPO1_T9, EPO2_T9, ETN1_T9, ETN2_T9), or both sites A and T9 (EPO1_EPO1, ETN1_ETN1, EPO2_EPO2, ETN1_ETN1, NP). After verification of targeted integration, three clones were chosen per cell design and were simultaneously grown in batch cultivation together with MCL. Cryovials with each biological replicate were thawed and passaged for one week before batch inoculation. For batch cultivation cells were seeded at 4×10^5 cells/mL in 40 mL CD CHO medium (Thermo Fisher Scientific), supplemented with 8 mM L-Glutamine and 2 µL/mL anti-clumping agent in 125 mL Erlenmeyer shake flasks, and were grown in a humidified incubator at 37°C, 5% CO_2 at 120 rpm. VCD and viability were monitored daily using the NucleoCounter NC-200 Cell Counter (ChemoMetec), samples for titer measurements were harvested daily, samples for RNA expression analysis were harvested on day 3. Cultures were discontinued on day 6. All cell lines were tested negative for mycoplasma contamination using MycoAlert kit (Lonza).

RNA-seq. Around 3 x 10^6 cells were harvested on day 3 in the exponential phase during batch cultivation. Cells were centrifuged at 200 g for 5 min, the pellet was completely resuspended in 1 mL TRIzol (Invitrogen) and stored at -80°C. RNA was extracted following the TRIzol manufacturer instructions. RNA concentrations were measured with Qubit (Thermo Fisher Scientific) and the quality was assessed with Fragment analyzer system (Agilent). All samples had RNA integrity number (RIN) 9.9-10. The total RNA

samples were prepared for sequencing using Illumina's TruSeq Stranded mRNA sample preparation kit, according to manufacturer's instructions. The samples were pooled and sequenced on Illumina's NextSeq 500 using NextSeq 500/550 High Output v2.5 (75 cycles) kit.

RNA-seq analysis. Quality of RNA-seq data was assessed using FastQC, adapter sequences and low-quality bases were trimmed using Trim Galore (Babraham Bioinformatics). Transcript quantification was performed using Salmon (v.1.1.0)[33] in mapping-based mode with default parameters against the reference transcriptome of *Cricetulus griseus* CriGri-PICR (GCF_003668045.1), 77-79% of the reads were mapped. Coding sequences for recombinant proteins (ETN, EPO) and landing pad genes (mCherry, tagBFP, NeoR, HygR) were added to the reference transcriptome as transgenes. Subsequent analyses of the count data were performed in R with Bioconductor packages tximport[34], edgeR[35]; packages tidyverse[36] and pheatmap[37] were used for data transformation and visualization. Count data from Salmon was imported to R using *tximport* function with "lengthScaledTPM" argument, generating transcript per million (TPM) values. Expressed genes were defined as having at least one TPM in at least three samples (biological replicates) and were retained for statistical analysis. PCA plots were based on the top 500 most variable genes, excluding transgenes.

Differential gene expression analysis. Differential gene expression analysis was performed using R package limma[38] to fit linear models to each of the genes with *voom* function, excluding transgenes from the dataset. Empirical Bayes model (*eBayes* function) was used to assess differential expression for each gene between the groups. Statistical tests for differences in expression levels were evaluated by comparing the sample groups (cell designs) consisting of three biological replicates. Genes with adjusted p-values lower than 0.05 (Benjamini-Hochberg FDR correction) were considered as significantly differentially expressed, no log fold change cut-off was used. KEGG pathway enrichment analysis of differentially expressed genes was performed with *kegga* function in limma package, using the set of genes that had human homologs. Gene ontology analysis was performed using R package Piano[39], based on the adjusted p-values and fold changes from the differential expression analysis. Gene set collection "goslim_generic Biological Process" was used. The Piano consensus score was based on

gene set statistics with mean, median, sum, stouffer and tailStrength. Gene sets that are significant in at least one directionality class ($p<0.05$) were evaluated.

Fuzzy clustering. Genes that displayed similar patterns of expression across the change in gene copy number were clustered by fuzzy c-means clustering using MFuzz package[40] in R. Clustering was performed on significantly differentially expressed genes. logTPM values of genes in biological replicates expressing the same recombinant protein at the same copy number were averaged, and the data were standardized so that each gene had a mean of 0 and st.dev of 1. Fuzzy clustering assigned each gene to one of the clusters (number of clusters was set to 4), and a membership score ranging from 0 to 1 was given to each gene. The core genes in each cluster were identified by specifying a cutoff on the membership score 0.8. KEGG pathway enrichment analysis was performed with *kegga* function for genes in the cluster, using the genes that had human homologs.

Acknowledgments

The authors would like to acknowledge Alexandra Hoffmeyer for the preparation of RNA-seq libraries and sequencing. The funding was provided by the Novo Nordisk Foundation (NNF10CC1016517, NNF16CC0020908).

References

(1) Walsh, G. (2018) Biopharmaceutical benchmarks 2018. *Nat. Biotechnol. 36*, 1136–1145.
(2) Templeton, N., and Young, J. D. (2018) Biochemical and metabolic engineering approaches to enhance production of therapeutic proteins in animal cell cultures. *Biochemical Engineering Journal*.
(3) Lee, J. S., Grav, L. M., Lewis, N. E., and Faustrup Kildegaard, H. (2015) CRISPR/Cas9-mediated genome engineering of CHO cell factories: Application and perspectives. *Biotechnol. J. 10*, 979–994.
(4) Gutierrez, J. M., Feizi, A., Li, S., Kallehauge, T. B., Hefzi, H., Grav, L. M., Ley, D., Baycin Hizal, D., Betenbaugh, M. J., Voldborg, B., Faustrup Kildegaard, H., Min Lee, G., Palsson, B. O., Nielsen, J., and Lewis, N. E. (2020) Genome-scale reconstructions of the mammalian secretory pathway predict metabolic costs and limitations of protein secretion. *Nat. Commun. 11*, 68.
(5) Jayaraj, G. G., Hipp, M. S., and Hartl, F. U. (2020) Functional Modules of the Proteostasis Network. *Cold Spring Harb. Perspect. Biol. 12*.
(6) Hetz, C., Zhang, K., and Kaufman, R. J. (2020) Mechanisms, regulation and functions of the unfolded protein response. *Nat. Rev. Mol. Cell Biol.*
(7) Seth, G., Charaniya, S., Wlaschin, K. F., and Hu, W.-S. (2007) In pursuit of a super

producer—alternative paths to high producing recombinant mammalian cells. *Current Opinion in Biotechnology*.
(8) Dietmair, S., Nielsen, L. K., and Timmins, N. E. (2012) Mammalian cells as biopharmaceutical production hosts in the age of omics. *Biotechnol. J. 7*, 75–89.
(9) Tamošaitis, L., and Smales, C. M. (2018) Meta-Analysis of Publicly Available Chinese Hamster Ovary (CHO) Cell Transcriptomic Datasets for Identifying Engineering Targets to Enhance Recombinant Protein Yields. *Biotechnol. J. 13*, e1800066.
(10) 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.
(11) Yusufi, F. N. K., Lakshmanan, M., Ho, Y. S., Loo, B. L. W., Ariyaratne, P., Yang, Y., Ng, S. K., Tan, T. R. M., Yeo, H. C., Lim, H. L., Ng, S. W., Hiu, A. P., Chow, C. P., Wan, C., Chen, S., Teo, G., Song, G., Chin, J. X., Ruan, X., Sung, K. W. K., Hu, W.-S., Yap, M. G. S., Bardor, M., Nagarajan, N., and Lee, D.-Y. (2017) Mammalian Systems Biotechnology Reveals Global Cellular Adaptations in a Recombinant CHO Cell Line. *Cell Syst 4*, 530–542.e6.
(12) Orellana, C. A., Marcellin, E., Palfreyman, R. W., Munro, T. P., Gray, P. P., and Nielsen, L. K. (2018) RNA-Seq Highlights High Clonal Variation in Monoclonal Antibody Producing CHO Cells. *Biotechnol. J. 13*, e1700231.
(13) Grav, L. M., Sergeeva, D., Lee, J. S., Marin de Mas, I., Lewis, N. E., Andersen, M. R., Nielsen, L. K., Lee, G. M., and Kildegaard, H. F. (2018) Minimizing Clonal Variation during Mammalian Cell Line Engineering for Improved Systems Biology Data Generation. *ACS Synth. Biol. 7*, 2148–2159.
(14) Carver, J., Ng, D., Zhou, M., Ko, P., Zhan, D., Yim, M., Shaw, D., Snedecor, B., Laird, M. W., Lang, S., Shen, A., and Hu, Z. (2020) Maximizing antibody production in a targeted integration host by optimization of subunit gene dosage and position. *Biotechnology Progress*.
(15) Zhang, L., Inniss, M. C., Han, S., Moffat, M., Jones, H., Zhang, B., Cox, W. L., Rance, J. R., and Young, R. J. (2015) Recombinase-mediated cassette exchange (RMCE) for monoclonal antibody expression in the commercially relevant CHOK1SV cell line. *Biotechnol. Prog. 31*, 1645–1656.
(16) Gaidukov, L., Wroblewska, L., Teague, B., Nelson, T., Zhang, X., Liu, Y., Jagtap, K., Mamo, S., Tseng, W. A., Lowe, A., Das, J., Bandara, K., Baijuraj, S., Summers, N. M., Lu, T. K., Zhang, L., and Weiss, R. (2018) A multi-landing pad DNA integration platform for mammalian cell engineering. *Nucleic Acids Res. 46*, 4072–4086.
(17) Sergeeva, D., Lee, G. M., Nielsen, L. K., and Grav, L. M. (2020) Multi-copy targeted integration for accelerated development of high-producing CHO cells. *ACS Synth. Biol. doi: 10.1021/acssynbio.0c00322*
(18) Pristovšek, N., Nallapareddy, S., Grav, L. M., Hefzi, H., Lewis, N. E., Rugbjerg, P., Hansen, H. G., Lee, G. M., Andersen, M. R., and Kildegaard, H. F. (2019) Systematic Evaluation of Site-Specific Recombinant Gene Expression for Programmable Mammalian Cell Engineering. *ACS Synth. Biol. 8*.
(19) Kol, S., Ley, D., Wulff, T., Decker, M., Arnsdorf, J., Schoffelen, S., Hansen, A. H., Jensen, T. L., Gutierrez, J. M., Chiang, A. W. T., Masson, H. O., Palsson, B. O., Voldborg, B. G., Pedersen, L. E., Kildegaard, H. F., Lee, G. M., and Lewis, N. E. (2020) Multiplex secretome engineering enhances recombinant protein production and purity. *Nat. Commun. 11*, 1908.
(20) Florin, L., Pegel, A., Becker, E., Hausser, A., Olayioye, M. A., and Kaufmann, H. (2009) Heterologous expression of the lipid transfer protein CERT increases therapeutic

protein productivity of mammalian cells. *J. Biotechnol. 141*, 84–90.
(21) Thuerauf, D. J., Morrison, L., and Glembotski, C. C. (2004) Opposing roles for ATF6alpha and ATF6beta in endoplasmic reticulum stress response gene induction. *J. Biol. Chem. 279*, 21078–21084.
(22) Thuerauf, D. J., Marcinko, M., Belmont, P. J., and Glembotski, C. C. (2007) Effects of the isoform-specific characteristics of ATF6 alpha and ATF6 beta on endoplasmic reticulum stress response gene expression and cell viability. *J. Biol. Chem. 282*, 22865–22878.
(23) Adamson, B., Norman, T. M., Jost, M., Cho, M. Y., Nuñez, J. K., Chen, Y., Villalta, J. E., Gilbert, L. A., Horlbeck, M. A., Hein, M. Y., Pak, R. A., Gray, A. N., Gross, C. A., Dixit, A., Parnas, O., Regev, A., and Weissman, J. S. (2016) A Multiplexed Single-Cell CRISPR Screening Platform Enables Systematic Dissection of the Unfolded Protein Response. *Cell 167*, 1867–1882.e21.
(24) Gass, J. N., Jiang, H.-Y., Wek, R. C., and Brewer, J. W. (2008) The unfolded protein response of B-lymphocytes: PERK-independent development of antibody-secreting cells. *Mol. Immunol. 45*, 1035–1043.
(25) Grant, G. D., and Cook, J. G. (2017) The Temporal Regulation of S Phase Proteins During G. *Adv. Exp. Med. Biol. 1042*, 335–369.
(26) Bertoli, C., Skotheim, J. M., and de Bruin, R. A. (2013) Control of cell cycle transcription during G1 and S phases. *Nat. Rev. Mol. Cell Biol. 14*.
(27) Ferguson, J., Bird, C., Wadhwa, M., and Burns, C. (2013) Detection of neutralizing antibodies to erythropoietin by inhibition of rHuEPO-stimulated EGR1 gene expression in the UT-7/EPO cell line. *J. Immunol. Methods 387*, 191–198.
(28) Cruz, C., and Houseley, J. (2014) Endogenous RNA interference is driven by copy number. *Elife 3*, e01581.
(29) Schneider, W. M., Chevillotte, M. D., and Rice, C. M. (2014) Interferon-Stimulated Genes: A Complex Web of Host Defenses. *Annu. Rev. Immunol. 32*, 513.
(30) Muerdter, F., Boryń, Ł. M., Woodfin, A. R., Neumayr, C., Rath, M., Zabidi, M. A., Pagani, M., Haberle, V., Kazmar, T., Catarino, R. R., Schernhuber, K., Arnold, C. D., and Stark, A. (2018) Resolving systematic errors in widely used enhancer activity assays in human cells. *Nat. Methods 15*, 141–149.
(31) Terenzi, F., deVeer, M. J., Ying, H., Restifo, N. P., Williams, B. R., and Silverman, R. H. (1999) The antiviral enzymes PKR and RNase L suppress gene expression from viral and non-viral based vectors. *Nucleic Acids Res. 27*, 4369–4375.
(32) Buccitelli, C., and Selbach, M. (2020) mRNAs, proteins and the emerging principles of gene expression control. *Nat. Rev. Genet.*
(33) Patro, R., Duggal, G., Love, M. I., Irizarry, R. A., and Kingsford, C. (2017) Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods 14*, 417–419.
(34) Soneson, C., Love, M. I., and Robinson, M. D. (2015) Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Res. 4*, 1521.
(35) Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics 26*, 139–140.
(36) Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller, K., Ooms, J., Robinson, D., Seidel, D., Spinu, V., Takahashi, K., Vaughan, D.,

Wilke, C., Woo, K., and Yutani, H. (2019) Welcome to the Tidyverse. *Journal of Open Source Software*.
(37) Kolde, R. pheatmap: Pretty Heatmaps. R package version 1.0.12. *https://CRAN.R-project.org/package=pheatmap*.
(38) Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., and Smyth, G. K. (2015) limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res. 43*, e47.
(39) Väremo, L., Nielsen, J., and Nookaew, I. (2013) Enriching the gene set analysis of genome-wide data by incorporating directionality of gene expression and combining statistical hypotheses and methods. *Nucleic Acids Research*.
(40) Futschik, M. E., and Carlisle, B. (2005) Noise-robust soft clustering of gene expression time-course data. *J. Bioinform. Comput. Biol. 3*, 965–988.

Conclusions and outlook

The development of precise genome editing techniques for mammalian cells enabled a potential paradigm shift in CHO cell line engineering. Traditional methods of CHO cell line generation based on random gene integration are now challenged by new methods based on targeted gene integration. This thesis aimed to develop and optimize targeted integration methods for improved production of biopharmaceuticals in CHO cells and encourage the shift to the cell line generation platforms that can accelerate and provide greater control in CHO cell line development.

The primary focus of the thesis was on the CRISPR/Cas9 and RMCE systems, which found their popularity in industrial applications in recent years[52,53]. In **Chapters 1 and 2**, we reviewed the tools for CRISPR/Cas editing of CHO cells and described the method for CRISPR/Cas9-mediated targeted gene integration. This method was then used in **Chapter 3**, where we combined CRISPR/Cas9 and RMCE to create a platform for accelerated CHO cell line generation. We showed that this platform minimizes clonal variation and enables robust comparative studies of CHO cells, which was demonstrated by the analysis of transcriptomes of CHO cells that produced different recombinant proteins. This platform has become the foundation for other comparative studies performed in our research group, including the systematic investigation of the interplay between the integration site and the expression cassette components[59] and validation of the mammalian secretory model[69]. In **Chapter 4**, we increased the productivity of cell lines using multi-copy targeted integration and optimal vector elements, reaching high titers and productivities. While the industry has described similar multi-copy targeted integration systems in broad terms[49,52,53], our study is the first Open Science system described in the literature. We have fully described targeted sites and genetic elements, and made vectors available from Addgene. Starting from this system, researchers will be able to do future studies using targeted integration in CHO with commercially relevant titers, thus increasing the impact on industry practice. Moreover, this study revealed the limitations of high-level protein production in CHO cells, which was linked

to the transcriptional bottleneck at higher copy numbers. This issue has not been identified in decades of study of protein expression in CHO when using random integration. In **Chapter 5**, we looked at the response to increased protein production in the multi-copy CHO cells using RNA-seq and elucidated common and protein-specific patterns of differential gene expression in the coding transcriptome. This study may help to advance the understanding of protein production in mammalian cells and give new ideas for rational CHO cell engineering for improved production of biopharmaceuticals.

Overall, the thesis proves that targeted gene integration is an advantageous method for CHO cell line generation that reduces genetic heterogeneity thus making cell line generation faster, more robust and predictable. We showed that the method is preferable when performing comparative studies of protein-producing CHO cell lines. We demonstrated that targeted integration can support high protein expression, making it valuable for the industrial production of therapeutic proteins as well as the study of CHO biology.

There are some limitations of targeted integration that have to be addressed in future studies to advance this technique. First, the integration efficiency of RMCE is low, especially when large multi-copy plasmids have to be inserted in the genome, as was shown in **Chapter 4**. This limitation is determined by two factors: delivery of donor DNA into the cell and efficiency of recombinase. The transfection efficiency of big plasmids into mammalian cells is known to decrease with the increase in plasmid size. So, a method to reduce the plasmid size such as minicircle DNA can be useful to increase RMCE efficiency of larger constructs[70,71]. A strategy of controlled concatemerization of plasmid might be developed in order to increase copy number while keeping the size of the plasmid small[58]. A change of recombinase transfection method from plasmid-based to protein-based might increase RMCE efficiency, similar to demonstrated improvement of CRISPR/Cas editing efficiency when Cas is transfected as ribonucleoprotein complex[72,73]. To increase recombinase efficiency, new recombination systems might be discovered or designed in the future[38]. Alternatively, nuclease-mediated targeted integration could be used.

The second limitation of the targeted integration strategy is specific productivity. Although in **Chapter 4** we have demonstrated that multi-copy integration and optimal vector design can improve it, a further increase in productivity is desired. A preferable

strategy would require not an increase in the copy number, but a manifold increase in the rate of transcription and translation of a single copy of GOI. It would diminish the importance of the first limitation related to targeted integration efficiency. The increase in transcription and translation might be achieved by the utilization of strong genetic elements or a hotspot genomic site for integration. It is doubtful that such hotspot genomic sites can be found empirically: a previous study has shown that 21 randomly selected genomic sites in CHO located in different genomic contexts have similar expression levels of a transgene[53]. The better strategy would be to design the optimal genomic background by bringing strong genetic elements within a vector or to rationally find genomic sites that can support high expression. Analysis of the structure of genetic regions of highly expressed genes in the human genome may help to select or design such sites in CHO cells. Yet, multi-copy integration is a more straightforward strategy to increase productivity than finding new genetic elements or sites, and could potentially be very successful in generating high-producing cell lines. However, as we showed in **Chapter 4**, a transcriptional limitation appears with the increase in copy number. Although the exact mechanism of the transcriptional bottleneck is still elusive, it has to be resolved in order to increase productivity from multi-copy targeted integration in the future.

The main advantage of targeted integration is the acceleration of CHO cell line development. In 2020 the speed of cell line generation became especially important due to the COVID-19 pandemic. It was recognized that traditional timelines of vaccine and monoclonal antibody development are too long to meet the urgent need for new medicines. This is the finest hour for targeted integration to show its promise in CHO cell generation. At least one biotech company, Regeneron, is using targeted integration to create COVID-19 treatment - a dual antibody cocktail against SARS-CoV-2, which entered phase 2/3 trials in July 2020. Such pace - 3 months from candidates selection to phase 1 trials[74] - is extraordinarily fast, considering that traditional mAb development pipeline lasts about 10-12 months[47]. The development of new vaccines can also be accelerated through the use of targeted integration. The COVID-19 vaccine development landscape is broad and includes various platform technologies, including recombinant proteins[75]. To evaluate targeted integration for the production of recombinant subunit vaccines, we used our RMCE systems created in **Chapter 3** and **Chapter 4** to express the

Sclamp vaccine developed at the University of Queensland (see **Appendix**). Sclamp is a fusion protein consisting of S1 subunit of SARS-CoV-2 spike protein and molecular clamp that stabilizes it in a trimeric form. In the **Appendix**, we demonstrated that targeted integration can be used for the generation of stable cell pools producing Sclamp within 3 weeks, allowing rapid production of vaccines. Altogether, targeted integration platforms can accelerate the development of new recombinant proteins, including the speed up the production of vaccines or mAbs. Hopefully, learning from the pandemic response, the future pipelines of CHO cell line development will be revised and shifted to the use of targeted integration allowing to reduce the time and resources.

Another advantage of targeted integration is the consistency of expression. In **Chapter 3** and **Chapter 4** we showed that different recombinant proteins can be stably expressed at similar levels with uniform cell phenotypes using RMCE systems. Thus, targeted integration helps to equalize and control the expression of proteins. This feature of targeted integration systems has already been acknowledged in 2006 when it was used by Symphogen for the expression of a polyclonal antibody cocktail comprising 25 mAbs. Similar productivity and growth of individual mAb-producing cell lines allowed to mix them in a single working cell bank, thus producing the antibody cocktail in a single cultivation[48]. The production of oligoclonal antibody cocktails is valuable for the creation of multivalent biopharmaceuticals. This strategy is utilized by Regeneron for the production of antibody cocktails against allergies and viral infections[76,77]. Antibody mixtures are also seen as promising antivenom therapeutics[78]. In addition, the consistency and control of expression provided by targeted integration are advantageous in the expression of bispecific or multispecific mAbs, where the tuning of ratios between light and heavy chains can improve antibody assembly and quality[52,79]. Thus, targeted integration is a method of choice for the production of complex biologics such as mAb mixtures or heterologous proteins where the ratio between the chains is important.

Targeted integration creates the platform for rational engineering of CHO cells by overexpression of effector genes towards improved product quality, protein secretion, metabolism and cell growth. Effector gene overexpression has shown some success in CHO engineering[61], although was not widely used in industry. This technique is not as straightforward as gene knockout, because gene expression levels have to be fine-tuned

in order to result in phenotypic changes. Their effect is often cell- and protein-specific, thus context-dependent strategies are required[80,81]. Also, the expression of multiple genes simultaneously might be needed for the improvement of protein expression[80,82]. Many previous studies of effector gene overexpression were performed in transient systems, and when transferred to stable protein-producing cell lines results were not reproducible[61]. Targeted integration enables stable, robust and controllable overexpression of genes, thus creating a new way to study effector gene overexpression, which can be directly evaluated in stable protein-producing cells eliminating the need of transient screening. In the **Appendix**, we described an ortho-dual-RMCE system consisting of two landing pads, one of which can be used for the expression of recombinant protein and the other for effector gene overexpression. This system can be applied in the future for stable or inducible expression of single as well as multiple effector genes in CHO cells to increase productivity or improve protein quality. Moreover, targeted integration enables robust library screening, reducing the noise originated from clonal variation, thus helping to define promising candidates in the screens more clearly. Besides screening of effector gene libraries[83], targeted integration can be used for non-viral CRISPR screens[84] or screens of antibody libraries[85,86].

The overall goal of CHO cell line development is the rapid, predictable and efficient generation of cell lines with stable phenotypes, high protein production and desired protein quality. Targeted integration can clearly advance traditional CHO cell line development in all these requirements, reducing time, resources and cost of delivering new biopharmaceuticals to the patients.

Appendix - Expression of SARS-CoV-2 Sclamp vaccine in CHO cells using targeted integration

Daria Sergeeva[1], Karen Kathrine Brøndum[1], Tune Wulff[1], Tae Kwang Ha[1], Daniel Duun[1], Trent Munro[2], Lise Marie Grav[1], Lars Keld Nielsen[1,2]

[1]The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Kgs. Lyngby, Denmark

[2]Australian Institute for Bioengineering and Nanotechnology, University of Queensland, Brisbane, Australia

Abstract

This report demonstrates the application of targeted gene integration for the production of the Sclamp vaccine in CHO cells. The Sclamp vaccine developed at the University of Queensland is a promising vaccine candidate against SARS-CoV-2, which entered clinical trials in July 2020. The objective of this study was to evaluate targeted integration strategies for the rapid expression of vaccine candidates in CHO cells in terms of speed, productivity, and protein quality. Our study shows that generation of stable CHO cell pools using targeted integration enables accelerated production of recombinant vaccines.

Researchers around the world are in a race for the fast development and production of vaccines against COVID-19 pandemic caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). The rapid spread of SARS-CoV-2 forced scientists to re-evaluate traditional timeframes for vaccine development, which typically require about 10 years of research and testing before reaching the public. Now under COVID-19 pressure, vaccine development pipelines are compressed to 12-18 months, resulting in the generation of over 100 vaccine candidates against SARS-CoV-2 within the first 6 months from the beginning of SARS-CoV-2 outbreak in January 2020. A wide range of vaccine development platforms are evaluated for COVID-19, including nucleic acids (DNA and RNA), virus-like particles, peptides, viral vectors, recombinant proteins, live attenuated virus and inactivated virus approaches. While some of the platforms are not currently the basis for licensed vaccines (e.g. nucleic acids), others, such as recombinant proteins, are already licensed as vaccines for other diseases[1–3].

A rapid vaccine response requires the acceleration of many vaccine development steps, including the production of pre-clinical and clinical material in lab scale and industrial scale. One of the technologies of vaccine production is expression of recombinant proteins in microbial or mammalian cells. Mammalian cells, such as Chinese hamster ovary (CHO) cells, are advantageous for the production of recombinant subunit vaccines that mimic viral surface proteins due to cells' ability to perform human-like post-translational modifications, especially N-glycosylation. To shorten the timelines of CHO cell line generation, targeted gene integration (TI) can be used instead of traditional random integration. TI avoids the need for tedious screening of cells and ensures long-term stability of CHO cell lines[4]. Using TI the CHO cell line development timeline can be reduced from approximately 4-6 months to 1-2 months. Moreover, the generation of cell pools instead of cell lines can compress the timeline even further for the fast delivery of protein candidates[5].

In this study, we evaluated targeted integration strategies for the expression of recombinant Sclamp vaccine in CHO cells. The Sclamp vaccine against SARS-CoV-2 was developed at the University of Queensland using a universal platform for stabilized subunit vaccine production - the molecular clamp[6]. The molecular clamp platform was previously used to develop vaccine candidates against various class I and class III

enveloped viruses including influenza, Ebola and MERS. The Sclamp vaccine is a recombinant fusion protein consisting of universal trimerization domain (molecular clamp) and the ectodomain S1 of the spike protein (S protein) of SARS-CoV-2, forming an equivalent of stable pre-fusion trimer of the S protein. The Sclamp vaccine induces a neutralizing antibody response against the S protein, thus preventing the SARS-CoV-2 uptake by cells via ACE2 receptor[7].

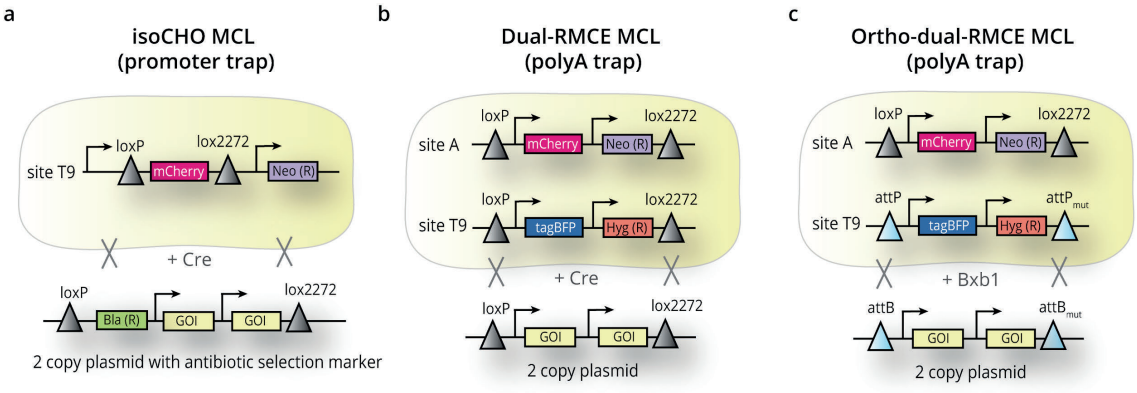


Figure 1. Master cell lines used for targeted integration. a. isoCHO MCL has a loxP/lox2272 landing pad for the integration of donor plasmids using Cre recombinase. **b.** Dual-RMCE MCL has two loxP/lox2272 landing pads in two genomic sites for simultaneous integration of donor plasmids using Cre recombinase. **c.** Ortho-dual-RMCE has two landing pads (loxP/lox2272 and attP/attP$_{mut}$) for ortogonal integration of donor plasmids using Cre or Bxb1 recombinase.

Here, we compared two targeted integration approaches for the production of the Sclamp vaccine: a) CHO cell pool generation using the isoCHO master cell line, b) CHO cell line generation using the dual-RMCE master cell line or ortho-dual-RMCE master cell line (Figure 1). These approaches are based on recombinase-mediated cassette exchange (RMCE). In the RMCE system, a master cell line (MCL) is generated first by precisely inserting a landing pad (fluorescence maker flanked by recombination sites) into the CHO genome using CRISPR/Cas9. Next, by co-transfection of a recombinase and a donor plasmid containing a gene of interest (GOI) flanked by recombination sites, the fluorescence marker is replaced by the GOI. Our previous studies show that RMCE systems can generate isogenic CHO cell lines with stable and uniform phenotypes[8] and can be used for the production of recombinant proteins at high levels[9]. In this study, we aimed to demonstrate the application of RMCE for fast vaccine production and compare

different strategies in terms of the development speed, quality, and quantity of the vaccine.

