Technical University of Denmark

DTU

Accelerated and Rational Design of Improved CHO Cell Factories

Grav, Lise Marie

Publication date:
2018

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

Link back to DTU Orbit

Citation (APA):
Grav, L. M. (2018). Accelerated and Rational Design of Improved CHO Cell Factories. Kgs. Lyngby : Novo Nordisk Foundation Center for Biosustainability.