In the first TI approach, we used isoCHO MCL for the generation of Sclamp-producing cell pools. Previously, we created isoCHO-EP MCL that has one mCherry-landing pad integrated into the genomic site T9 of CHO-S cell[8]. This landing pad has EF1a promoter outside loxP/lox2272 recombination sites, creating a promoter trap for GOI (Figure 1a). To generate cell pools producing Sclamp or His-tagged Sclamp (Sclamp-His), we built donor plasmids that have a promoter-less blasticidin selection marker and two copies of Sclamp gene driven by 100RPU.2 promoters. These plasmids were integrated into isoCHO-EP MCL by RMCE with efficiency 0.13-0.16% (% of mCherry-negative cells), and blasticidin selection was started 4 days after transfection. After 13 days of selection, 91-95% of cells in the cell pools were mCherry-negative, indicating the integration of Sclamp expression cassettes into genomic site T9 (Figure 2a). isoCHO-Sclamp and isoCHO-SclampHis cell pools were grown in the batch culture to evaluate if the His-tag affects the productivity and quality of Sclamp vaccine. Cells displayed growth and expression of Sclamp and Sclamp-His (Figure 2b-d), thus no effect of His-tag was observed. According to native PAGE, Sclamp vaccine was predominantly expressed in trimeric form with a size around 840 kDa. However, we saw a sign of product degradation by the appearance of the second Sclamp band with size ~740 kDa on day 5 of batch culture (Figure 2d). The Sclamp monomers were approximately 280 kDa according to SDS-PAGE (Figure 2c), of which only 143 kDa is accounted for by the amino acid sequence, indicating a large extent of post-translational modifications. The Sclamp monomer has 24 predicted N-glycosylation sites and after the removal of N-glycans by PNGase F treatment the size of the protein shifted to around 160 kDa, confirming abundant N-glycosylation of Sclamp (Figure 2e). The protein sequence of deglycosylated Sclamp was verified by mass spectrometry (sequence coverage 64%). Among 24 predicted N-glycosylation sites, 9 sites were confirmed to be glycosylated, 5 sites were not glycosylated and 10 sites were not covered by mass spectrometry analysis. Thus, we showed that targeted integration can be used for the generation of stable CHO cell pools producing glycosylated Sclamp vaccine within 3 weeks from cell transfection.

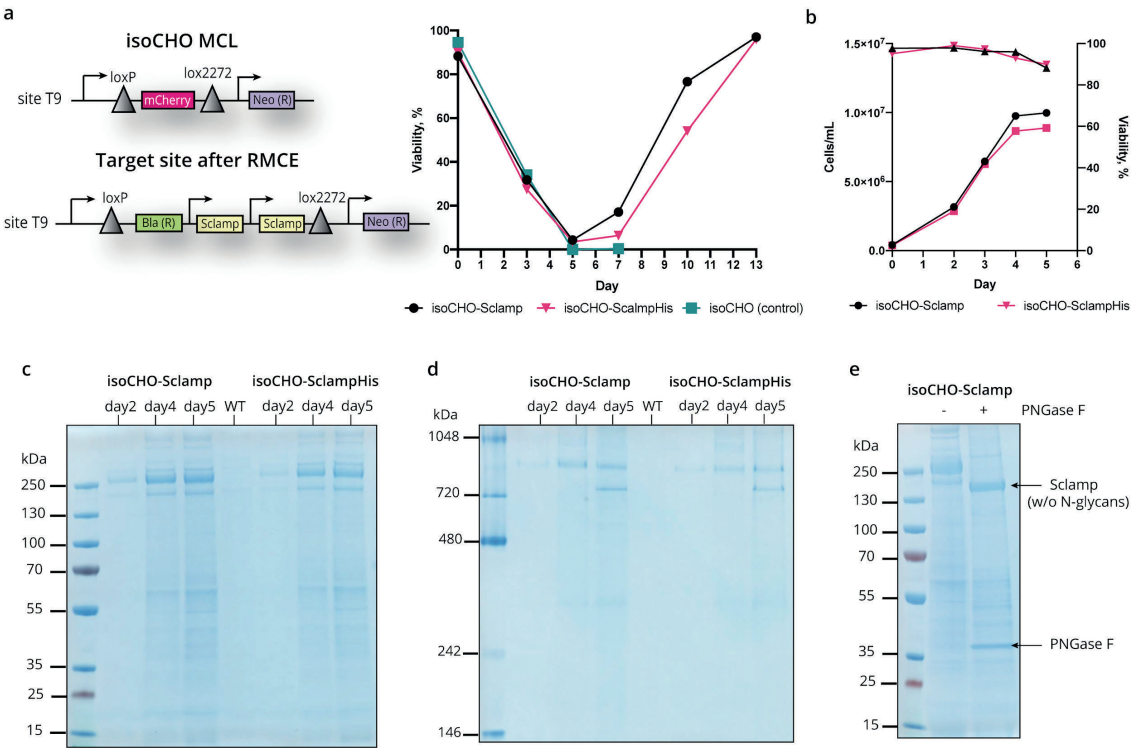


Figure 2. Sclamp production in isoCHO cell pools. a. isoCHO MCL was used for the generation of stable cell pools by antibiotic selection. Two copy plasmid with blasticidin resistance gene was integrated into genomic site T9 by Cre-mediated RMCE, cells were selected in the media with antibiotic for 13 days. **b.** Batch cultivation of isoCHO-Sclamp and isoCHO-SclampHis cell pools after antibiotic selection (viable cell density and viability are shown). **c.** SDS-PAGE of supernatants from batch cultivation (day 2, 4, 5), showing monomers of Sclamp and Sclamp-His. WT is a supernatant of CHO-S wild-type cells not producing Sclamp. **d.** Native PAGE of supernatants from batch cultivation (day 2, 4, 5), showing trimers of Sclamp and Sclamp-His. WT is a supernatant of CHO-S wild-type cells not producing Sclamp. **e.** SDS-PAGE of Sclamp supernatant (day 4) without or with PNGase F treatment.

In the second TI approach, we used dual-RMCE MCLs for the generation of cell lines producing Sclamp or Sclamp-His. We previously created dual-RMCE MCL BP5C9 that has two loxP/lox2272 landing pads: an mCherry-landing pad in the genomic site A and a tagBFP-landing pad in genomic site T9 of CHO-S cell[9] (Figure 1b). These landing pads have a BGH poly(A) tail outside the recombination sites, thus creating a poly(A)-trap for the GOI. The presence of two landing pads with the same loxP/lox2272 recombination sites can support the integration of donor plasmids into one of each genomic site or two sites simultaneously. To generate cell lines producing Sclamp or Sclamp-His, we created

donor plasmids with two copies of the GOI under 100RPU.2 promoters. We co-transfected Cre recombinase and donor plasmids and performed FACS to select the cells that had exchanged fluorescent markers to GOI. The RMCE efficiency of single-site integration was 0.11% (% of tagBFP-negative cells), but no double-negative cells with simultaneous integration into two sites were observed, probably due to the large size of two-copy insert (10 kb). tagBFP-negative/mCherry-positive cells that had exchanged the fluorescent marker in the site T9 were single-cell sorted, expanded and verified by genomic PCR and copy number analysis (Figure 3a). Thus, stable subclones of dual-RMCE MCL with two copies of Sclamp integrated in the genomic site T9 were generated within 6 weeks from cell transfection.

Analogous to loxP/lox2272 dual-RMCE MCL, we developed an orthogonal dual-RMCE MCL (ortho-dual-RMCE MCL), which has two landing pads for orthogonal recombinases Cre and Bxb1. Ortho-dual-RMCE MCL BP6A6 has loxP/lox2272 mCherry-landing pad in site A for Cre-mediated RMCE and attP/attP$_{mut}$ tagBFP-landing pad in site T9 for Bxb1-mediated RMCE (Figure 1c). This system allows us to independently insert GOIs into one genomic site or the other, while using the other site for screening of effector genes in CHO cells. We used this ortho-dual-RMCE MCL to generate Sclamp-producing cell lines. We created donor plasmids with attB/attB$_{mut}$ recombination sites for Bxb1-mediated RMCE in the site T9. These donor plasmids had two copies of Sclamp or Sclamp-His under 100RPU.2 promoters. After the co-transfection of Bxb1 recombinase and donor plasmids into ortho-dual-RMCE MCL, 0.03% cells were tagBFP-negative, indicating the integration of Sclamp expression cassettes into genomic site T9. We sorted these tagBFP-negative/mCherry-positive single cells by FACS, expanded clones and verified targeted integration by genomic PCR and copy number analysis (Figure 3b). Thus, within 6 weeks we generated stable cell lines that had two copies of Sclamp inserted into genomic site T9 of ortho-dual-RMCE MCL.

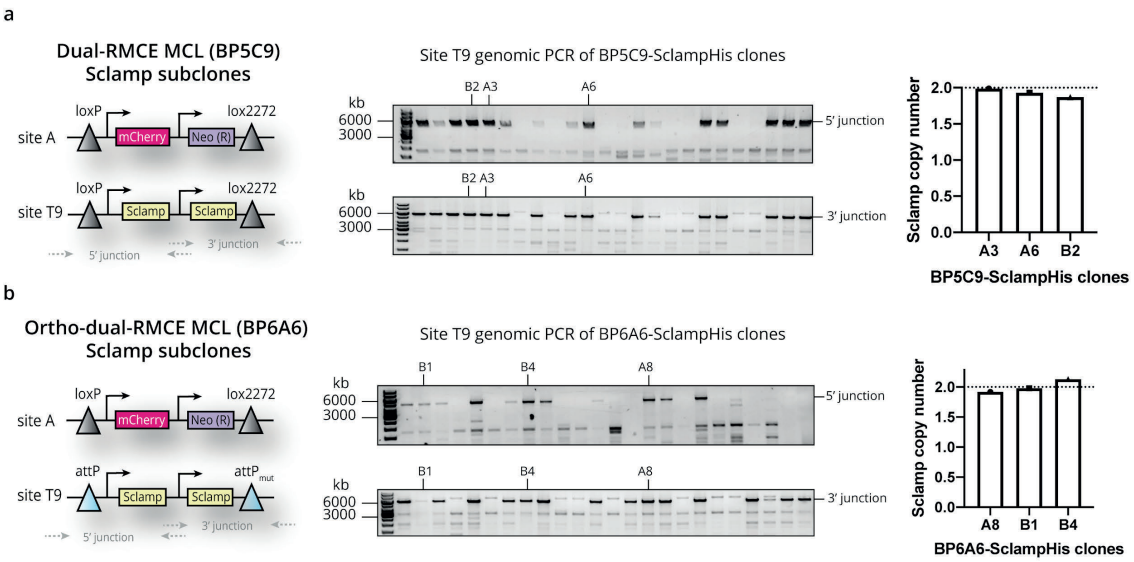


Figure 3. Generation of Sclamp-producing cell lines using dual-RMCE and ortho-dual-RMCE MCLs. a. Two copy Sclamp or Sclamp-His plasmid was integrated into site T9 of dual-RMCE MCL BP5C9. Clones were verified by PCR amplification of the targeted region and copy number analysis (only Sclamp-His clones are shown). **b.** Two copy Sclamp or Sclamp-His plasmid was integrated into site T9 of ortho-dual-RMCE MCL BP6A6. Clones were verified by PCR amplification of the targeted region and copy number analysis (only Sclamp-His clones are shown).

Next, we evaluated the productivity of cell pools and cell lines generated by different targeted integration strategies. We cultivated isoCHO-SclampHis cell pool and three biological replicates of dual-RMCE and ortho-dual-RMCE subclones producing Sclamp-His in the batch culture. We saw that cell lines and the cell pool had similar growth and production of Sclamp-His (Figure 4a,b). To measure vaccine concentration in the supernatant, we developed a biolayer interferometry assay using anti-penta-His sensors. As a standard for the assay, we purified Sclamp-His from isoCHO-SclampHis cell pool using immobilized metal ion affinity chromatography (IMAC). The elution fraction contained ~50% of full Sclamp-His based on the analysis of purity (Figure 4c), so we corrected the concentration of the standard accordingly. We observed that isoCHO cell pool, dual-RMCE and ortho-dual-RMCE cell lines with two copies of Sclamp-His produce a similar amount of vaccine with specific productivity 3-4 pg/cell/day (pcd), reaching titer around 40-50 mg/L on day 5 (Figure 4d).

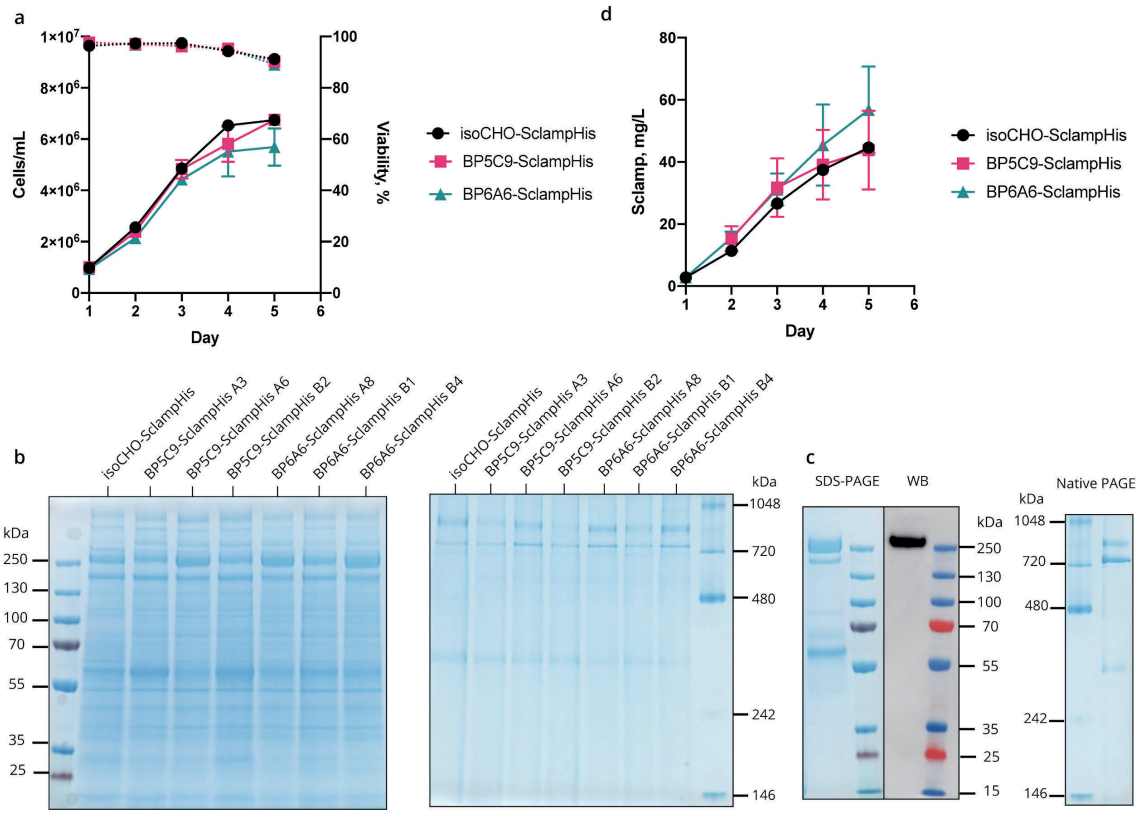


Figure 4. Batch cultivation of Sclamp-His cell pool and cell lines. isoCHO-SclampHis cell pool and three subclones of each dual-RMCE (BP5C9) and ortho-dual-RMCE (BP6A6) master cell lines were grown in batch culture. **a.** Viable cell density and viability (shown mean and SD for biological replicates). **b.** SDS-PAGE and native PAGE of culture supernatants on day 5. **c.** Analysis of Sclamp-His purity after IMAC. The purity of eluted fraction was analysed by SDS-PAGE, western blot with anti-His antibodies and native PAGE. **d.** Titer of Sclamp-His during batch cultivation measured by biolayer interferometry (shown mean and SD for biological replicates).

In conclusion, we showed that both TI strategies (cell pool generation and cell line generation) are applicable for the expression of Sclamp vaccine and can produce equal amounts of Sclamp when two copies of the GOI are inserted into the same genomic site. Cell pool generation was two times faster than cell line generation and required less effort and facilities, thus, it is more preferable when the speed of product development is important. The simultaneous integration of four Sclamp copies into two genomic sites of dual-RMCE MCL BP5C9 was not possible due to the large size of the donor plasmid and low integration efficiency, thus limiting the advantage of this system for accelerated high-producing cell line generation for Sclamp vaccine production. Sclamp specific productivity in dual-RMCE cell lines was lower (3-4 pcd) than the productivity of erythropoietin- or etanercept-producing cell lines generated by the same strategy

previously (7-9 pcd)[9], which indicates the difficulty of Sclamp expression. The observed low productivity of TI-generated cells may not be enough for the large-scale production of vaccines, so cell engineering and bioprocess development are necessary to meet the manufacturing demand. For the cell engineering, ortho-dual-RMCE cell lines generated in this study can be used in the future to overexpress effector genes that can increase vaccine production.

Spike protein of SARS-CoV-2 has large trimeric metastable structure and is generally considered as difficult to produce recombinantly[10]. The stabilized variant of S protein S-2P, which is encoded in vaccine candidate mRNA-1273 created by Moderna and National Institute of Allergy and Infectious Diseases[11], is expressed with a low titer of ~0.5 mg/L in Freestlyle 293-F cells[12], which is not cost-effective for large-scale production of vaccine in the form of recombinant protein. An introduction of mutations in the S2 subunit of spike protein allowed the creation of the HexaPro variant with improved stability and higher expression, reaching ~10 mg/L in FreeStyle 293-F or 32 mg/L in ExpiCHO cells[10]. Here we showed that stabilized vaccine Sclamp can be produced at expression levels of 40-50mg/L in stable CHO cells (batch culture), at similar levels to HexaPro expression in transient ExpiCHO cells (fed-batch culture).

In vaccine development, the fastest possible timeline is a necessity, there is no time to be wasted. Speed up production of vaccines by stable TI cell pools allows to generate the material for both pre-clinical and clinical studies in a single cell development procedure, enabling consistent production of vaccines for both discovery and development campaigns. To further accelerate vaccine production, stable TI cell pools can replace transient transfections of vaccine candidates, which is commonly used to generate the first material for the assessment of candidates. Moreover, TI might help to equalize the expression of different vaccine candidates, ensuring that no candidate will be turned down because of cell development failures caused by unpredictability of random integration. Thus, the use of targeted integration for cell pool generation permits the fast delivery of the recombinant vaccines by shortening the overall cell development timeline and reducing resource spend.

Methods

Plasmid cloning. Sclamp sequence encoding amino acid residues 1-679 of SARS-CoV-2 S protein (Uniprot P0DTC2) and molecular clamp residues 680-1276 were codon-optimized for *Cricetulus griseus* and synthetized by GeneArt (Thermo Fisher Scientific), including consensus Kozak sequence CGCCACC. TEV-His6 tag (ENLYFQGHHHHHH) was added at C-terminus of Sclamp to create Sclamp-His. Plasmids made in the study are listed in Table 1. RMCE donor plasmids were generated by USER cloning in Stbl3 competent cells (Thermo Fisher Scientific), verified by Sanger sequencing and restriction analysis. Plasmids used for the amplification of genetic elements have been described previously[9], blasticidin resistance gene was amplified from Addgene plasmid #17492. Cre recombinase plasmid OGS591 (Sigma-Aldrich) was used for Cre-mediated RMCE. Bxb1 recombinase sequence described in[13] was codon-optimized for *Cricetulus griseus*, synthesized as gBlock (IDT) and cloned into plasmid CMV-Bxb1-BGHpA by USER assembly. attB/attB$_{mut}$ sites were GGCCGGCTTGTCGACGACGGCGGTCTCCGTCGTCAGGATCATCCGG and GGCCGGCTTGTCGACGACGGCGGACTCCGTCGTCAGGATCATCCGG, respectively[14]. Landing pad plasmid PL0625 for ortho-dual-RMCE cell line was created by USER cloning, plasmids used for the amplification of genetic elements were described previously[9]. attP/attP$_{mut}$ sites were GTGGTTTGTCTGGTCAACCACCGCGGTCTCAGTGGTGTACGGTACAAACCCA and GTGGTTTGTCTGGTCAACCACCGCGGACTCAGTGGTGTACGGTACAAACCCA, respectively[14]. For the transfection in CHO cells, plasmids were purified using NucleoBond Xtra Midi EF kit (Macherey-Nagel) according to manufacturer's instructions.

Table 1. Plasmids created in the study

Plasmid name	**Features**
PL0877	pJ204_loxP_100RPU.2_Sclamp_BGHpA_HS4_LO1ex_HS4_100RPU.2_Sclamp_lox2272
PL0878	pJ204_loxP_100RPU.2_SclampHis_BGHpA_HS4_LO1ex_HS4_100RPU.2_SclampHis_lox2272

PL0879	pJ204_attB_100RPU.2_Sclamp_BGHpA_HS4_LO1ex_HS4_100RPU.2_Sclamp_attBmut
PL0880	pJ204_attB_100RPU.2_Sclamp_BGHpA_HS4_LO1ex_HS4_100RPU.2_Sclamp_attBmut
PL0881	pJ204_loxP_BlaR_HS4_LO1ex_HS4_100RPU.2_Sclamp_BGHpA_HS4_LO1ex_HS4_100 RPU.2_Sclamp_lox2272
PL0882	pJ204_loxP_BlaR_HS4_LO1ex_HS4_100RPU.2_SclampHis_BGHpA_HS4_LO1ex_HS4_1 00RPU.2_SclampHis_lox2272
PL0625	pJ204_5'HA(siteT9)_attP_EF1a_tagBFP_BGHpA_SV40_HygR_attPmut_BGHpA_3'HA(sit eT9)_CMV_ZsGreen1-DR_BGHpA
PL0616	pJ204_CMV_Bxb1_BGHpA

Cell cultivation. CHO-S cells (Thermo Fisher Scientific) were maintained in CD CHO medium supplemented with 8 mM L-Glutamine (Thermo Fisher Scientific) and cultivated in 125 mL Erlenmeyer shake flasks (Corning), incubated at 37°C, 5% CO_2 at 120 rpm in humidified incubator and passaged every 2−3 days. Viable cell density and viability were monitored using the NucleoCounter NC-200 Cell Counter (ChemoMetec), using Via1-Cassettes and "Viability and Cell Count Method 2" assay.

Master cell lines. Generation of isoCHO-EP MCL was described previously in[8] . Dual-RMCE MCL BP5C9 was previously described in[9]. Ortho-dual-RMCE MCL BP6A6 were generated by CRISPR/Cas9-mediated targeted integration of plasmid PL0625 into genomic site T9 of MP3 MCL (described in[9]), similarly to generation of dual-RMCE BP5C9 master cell line.

Stable isoCHO cell pools generation. 1 × 10^6 cells/mL isoCHO-EP MCL was transfected with RMCE donor plasmid PL0881 (Sclamp) or PL0882 (Sclamp-His) and Cre recombinase in a ratio 3:1 using FreeStyle Max transfection reagent. 4 days after transfection, blasticidin selection was started by exchanging the media to CD CHO with 8 mM Gln and 5 µg/mL blasticidin (Thermo Fisher Scientific). Cells were passaged in selection media every 2-3 days for 13 days, before cell viability reached >90%.

Throughout the selection process, cells were grown in suspension at 37°C, 5% CO_2, 120 rpm. Integration efficiency and selection efficiency were determined using MACSQuant Analyzer flow cytometer (Miltenyi Biotec) by measuring % of mCherry-negative cells in the beginning and the end of antibiotic selection, using isoCHO-EP MCL as gating control.

Stable dual-RMCE and ortho-dual-RMCE cell lines generation. Dual-RMCE MCL BP5C9 or ortho-dual-RMCE BP6A6 at concentration 1×10^6 cells/mL were transfected with RMCE donor plasmids (PL0877-PL0880) and Cre-recombinase plasmid in 3:1 ratio using FreeStyle MAX transfection reagent (Thermo Fisher Scientific). Cells were passed two times after transfection and then were FACS sorted on FACSJazz cell sorter (BD Biosciences) as described previously[9]. tagBFP-negative/mCherry-positive cells were single-sorted, selecting the clones with RMCE in genomic site T9. Subsequently, clones were expanded and verified by PCR of the targeted regions and copy number analysis

Verification of cell lines by genomic PCR and copy number analysis. Targeted integration was confirmed by PCR amplification of genomic DNA of targeted locus using specific primers binding outside of the site T9 and primers specific to the LO1ex linker located between two copies of Sclamp expression cassettes. PCR conditions were described previously[9]. Copy number analysis was performed by digital PCR using TaqMan assays for endogenous C1GALT1C1 (Cosmc) gene and Sclamp gene. Digital PCR procedure and Cosmc TaqMan assay sequences were described previously[9]. Sclamp Taqman assay sequences were the following: forward primer GTGACACAGCGGAACTTCTA, reverse primer GGGTCGTACACGGTATTGTT, FAM-MGB probe TCACCACCGACAACACCTTTGTGT.

Batch cultivation. Cells were seeded at 4×10^5 cells/mL in 40 mL CD CHO medium, supplemented with 8 mM L-Glutamine and 2 µL/mL anti-clumping agent in 125 mL Erlenmeyer shake flasks. Cells were incubated in a humidified incubator at 37°C, 5% CO2 at 120 rpm. VCD and viability were monitored daily using the NucleoCounter NC-200 Cell Counter.

PAGE and Western blot. SDS-PAGE of culture supernatants was performed in reduced conditions using Novex 4-12% Bis-Tris protein gels with MOPS buffer. For native PAGE, proteins were separated on 7% Tris-acetate gels with Novex Tris-Glycine Native

Running Buffer (Thermo Fisher Scientific). For western blot analysis, proteins after SDS-PAGE were transferred to iBlot2 nitrocellulose membranes using iBlot2 dry blotting system (Thermo Fisher Scientific). PentaHis HRP Conjugate Kit (Qiagen) was used for detection of Sclamp-His according to manufacturer's instructions. Membrane was visualized using Amersham ECL Prime Western Blotting Detection Reagent on Amersham Imager 600 (GE).

Mass spectrometry. Supernatant of isoCHO-Sclamp from day 4 of batch cultivation was concentrated using Amicon Ultra 30K (Merck Millipore) and treated with PNGase F (NEB) for 16 hours at 37°C. Deglycosylated Sclamp was extracted after SDS-PAGE, reduced by DTT, alkylated with iodoacetamide and digested with trypsin. Peptide mapping was performed on Orbitrap Exploris 480 (Thermo Fisher Scientific) instrument coupled to an CapLC system (Thermo Fisher scientific). The sample was captured on a precolumn (µ-precolumn C18 PepMap 100, 5µm, 100Å) at a flow of 10 µl/min and then the peptides were separated on an 15 cm C18 easy spray column (PepMap RSLC C18 2µm, 100Å, 150 µmx15cm) at a flow of 1.2 µl/min in a gradient 4-76% acetonitrile in water over a total of 16 minutes. Mass spectrometer was operated in data dependent mode using the following settings: MS-level scans were performed with Orbitrap resolution set to 120,000; AGC Target 3.0e6; maximum injection time 50 ms; intensity threshold 5.0e3; dynamic exclusion 25 sec. Data dependent MS2 selection was performed in Top 20 Speed mode with HCD collision energy set to 28% (AGC target 1.0e4, maximum injection time 30 ms, Isolation window 1.3 m/z) and a resolution of 30,000. The data was analysed in Proteome discoverer 2.4 using the following settings: Fixed modifications: Carbamidomethyl (C) and Variable modifications: oxidation of methionine residues and deamidation of Asparagine; Trypsin as enzyme with one missed cleavage allowance. Sequence of Sclamp and reference CHO proteome UP000001075 was used for the search in the protein database. The sites that had N-glycans were determined by the conversion of Asn to Asp on the predicted N-glycosylation sites.

Sclamp-His purification. Sclamp-His was purified from clarified supernatant using HisTrap FF 1 mL column (GE). After washing the column with a binding buffer (20 mM sodium phosphate, 0.5 M NaCl, 10mM imidazole, pH 7.4), Sclamp-His was eluted with the elution buffer (20 mM sodium phosphate, 0.5 M NaCl, 500 mM imidazole, pH

7.4) (isocratic elution). Fractions containing Sclamp-His were pooled, and buffer was exchanged to PBS by Amicon Ultra 50K (Merck Millipore). The purity of the protein was assessed by SDS-PAGE, western blot and native PAGE. Purified Sclamp-His was stored at -80°C.

Titer measurement by biolayer interferometry. Biolayer interferometry assays were performed on an Octet Red (ForteBio) instrument at 30°C with shaking at 1000 rpm. Anti-penta-HIS (HIS1K) biosensors were hydrated in PBS for 10 min and incubated in 1ug/mL of biocytin for 10 min prior measurements. Supernatants or purified Sclamp-His spiked in spent CD CHO media at different concentrations (3.75-120 µg/mL) were diluted 1:1 in Kinetics buffer (Fortebio). After equilibration in PBS for 60 s, sample measurements were taken for 200 s in basic quantification mode. The baseline (spent SD CHO media) was subtracted, and binding rate was calculated using ForteBio data analysis software.

Authors contribution

D.S. designed experiments, performed plasmid cloning, cell line and cell pool generation, developed biolayer interferometry assay, analysed the data and wrote the report. K.K.B. performed FACS. T.W. performed mass spectrometry analysis. T.K.H. performed batch cultivation. D.D. purified Sclamp-His. T.M. reviewed the report. L.M.G. and L.K.N. supervised the project and reviewed the report.

Funding

The project was supported by the funding from the Novo Nordisk Foundation (NNF10CC1016517, NNF16CC0020908).

References

(1) Amanat, F., and Krammer, F. (2020) SARS-CoV-2 Vaccines: Status Report. *Immunity 52*, 583–589.
(2) Hodgson, J. (2020) The pandemic pipeline. *Nature Biotechnology*.
(3) Thanh Le, T., Andreadakis, Z., Kumar, A., Gómez Román, R., Tollefsen, S., Saville, M., and Mayhew, S. (2020) The COVID-19 vaccine development landscape. *Nat. Rev. Drug Discov. 19*, 305–306.
(4) 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.
(5) Scarcelli, J. J., Shang, T. Q., Iskra, T., Allen, M. J., and Zhang, L. (2017) Strategic deployment of CHO expression platforms to deliver Pfizer's Monoclonal Antibody Portfolio. *Biotechnology Progress*.
(6) Chappell K, Watterson D, Young P. Rapid response pipeline for stabilized subunit vaccines. *https://dc.engconfintl.org/vt_vii/111/*.
(7) (2020, April 29) UQ COVID-19 vaccine shown to induce potent protective response in pre-clinical trials. *https://www.uq.edu.au/news/article/2020/04/uq-covid-19-vaccine-shown-induce-potent-protective-response-pre-clinical-trials*.
(8) Grav, L. M., Sergeeva, D., Lee, J. S., Marin de Mas, I., Lewis, N. E., Andersen, M. R., Nielsen, L. K., Lee, G. M., and Kildegaard, H. F. (2018) Minimizing Clonal Variation during Mammalian Cell Line Engineering for Improved Systems Biology Data Generation. *ACS Synth. Biol. 7*, 2148–2159.
(9) Sergeeva, D., Lee, G. M., Nielsen, L. K., and Grav, L. M. (2020) Multi-copy targeted integration for accelerated development of high-producing CHO cells. *ACS Synth. Biol. 2020. doi: 10.1021/acssynbio.0c00322*
(10) Hsieh, C.-L., Goldsmith, J. A., Schaub, J. M., DiVenere, A. M., Kuo, H.-C., Javanmardi, K., Le, K. C., Wrapp, D., Lee, A. G., Liu, Y., Chou, C.-W., Byrne, P. O., Hjorth, C. K., Johnson, N. V., Ludes-Meyers, J., Nguyen, A. W., Park, J., Wang, N., Amengor, D., Lavinder, J. J., Ippolito, G. C., Maynard, J. A., Finkelstein, I. J., and McLellan, J. S. (2020) Structure-based design of prefusion-stabilized SARS-CoV-2 spikes. *Science*.
(11) Jackson, L. A., Anderson, E. J., Rouphael, N. G., Roberts, P. C., Makhene, M., Coler, R. N., McCullough, M. P., Chappell, J. D., Denison, M. R., Stevens, L. J., Pruijssers, A. J., McDermott, A., Flach, B., Doria-Rose, N. A., Corbett, K. S., Morabito, K. M., O'Dell, S., Schmidt, S. D., Swanson, P. A., 2nd, Padilla, M., Mascola, J. R., Neuzil, K. M., Bennett, H., Sun, W., Peters, E., Makowski, M., Albert, J., Cross, K., Buchanan, W., Pikaart-Tautges, R., Ledgerwood, J. E., Graham, B. S., Beigel, J. H., and mRNA-1273 Study Group. (2020) An mRNA Vaccine against SARS-CoV-2 - Preliminary Report. *N. Engl. J. Med.*
(12) Wrapp, D., Wang, N., Corbett, K. S., Goldsmith, J. A., Hsieh, C.-L., Abiona, O., Graham, B. S., and McLellan, J. S. (2020) Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science 367*, 1260–1263.
(13) Duportet, X., Wroblewska, L., Guye, P., Li, Y., Eyquem, J., Rieders, J., Rimchala, T., Batt, G., and Weiss, R. (2014) A platform for rapid prototyping of synthetic gene networks in mammalian cells. *Nucleic Acids Res. 42*, 13440–13451.
(14) Inniss, M. C., Bandara, K., Jusiak, B., Lu, T. K., Weiss, R., Wroblewska, L., and Zhang, L. (2017) A novel Bxb1 integrase RMCE system for high fidelity site-specific integration of mAb expression cassette in CHO Cells. *Biotechnology and Bioengineering*.

Bibliography

Introduction, conclusion and outlook

(1) Walsh, G. (2018) Biopharmaceutical benchmarks 2018. *Nat. Biotechnol. 36*, 1136–1145.
(2) Life Sciences Foundation. (2015) Vital tools: a brief history of CHO cells. *http://biomanufacturing.org/uploads/files/547998065159985597-cho-history.pdf*.
(3) Noh, S. M., Sathyamurthy, M., and Lee, G. M. (2013) Development of recombinant Chinese hamster ovary cell lines for therapeutic protein production. *Current Opinion in Chemical Engineering*.
(4) 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.
(5) Wigler, M., Silverstein, S., Lee, L. S., Pellicer, A., Cheng, Y. c., and Axel, R. (1977) Transfer of purified herpes virus thymidine kinase gene to cultured mouse cells. *Cell 11*, 223–232.
(6) Pellicer, A., Robins, D., Wold, B., Sweet, R., Jackson, J., Lowy, I., Roberts, J. M., Sim, G. K., Silverstein, S., and Axel, R. (1980) Altering genotype and phenotype by DNA-mediated gene transfer. *Science 209*, 1414–1422.
(7) Robins, D. M., Ripley, S., Henderson, A. S., and Axel, R. (1981) Transforming DNA integrates into the host chromosome. *Cell 23*, 29–39.
(8) Colbère-Garapin, F., Horodniceanu, F., Kourilsky, P., and Garapin, A. C. (1981) A new dominant hybrid selective marker for higher eukaryotic cells. *J. Mol. Biol. 150*, 1–14.
(9) Southern, P. J., and Berg, P. (1982) Transformation of mammalian cells to antibiotic resistance with a bacterial gene under control of the SV40 early region promoter. *J. Mol. Appl. Genet. 1*, 327–341.
(10) Mortensen, R. M., and Kingston, R. E. (2009) Selection of transfected mammalian cells. *Curr. Protoc. Mol. Biol. Chapter 9*, Unit9.5.
(11) Nunberg, J. H., Kaufman, R. J., Schimke, R. T., Urlaub, G., and Chasin, L. A. (1978) Amplified dihydrofolate reductase genes are localized to a homogeneously staining region of a single chromosome in a methotrexate-resistant Chinese hamster ovary cell line. *Proceedings of the National Academy of Sciences*.
(12) Kaufman, R. J., and Sharp, P. A. (1982) Amplification and expression of sequences cotransfected with a modular dihydrofolate reductase complementary dna gene. *J. Mol. Biol. 159*, 601–621.
(13) Cockett, M. I., Bebbington, C. R., and Yarranton, G. T. (1990) High level expression of tissue inhibitor of metalloproteinases in Chinese hamster ovary cells using glutamine synthetase gene amplification. *Biotechnology 8*, 662–667.
(14) Fan, L., Frye, C. C., and Racher, A. J. (2013) The use of glutamine synthetase as a selection marker: recent advances in Chinese hamster ovary cell line generation processes. *Pharmaceutical Bioprocessing*.
(15) Noh, S. M., Shin, S., and Lee, G. M. (2018) Comprehensive characterization of

glutamine synthetase-mediated selection for the establishment of recombinant CHO cells producing monoclonal antibodies. *Sci. Rep. 8*, 5361.
(16) Smithies, O., Gregg, R. G., Boggs, S. S., Koralewski, M. A., and Kucherlapati, R. S. (1985) Insertion of DNA sequences into the human chromosomal beta-globin locus by homologous recombination. *Nature 317*, 230–234.
(17) Thomas, K. R., and Capecchi, M. R. (1987) Site-directed mutagenesis by gene targeting in mouse embryo-derived stem cells. *Cell 51*, 503–512.
(18) Mansour, S. L., Thomas, K. R., and Capecchi, M. R. (1988) Disruption of the proto-oncogene int-2 in mouse embryo-derived stem cells: a general strategy for targeting mutations to non-selectable genes. *Nature 336*, 348–352.
(19) Capecchi, M. R. (2007) Gene targeting 1977-present. *https://www.nobelprize.org/uploads/2018/06/capecchi_lecture.pdf*.
(20) Rouet, P., Smih, F., and Jasin, M. (1994) Expression of a site-specific endonuclease stimulates homologous recombination in mammalian cells. *Proc. Natl. Acad. Sci. U. S. A. 91*, 6064–6068.
(21) Moehle, E. A., Rock, J. M., Lee, Y.-L., Jouvenot, Y., DeKelver, R. C., Gregory, P. D., Urnov, F. D., and Holmes, M. C. (2007) Targeted gene addition into a specified location in the human genome using designed zinc finger nucleases. *Proceedings of the National Academy of Sciences*.
(22) Gaj, T., Gersbach, C. A., and Barbas, C. F., 3rd. (2013) ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends Biotechnol. 31*, 397–405.
(23) Anzalone, A. V., Koblan, L. W., and Liu, D. R. (2020) Genome editing with CRISPR–Cas nucleases, base editors, transposases and prime editors. *Nature Biotechnology*.
(24) Suzuki, K., Tsunekawa, Y., Hernandez-Benitez, R., Wu, J., Zhu, J., Kim, E. J., Hatanaka, F., Yamamoto, M., Araoka, T., Li, Z., Kurita, M., Hishida, T., Li, M., Aizawa, E., Guo, S., Chen, S., Goebl, A., Soligalla, R. D., Qu, J., Jiang, T., Fu, X., Jafari, M., Esteban, C. R., Berggren, W. T., Lajara, J., Nuñez-Delicado, E., Guillen, P., Campistol, J. M., Matsuzaki, F., Liu, G.-H., Magistretti, P., Zhang, K., Callaway, E. M., Zhang, K., and Belmonte, J. C. I. (2016) In vivo genome editing via CRISPR/Cas9 mediated homology-independent targeted integration. *Nature 540*, 144–149.
(25) Roth, T. L., Puig-Saus, C., Yu, R., Shifrut, E., Carnevale, J., Li, P. J., Hiatt, J., Saco, J., Krystofinski, P., Li, H., Tobin, V., Nguyen, D. N., Lee, M. R., Putnam, A. L., Ferris, A. L., Chen, J. W., Schickel, J.-N., Pellerin, L., Carmody, D., Alkorta-Aranburu, G., Del Gaudio, D., Matsumoto, H., Morell, M., Mao, Y., Cho, M., Quadros, R. M., Gurumurthy, C. B., Smith, B., Haugwitz, M., Hughes, S. H., Weissman, J. S., Schumann, K., Esensten, J. H., May, A. P., Ashworth, A., Kupfer, G. M., Greeley, S. A. W., Bacchetta, R., Meffre, E., Roncarolo, M. G., Romberg, N., Herold, K. C., Ribas, A., Leonetti, M. D., and Marson, A. (2018) Reprogramming human T cell function and specificity with non-viral genome targeting. *Nature 559*, 405–409.
(26) Canaj, H., Hussmann, J. A., Li, H., Beckman, K. A., Goodrich, L., Cho, N. H., Li, Y. J., Santos, D. A., McGeever, A., Stewart, E. M., Pessino, V., Mandegar, M. A., Huang, C., Gan, L., Panning, B., Huang, B., Weissman, J. S., and Leonetti, M. D. Deep profiling reveals substantial heterogeneity of integration outcomes in CRISPR knock-in experiments.
(27) Levin, H. L., and Moran, J. V. (2011) Dynamic interactions between transposable elements and their hosts. *Nat. Rev. Genet. 12*, 615–627.
(28) Balasubramanian, S., Peery, R. B., Minshull, J., Lee, M., White, R., Kelly, R. M., and Barnard, G. C. (2018) Generation of High Expressing Chinese Hamster Ovary Cell Pools

Using the Leap-In Transposon System. *Biotechnol. J. 13*, e1700748.
(29) Balasubramanian, S., Rajendra, Y., Baldi, L., Hacker, D. L., and Wurm, F. M. (2016) Comparison of three transposons for the generation of highly productive recombinant CHO cell pools and cell lines. *Biotechnol. Bioeng. 113*, 1234–1243.
(30) Yoshida, J., Akagi, K., Misawa, R., Kokubu, C., Takeda, J., and Horie, K. (2017) Chromatin states shape insertion profiles of the piggyBac, Tol2 and Sleeping Beauty transposons and murine leukemia virus. *Sci. Rep. 7*, 43613.
(31) Yant, S. R., Huang, Y., Akache, B., and Kay, M. A. (2007) Site-directed transposon integration in human cells. *Nucleic Acids Res. 35*, e50.
(32) Owens, J. B., Urschitz, J., Stoytchev, I., Dang, N. C., Stoytcheva, Z., Belcaid, M., Maragathavally, K. J., Coates, C. J., Segal, D. J., and Moisyadi, S. (2012) Chimeric piggyBac transposases for genomic targeting in human cells. *Nucleic Acids Res. 40*, 6978–6991.
(33) Luo, W., Galvan, D. L., Woodard, L. E., Dorset, D., Levy, S., and Wilson, M. H. (2017) Comparative analysis of chimeric ZFP-, TALE- and Cas9-piggyBac transposases for integration into a single locus in human cells. *Nucleic Acids Res. 45*, 8411–8422.
(34) Kovač, A., Miskey, C., Menzel, M., Grueso, E., Gogol-Döring, A., and Ivics, Z. RNA-guided Retargeting of Sleeping Beauty Transposition in Human Cells.
(35) Rubens, J. R., Von Maltzahn, G. A., Citorik, R. J., Steinberg, B. E., and Li, D. (2020, April 9) Methods and compositions for modulating a genome. *US Patent*.
(36) Strecker, J., Ladha, A., Gardner, Z., Schmid-Burgk, J. L., Makarova, K. S., Koonin, E. V., and Zhang, F. (2019) RNA-guided DNA insertion with CRISPR-associated transposases. *Science 365*, 48–53.
(37) Klompe, S. E., Vo, P. L. H., Halpin-Healy, T. S., and Sternberg, S. H. (2019) Transposon-encoded CRISPR-Cas systems direct RNA-guided DNA integration. *Nature 571*, 219–225.
(38) Olorunniji, F. J., Rosser, S. J., and Stark, W. M. (2016) Site-specific recombinases: molecular machines for the Genetic Revolution. *Biochem. J 473*, 673–684.
(39) Sauer, B., and Henderson, N. (1990) Targeted insertion of exogenous DNA into the eukaryotic genome by the Cre recombinase. *New Biol. 2*, 441–449.
(40) O'Gorman, S., Fox, D. T., and Wahl, G. M. (1991) Recombinase-mediated gene activation and site-specific integration in mammalian cells. *Science 251*, 1351–1355.
(41) Merrick, C. A., Zhao, J., and Rosser, S. J. (2018) Serine Integrases: Advancing Synthetic Biology. *ACS Synth. Biol. 7*, 299–310.
(42) Branda, C. S., and Dymecki, S. M. (2004) Talking about a revolution: The impact of site-specific recombinases on genetic analyses in mice. *Dev. Cell 6*, 7–28.
(43) Jusiak, B., Jagtap, K., Gaidukov, L., Duportet, X., Bandara, K., Chu, J., Zhang, L., Weiss, R., and Lu, T. K. (2019) Comparison of Integrases Identifies Bxb1-GA Mutant as the Most Efficient Site-Specific Integrase System in Mammalian Cells. *ACS Synth. Biol. 8*, 16–24.
(44) Thompson, D. B., Aboulhouda, S., Hysolli, E., Smith, C. J., Wang, S., Castanon, O., and Church, G. M. (2018) The Future of Multiplexed Eukaryotic Genome Engineering. *ACS Chem. Biol. 13*, 313–325.
(45) Nivina, A., Grieb, M. S., Loot, C., Bikard, D., Cury, J., Shehata, L., Bernardes, J., and Mazel, D. (2020) Structure-specific DNA recombination sites: Design, validation, and machine learning–based refinement. *Science Advances*.
(46) Chaikind, B., Bessen, J. L., Thompson, D. B., Hu, J. H., and Liu, D. R. (2016) A programmable Cas9-serine recombinase fusion protein that operates on DNA

sequences in mammalian cells. *Nucleic Acids Res. 44*, 9758–9770.
(47) Kelley, B. (2020) Developing therapeutic monoclonal antibodies at pandemic pace. *Nat. Biotechnol. 38*, 540–545.
(48) Wiberg, F. C., Rasmussen, S. K., Frandsen, T. P., Rasmussen, L. K., Tengbjerg, K., Coljee, V. W., Sharon, J., Yang, C.-Y., Bregenholt, S., Nielsen, L. S., Haurum, J. S., and Tolstrup, A. B. (2006) Production of target-specific recombinant human polyclonal antibodies in mammalian cells. *Biotechnol. Bioeng. 94*, 396–405.
(49) Crawford, Y., Zhou, M., Hu, Z., Joly, J., Snedecor, B., Shen, A., and Gao, A. (2013) Fast identification of reliable hosts for targeted cell line development from a limited-genome screening using combined φC31 integrase and CRE-Lox technologies. *Biotechnology Progress*.
(50) Zhang, L., Inniss, M. C., Han, S., Moffat, M., Jones, H., Zhang, B., Cox, W. L., Rance, J. R., and Young, R. J. (2015) Recombinase-mediated cassette exchange (RMCE) for monoclonal antibody expression in the commercially relevant CHOK1SV cell line. *Biotechnol. Prog. 31*, 1645–1656.
(51) Inniss, M. C., Bandara, K., Jusiak, B., Lu, T. K., Weiss, R., Wroblewska, L., and Zhang, L. (2017) A novel Bxb1 integrase RMCE system for high fidelity site-specific integration of mAb expression cassette in CHO Cells. *Biotechnol. Bioeng. 114*, 1837–1846.
(52) Carver, J., Ng, D., Zhou, M., Ko, P., Zhan, D., Yim, M., Shaw, D., Snedecor, B., Laird, M. W., Lang, S., Shen, A., and Hu, Z. (2020) Maximizing antibody production in a targeted integration host by optimization of subunit gene dosage and position. *Biotechnology Progress*.
(53) Gaidukov, L., Wroblewska, L., Teague, B., Nelson, T., Zhang, X., Liu, Y., Jagtap, K., Mamo, S., Tseng, W. A., Lowe, A., Das, J., Bandara, K., Baijuraj, S., Summers, N. M., Lu, T. K., Zhang, L., and Weiss, R. (2018) A multi-landing pad DNA integration platform for mammalian cell engineering. *Nucleic Acids Res. 46*, 4072–4086.
(54) Chen, G., Babb, R., and Fandl, J. P. (2018, November 15) Enhanced expression and stability regions. *US Patent*.
(55) Feary, M., Young, R. J., Moffat, M., Casperson, G. F., Jones, H. L., and Zhang, L. (2020, January 2) Multi-site specific integration cells for difficult to express proteins. *US Patent*.
(56) Ng, C. K. D., Crawford, Y. G., Shen, A., Zhou, M., Snedecor, B. R., Misaghi, S., and Gao, A. E. (2019, August 15) Targeted integration of nucleic acids. *World Patent*.
(57) O'callaghan, P. M., Bevan, S., Young, R., Fraser, P., and Zhang, L. (2020, April 9) Ssi cells with predictable and stable transgene expression and methods of formation. *World Patent*.
(58) Hamaker, N. K., and Lee, K. H. (2018) Site-specific integration ushers in a new era of precise CHO cell line engineering. *Current Opinion in Chemical Engineering*.
(59) Pristovšek, N., Nallapareddy, S., Grav, L. M., Hefzi, H., Lewis, N. E., Rugbjerg, P., Hansen, H. G., Lee, G. M., Andersen, M. R., and Kildegaard, H. F. (2019) Systematic Evaluation of Site-Specific Recombinant Gene Expression for Programmable Mammalian Cell Engineering. *ACS Synth. Biol. 8*, 758–774.
(60) Simon Fisher, Kerstin Otte. (2019) CHO Cell Engineering for Improved Process Performance and Product Quality, in *Cell Culture Engineering: Recombinant Protein Production* (Gyun Min Lee, Helene Faustrup Kildegaard, Ed.).
(61) Hansen, H. G., Pristovšek, N., Kildegaard, H. F., and Lee, G. M. (2017) Improving the secretory capacity of Chinese hamster ovary cells by ectopic expression of effector genes: Lessons learned and future directions. *Biotechnol. Adv. 35*, 64–76.

(62) Lee, J. S., Grav, L. M., Lewis, N. E., and Faustrup Kildegaard, H. (2015) CRISPR/Cas9-mediated genome engineering of CHO cell factories: Application and perspectives. *Biotechnol. J. 10*, 979–994.
(63) Hilliard, W., MacDonald, M. L., and Lee, K. H. (2020) Chromosome-scale scaffolds for the Chinese hamster reference genome assembly to facilitate the study of the CHO epigenome. *Biotechnology and Bioengineering*.
(64) Yusufi, F. N. K., Lakshmanan, M., Ho, Y. S., Loo, B. L. W., Ariyaratne, P., Yang, Y., Ng, S. K., Tan, T. R. M., Yeo, H. C., Lim, H. L., Ng, S. W., Hiu, A. P., Chow, C. P., Wan, C., Chen, S., Teo, G., Song, G., Chin, J. X., Ruan, X., Sung, K. W. K., Hu, W.-S., Yap, M. G. S., Bardor, M., Nagarajan, N., and Lee, D.-Y. (2017) Mammalian Systems Biotechnology Reveals Global Cellular Adaptations in a Recombinant CHO Cell Line. *Cell Syst 4*, 530–542.e6.
(65) Orellana, C. A., Marcellin, E., Palfreyman, R. W., Munro, T. P., Gray, P. P., and Nielsen, L. K. (2018) RNA-Seq Highlights High Clonal Variation in Monoclonal Antibody Producing CHO Cells. *Biotechnol. J. 13*, e1700231.
(66) Chang, M. M., Gaidukov, L., Jung, G., Tseng, W. A., Scarcelli, J. J., Cornell, R., Marshall, J. K., Lyles, J. L., Sakorafas, P., Chu, A.-H. A., Cote, K., Tzvetkova, B., Dolatshahi, S., Sumit, M., Mulukutla, B. C., Lauffenburger, D. A., Figueroa, B., Jr, Summers, N. M., Lu, T. K., and Weiss, R. (2019) Small-molecule control of antibody N-glycosylation in engineered mammalian cells. *Nat. Chem. Biol. 15*, 730–736.
(67) Tadauchi, T., Lam, C., Liu, L., Zhou, Y., Tang, D., Louie, S., Snedecor, B., and Misaghi, S. (2019) Utilizing a regulated targeted integration cell line development approach to systematically investigate what makes an antibody difficult to express. *Biotechnol. Prog. 35*, e2772.
(68) Lee, Y., Kwak, J. M., and Lee, J. S. (2020) Endogenous p21-Dependent Transgene Control for CHO Cell Engineering. *ACS Synth. Biol. 9*, 1572–1580.
(69) Gutierrez, J. M., Feizi, A., Li, S., Kallehauge, T. B., Hefzi, H., Grav, L. M., Ley, D., Baycin Hizal, D., Betenbaugh, M. J., Voldborg, B., Faustrup Kildegaard, H., Min Lee, G., Palsson, B. O., Nielsen, J., and Lewis, N. E. (2020) Genome-scale reconstructions of the mammalian secretory pathway predict metabolic costs and limitations of protein secretion. *Nat. Commun. 11*, 68.
(70) Almeida, A. M., Queiroz, J. A., Sousa, F., and Sousa, Â. (2020) Minicircle DNA: The Future for DNA-Based Vectors? *Trends Biotechnol.*
(71) Wang, X., Kawabe, Y., Hada, T., Ito, A., and Kamihira, M. (2018) Cre-Mediated Transgene Integration in Chinese Hamster Ovary Cells Using Minicircle DNA Vectors. *Biotechnol. J. 13*, e1800063.
(72) Zuris, J. A., Thompson, D. B., Shu, Y., Guilinger, J. P., Bessen, J. L., Hu, J. H., Maeder, M. L., Joung, J. K., Chen, Z.-Y., and Liu, D. R. (2015) Cationic lipid-mediated delivery of proteins enables efficient protein-based genome editing in vitro and in vivo. *Nat. Biotechnol. 33*, 73–80.
(73) Liang, X., Potter, J., Kumar, S., Ravinder, N., and Chesnut, J. D. (2017) Enhanced CRISPR/Cas9-mediated precise genome editing by improved design and delivery of gRNA, Cas9 nuclease, and donor DNA. *J. Biotechnol. 241*.
(74) Regeneron Begins First Clinical Trials of Anti-Viral Antibody Cocktail REGN-COV2 for the Treatment and Prevention of COVID-19.
(75) Thanh Le, T., Andreadakis, Z., Kumar, A., Gómez Román, R., Tollefsen, S., Saville, M., and Mayhew, S. (2020) The COVID-19 vaccine development landscape. *Nat. Rev. Drug Discov. 19*, 305–306.

(76) Pascal, K. E., Dudgeon, D., Trefry, J. C., Anantpadma, M., Sakurai, Y., Murin, C. D., Turner, H. L., Fairhurst, J., Torres, M., Rafique, A., Yan, Y., Badithe, A., Yu, K., Potocky, T., Bixler, S. L., Chance, T. B., Pratt, W. D., Rossi, F. D., Shamblin, J. D., Wollen, S. E., Zelko, J. M., Carrion, R., Jr, Worwa, G., Staples, H. M., Burakov, D., Babb, R., Chen, G., Martin, J., Huang, T. T., Erlandson, K., Willis, M. S., Armstrong, K., Dreier, T. M., Ward, A. B., Davey, R. A., Pitt, M. L. M., Lipsich, L., Mason, P., Olson, W., Stahl, N., and Kyratsous, C. A. (2018) Development of Clinical-Stage Human Monoclonal Antibodies That Treat Advanced Ebola Virus Disease in Nonhuman Primates. *J. Infect. Dis. 218*, S612–S626.
(77) Orengo, J. M., Radin, A. R., Kamat, V., Badithe, A., Ben, L. H., Bennett, B. L., Zhong, S., Birchard, D., Limnander, A., Rafique, A., Bautista, J., Kostic, A., Newell, D., Duan, X., Franklin, M. C., Olson, W., Huang, T., Gandhi, N. A., Lipsich, L., Stahl, N., Papadopoulos, N. J., Murphy, A. J., and Yancopoulos, G. D. (2018) Treating cat allergy with monoclonal IgG antibodies that bind allergen and prevent IgE engagement. *Nat. Commun. 9*, 1421.
(78) Casewell, N. R., Jackson, T. N. W., Laustsen, A. H., and Sunagar, K. (2020) Causes and Consequences of Snake Venom Variation. *Trends Pharmacol. Sci. 41*, 570–581.
(79) Blanco, N., Williams, A. J., Tang, D., Zhan, D., Misaghi, S., Kelley, R. F., and Simmons, L. C. (2020) Tailoring translational strength using Kozak sequence variants improves bispecific antibody assembly and reduces product-related impurities in CHO cells. *Biotechnology and Bioengineering*.
(80) Cartwright, J. F., Arnall, C. L., Patel, Y. D., Barber, N. O. W., Lovelady, C. S., Rosignoli, G., Harris, C. L., Dunn, S., Field, R. P., Dean, G., Daramola, O., Gibson, S. J., Peden, A. A., Brown, A. J., Hatton, D., and James, D. C. (2020) A platform for context-specific genetic engineering of recombinant protein production by CHO cells. *J. Biotechnol. 312*, 11–22.
(81) Brown, A. J., Gibson, S. J., Hatton, D., Arnall, C. L., and James, D. C. (2019) Whole synthetic pathway engineering of recombinant protein production. *Biotechnol. Bioeng. 116*, 375–387.
(82) Eisenhut, P., Klanert, G., Weinguny, M., Baier, L., Jadhav, V., Ivansson, D., and Borth, N. (2018) A CRISPR/Cas9 based engineering strategy for overexpression of multiple genes in Chinese hamster ovary cells. *Metab. Eng. 48*, 72–81.
(83) Matreyek, K. A., Stephany, J. J., Chiasson, M. A., Hasle, N., and Fowler, D. M. (2020) An improved platform for functional assessment of large protein libraries in mammalian cells. *Nucleic Acids Res. 48*, e1.
(84) Xiong, K., la Cour Karottki, K. J., Hefzi, H., Li, S., Grav, L. M., Li, S., Spahn, P., Lee, J. S., Lee, G. M., Kildegaard, H. F., Lewis, N. E., and Pedersen, L. E. (2020, May 19) Using targeted genome integration for virus-free genome-wide mammalian CRISPR screen.
(85) Parthiban, K., Perera, R. L., Sattar, M., Huang, Y., Mayle, S., Masters, E., Griffiths, D., Surade, S., Leah, R., Dyson, M. R., and McCafferty, J. (2019) A comprehensive search of functional sequence space using large mammalian display libraries created by gene editing. *MAbs 11*, 884–898.
(86) Mason, D. M., Weber, C. R., Parola, C., Meng, S. M., Greiff, V., Kelton, W. J., and Reddy, S. T. (2018) High-throughput antibody engineering in mammalian cells by CRISPR/Cas9-mediated homology-directed mutagenesis. *Nucleic Acids Res. 46*.

Supporting information for Chapter 3

Minimizing clonal variation during mammalian cell line engineering for improved systems biology data generation

Lise Marie Grav[1], Daria Sergeeva[1], Jae Seong Lee[1,2], Igor Marin de Mas[1], Nathan E. Lewis[3, 4], Mikael Rørdam Andersen[5], Lars Keld Nielsen[1,6], Gyun Min Lee[1,7], Helene Faustrup Kildegaard[1]

[1]The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Kgs. Lyngby, Denmark

[2]Department of Molecular Science and Technology, Ajou University, Suwon, Republic of Korea

[3]Department of Pediatrics, University of California, San Diego, United States

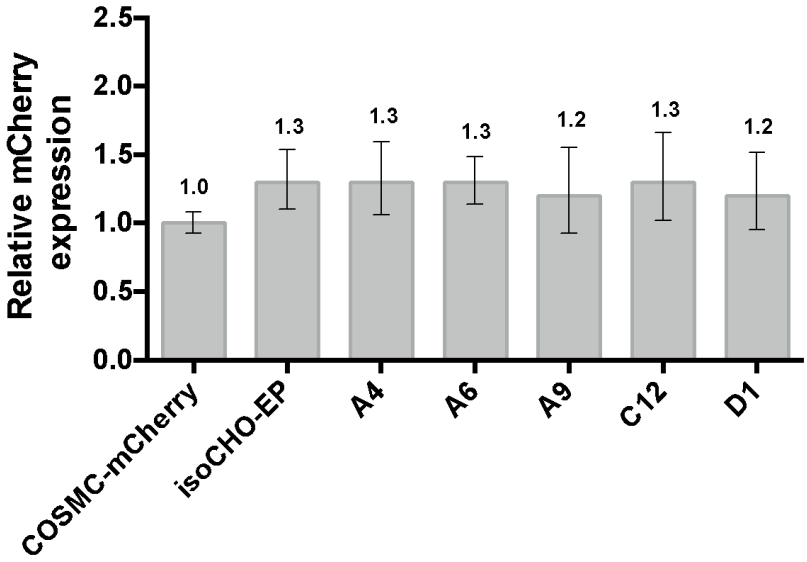
[4]The Novo Nordisk Foundation Center for Biosustatainability, University of California, San Diego, United States

[5]Department of Biotechnology and Biomedicine, Technical University of Denmark, Kgs. Lyngby, Denmark

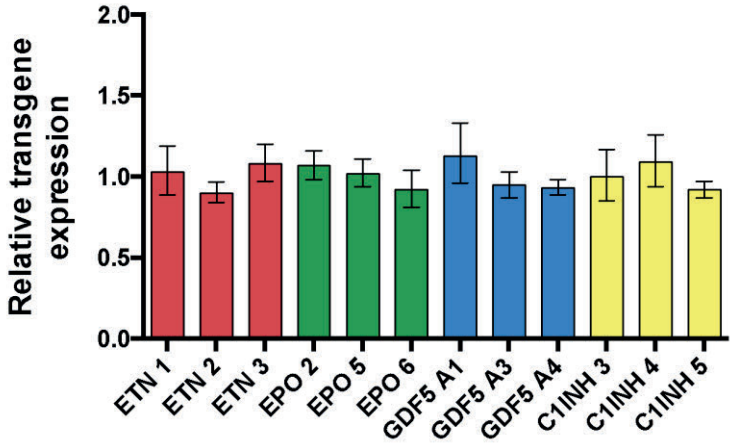
[6]Australian Institute for Bioengineering and Nanotechnology, University of Queensland, Brisbane, Australia

[7]Department of Biological Sciences, KAIST, Daejeon, Republic of Korea

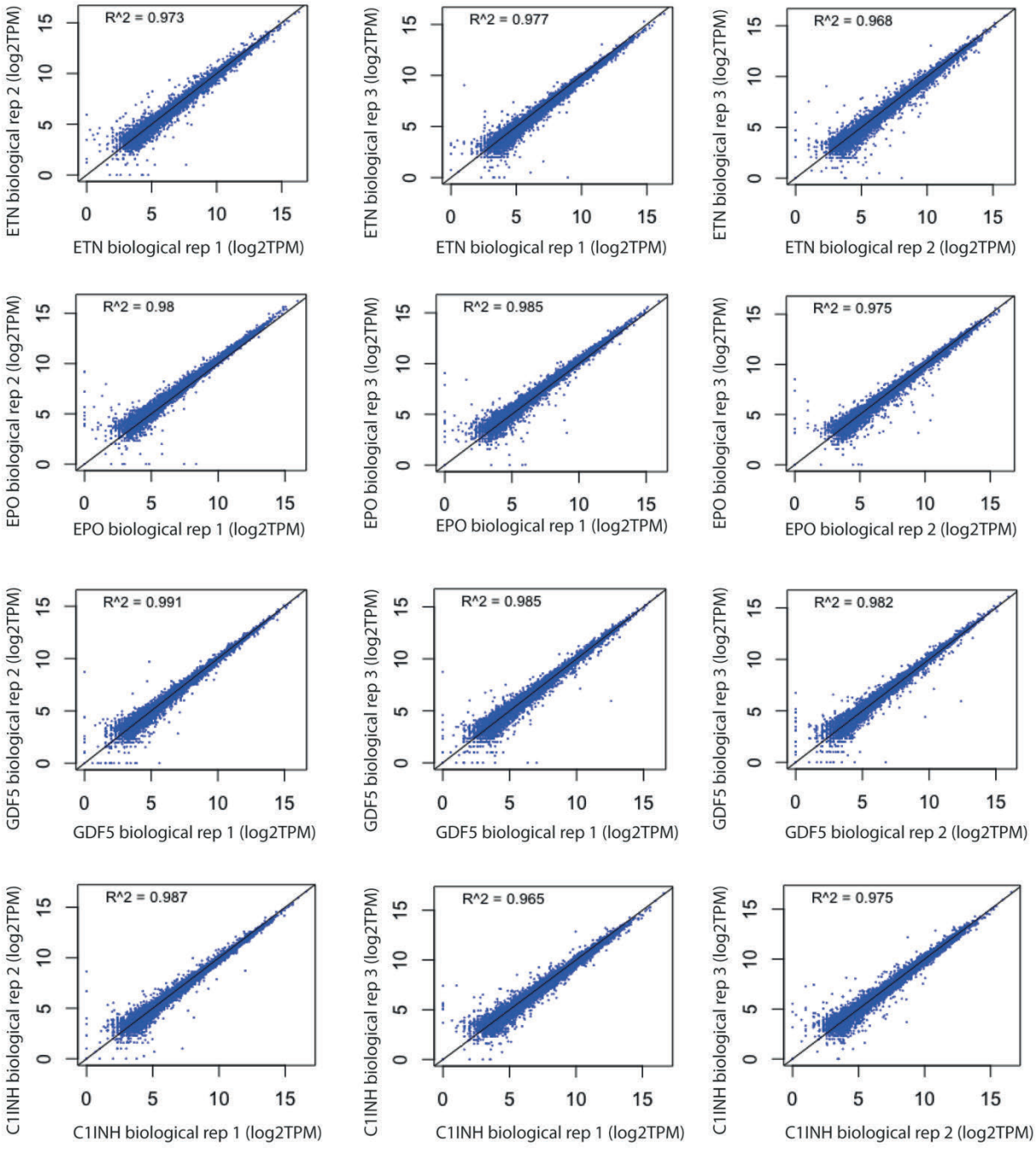
Supporting Figures



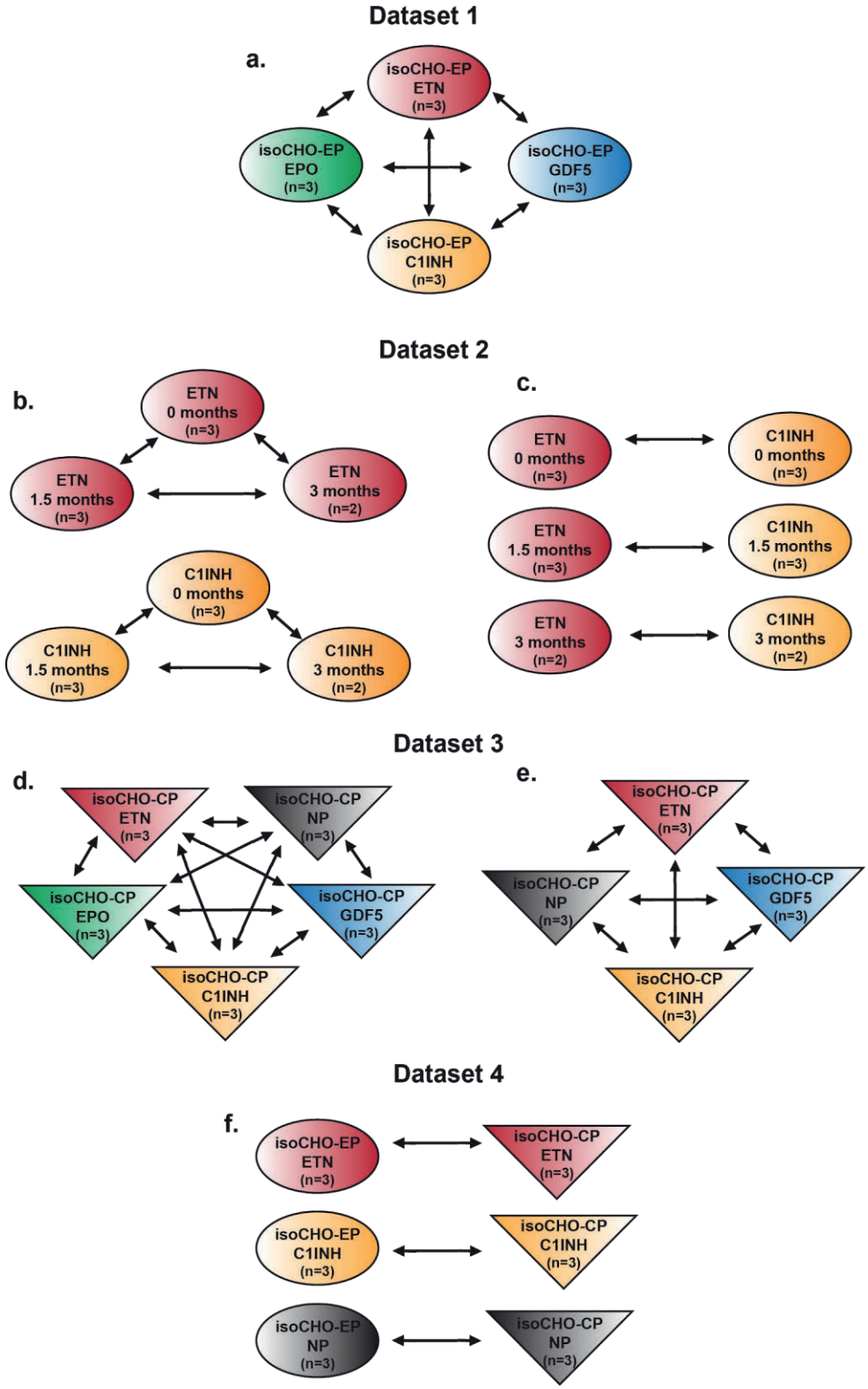
Supporting Figure S1 Relative mCherry expression levels of the panel of mCherry-EP clones, compared to COSMC-mCherry clone, generated in previous study[7]**.** Error bars represent the standard deviations of technical replicates (n≥3).



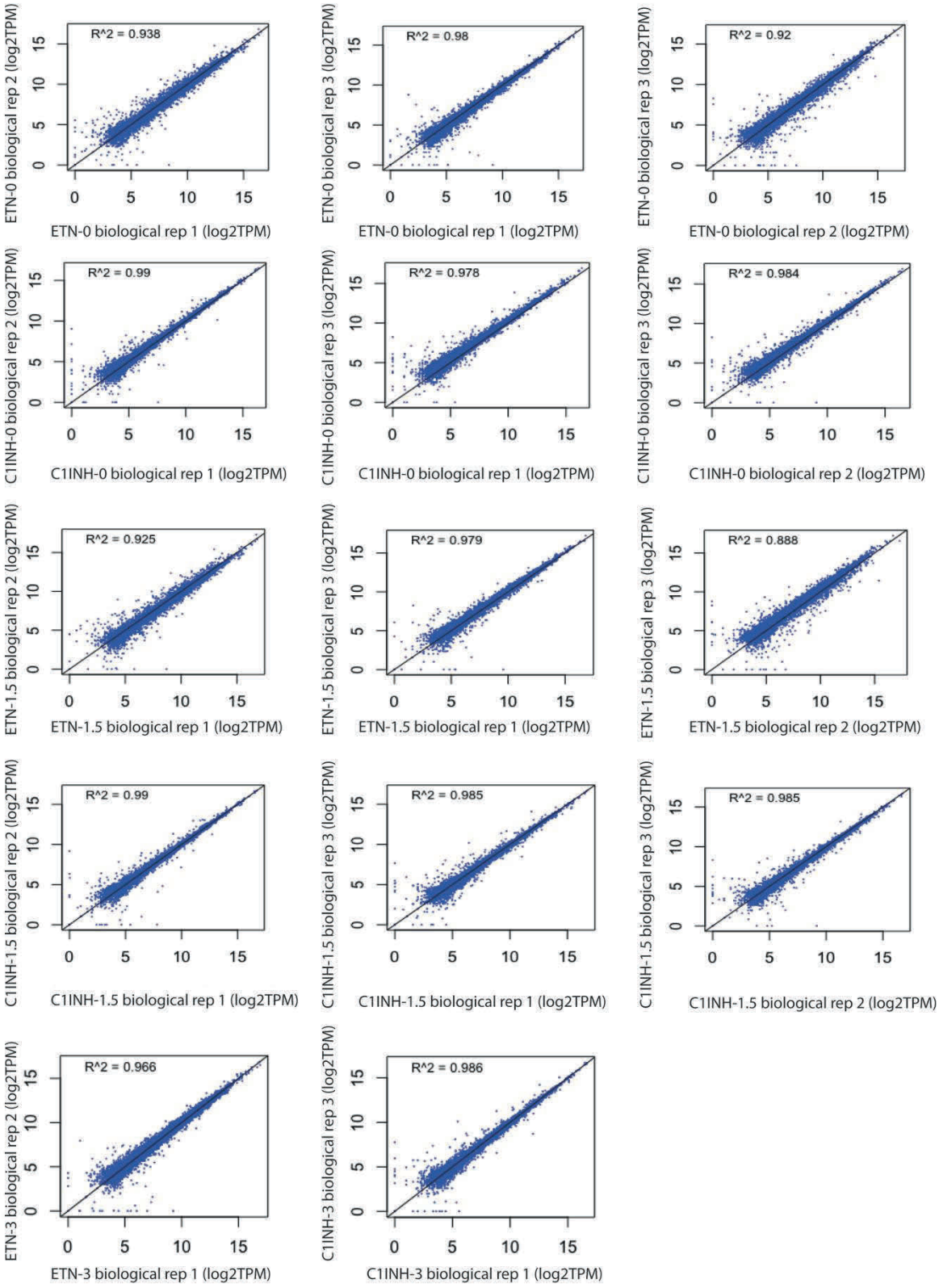
Supporting Figure S2 Relative levels of transgene expression in isoCHO-EP-derived subclones, measured by qRT-PCR, normalized to average value of corresponding isoCHO-EP subclones. The error bars represent the standard deviations of technical replicates (n=3).



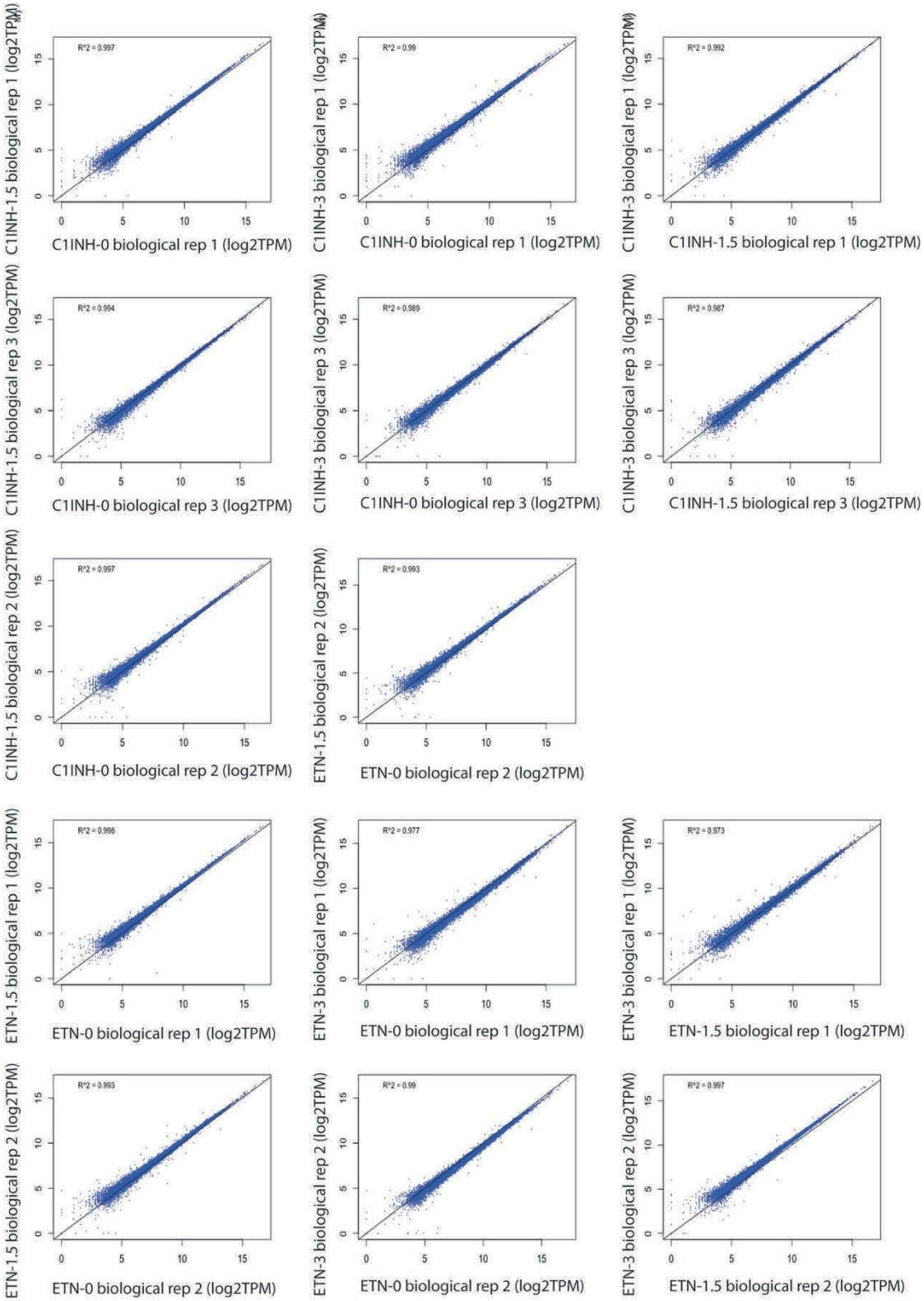
Supporting Figure S3 Comparison of expression values for pairs of replicates in dataset 1, measured in log2TPM, the spearman correlation coefficient (R^2) is denoted in the top left corner for each comparison.



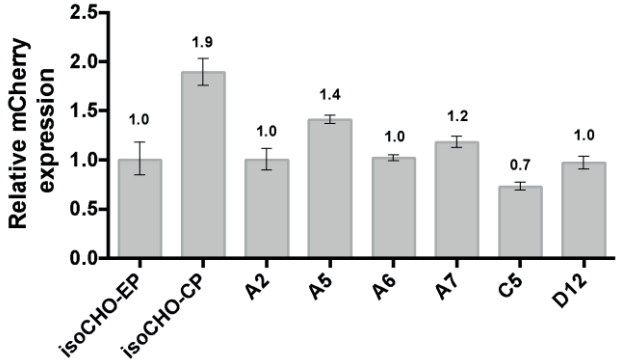
Supporting Figure S4 Overview of differential expression analysis datasets.



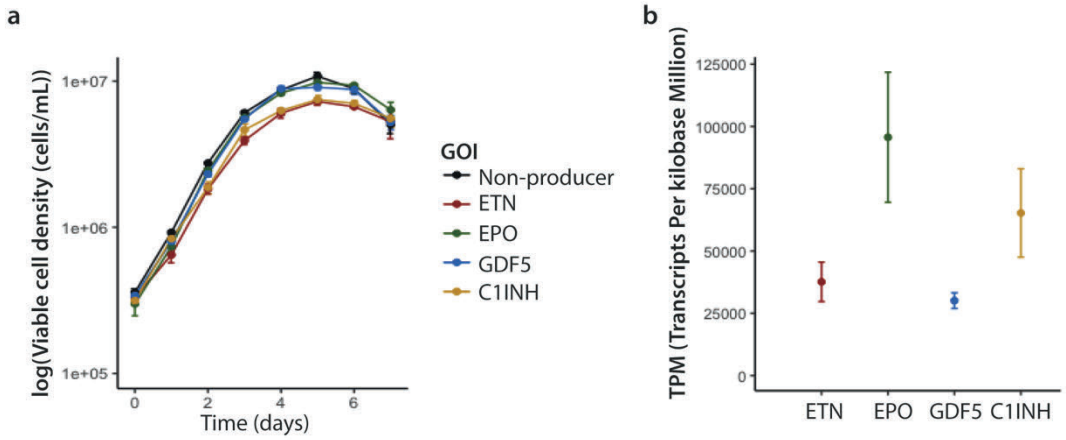
Supporting Figure S5 Comparison of expression values for pairs of replicates in dataset 2, measured in log2TPM, the spearman correlation coefficient (R^2) is denoted in the top left corner for each comparison. The number following the name of the gene of interest denotes the number of months in culture for the particular sample.



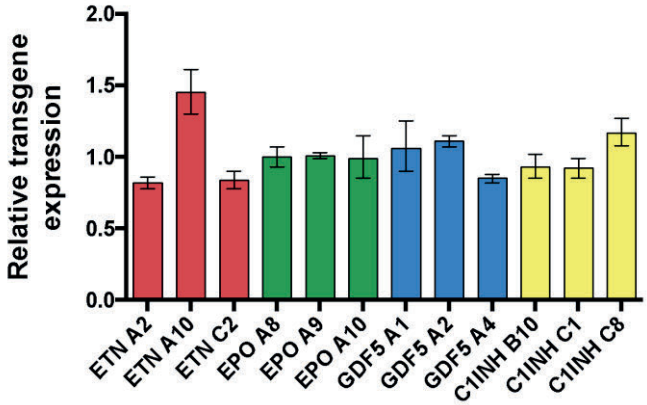
Supporting Figure S6 Comparison of expression values within biological replicates measured at different time points, measured in log2TPM, the spearman correlation coefficient (R^2) is denoted in the top left corner for each comparison. The number following the name of the gene of interest denotes the sample time point (in months).



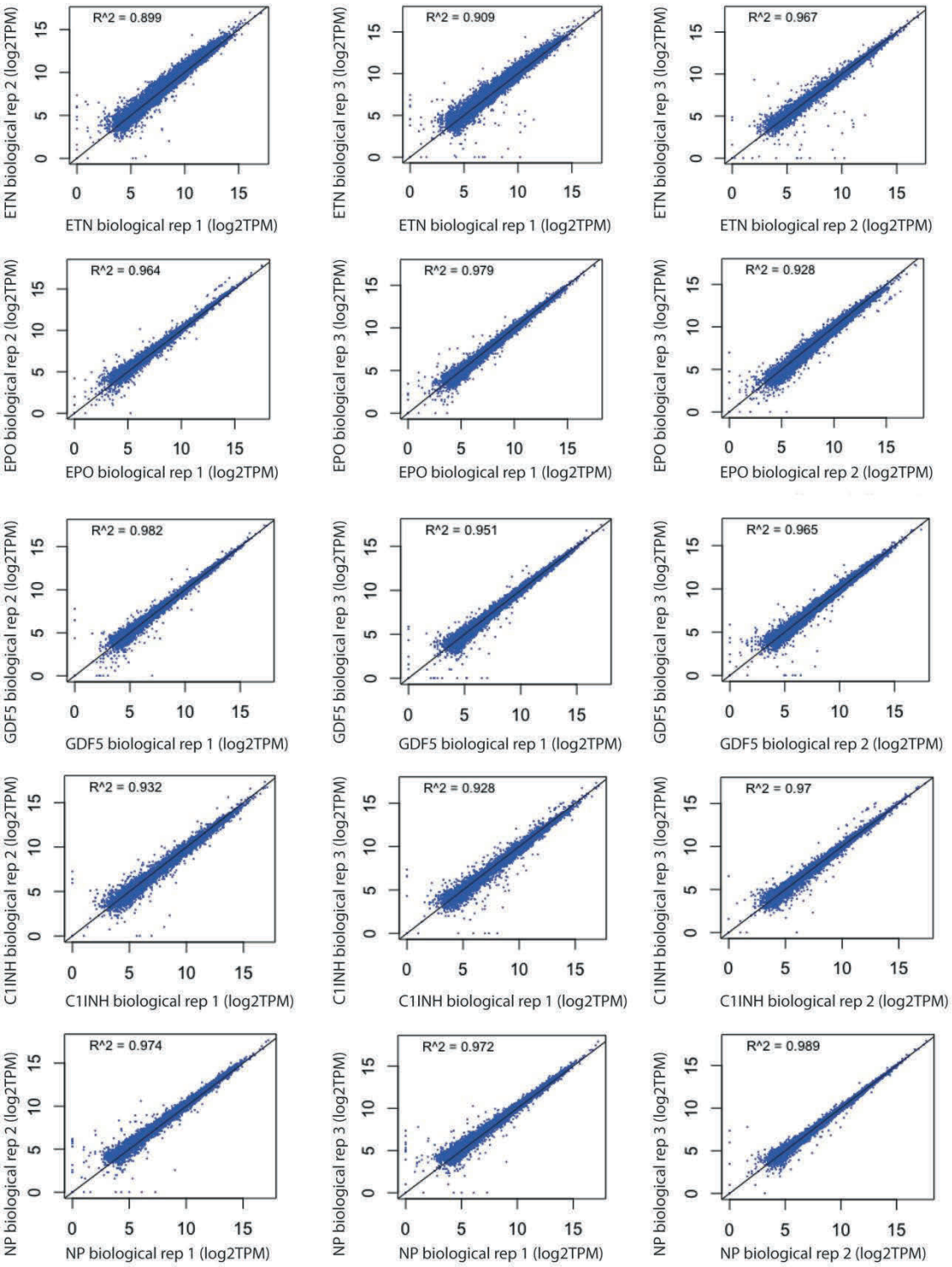
Supporting Figure S7. Relative mCherry expression levels of the panel of mCherry-CP clones, normalized to isoCHO-EP**.** Error bars represent the standard deviations of technical replicates (n≥3).



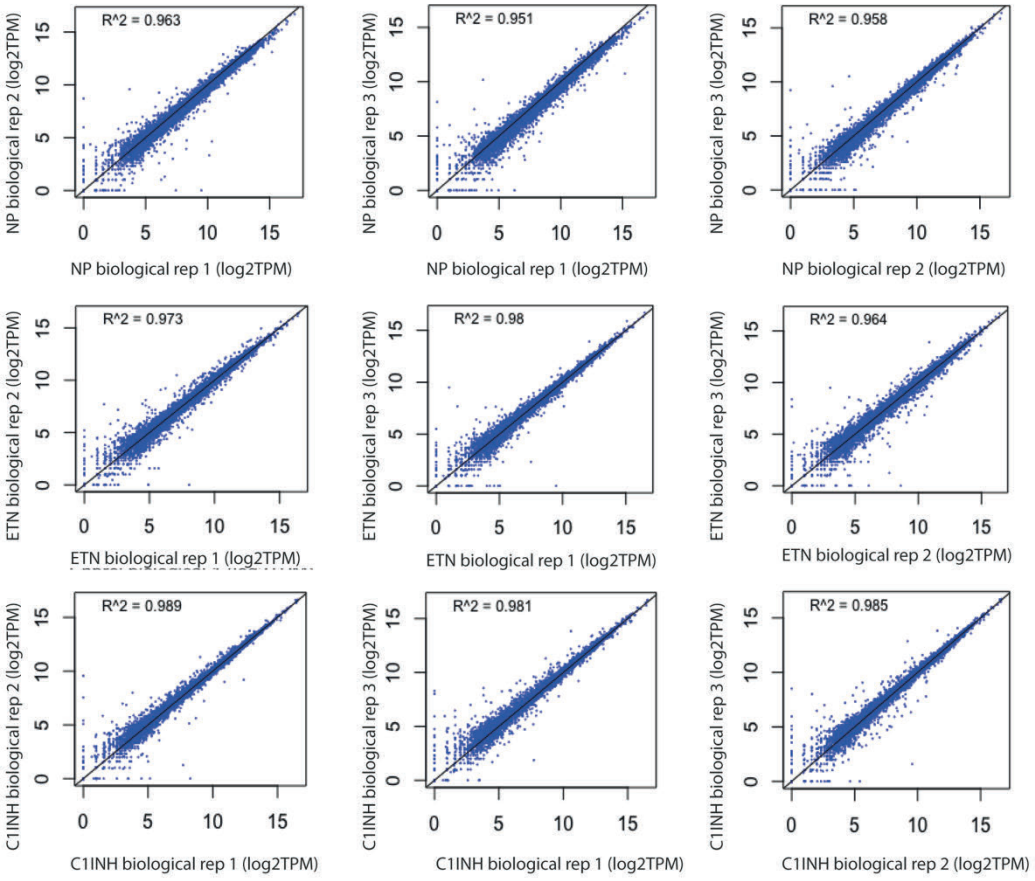
Supporting Figure S8. Phenotypes of isoCHO-CP subclones. (a) Viable cell densities of isoCHO-CP subclones expressing ETN, EPO, GDF5 or C1INH. The error bars of each line represent the standard deviations of three isogenic subclones expressing the same GOI (n=3). **(b)** Relative levels of transgene expression, as measured in transcripts per kilobase million (TPM) of ETN, EPO, GDF5 or C1INH. The error bars represent the standard deviations of three isogenic clones expressing the same GOI (n=3).



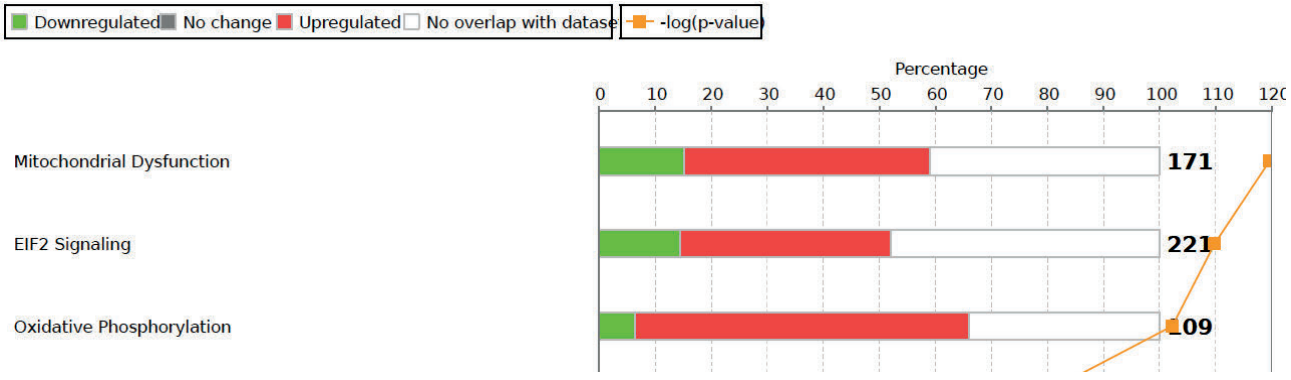
Supporting Figure S9 Relative levels of transgene expression in isoCHO-CP-derived subclones, measured by qRT-PCR, normalized to average value of corresponding isoCHO-CP subclones. The error bars represent the standard deviations of technical replicates (n=3).



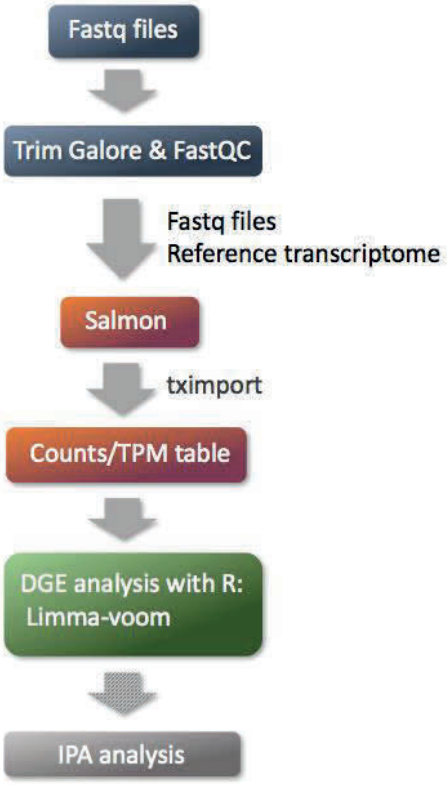
(continues on the next page)



Supporting Figure S10 Comparison of expression values for a pair of replicates in dataset 3 and 4, measured in log2TPM, the spearman correlation coefficient (R^2) is denoted in the top left corner for each comparison. Row 1-5 are isoCHO-CP clones, and row 6-8 are isoCHO-EP clones.



Supporting Figure S11 Top three pathways from the canonical pathway enrichment analysis.



Supporting Figure S12 Overview of the transcriptomic analysis pipeline.

Supporting Tables

Supporting Table S1. Fluorescence level analysis of mCherry-EP clones. Data shows mean value of two replicates.

Clone	CHO-S WT	A4	A6	A9	C12	D1	isoCHO-EP
mCherry positive population	0,0%	99,7%	100,0%	99,8%	99,8%	99,8%	100,0%
mCherry intensity	4	138	133	136	128	128	158

Supporting Table S2. Fluorescence level analysis of mCherry-CP clones. Data shows mean value of two replicates.

Clone	CHO-S WT	A2	A5	A6	A7	C5	D12	isoCHO-CP
mCherry positive population	0,2%	97,6%	92,0%	97,2%	87,4%	93,4%	94,4%	93,2%
mCherry intensity	6	139	165	160	148	188	136	205

Supporting Table S3. Nucleotide sequence of CHO codon-optimized human GDF5 gene.

ATGAGACTGCCCAAGCTGCTGACCTTCCTGCTGTGGTATCTGGCCTGGCTGGACCTGGAATTCATCTGCACCGTGCTGGGCGCTCCCGATCTGG
GACAGAGGCCTCAGGGAACCAGACCCGGACTGGCTAAGGCCGAGGCCAAAGAGAGGCCTCCCCTGGCCAGAAACGTGTTCAGACCTGGCGG
CCACTCTTACGGCGGAGGCGCCACCAATGCCAACGCCAGAGCTAAGGGCGGCACCGGACAGACAGGTGGCCTGACCCAGCCTAAGAAGGAC
GAGCCCAAGAAGCTGCCTCCTAGACCAGGCGGCCCTGAGCCTAAGCCTGGACATCCTCCACAGACCAGACAGGCCACCGCCAGAACCGTGAC
CCCTAAGGGACAGCTGCCTGGCGGAAAGGCCCCTCCTAAGGCTGGCTCTGTGCCCTCCAGCTTTCTGCTGAAGAAGGCCAGAGAGCCTGGCCC
CCCTAGAGAGCCCAAAGAGCCCTTCAGACCCCCCCCTATCACCCCCCACGAGTACATGCTGTCCCTGTACCGGACCCTGTCTGACGCCGATCGG
AAGGGCGGAAACTCCTCCGTGAAGCTGGAAGCCGGCCTGGCCAACACCATCACCAGCTTCATCGACAAGGGCCAGGACGACAGGGGTCCCGT
CGTGCGGAAGCAGAGATACGTGTTCGACATCTCCGCCCTGGAAAAGGACGGCCTGCTGGGAGCCGAGCTGCGGATCCTGAGAAAGAAGCCTT
CCGACACCGCCAAGCCTGCTGCTCCTGGCGGAGGTAGAGCTGCCCAGCTGAAGCTGTCCAGCTGCCCTTCTGGCAGACAGCCTGCCGCTCTGC
TGGATGTGCGATCTGTGCCAGGACTGGACGGCTCCGGATGGGAGGTGTTCGATATCTGGAAGCTGTTCCGCAACTTCAAGAACTCCGCCCAGC
TGTGCCTGGAACTGGAAGCTTGGGAGAGGGGCAGAGCCGTGGATCTGAGAGGCCTGGGCTTCGACAGAGCCGCTAGACAGGTGCACGAGA
AGGCCCTGTTTCTGGTGTTCGGCCGGACCAAGAAGCGGGACCTGTTCTTCAACGAGATCAAGGCCAGATCCGGCCAGGATGACAAGACCGTG
TACGAGTACCTGTTCTCCCAGCGGCGGAAGCGGAGAGCCCCTCTGGCTACAAGACAGGGCAAGCGGCCCTCCAAGAACCTGAAGGCCCGGTG
CTCTAGAAAGGCCCTGCACGTGAACTTCAAGGACATGGGCTGGGACGACTGGATCATTGCCCCCCTGGAATACGAGGCCTTCCACTGCGAGG
GCCTGTGCGAGTTCCCTCTGAGATCCCACCTGGAACCCACCAACCACGCCGTGATCCAGACCCTGATGAACTCCATGGACCCCGAGTCCACCCC
CCCTACCTGTTGTGTGCCTACCCGGCTGTCCCCCATCTCCATCCTGTTCATCGACTCCGCCAACAACGTGGTGTACAAGCAGTACGAGGACATG
GTGGTGGAATCCTGCGGCTGCCGGTGA

Supporting Table S4. Nucleotide sequences of the 750 bp homology arms of the donor plasmid

Nucleotide sequence of 5' homology arm
GATAAACTATTCTTCATTTGCCAGGGGTGAACGAAACTCCAGTGCACATTTCTCAGCTCTATAGCTGAGAGATACCTCATAAAACATAGGATAC TTAGGTAAATTTGAGTTTCAGGTGAAGATTTTTTAAGTATAATCATGTCTTGCACAATGTTTGGGATTCATTTTTATTGATTAAATCTGACAGTCC TTTTTTTTTCTTTCCTTTAAAAAACTTTTTTTTTTTTTTTAAAGACAGGGTCTTATGTATCCTAAGGTGACCTTGAACTTGCCTTGTAGATGAGTCTG GCTTTGAGCCCCTAATACCTACCTCCACCTCTCAAATCCTAGGATTACAGGTGTATACCACTCTACCCATTTGACAATTCTACTTTTGAGCCATCT TAAAGGTAAAAAGTACACTTATTTATCATATGCCTGTGTGTGGAAGTCAGAACAACTTGTGGTAGTTGGTTCTTTCCTTCATGGGGGCCCTGGG GGTTAGACTCCAATTGTCAGGCTTGCTGGTAAATGCCTTATTACCACCTTGATGTGGAGGTACAAATAATGATTCACAGGACTAGGTTCAAATA

TTAAATATAATTTATTCCGGGGAGAAGTCACTTACAGAAGAAGCAGACAGCTGCTGCAGGTTCCAGCTCTAAGGATCCGCACAGGCTACTCCC
TGTGGCCCAACCAGACGGAAGTGAGAGCCACCTGCTGGGACCCCAGACCTAAACTCACTCCCAGTCAGGTCCTCACCTGTAACCAC

Nucleotide sequence of 3' homology arm

TGGATATCTCACTACACTTTGATGGCCCAATCTTTTTTTAAAAGCATATATTTATTGAATAAAATATGAGTTCATATTGGCATCCTCATGCATGTC
TGGCAAGTACTTGGCTATTATTCACTTCCTCCTCTAAATCCCATCCATTCTTGGGTCCATAAACCTATCATCTTAGTTTCTCCTCATTTCAAAAGAA
ATTAGGAAACAGAAGTTTGTAACAGGAAATGGATTTATTTATAGTTTTGCCACTGGGGAGAAAAAATAACTCCATCCTGTTCTCTAACATGCTG
GTAGTCTTTTACAGTTTATTGAAAACTATGAGACGGGATGGAGAGGTGGTTCAGAGGTTAAGAGCAATGGCTGTTCTTCCAGAGGTCCTGAGT
TCAATTCCCAGCAACCACATAGTAGCTCACAGCTATCTGTAATGAGATCTGGTGTGCAAGCATACATGCCGGCAGAACACTGTATACAAAATAA
ATAAATCTTAAAAAAAAAAAGAAAACTATGAGACAAAGACAAGAAGGTATGTAGCTTTGCTGTGAGACAGTTCAACATGGCTTTTCTGGGGTT
GTGAGGAGGTCACTAGTTTCATATGACCTGCAGTATCTCAGTTCTCAGCAGTTCTTGTTTAGCATGGAGGGGGTCTCTCTTTACAATTTTGCCAT
GGTTCCAGCCCTTTCTCAAGCATGGAATCACTGCTATATTAGTGTCCTCTTTGTTATGGGGTAAAAACCCAGAGACCATTCAAGCAAC

Supporting Table S5. Primers for plasmids construction, junction and insert PCR.

sgRNA vector construction (for primer annealing)

Primer name	**Sequence (5' -3')**
sgRNA T9 fwd	GGAAAGGACGAAACACCGCCCACTACAGGTTCGGGCGGTTTTAGAGCTAGAAAT
sgRNA T9 rev	CTAAAACcgcccgaacctgtagtgggCGGTGTTTCGTCCTTTCCACAAGATAT

*Target sequence marked in blue

lox-mCherryOri vector construction (USER primers)

Primer name	**Sequence (5' -3')**
LoxP-kozak_mcherry_LA_fwd*	agtcggtgUATAACTTCGTATAGCATACATTATACGAAGTTATCGCCACCATGGTGAGCA
Lox2272-mcherry_O4_rev*	AGACTGTGUataacttcgtataaagtatcctatacgaagttatCTACTTGTACAGCTCGT
BGH pA_O4_fwd	ACACAGTCUCTGTGCCTTCTAGTTGCC
BGH pA_O5_rev	ACGCAAGUCCATAGAGCCCACCGCAT

*Lox sequences are marked in green.

Landing pad vectors construction (USER primers)

Primer name	**Sequence (5' -3')**
EF-1a_LB_fwd	aagcagcgUGTGAGGCTCCGGTGCCC
EF-1a_LC_rev	atgacgtcUTCACGACACCTGAAATGGAA
LinkB-mCMVenhancer-fwd	AAGCAGCGUGAGTCAATGGGAAAAACC

HTLV5'UTR-linkC-Rev	ATGACGTCUGTAGGCGCCGGTCACA
LoxP_LC_fwd	agacgtcaUATAACTTCGTATAGCATACAT
BGH pA_O2_rev	ATCGCACUccatagagcccaccgcatcc
Marker NeoR_O2_fwd	AGTGCGAUCTGTGGAATGTGTGTCAGTT
Marker NeoR_LD_rev	actcagaccUcagacatgataagatacattg
CMV_O1_fwd	ACGTCGCUGTTGACATTGATTATTGACT
BGH pA_O5_rev	ACGCAAGUccatagagcccaccgcatcc
pJ204 backbone_O5_fwd	ACTTGCGUAGTGAGTCGAATAAGGGCGACACAAA
pJ204 backbone_LA_rev	acaccgacUGAGTCGAATAAGGGCGACACCCCA
T9 5' arm_750bp_LA_fwd	agtcggtgUGATAAACTATTCTTCATTTGC
T9 5' arm_750bp_LB_rev	acgctgctUGTGGTTACAGGTGAGGACC
T9 3' arm_750bp_LD_fwd	aggtctgagUTGGATATCTCACTACACTTTG
T9 3' arm_750bp_O1_rev	AGCGACGUGTTGCTTGAATGGTCTCTGG

RMCE donor vectors construction (USER primers)

Primer name	**Sequence (5’ -3’)**
LoxP-kozak_EPO_LA_fwd	agtcggtgUATAACTTCGTATAGCATACATTATACGAAGTTATCGCCACCATGGGAGTGC
Lox2272-EPO_O5_rev	ACGCAAGUataacttcgtataaagtatcctatacgaagttatTCATCTATCGCCGGTCC
LoxP-kozak_ETN_LA_fwd	agtcggtgUATAACTTCGTATAGCATACATTATACGAAGTTATAGCACCATGGCGCCCGT
Lox2272-ETN_O5_rev	ACGCAAGUataacttcgtataaagtatcctatacgaagttatTTATCATTTACCCGGAG
LoxP-kozak_C1INH_LA_fwd	agtcggtgUATAACTTCGTATAGCATACATTATACGAAGTTATAGCACCATGGCCAGCAG
Lox2272-C1INH_O5_rev	ACGCAAGUataacttcgtataaagtatcctatacgaagttatTCAGGCTCTGGGGTCGTA
LoxP-kozak_GDF5_LA_fwd	AGTCGGTGUATAACTTCGTATAGCATACATTATACGAAGTTATGCCACCATGAGACTGCC
Lox2272-GDF5_O5_rev	ACGCAAGUataacttcgtataaagtatcctatacgaagttatTCACCGGCAGCCGCAG

Junction PCR

Primer name	**Sequence (5’ -3’)**
T9 5’ junction genomic fwd	GCATGCACAGAGAGGGACAT
T9 3’ junction genomic rev	CCCTCTGCAACTGCTAACCA
Neo(R) junction fwd	CTGGACGAAGAGCATCAGGG

EF1-a junction rev	ATCCTGGCCCGCATTTACAA
mCMV enh junction rev	TGGCTTACCTCCCATTGACC

Insert PCR

Primer name	**Sequence (5' -3')**
EF1a-mcherry junction fwd	CCTCAGACAGTGGTTCAAAGT
BGH pA rev	AGATGGCTGGCAACTAGAAG
min-hEF1a fwd	GGAGAACCGTATATAAGTGCAGTAG

Supporting Table S6. qRT-PCR primers and probes.

qRT-PCR primers (SYBR Green assay)

Gene	**Fwd primer**	**Rev primer**
mCherry	AGGACGGCGAGTTCATCTA	CCCATGGTCTTCTTCTGCATTA
GAPDH	TTGTCATCAACGGGAAGG	GTGAAGACGCCAGTAGATT

TaqMan assays

Gene	**Fwd primer**	**Rev primer**	**Probe**	**Dye**
Fkbp1a	CTCTCGGGACAGAAACAAGC	GACCTACACTCATCTGGGCTAC	ATGCTAGGCAAGCAGGAGGTGATC	VIC-MGB
Gnb1	CCATATGTTTCTTTCCCAATGGC	AAGTCGTCGTACCCAGCAAG	ACTGGTTCAGACGATGCTACGTGC	ABY-MGB
ETN	CAGCCGGAGAACAACTACAA	CATCACGGAGCATGAGAAGA	TACAGCAAGCTCACCGTGGACAAG	FAM-MGB
EPO	CTGGAAAGATACCTGCTGGAAG	AGGCGTAGAAGTTCACTTTGG	CCAAAGAGGCCGAGAACATCACCA	FAM-MGB
GDF5	GTGATCCAGACCCTGATGAAC	GTCGATGAACAGGATGGAGATG	TACCTGTTGTGTGCCTACCCGG	FAM-MGB
C1INH	GGATGGAGCCCTTTCACTTTA	GGATGACCAGGCTCAGATTATG	TCATCGACCAGACCCTGAAGGCTA	FAM-MGB

Supporting Table S7. RNA-seq datasets.

Dataset	**Samples**	**Sequencing kit**
1	**isoCHO-EP**: ETN (n=3), EPO (n=3), GDF5 (n=3), C1INH (n=3)	mid-output

2	**isoCHO-EP**: ETN-0 months (n=3), ETN-1.5 months (n=3), ETN-3 months (n=2), C1INH-0 months (n=3), C1INH-1.5 months (n=3), C1INH-3 months (n=2)	high-output
3	**isoCHO-CP**: ETN (n=3), EPO (n=3), GDF5 (n=3), C1INH (n=3), non-producer (n=3)	high-output
4	**isoCHO-EP**: ETN (n=3), C1INH (n=3), non-producer (n=3) **isoCHO-CP**: ETN (n=3), C1INH (n=3), non-producer (n=3)	high-output

Supporting information for Chapter 4

Multi-copy targeted integration for accelerated development of high-producing CHO cells

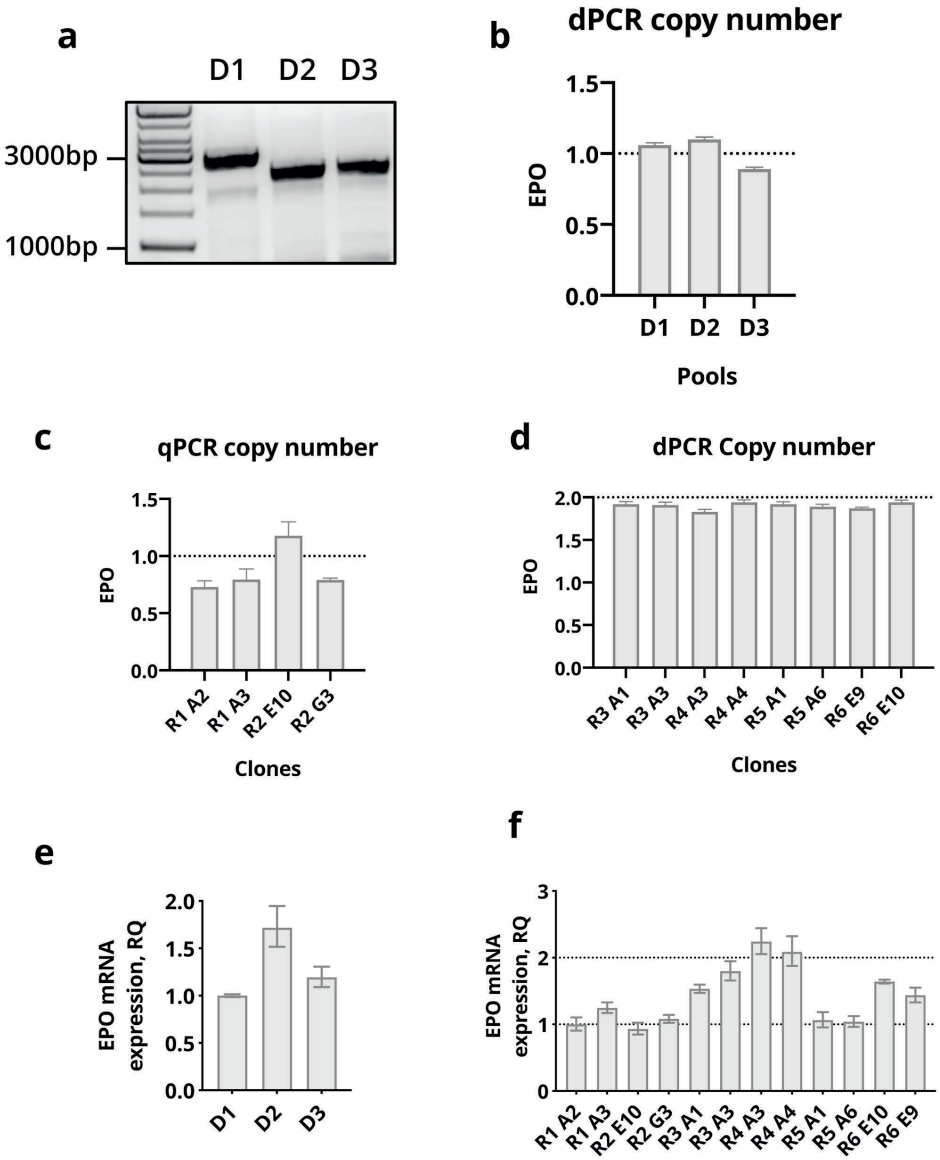
Daria Sergeeva[1], Gyun Min Lee[1,2], Lars Keld Nielsen[1,3], Lise Marie Grav[1*]

[1]The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Kgs. Lyngby, Denmark

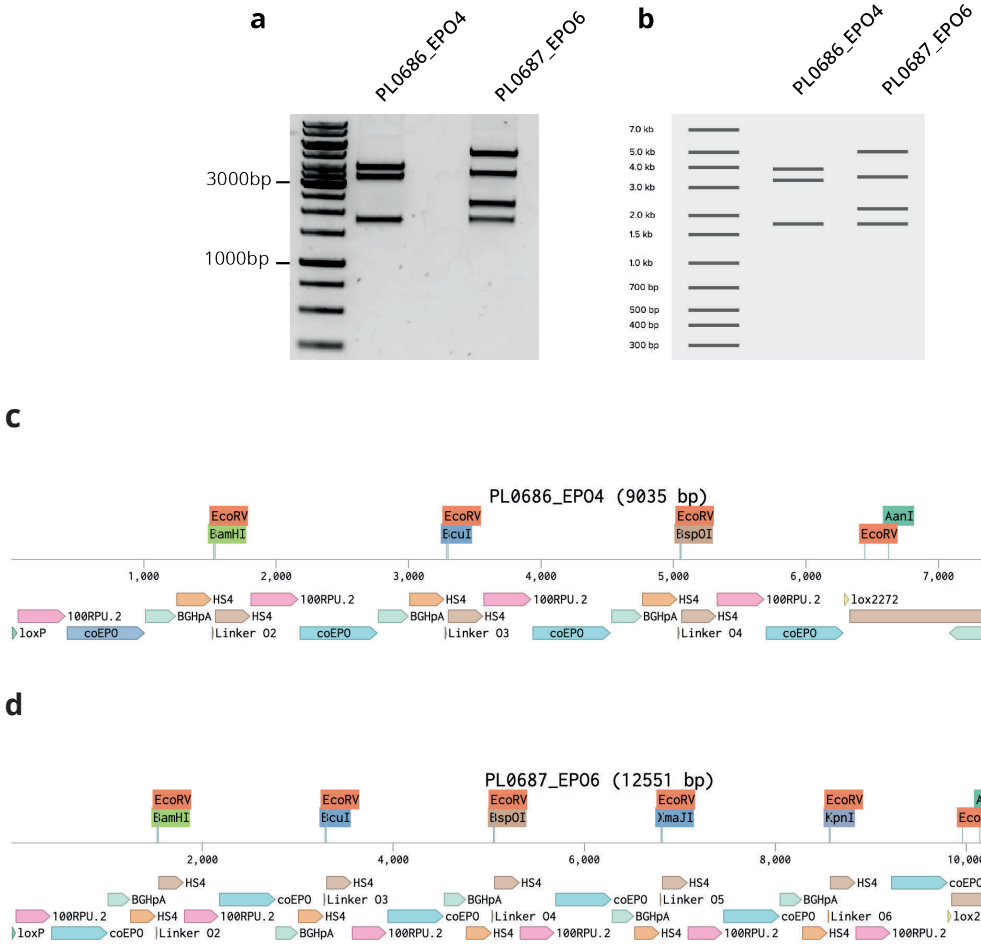
[2]Department of Biological Sciences, KAIST, Daejeon, Republic of Korea

[3]Australian Institute for Bioengineering and Nanotechnology, University of Queensland, Brisbane, Australia

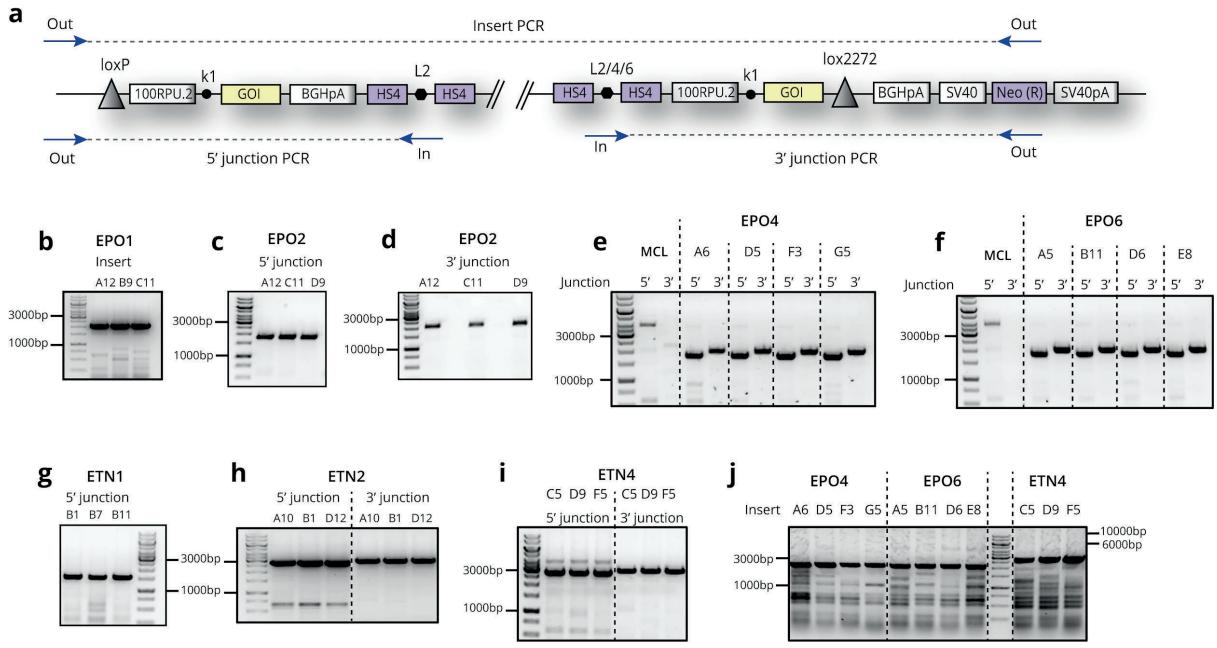
I. Supporting Figures



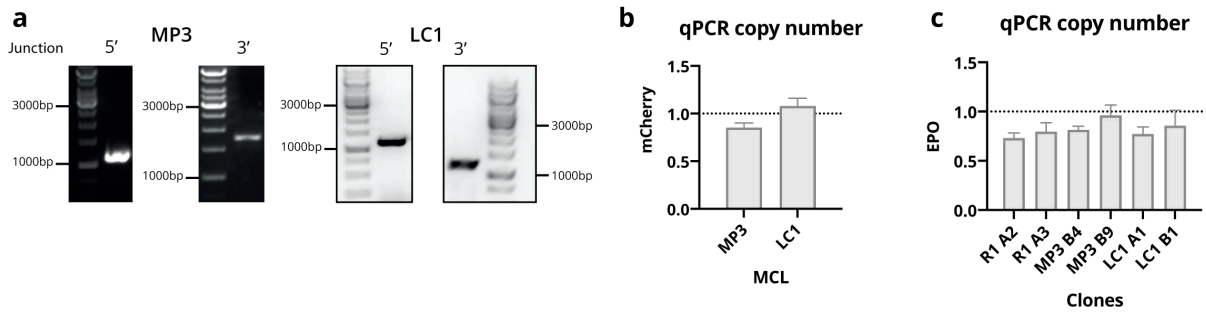
Supporting Figure 1. Verification of D1-D3 cell pools and R1-R6 clones. **a**. Insert PCR of D1-D3 pools (primer set I). **b**. Copy number analysis of cell pools D1-D3. **c**. Copy number analysis of clones R1 and R2. **d**. Copy number analysis of clones R3-R6. **e.** EPO mRNA expression of cell pools D1-D3, showing mean and SD for technical replicates. **f**. EPO mRNA expression of clones R1-R6, showing mean and SD for technical replicates.



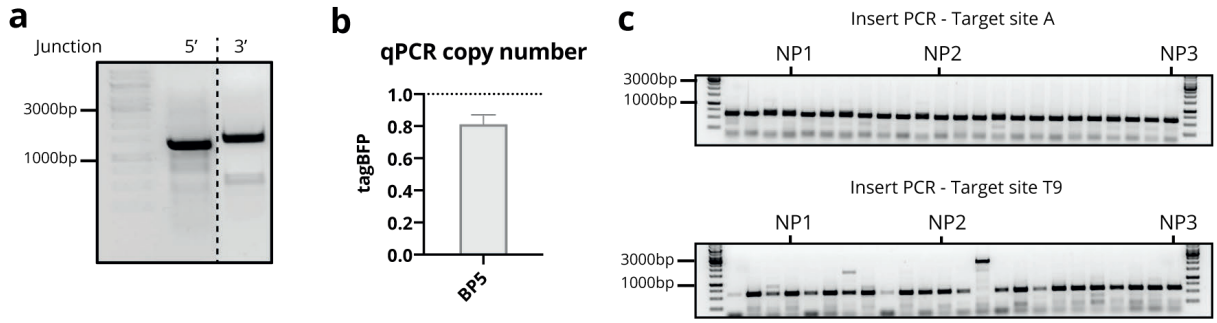
Supporting Figure 2. Multi-copy plasmids verification (EPO4, EPO6). **a**. *in vitro* plasmid digestion. **b**. *in silico* plasmid digestion. **c**. PL0686_EPO4 plasmid map with restriction sites. **d**. PL0687_EPO6 plasmid map with restriction sites. Plasmid PL0686_EPO4 was digested with AanI, BamHI, BcuI restriction enzymes, PL0687_EPO6 was digested with BamHI, BcuI, XmaJI, ScaI.



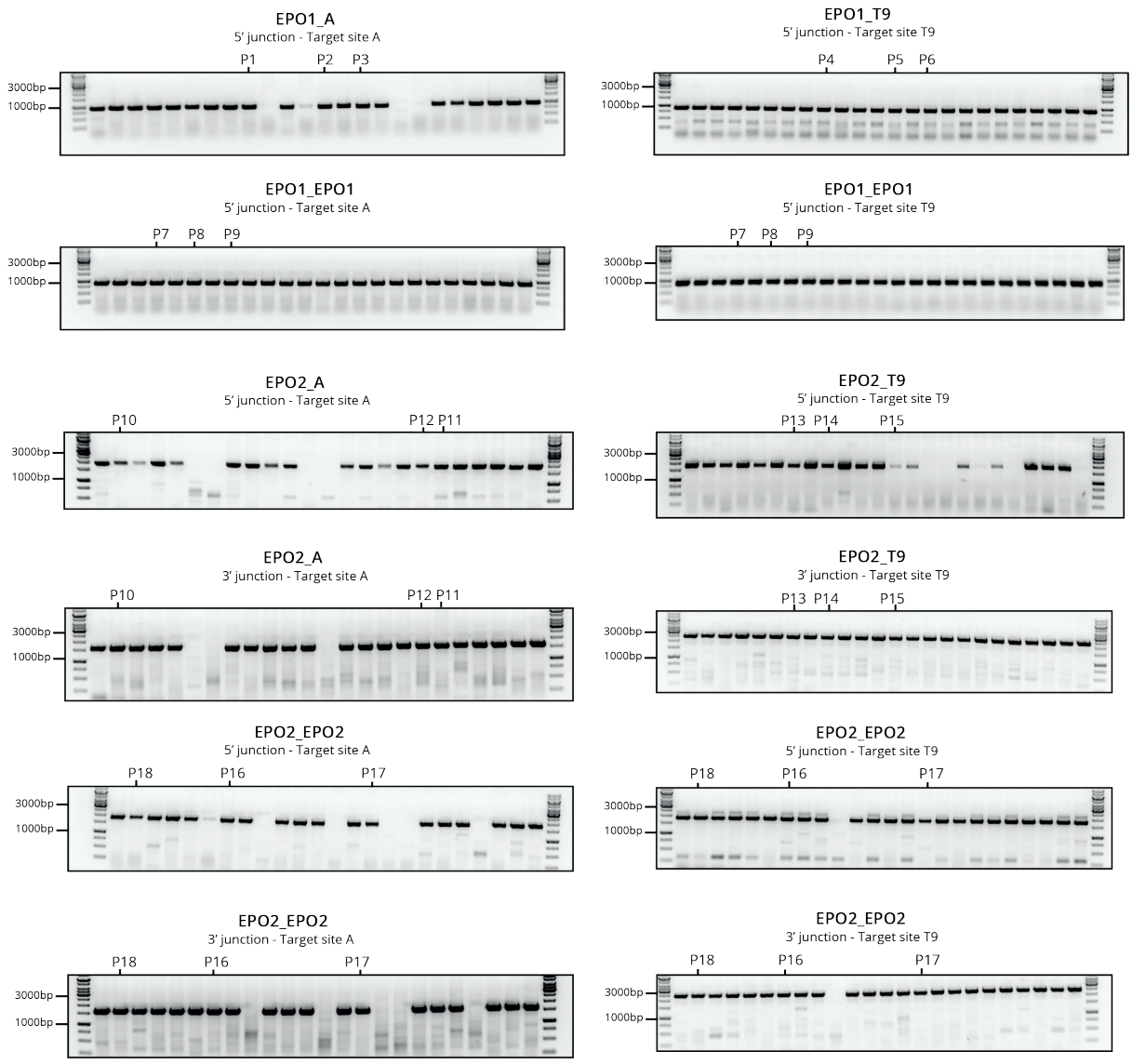
Supporting Figure 3. Verification of EPO1-EPO6 and ETN1-ETN4 clones. **a**. Schematic representation of insert and 5′ and 3′ junction PCR setup used for verification of multi-copy clones. **b**. Insert PCR of EPO1 clones (primer set VIII). **c**. 5′ junction PCR of EPO2 clones (primer set II). **d**. 3′ junction PCR of EPO2 clones (primer set III). **e**. 5′ and 3′ junction PCR of EPO4 clones and AD1 MCL (primer sets IV and V). **f**. 5′ and 3′ junction PCR of EPO6 clones and AD1 MCL (primer sets IV and VI). **g**. 5′ junction PCR of ETN1 clones (primer set VII). **h**. 5′ and 3′ junction PCR of ETN2 clones (primer sets II and III). **i**. 5′ and 3′ junction PCR of ETN4 clones (primer sets IV and V). **j**. Insert PCR of EPO4, EPO6 and ETN4 clones (primer set VIII). Expected amplicon sizes for EPO4 - 7.4 kb, EPO6 - 10.9 kb, ETN4 - 10.9 kb.



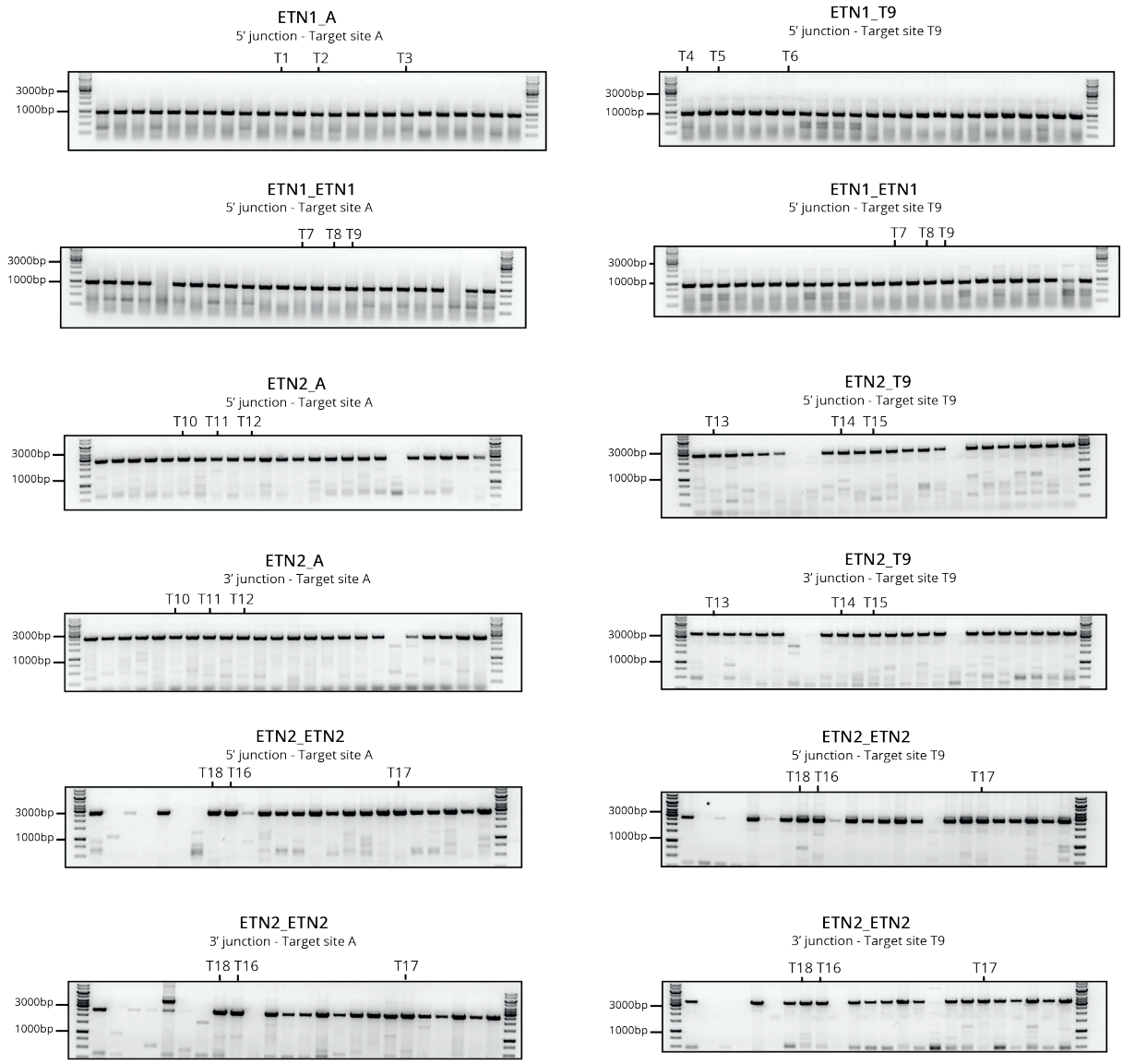
Supporting Figure 4. Verification of MP3 and LC1 master cell lines and clones AD1_EPO1(R1), MP3_EPO1 and LC1_EPO1. **a**. 5′ and 3′ junction PCR of MP3 and LC1 MCLs (primer sets IX, X, XI). **b**. Copy number analysis of MP3 and LC1 MCLs. **c**. Copy number analysis of AD1_EPO1(R1), MP3_EPO1 and LC1_EPO1 clones.



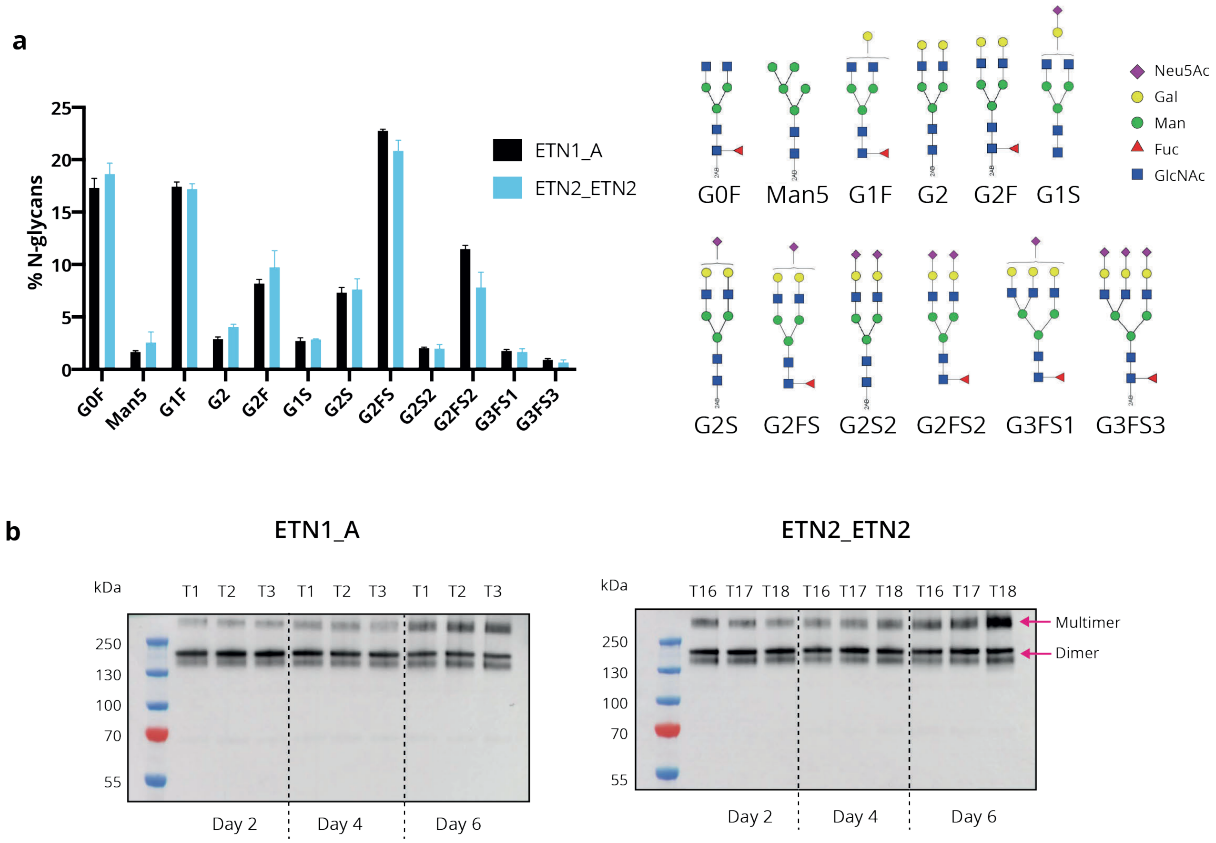
Supporting Figure 5. Verification of dual-RMCE BP5 master cell line and non-producer (NP) clones. **a**. 5′ and 3′ junction PCR of dual-RMCE MCL BP5 (primer sets XII, XIII). **b**. Copy number analysis of BP5. **c**. Verification of NP clones by insert PCR (primer sets XX, XXI).



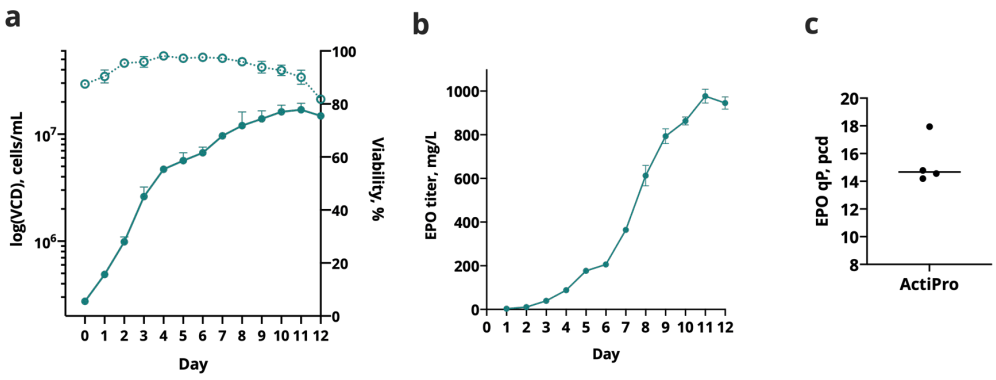
Supporting Figure 6. Verification of dual-RMCE EPO clones P1-P18 by 5' and 3' junction PCR. Primer sets are listed in Supporting Table 5.



Supporting Figure 7. Verification of dual-RMCE ETN clones T1-T18 by 5' and 3' junction PCR. Primer sets are listed in Supporting Table 5.



Supporting Figure 8. Analysis of ETN protein quality by N-glycan analysis and Western blot. **a**. N-glycan profile of ETN samples from batch cultivation of ETN1_A and ETN2_ETN2 clones, showing mean with SD for three biological replicates. **b**. Western blot of batch culture supernatants from day 2, day 4 and day 6 for ETN1_A and ETN2_ETN2 clones.



Supporting Figure 9. Fed-batch cultivation of two EPO2_EPO2 clones in ActiPro media, 2 technical replicates per clone. **a**. Viable cell densities (VCD) and viability, showing mean with SD for biological replicates. **b.** EPO titer, showing mean with SD for biological replicates. **c**. EPO specific productivity, showing median value for two clones with two technical replicates.

II. Supporting Tables

Supporting Table 1. RMCE integration efficiency for cell designs EPO1-EPO6 and ETN1-ETN6. "mCherry-negative" is a percent of mCherry-negative cells after RMCE transfection analyzed by FACS. "PCR positive clones" is a percent of mCherry-negative clones that were verified to have targeted integration by insert or 5′ and 3′ junction PCR.

Cell Design	Insert size (kb)	mCherry-negative, %	PCR positive clones	Cell Design	Insert size (kb)	mCherry-negative, %	PCR positive clones
EPO1	0.97	1.8	92% (44/48)	ETN1	1.86	2.59	96% (46/48)
EPO2	2.71	0.18	83% (40/48)	ETN2	4.50	0.35	58% (28/48)
EPO4	6.25	0.11	31% (15/48)	ETN4	9.81	0.06	29% (14/48)
EPO6	9.76	0.06	19% (9/48)	ETN6	15.1	0	0

Supporting Table 2. RMCE integration efficiency for dual-RMCE EPO and ETN clones. "RMCE" is a percent of mCherry-negative/tagBFP-positive, mCherry-positive/tagBFP-negative or mCherry-negative/tagBFP-negative cells after RMCE transfection analyzed by FACS. "PCR positive clones" is a percent of clones that were verified to have recombinase-mediated integration by PCR analysis. For EPO1_EPO1, ETN1_ETN1, EPO2_EPO2, and ETN2_ETN2 the percent of PCR positive clones represent positive clones with integration in both target sites (site A, site T9).

Cell Design	RMCE, %	PCR positive clones	Cell Design	RMCE, %	PCR positive clones
EPO1_A	3.1	90% (43/48)	ETN1_A	3.7	96% (46/48)
EPO1_T9	2.2	98% (47/48)	ETN1_T9	2.4	100% (48/48)
EPO1_EPO1	0.57	100% (48/48, 48/48)	ETN1_ETN1	0.63	92% (45/48; 47/48)
EPO2_A	0.59	83% (40/48)	ETN2_A	0.89	88% (42/48)
EPO2_T9	0.6	65% (31/48)	ETN2_T9	0.52	88% (42/48)
EPO2_EPO2	0.01	69% (35/48; 44/48)	ETN2_ETN2	0.01	83% (40/48; 40/48)

Supporting Table 3. Plasmids used and created in this study

Landing pad plasmids

Plasmid ID	Plasmid name	Features (5'→3')	Plasmid info
PL0627	pJ204-loxP-EF1a-mCherry-NeoR-lox2272_BGHpA-siteA-HDR	5'HA(siteA)_loxP_EF1a_mCherry_BGHpA_SV40_NeoR_lox2272_BGHpA_3'HA(siteA)_CMV_ZsGreen1-DR_BGHpA	Created in this study, Addgene ID 154826
PL0586	pJ204-loxP-EF1a-mCherry-NeoR-lox2272-SV40pA-siteA-HDR	5'HA(siteA)_loxP_EF1a_mCherry_BGHpA_SV40_NeoR_lox2272_SV40pA_3'HA(siteA)_CMV_ZsGreen1-DR_BGHpA	Created in this study
PL0626	pJ204-loxP-EF1a-tagBFP-HygR-lox2272-BGHpA-siteT9-HDR	5'HA(siteT9)_loxP_EF1a_tagBFP_BGHpA_SV40_HygR_lox2272_BGHpA_3'HA(siteT9)_CMV_ZsGreen1-DR_BGHpA	Created in this study, Addgene ID 154827

RMCE donor plasmids

Plasmid ID	Plasmid name	Features (5'→3')	Plasmid info
PL0516	D1	pJ204_loxP_CP_k1_**coEPO**_lox2272	Described in Pristovšek et al.[1]
PL0588	D2 (pJ204-loxP-100RPU.2-coEPO-lox2272)	pJ204_loxP_100RPU.2_k1_**coEPO**_lox2272	Created in this study, Addgene ID 154828
PL0589	D3	pJ204_loxP_100RPU.2_k2_**coEPO**_lox2272	Created in this study
PL0588	R1	Same as D2	
PL0633	R2	pJ204_loxP_100RPU.2_k1_**coEPO**_WPRE_lox2272	Created in this study
PL0634	R3	pJ204_loxP_100RPU.2_k1_**coEPO**_BGHpA_100RPU.2_k1_**coEPO**_lox2272	Created in this study
PL0635	R4 (pJ204-loxP-100RPU.2-coEPO-BGHpA-cHS4-100RPU.2-coEPO-lox2272)	pJ204_loxP_100RPU.2_k1_**coEPO**_BGHpA_HS4_LO1ex_HS4_100RPU.2_k1_**coEPO**_lox2272	Created in this study, Addgene ID 154829

PL0636	R5	pJ204_loxP_100RPU.2_k1_**coEPO**-WPRE_BGHpA_100RPU.2_k1_**coEPO**_WPRE_lox2272	Created in this study
PL0637	R6	pJ204_loxP_100RPU.2_k1_**coEPO**-WPRE_BGHpA_HS4_LO1ex_HS4_100RPU.2_k1_**coEPO**_WPRE_lox2272	Created in this study
PL0588	EPO1	Same as R1	
PL0635	EPO2	Same as R4	
PL0686	EPO4	pJ204_loxP_100RPU.2_k1_**coEPO**_BGHpA_HS4_LO2ex_HS4_100RPU.2_k1_**coEPO**_BGHpA_HS4_LO3ex_HS4_100RPU.2_k1_**coEPO**_BGHpA_HS4_LO4ex_HS4_100RPU.2_k1_**coEPO**_lox2272	Created in this study
PL0687	EPO6	pJ204_loxP_100RPU.2_k1_**coEPO**_BGHpA_HS4_LO2ex_HS4_100RPU.2_k1_**coEPO**_BGHpA_HS4_LO3ex_HS4_100RPU.2_k1_**coEPO**_BGHpA_HS4_LO4ex_HS4_100RPU.2_k1_**coEPO**_BGHpA_HS4_LO5ex_HS4_100RPU.2_k1_**coEPO**_BGHpA_HS4_LO6ex_HS4_100RPU.2_k1_**coEPO**_lox2272	Created in this study
PL0665	ETN1 (pJ204-loxP-100RPU.2-ETN-lox2272)	pJ204_loxP_100RPU.2_k1_**ETN**_lox2272	Created in this study, Addgene ID 154833
PL0688	ETN2 (pJ204-loxP-100RPU.2-ETN-BGHpA-cHS4-100RPU.2-ETN-lox2272)	pJ204_loxP_100RPU.2_k1_**ETN**_BGHpA_HS4_LO1ex_HS4_100RPU.2_k1_**ETN**_lox2272	Created in this study, Addgene ID 154834
PL0689	ETN4	pJ204_loxP_100RPU.2_k1_**ETN**_BGHpA_HS4_LO2ex_HS4_100RPU.2_k1_**ETN**_BGHpA_HS4_LO3ex_HS4_100RPU.2_k1_**ETN**_BGHpA_HS4_LO4ex_HS4_100RPU.2_k1_**ETN**_lox2272	Created in this study
PL0690	ETN6	pJ204_loxP_100RPU.2_k1_**ETN**_BGHpA_HS4_LO2ex_HS4_100RPU.2_k1_**ETN**_BGHpA_HS4_LO3ex_HS4_100RPU.2_k1_**ETN**_BGHpA_HS4_LO4ex_HS4_100RPU.2_k1_**ETN**_BGHpA_HS4_LO5ex_HS4_100RPU.2_k1_**ETN**_BGHpA_HS4_LO6ex_HS4_100RPU.2_k1_**ETN**_lox2272	Created in this study
PL0569	3xStop_SPA	pJ204_loxP_taatagtgaataaaatatctttattttcattacatctgtgtgttggttttttgtgtg_lox2272	Created in this study

Other plasmids

Plasmid ID	Plasmid name	Features (5'→3')	Plasmid info
PL0209	Cre recombinase	PSF-CMV-CRE	Sigma-Aldrich, cat. no.:OGS591
PL0120	pJ607-CMV-GFP-2A-Cas9	pJ607_CMV_eGFP_2A_Cas9 (codon-optimized)	Described in Grav et al.[2] Addgene ID 154830
PL0458	pRSFDuet1-sgRNA(siteA)	pRSFDuet1_U6_sgRNA(siteA)	Described in Pristovšek et al.[1] Addgene ID 154831
PL0184	pRSFDuet1-sgRNA(siteT9)	pRSFDuet1_U6_sgRNA_(siteT9)	Described in Pristovšek et al.[1] Addgene ID 154832

Supporting Table 4. List of extended USER linkers with restriction sites (restriction sites in bold, USER linkers in lower case).

USER linkers

Linker name	Linker-specific restriction sites	Sequence (5'→3')
L1	ClaI, EcoRV	**ATCGAT**acgtcgct**GATATC**
L2	ClaI, BamHI, EcoRV	**ATCGAT**ACGTagtgcgat**GGATCC**CGCT**GATATC**
L3	ClaI, BcuI, EcoRV	**ATCGAT**ACGTaccagcgct**ACTAGT**CGCT**GATATC**
L4	ClaI, BspOI, EcoRV	**ATCGAT**ACGTacacagtct**GCTAGC**CGCT**GATATC**
L5	ClaI, XmaJI, EcoRV	**ATCGAT**ACGTacttgcgt**CCTAGG**CGCT**GATATC**
L6	ClaI, KpnI, EcoRV	**ATCGAT**ACGTattaagct**GGTACC**CGCT**GATATC**

Supporting Table 5. Primers used for verification of targeted integration

Primer set	PCR	Cell designs	Forward primer (5'→3')	Reverse primer (5'→3')
I	Insert	D1, D2, D3, R1, R2	PR1720 (site A, 5'OUT): AGTTGGAGACTCGGCAAGAG	PR0775 (NeoR, 3'OUT): TCGCTTGGTGGTCGAATG
II	5' junction	R3, R4, R5, R6, EPO2, ETN2, EPO2_A, EPO2_EPO2(site A), ETN2_A, ETN2_ETN2(siteA)	PR1720 (site A, 5'OUT): AGTTGGAGACTCGGCAAGAG	PR2537 (LO1ex, 3'IN): GATATCAGCGACGTATCGAT
III	3' junction	R3, R4, R5, R6, EPO2, ETN2, EPO2_A, EPO2_EPO2(site A), ETN2_A, ETN2_ETN2(siteA)	PR2536 (LO1ex, 5'IN): ATCGATACGTCGCTGATATC	PR1289 (NeoR, 3'OUT): GTGCCCAGTCATAGCCGAAT
IV	5' junction	EPO4, EPO6, ETN4	PR1720 (site A, 5'OUT): AGTTGGAGACTCGGCAAGAG	PR2797 (LO2ex, 3'IN): CGGGATCCATCGCACT
V	3' junction	EPO4, ETN4	PR2800 (LO4ex, 3'IN): ACACAGTCTGCTAGC	PR1289 (NeoR, 3'OUT): GTGCCCAGTCATAGCCGAAT
VI	3' junction	EPO6	PR2757 (LO6ex, 3'IN): ATTAAGCTggtaccCGCTGATATC ATTTAAATGGG	PR1289 (NeoR, 3'OUT): GTGCCCAGTCATAGCCGAAT
VII	5' junction	ETN1	PR1720 (site A, 5'OUT): AGTTGGAGACTCGGCAAGAG	PR0143 (ETN, 3'IN): GGTGGGCATGTGTGAGTTTTG
VIII	Insert	EPO1, EPO4, EPO6, ETN4	PR2845 (site A, 5'OUT): AGTGAGGGCAAGATGGAAGAC C	PR2847 (NeoR, 3'OUT): CTTCAGTGACAACGTCGAGCAC
IX	5' junction	MP3, LC1	PR1865 (site A, 5'OUT): GCCGCATGACCTTGTTCAAA	PR1163 (EF1a, 3'IN): AAGGGCCATAACCCGTAAAG
X	3' junction	MP3	PR0254 (NeoR, 3'IN): CTGGACGAAGAGCATCAGGG	PR1866 (site A, 3'OUT): TCTGGCACAAGATGTAATGCTG
XI	3' junction	LC1	PR0255 (SV40pA, 3'IN): GCATTCTAGTTGTGGTTTGTCC A	PR1866 (site A, 3'OUT): TCTGGCACAAGATGTAATGCTG
XII	5' junction	BP5	PR0890 (site T9, 5'OUT): GCATGCACAGAGAGGGACAT	PR1163 (EF1a, 3'IN): AAGGGCCATAACCCGTAAAG

XIII	3’ junction	BP5	PR02300 (HygR, 3’IN): TGTCGGGCGTACACAAATC	PR0891 (site T9, 3’OUT): CCCTCTGCAACTGCTAACCA
XIV	5’ junction	EPO1_A, EPO1_EPO1(site A)	PR1720 (site A, 5’OUT): AGTTGGAGACTCGGCAAGAG	PR1170 (EPO, 3’IN): GTCGGGCACGGTGATATTCT
XV	5’ junction	EPO1_T9, EPO1_EPO1(site T9)	PR1862 (site T9, 5’OUT): TCAGGCTTGCTGGTAAATGC	PR1170 (EPO, 3’IN): GTCGGGCACGGTGATATTCT
XVI	5’ junction	ETN1_A, ETN1_ETN1(site A)	PR1720 (site A, 5’OUT): AGTTGGAGACTCGGCAAGAG	PR1171 (ETN, 3’IN): GAGCTGGGTGTATGTGCTGT
XVII	5’ junction	ETN1_T9, ETN1_ETN1(site T9)	PR1862 (site T9, 5’OUT): TCAGGCTTGCTGGTAAATGC	PR1171 (ETN, 3’IN): GAGCTGGGTGTATGTGCTGT
XVIII	5’ junction	EPO2_T9, EPO2_EPO2(site T9), ETN2_T9, ETN2_ETN2(site T9)	PR1862 (site T9, 5’OUT): TCAGGCTTGCTGGTAAATGC	PR2537 (LO1ex, 3’IN): GATATCAGCGACGTATCGAT
XIX	3’ junction	EPO2_T9, EPO2_EPO2(site T9), ETN2_T9, ETN2_ETN2(site T9)	PR2536 (LO1ex, 5’IN): ATCGATACGTCGCTGATATC	PR0891 (site T9, 3’OUT): CCCTCTGCAACTGCTAACCA
XX	Insert	NP (site A)	PR1720 (site A, 5’OUT): AGTTGGAGACTCGGCAAGAG	PR1167 (BGHpA, 3’OUT): CCTGCTATTGTCTTCCCAATCC
XXI	Insert	NP (site T9)	PR1862 (site T9, 5’OUT): TCAGGCTTGCTGGTAAATGC	PR1167 (BGHpA, 3’OUT): CCTGCTATTGTCTTCCCAATCC

Supporting Table 6. TaqMan assays

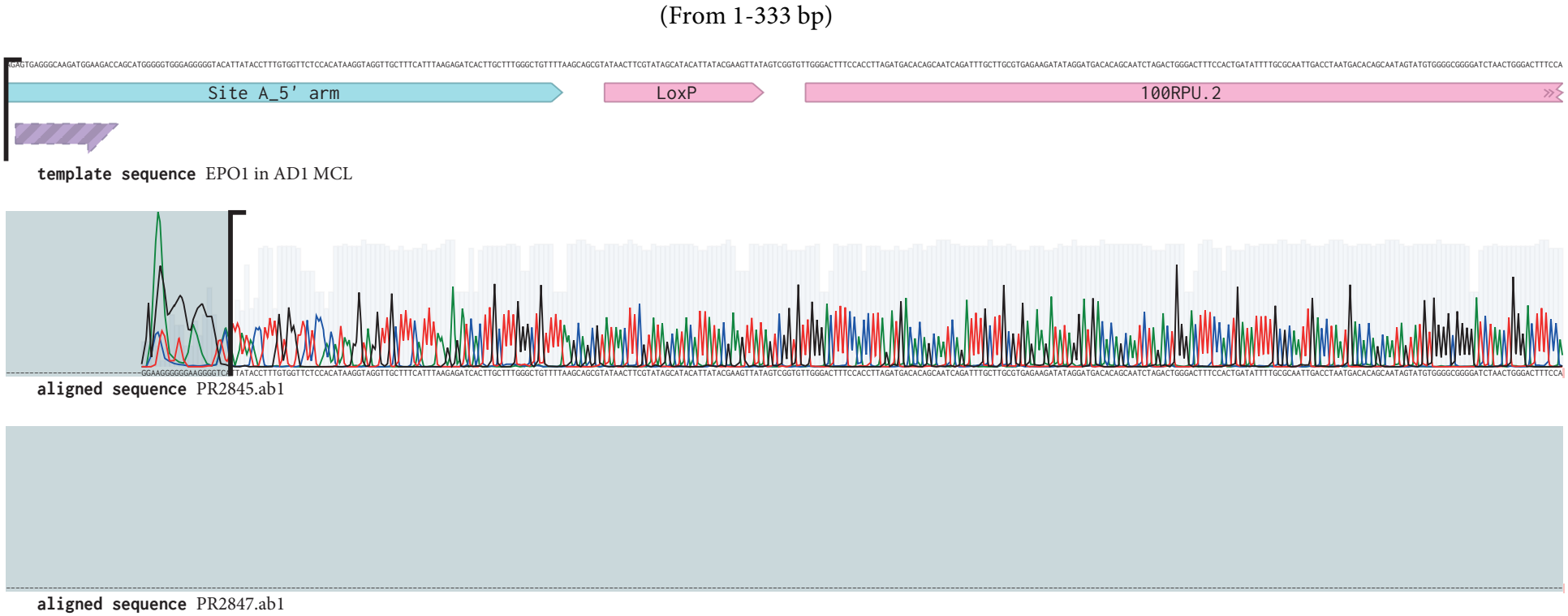
Gene	Forward primer (5'→3')	Reverse primer (5'→3')	Probe (5'→3')	Dye
Cosmc	ACCCGAACCAG GTAGTAGAA	ACATGTCCAAAGG CCCTAAG	AGTGACAGCCATATTGGAACAGCAT CC	VIC-MGB
mCherry	GACTACTTGAA GCTGTCCTTCC	CGCAGCTTCACCT TGTAGAT	TTCAAGTGGGAGCGCGTGATGAA	FAM-MGB
tagBFP	GATCCAAGAAA CCCGCTAAGA	CGACGTAGGTCTC GTTGTTG	TGCCTGGCGTCTACTATGTGGACTA	FAM-MGB
ETN	CAGCCGGAGAA CAACTACAA	CATCACGGAGCAT GAGAAGA	TACAGCAAGCTCACCGTGGACAAG	FAM-MGB
EPO	CTGGAAAGATA CCTGCTGGAAG	AGGCGTAGAAGTT CACTTTGG	CCAAAGAGGCCGAGAACATCACCA	FAM-MGB
Fkbp1a	CTCTCGGGACA GAAACAAGC	GACCTACACTCAT CTGGGCTAC	ATGCTAGGCAAGCAGGAGGTGATC	VIC-MGB
Gnb1	CCATATGTTTCT TTCCCAATGGC	AAGTCGTCGTACC CAGCAAG	ACTGGTTCAGACGATGCTACGTGC	ABY-MGB

Supporting References

(1) Pristovšek, N., Nallapareddy, S., Grav, L. M., Hefzi, H., Lewis, N. E., Rugbjerg, P., Hansen, H. G., Lee, G. M., Andersen, M. R., and Kildegaard, H. F. (2019) Systematic Evaluation of Site-Specific Recombinant Gene Expression for Programmable Mammalian Cell Engineering. *ACS Synth. Biol. 8*, 758–774.
(2) Grav, L. M., Lee, J. S., Gerling, S., Kallehauge, T. B., Hansen, A. H., Kol, S., Lee, G. M., Pedersen, L. E., and Kildegaard, H. F. (2015) One-step generation of triple knockout CHO cell lines using CRISPR/Cas9 and fluorescent enrichment. *Biotechnol. J. 10*, 1446–1456.

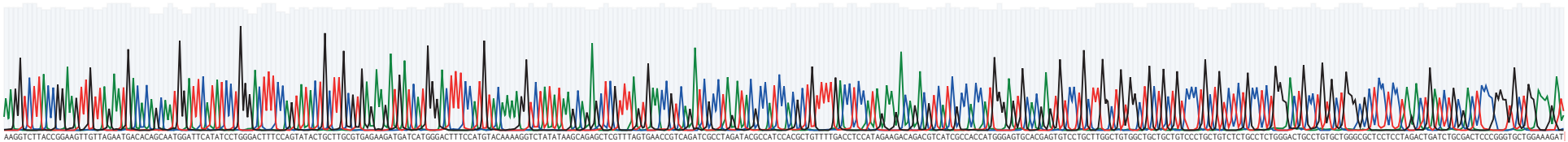
Supporting Sequence Data

Supporting Sequence Data 1. Artifact PCR product sequence alignment. Example for clone EPO6 E8, amplicon was isolated from gel (Supporting Figure 3j) and sequenced using primers PR2845 and PR2847



(From 334-999 bp)

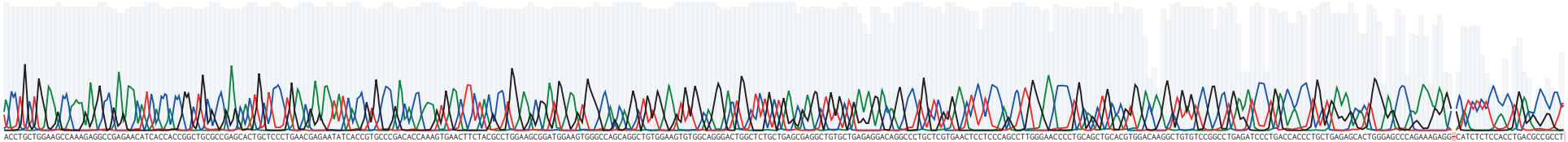
template sequence EPO1 in AD1 MCL



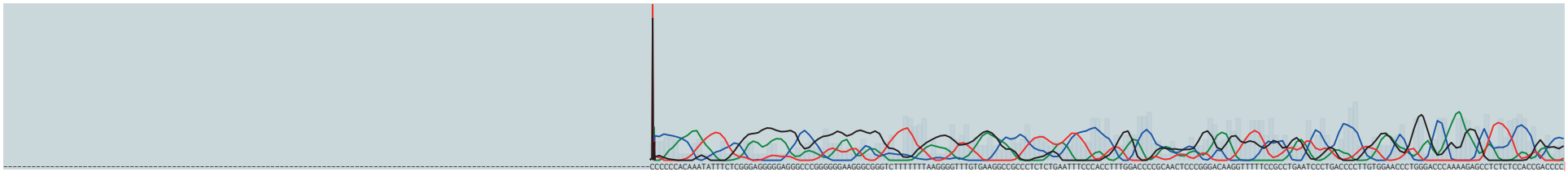
aligned sequence PR2845.ab1

aligned sequence PR2847.ab1

template sequence EPO1 in AD1 MCL

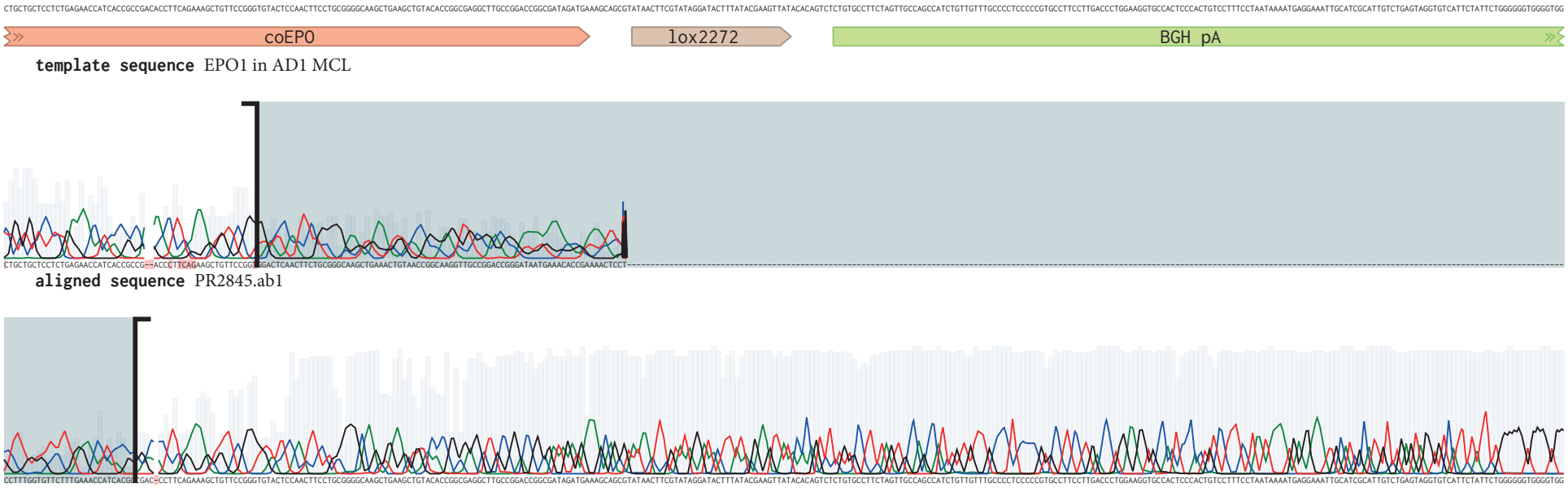


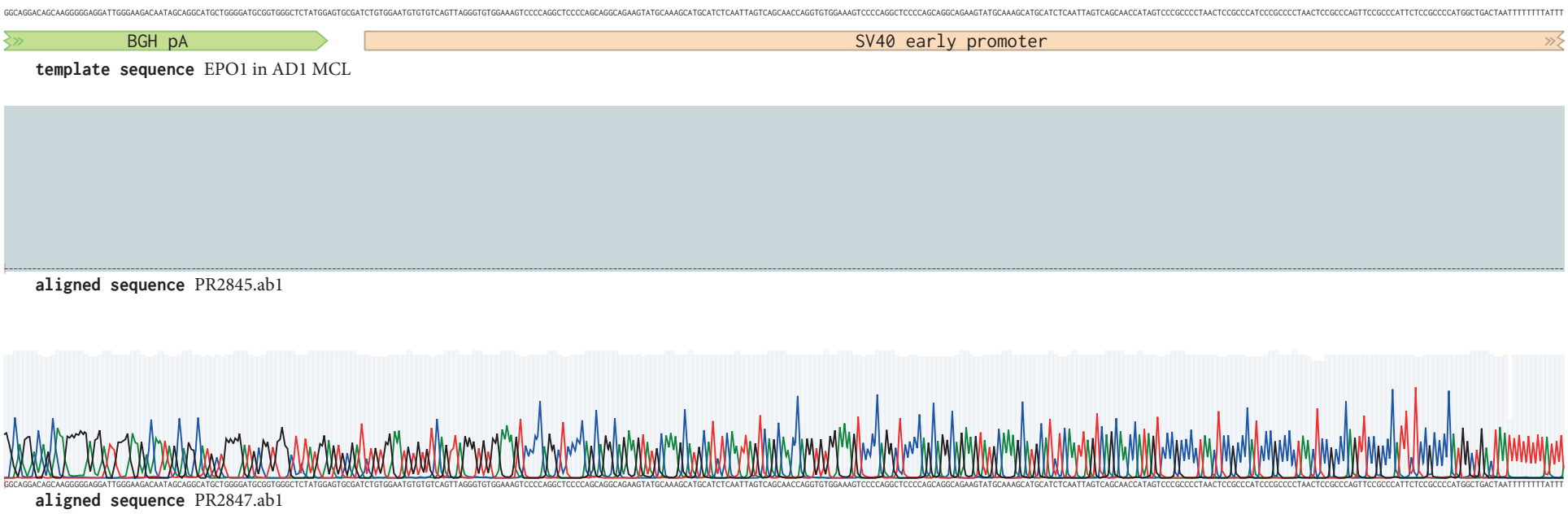
aligned sequence PR2845.ab1



aligned sequence PR2847.ab1

(From 1000-1665 bp)

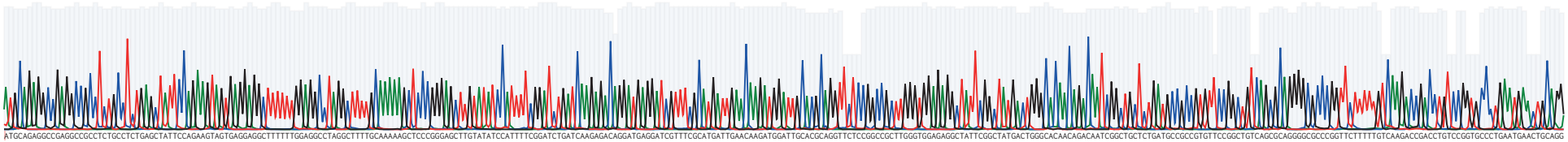




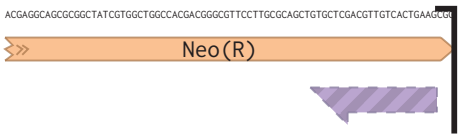
(From 1666-2076 bp)

template sequence EPO1 in AD1 MCL

aligned sequence PR2845.ab1

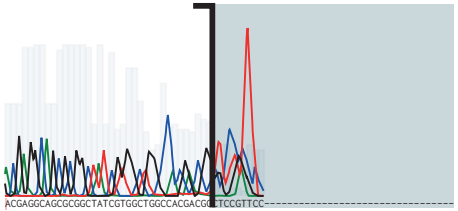


aligned sequence PR2847.ab1



template sequence EPO1 in AD1 MCL

aligned sequence PR2845.ab1



aligned sequence PR2847.ab1)

Supporting information for Chapter 5

Transcriptional response to recombinant protein production in isogenic multi-copy CHO cells

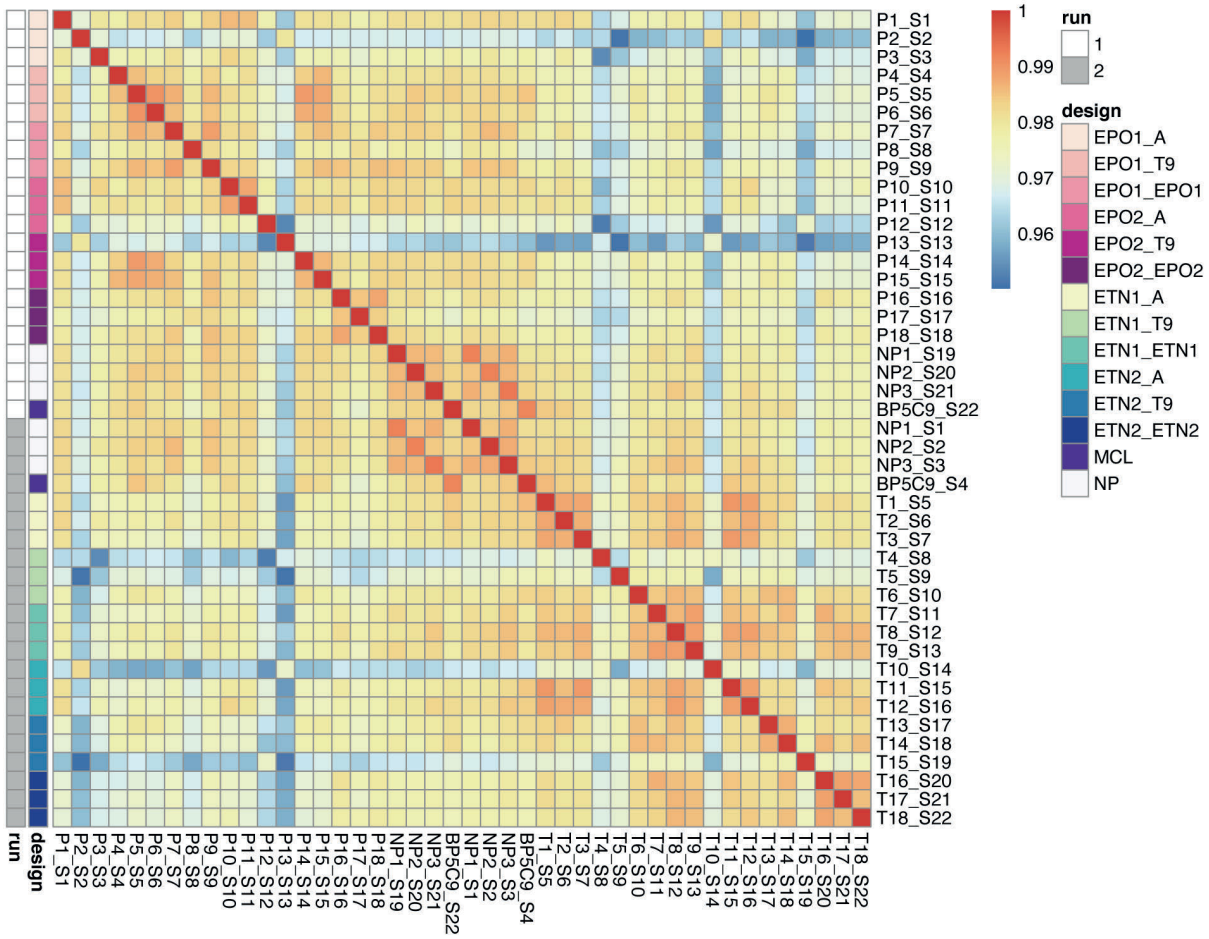
Daria Sergeeva[1], Lise Marie Grav[1], Gyun Min Lee[1,2], Lars Keld Nielsen[1,3]

[1]The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Kgs. Lyngby, Denmark

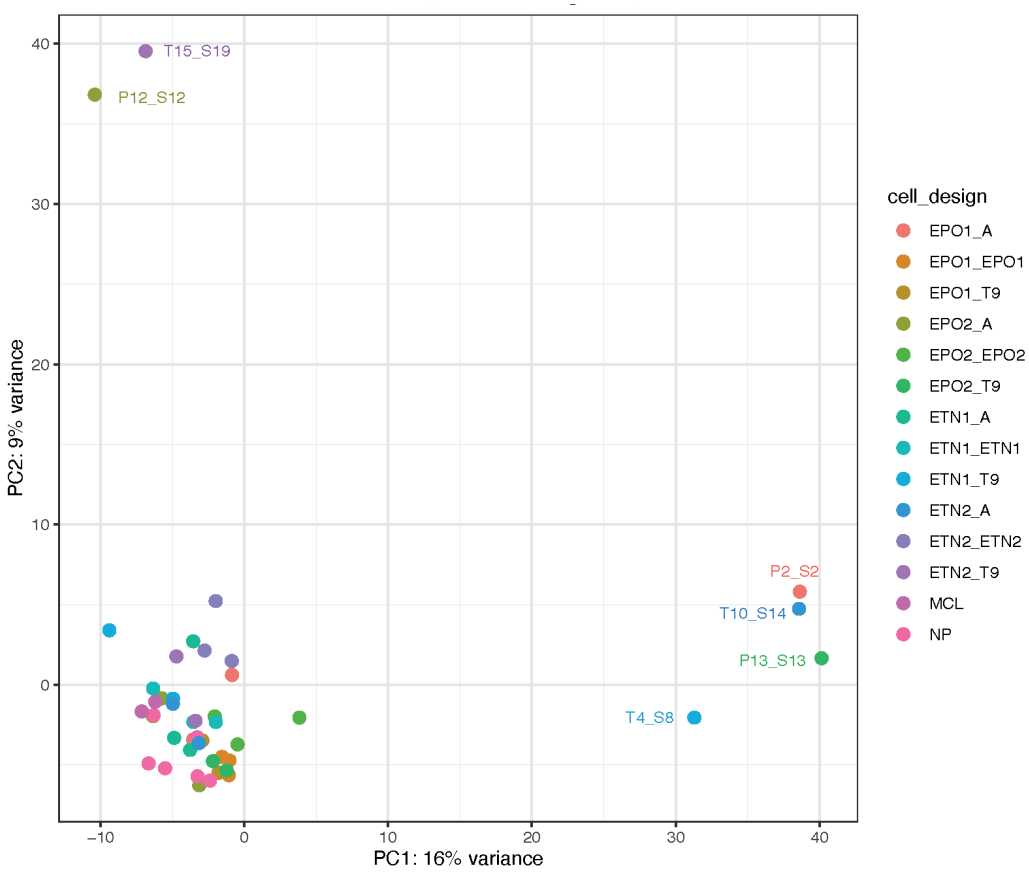
[2]Department of Biological Sciences, KAIST, Daejeon, Republic of Korea

[3]Australian Institute for Bioengineering and Nanotechnology, University of Queensland, Brisbane, Australia

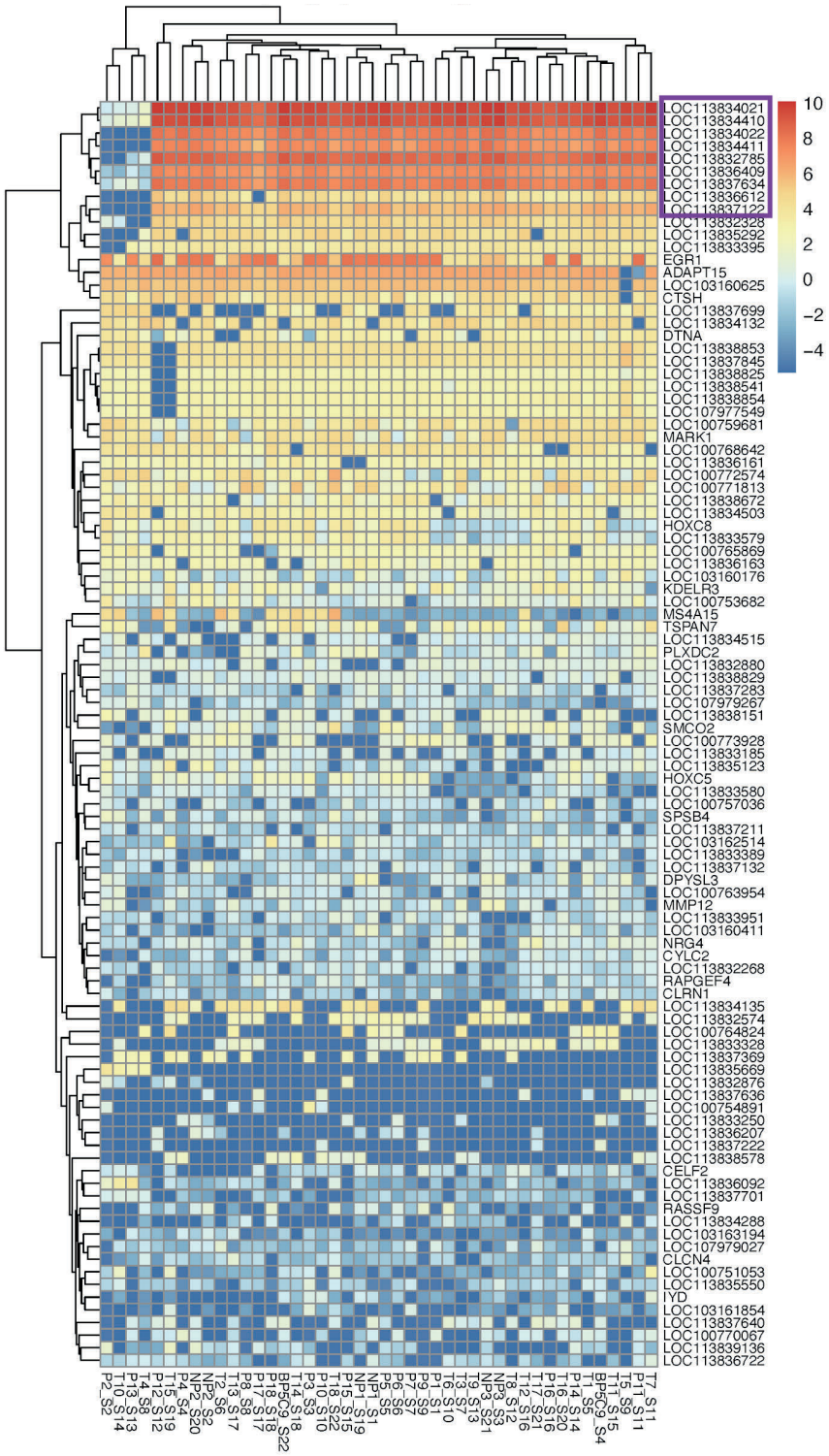
I. Supporting Figures



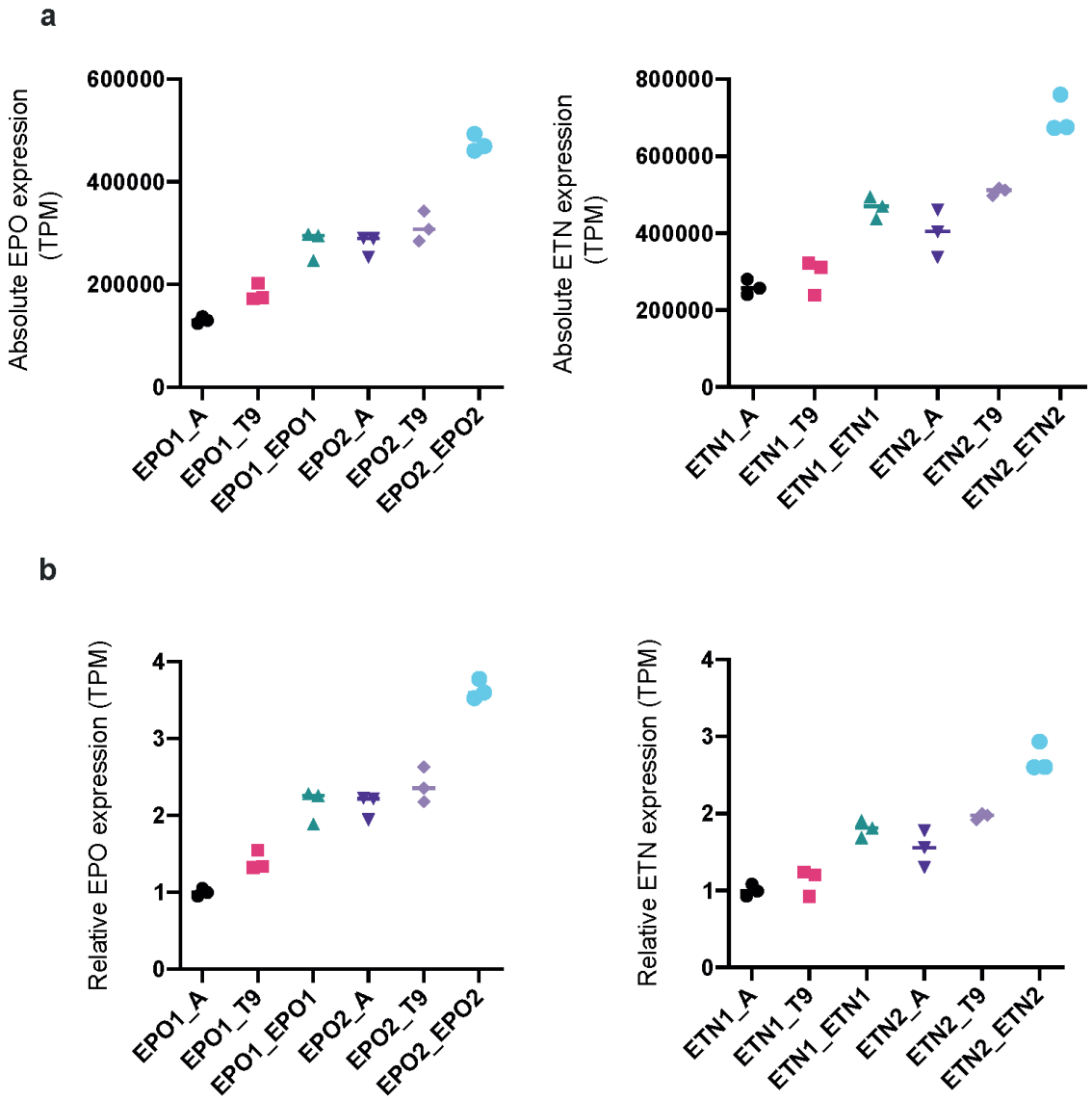
Supporting Figure 1. Pairwise Pearson's r correlation coefficients between all samples (n=12380 genes).



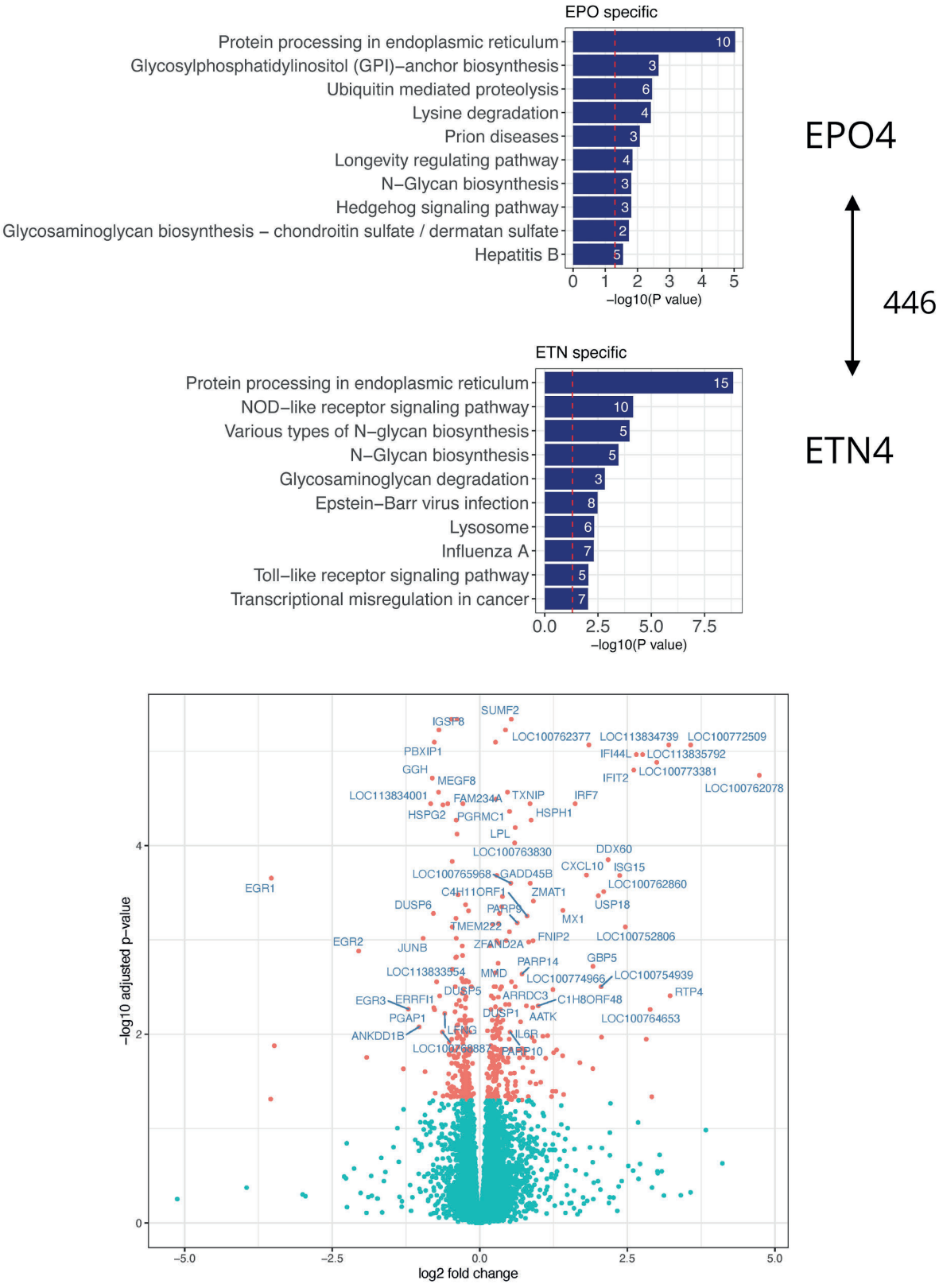
Supporting Figure 2. Principal component analysis (PCA) based on top 500 variable transcripts (Chinese hamster), excluding transgenes.



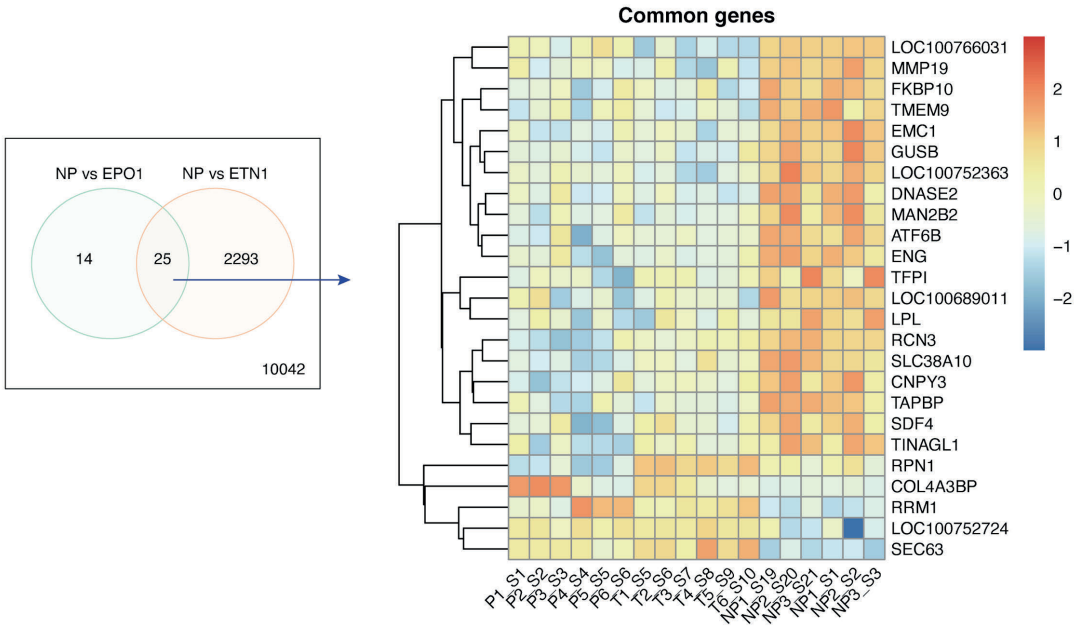
Supporting Figure 3. Top 100 variable transcripts (Chinese hamster), excluding transgenes. Framed genes are described in Supporting Table 1.



Supporting Figure 4. EPO and ETN expression in multi-copy isogenic CHO cells. **a.** Absolute TPM values, showing median value for three biological replicates. **b.** Relative TPM values normalized to EPO1_A or ETN1_A, showing median value for three biological replicates.



Supporting Figure 5. Protein-specific response to recombinant protein production. Top KEGG enriched pathways and volcano plot EPO4 vs ETN4 (differentially expressed genes with p-value <0.05 and no LFC cut-off are depicted in red. Genes with LFC ≥ 0.5, p-value < 0.01 are labeled)



Supporting Figure 6. Common genes between comparisons NP vs EPO1 and NP vs ETN1

II. Supporting Tables

Supporting Table 1. Top variable Chinese hamster genes cross-mapping

Gene symbol	Annotation	Cross-mapping genes	Gene Location
LOC113832785	bromodomain-containing protein 4-like	LOC113834411	>NW_020822409.1 NW_020822409.1:177058..182946 (+ strand) class=gene length=5889
LOC113834410	MLV-related proviral Env polyprotein-like	LOC113834022, LOC113837122	>NW_020822468.1 NW_020822468.1:36485940..36488076 (+ strand) class=gene length=2137
LOC113834411	bromodomain-containing protein 4-like	LOC113832785	>NW_020822468.1 NW_020822468.1:36480329..36482755 (+ strand) class=gene length=2427
LOC113834021	LTR	LOC113836409	>NW_020822458.1 NW_020822458.1:12499334..12508758 (+ strand) class=mRNA length=9425
LOC113834022	MLV-related proviral Env polyprotein-like	LOC113834410, LOC113837122	>NW_020822458.1 NW_020822458.1:12508902..12510881 (+ strand) class=gene length=1980
LOC113836409	LTR	LOC113834021	>NW_020822601.1 NW_020822601.1:41266834..41272634 (- strand) class=gene length=5801
LOC113836612	LTR	LOC113837634	>NW_020822608.1 NW_020822608.1:26337432..26340296 (+ strand) class=gene length=2865
LOC113837122	MLV-related proviral Env polyprotein-like	LOC113834410, LOC113834022	>NW_020822636.1 NW_020822636.1:8034613..8037261 (- strand) class=gene length=2649
LOC113837634	LTR	LOC113836612	>NW_020822686.1 NW_020822686.1:213752..215445 (+ strand) class=gene length=1694

Supporting Table 2. EPO fuzzy clusters core genes (membership score >0.8)

Cluster 1 - linear up	Cluster 2 - linear down	Cluster 3 - non-linear up	Cluster 4 - non-linear down
SPATA2	ASAH1	TEDC2	FIP1L1
HDLBP	LAMP1	NKAP	ZCCHC4
SLC39A14	ARF3	GPN2	CIAO3
SEC61B	RRAS	INTS11	LOC113832678
SLF2	HPSE	ABCF1	LOC113830720
STRIP1	RNASET2	EIF2B1	LOC103159776
SEL1L	C1R	CDKN2AIPNL	LOC113835495
CROT	WDR1	LOC100772005	ZNF606
NEDD4	PIKFYVE	PDSS1	LOC103159102
DDB1	LOC100770246	DDX56	LOC113836598
PGGHG	CEP72	GTF2F1	UBR4
CDR2L	VPS25	METTL2A	RPL23
LOC113831191	CTSO	CHAF1B	ZFC3H1
EHMT1	LOC113834741	LYSMD2	BCL2L11
GTF2H4	SUN2	SLC39A7	ATG12
PGM3	NCSTN	WRB	KDM4C
AMPD2	CAPZB	MRPL20	DPF2
CACTIN	LOC100758361	AP2A2	ZFAND5
CARMIL1	NAGA	ELP5	CDK19
TUBA1A	ACADL	SCFD2	CCNDBP1
MBNL3	TGFBR3	KHSRP	NARF
SON	DARS	LOC113834671	LOC113834936
LOC113837284	APP	DCPS	LOC100771943
BRPF1	C2H1ORF159	POLD3	USP8
PAF1	ZSWIM3	TTC27	LOC113832264
CIRBP	CD164	YWHAH	SMIM14
LRRC8E	PLIN2	DROSHA	CD68
SCARA3	B2M	MTAP	FKBP1B
TRABD	MRPL39	SPTLC2	DZIP3
CTCF	CR1L	FANCI	CLUAP1
RAB8A	GLB1	TK1	KMT2E
CTNNBIP1	PLD3	SF3A2	LOC100753707
SLC35B1	FAM3A	P2RY2	LOC103159266
ASPSCR1	SEL1L3	PRKAG1	MPC2
CENPB	SALL3	ELOA	ZBTB26
LOC100752430	SORT1	LOC103159745	LOC113833954
INTS1	RCBTB2	EIF5A	LOC113833575
COPE	TXNDC12	PIDD1	LOC113835331
CCDC85B	KLF11	SLC41A1	FOSL2
SNX4	C1RL	FUS	RLF
SSBP3	TCTN1	PRMT5	LOC103163417
PDE4A	LOC100768006	GJB3	CLCN6
ABCA9	AIG1	STMN1	PIK3CG

OSTC	CEP83	EXO1	SRP72
CALM1	ATRAID	NFIC	LOC113830879
ATP2C1	PEX10	DCK	POLR3E
TP53RK	LOC107977391	NT5DC2	LOC100759458
MYO18A	LOC100763497	RAN	ABHD4
HSP90B1	OGT	TOMM5	VAMP4
SRPRA	LOC113835544	POLD1	LOC103160752
DCAF1	TMEM62	RBM10	TMEM87A
LOC113834516	ERMP1	NIP7	ARHGAP25
ZNF512B	LOC100756663	TBL3	LOC103160179
ALG12	LOC113833600	UBL4A	STYXL1
SFXN2	EMC7	TMEM201	LOC100768184
PTEN	ARRDC3	AKT1	NUB1
ARFIP1	RABEPK	VASH1	PNRC1
PRR22	LOC100773835	SAE1	NCOR1
LTA4H	UNC5A	BRF1	RNF169
CUNH12ORF43	SMPD1	EIF4A3	PDE10A
ENTPD7	CAT	DPH2	NDUFB4
LOC103158569	ADPGK	TIMELESS	CLCN3
HERPUD1	S100A5	MAP3K6	RPS15A
FAM208B	EXOC1	RECQL4	ATXN3
UTP20	GLG1	G3BP1	LOC103160550
	GUSB	TRAIP	ACBD4
	MFGE8	RRP1	LOC107979012
	ITFG1	IFRD2	LOC100760401
	SCARB1	MLLT1	LOC103160687
	SSH2	CLUH	ZC3H6
	HMGCS1	PPIL1	LOC103162488
	TUFT1	ANAPC5	TCEANC
	LOC100756951	POLR1C	LOC113833606
	LOC100756505	LOC113837470	LOC100762293
	SLC25A20	CHST12	LOC100756292
	LOC100773545	CHEK2	CHMP2B
	GANAB	RFC2	UNC5B
	LOC100752363	PSMG3	PDE7A
	TMBIM6	SNRPA	ABHD18
	NDUFA4	ING5	LOC107979868
	FKBP5	MCM2	LOC100757671
	FYB1	SMG5	LOC103161306
	USP4	CUNH3ORF38	CEP120
	TCTN3	CDK2AP1	LOC107978541
	ST13	ZNF283	ITSN2
	TMEM9	RFC5	ATP6V1H
	PDIA5	ANKLE1	TMEM134
	ZWILCH	CASTOR1	NDUFS4
	GALNS	C6H19ORF48	LOC100755586
	SLC41A3	ROR1	ZYG11B

		CHAF1A	POR
		GINS1	LOC113835897
		TXNDC5	POLR3GL
		EIF4H	PGAP2
		LPIN3	ATF4
		GATAD2A	NUDT12
		PCLAF	LOC103163979
			GIPC1
			BNIP3L
			ACADSB
			ARMCX3
			LOC113833492
			FAM193B
			LOC107979026
			CALCOCO2
			AP3M1
			LOC100772630
			LOC103164274
			PLTP
			IDH3B
			TM9SF4
			TRIT1
			LETM2
			ST3GAL3
			CCNH
			AMN1
			DMTF1
			LOC103163700
			LOC100765709
			RPL5
			MIB2
			NFE2L1
			ZC3HAV1
			LOC103161105
			NDUFA10
			MOCS2
			LOC107978216
			EDRF1
			GZF1
			LOC100753687
			LOC103164243
			LOC113833165
			FAM135A
			CCDC90B
			UBE2H

Supporting Table 3. ETN fuzzy clusters core genes (membership score >0.8)

Cluster 1 - linear up	Cluster 2 - linear down	Cluster 3 - non-linear up	Cluster 4 - non-linear down
MCF2L	GSDME	TTF2	TXNDC12
LOC113833691	LOC100756022	STAG1	TEX264
INTS3	AKAP9	EIF4H	TNFSF9
LOC113832274	METTL23	TRIP13	GALNS
ARID5A	QSOX2	LOC100771643	ACSS2
LOC100762252	SURF2	CEP162	ZFR2
SLC39A14	RALGAPA1	MECR	P3H3
UBL4A	ZC3H6	CCDC86	IGSF8
PPIP5K1	RLF	INKA2	CIAO2B
LOC103161085	TMEM143	SMG7	SMIM14
SPCS2	CWC27	EMC3	LOC107978793
LOC100754853	TSGA10	COIL	HEATR5B
TMEM50B	UROS	EIF6	P3H1
SHC1	LOC107979212	ICMT	GSS
CMTM3	FUCA1	C1H3ORF67	UBE2W
CTNNBIP1	FLOT2	SAE1	UBC
SELENOS	INSIG1	RPS2	CCDC40
NNMT	LOC113834421	MCM7	BST2
GCC2	LOC103161022	RAN	LOC113836427
MLEC	ASIP	ATP1B3	ARID5B
CSRP1	USP6NL	PDSS1	LOC103159523
RNF34	KIF1BP	TMEM209	PBXIP1
OTUD4	POLR3GL	AP2M1	DKK3
DDX24	KCMF1	AKAP11	ITGAV
NFIA	LOC113833964	CTCF	USP4
FNDC3B	VPS36	RNASEH2A	LOC100764834
LOC100759692	PDE9A	SENP1	MESD
UACA	LOC113835934	NT5DC2	CTSO
DDX56	GRHPR	POLE	LCORL
PPM1B	TENT5D	TRA2B	PDCD7
LOC113837384	LPIN1	GOT2	ZC3H3
ADAR	DAP3	CNOT9	CRIPT
TRAF3IP1	ATXN7L1	POLR2D	SLC35B3
GNAI2	LOC113833361	LOC100773089	NDUFC1
OGDH	MIB2	GPS1	UGGT1
MANF	LOC100773172	FAM133B	OGFOD1
TMEM185B	APMAP	RPS27	GPR132
DNAJB11	LOC100773619	LTN1	ICAM5
FKBP2	TRNAU1AP	DDX23	GGT7
CNDP2	PRDM5	DALRD3	FAM234A
PDK3	RFX5	SENP3	MIEN1
SGMS2	LSS	LOC100763852	NIPA1
NCBP2	SPATA1	DIS3	MAFG

DNAJC15	VPS8	HSP90AA1	EVC2
MAGT1	LOC103159488	RAD9A	ATP6V0B
LOC103158614	MAMDC2	CSNK1A1	RILPL1
MAZ	GIPC1	SMG5	FCGRT
VCP	FOS	KNOP1	ZC3HAV1
CUNH12ORF43	LOC113833160	PSMB3	ZBTB20
E2F3	SEMA3A	SRSF7	P3H2
LOC100770709	LOC100758361	TAS1R1	FAM177A1
GMPPB	ARFIP2	CD2BP2	LOC100773966
SMAD3	B4GALT3	SUZ12	VSTM4
SSB	C1R	MCM6	LOC100768787
USP1	LOC113831708	PRMT5	TGM2
PIK3AP1	RPL5	NIP7	FXR2
ECHS1	ACO2	ARHGAP1	CHMP5
RHNO1	SDHC	MAP3K7	IRX5
B4GALT1	SETD3	PCYT1A	TMEM9
PITPNA	OGT	NAP1L4	STXBP1
HAUS7	ARSG	AEN	SYNE1
NOB1	SDHA	TJP2	SENP7
PRKACA	MPC2	SUPT5H	LOC103163630
PDIA6	SSU72	ST7	GLG1
ATF5	TGIF1	SNX4	LPP
HMGA2	DARS	MDM1	LOC100768249
ZNF629	CNOT6L	TEDC2	C3H16ORF87
KTI12	MAP3K13	PACS2	SBSN
SYVN1	CTNNAL1	NUP107	AMT
GNL1	ACP1	TSN	IGBP1
LRIF1	LOC113832678	CASC4	UXS1
LOC100758641	ZNF346	THADA	HSPG2
NUS1	ATRAID	NOM1	DNASE2
RBBP7	FIP1L1	TCOF1	KPNA6
CMTM4	MINDY2	HMGB1	LOC100766031
CRK	SNW1	THOP1	CUX1
FEM1A	MICU2	G3BP1	MRPL34
PRIM2	LOC107977406	LOC100761779	UNC5B
POLK	GAK	PELP1	NICN1
PCBP1	PLCXD3	HNRNPD	FEZ2
FAM208B	MAP4K4	SRRT	LMAN1
HSPA5	LOC100769106	POLR2B	ARSA
UNC5C	LOC100762293	CUNH7ORF50	GLA
LOC100765042	MFSD14A	FBXL3	DNAJC17
C2H18ORF21	LRPAP1	TOMM22	JUNB
LOC100773147	MYO6	KPNB1	LAMC1
CHST12	NCOR1	HNRNPDL	SLC22A17
ZNF3	LOC113834334	NUP214	LOC103159711
LOC113836166	PDK1	ASB13	MICU3
PDIA4	DBP	LOC100766272	ABHD4

LOC100770769	MORN4	NUDC	COG5
CARNS1	B3GALNT2	SLC35E4	MYO5A
HMOX2	ERMP1	SHMT1	
SEC11C	LOC107977431	NUP54	
NUDT2	ARHGAP25	ASF1B	
ARSJ	C3H11ORF58	PRIM1	
XBP1	CUNH2ORF68	LEMD3	
MRPL40	HECA	MGAT1	
PSMA3	SIN3B	CORO1C	
	PRRX1	RBMX	
	LOC113833304	TMEM109	
	RNF41	HNRNPH1	
	TFRC	PPP1R8	
	NDUFB5	EXO1	
	ELP1	IKBIP	
	FERMT2	LOC100759640	
	LOC113835613	ARHGEF37	
	ZMAT5	PRPF19	
	CNTRL	RPS20	
	PI4KA	TOMM40	
	NAT9	PAQR4	
	LOC113833575	ATP5F1B	
	ZC3H14	DTL	
	RNF169	AHCY	
	LOC100768142	INTU	
	PDE7A	PHB2	
	SENP6	UHRF1	
	LOC113832892	LOC100759021	
	LOC113837827	CC2D1B	
	LOC113835899	AUP1	
	PTPN23	FDXR	
	LOC113836011	TTC28	
	VAC14	DSCC1	
	ARRB1	PPIP5K2	
	STXBP5	MCM4	
	EXOC4	TIGD3	
	ZBED6	UNG	
	NEK1	NUP160	
	USP3	HNRNPM	
	SIAE	WRAP53	
	NNT	TRAPPC10	
	BNIP3	LOC100752724	
	IQCC	UBE2J2	
	MRPL33	CDCA3	
	RAD23A	TRMO	
	MRVI1	POLR2E	
	ARL4A	HEATR1	

	USP8	NASP	
	LOC100770382	PRMT1	
	CLYBL	SF3A3	
	CRTC3	MNS1	
	DZIP3	ROR1	
	MPLKIP	XPO1	
	MSMO1	ZGRF1	
	ZCCHC4	ENTPD4	
	WASHC3	MYBL2	
	SSR2	DGKZ	
	TMEM94	MIOS	
	MON2	MAD2L1	
	CUNH15ORF61	PDS5A	
	LOC107978151	SGTA	
	NDUFS1	RANBP6	
	SPG7	LOC100754792	
	TNKS	LOC103158544	
	PEX19	TMPPE	
	CSNK1G2	MBOAT1	
	GNA15	SIVA1	
	MKRN3	AAAS	
		STT3A	
		MTA2	
		USP7	

Supporting Table 4. Common genes in fuzzy clusters (no membership score cut-off)

Cluster 1 - linear up	Cluster 2 - linear down	Cluster 3 - non-linear up	Cluster 4 - non-linear down
SLC39A14	ASAH1	TEDC2	CIAO3
SEC61B	C1R	GPN2	LOC103159102
SLF2	WDR1	INTS11	ZFC3H1
NEDD4	CEP72	ABCF1	CCNDBP1
PGGHG	VPS25	PDSS1	NARF
CDR2L	SUN2	CHAF1B	LOC113832264
LOC113831191	NCSTN	AP2A2	SMIM14
CACTIN	CAPZB	ELP5	LOC100759458
CARMIL1	LOC100758361	KHSRP	ABHD4
TUBA1A	DARS	LOC113834671	LOC103160179
SON	C2H1ORF159	POLD3	LOC103160550
PAF1	PLIN2	FANCI	LOC103160687
LRRC8E	MRPL39	TK1	UNC5B
TRABD	GLB1	PRKAG1	ABHD18
RAB8A	FAM3A	EIF5A	ZYG11B
CTNNBIP1	SEL1L3	PIDD1	CALCOCO2
SLC35B1	SALL3	FUS	PLTP
ASPSCR1	SORT1	PRMT5	TRIT1
CCDC85B	RCBTB2	STMN1	LETM2
SSBP3	LOC100768006	EXO1	AMN1
OSTC	CEP83	NT5DC2	DMTF1
CALM1	ATRAID	RAN	ZC3HAV1
SRPRA	LOC100763497	TOMM5	EDRF1
LOC113834516	OGT	NIP7	RNF114
SFXN2	LOC113835544	TBL3	LOC113831885
PTEN	TMEM62	SAE1	HSPB8
CUNH12ORF43	ERMP1	EIF4A3	LOC100753467
LOC103158569	LOC100756663	RECQL4	AQP1
HERPUD1	EMC7	G3BP1	TOLLIP
FAM208B	UNC5A	RRP1	ZNF329
SAFB2	ADPGK	IFRD2	RHOQ
JADE2	S100A5	MLLT1	APPL1
PDIA4	GUSB	PPIL1	ABCA7
LOC100758550	MFGE8	ANAPC5	LOC103163029
EIF2AK3	ITFG1	RFC2	AVPI1
KCNQ5	SCARB1	PSMG3	LGI4
TMEM50B	HMGCS1	SNRPA	LOC113837182
SZRD1	LOC100756505	MCM2	DENND1A
CPNE1	GANAB	SMG5	LOC103159523
PDIA6	LOC100752363	RFC5	KIAA0232
ILDR2	PDIA5	ANKLE1	LOC113833349
CDK2AP2	ZWILCH	C6H19ORF48	GDI1
SELENOS	SPINDOC	ROR1	DDX17

SURF4	EMC4	CHAF1A	ZFYVE27
LRRC42	MSMO1	EIF4H	TSSK4
ATP6V0A2	LOC100761946	GATAD2A	ARID4B
HGS	SIAE	LOC100760727	UFSP2
ERC1	ARRB1	CEP162	MMP24OS
MANF	FKBP9	NOLC1	ATL2
NSMF	OGFOD3	TWF2	LOC103164340
SLC41A2	GPHR	MCM7	LOC100768249
DNAJC3	RAPGEF1	CDCA5	CFAP36
SEC23IP	LOC103161734	CHTF18	LOC100753357
SYVN1	APLP2	CPSF1	EEF1A2
SEC24D	CTSZ	PAQR4	LOC113835325
SETX	ERLEC1	CCT5	NPPB
LPAR1	ANKRD34A	E2F7	DSTN
MTMR1	CD82	LOC100769195	ANKRD13B
PDCD5	ERLIN1	CHRNB1	SPG11
PDHB	NDUFB5	MCM3	TMEM87B
NOB1	LRPAP1	PSME3	PRR13
PARG	BTD	TRIR	LOC100752946
LOC100751546	TMED1	AHCY	PSAP
IPO13	APMAP	AIFM1	LOC103162162
DDX39B	LOC103161380	CDCA4	FAM234A
FICD	BCAP31	E2F1	ZC3H3
SSB	PCOLCE2	PRPF19	MEF2D
CMTM3	LOC100750822	SEPHS2	PHF10
SLC33A1	GTDC1	MYO19	IFNAR1
MIDN	MAN2B2	LARP1	HDHD2
FBXO33	CNOT6L	HAUS5	SGPL1
SLC12A9	ZNF608	PA2G4	FAM177A1
MORF4L2	INSIG1	DNMT1	OPTN
SH3PXD2B	NDUFS1	ORC1	GIT2
TTPAL	LOC100759732	UHRF1	ING4
NANS	SELENOH	HLCS	LOC100753964
SELENOK	SUCLG1	CDC7	PDCD7
DXO	TMEM140	HECTD3	EAPP
TRAF3IP1	TMPRSS6	CTBP1	SLC22A17
SAMD1	LOC107977970	BAG5	FXR2
PIK3R2	PKM	FEN1	ALDOA
FHOD1	UBL5	GOT2	RDH11
BZW1	WDR5B	PELP1	LOC100752268
PPP1R14B	PRDX4	DTYMK	GTPBP2
LOC113833691	MRPS33	ARHGEF10	PHKA2
HSPA5	LPCAT3	KPNB1	STXBP1
SH3BP1	IFRD1	STK11	KLF9
MCF2L	SDF4	TCF19	LOC103163422
LOC100762413	LOC103161891	RPS6KA4	FTH1
NUP62	TMED4	THRAP3	CHFR

PEX5L	CENPA	FDXR	MAPK8IP3
KCNAB2	LPXN	TRIM47	LOC100771461
SEPT11	ATF6B	JPT2	KANSL1L
FEM1A	CEP57L1	IVNS1ABP	LOC103159711
PRPF3	LOC100754872	PAICS	ZFR2
LOC100770709	GINM1	DALRD3	GRINA
LOC100771739	EMC1	DHX9	ZC3H11A
RRS1	LOXL3	DTL	RAB6A
PWWP3A	EEPD1	PIN1	PWWP2A
TMEM185B	LOC113833361	ALKBH7	LOC107978829
AMD1	PTGFRN	SRSF3	VPS37A
XBP1	TFPI	SRSF7	LOC100765686
LOC103160811	GPC1	ZMPSTE24	CLIP1
MROH1	MAMDC2	THOP1	LAMP2
RGL2	TFRC	YKT6	HOXB6
RHOG	QSOX1	PRIM1	LEMD2
ADAMTSL5	SSR2	LOC113833500	ATP6V1D
N4BP2	ECH1	THADA	GSTP1
ANAPC2	NHLRC3	MAGOH	GPR132
PAK3	SIRT3	NUP93	LOC113832375
SYNGAP1	TBC1D10C	POLE	CATSPERG
SLC27A4	RNASEH2C	AAAS	LCORL
EML6	TOMM6	TMED2	NDUFC1
CRELD2	LOC113834948	NUDC	LOC103161071
WRNIP1	M6PR	DLST	CFAP58
CUNH15ORF39	PARVA	PSMA7	KATNA1
ZNF605	ERO1A	ZFP41	LOC100766583
FZD1	LGALS3	IPO4	CTSA
SVBP	AGA	RRP12	STIM2
SDF2L1	GNPTG	EFTUD2	AMT
PPM1B	LEKR1	RANBP1	LOC113833851
EDEM1	CUNH2ORF68	ARHGAP39	LOC113838144
HYOU1	LOC103163310	MYOF	YPEL5
DCTN1	TLK2	FAM3C	LOC113837038
CASKIN2	CADM1	NPLOC4	LOC100761572
LOC103159983	LOC103159590	GPS1	MYZAP
RPL7L1	GNA15	ABCF2	LOC100772170
LOC103162032	STRBP	NAA50	CD59
SHROOM3	DGUOK	PAK1IP1	NDFIP1
USP48	JCHAIN	MCM5	CUX1
SHE	VPS36	DAXX	
USO1	MBTPS1	UNG	
TRAF5	FUCA1	PSMF1	
RASA2	PDE9A	CBX1	
SUB1	LPIN1	WDHD1	
LOC100764369	RMDN1	RAD54L	
SMTN	RNLS	TRIP13	

NUCB1	ARL4A	CCT6A	
LOC100770325	CRIM1	PTBP1	
MAPK1IP1L	CHID1	UBA2	
NUS1	DBP	LMF2	
PLEKHG3	MAOB	PPM1G	
LOC113836685	MANBAL	CLEC16A	
CLP1	RBBP9	ABCB8	
BRCA1	SIN3B	NACC1	
NR2F6	UROS	GRK6	
INTS3	ERN1	MAP6	
SKP2	MAP4K4	PAFAH1B3	
DIXDC1	LOC103163253	ANAPC15	
ATAD1	KANK3	TIMM13	
LOC100762252	LOC100758278	ZNF367	
HMOX2	PRR11	ERH	
GTPBP3	PRSS27	YBX1	
SMCHD1		BOP1	
GMPPB		DHFR	
LOC100753789		JAK3	
COL4A3BP		CIT	
CEP131		EIF3L	
URB2		RNASEH2A	
DNAJB11		MCM10	
FKBP2		THOC5	
TBCCD1		CCDC86	
CABLES2		TEX261	
RBBP4		HELLS	
PDIA3		DNAAF5	
SOCS1		LOC100761779	
SGMS2		RAP2B	
TNS2		CNOT9	
LOC100760261		SUPT5H	
IL22RA1		HEXIM1	
FAM171A1		MCM4	
CDC6		LRRC59	
PSORS1C2		RPA2	
CRK		FAM168B	
ZNHIT3		SET	
UHMK1		CDCA7	
ATXN7L3		SLC25A10	
WDR81		SGTA	
GEMIN4		PPP1R8	
ATXN7L3B		E2F2	
AIMP2		SRRT	
		CSE1L	
		BEND3	
		TTF2	

		SNRPD2	
		PTP4A3	
		RBL1	
		TSN	
		LMNB2	
		KIAA0100	
		WDR90	
		HEATR1	
		NUP85	
		YWHAB	
		ATAD5	
		NASP	
		MCM6	
		NONO	
		MTX1	
		TCF3	
		DEK	
		TOMM40	
		FSCN1	
		TMX1	
		ALYREF	
		DBR1	
		SEC24A	
		LMNB1	
		LOC100756717	
		PRPF4	
		NOP58	
		TOPBP1	
		HMGB1	
		FRMD8	
		IRAK1	
		LOC100763852	
		LTN1	
		ACTN4	
		LOC100766053	
		POLD2	
		C1H3ORF67	
		DSCC1	
		BRD2	
		TWNK	
		MRPL28	
		WDR82	
		U2AF1	
		TMEM97	
		LOC100755061	
		DCAKD	
		DHX37	

		INKA2	
		SENP1	
		GLTP	
		RPS2	
		LOC100763157	
		DDX46	
		TPRG1L	
		CCT2	
		SIVA1	
		NUP160	
		MYBL2	
		NCL	

Technical
University of
Denmark

The Novo Nordisk Foundation Center for Biosustainability
biosustain.dtu.dk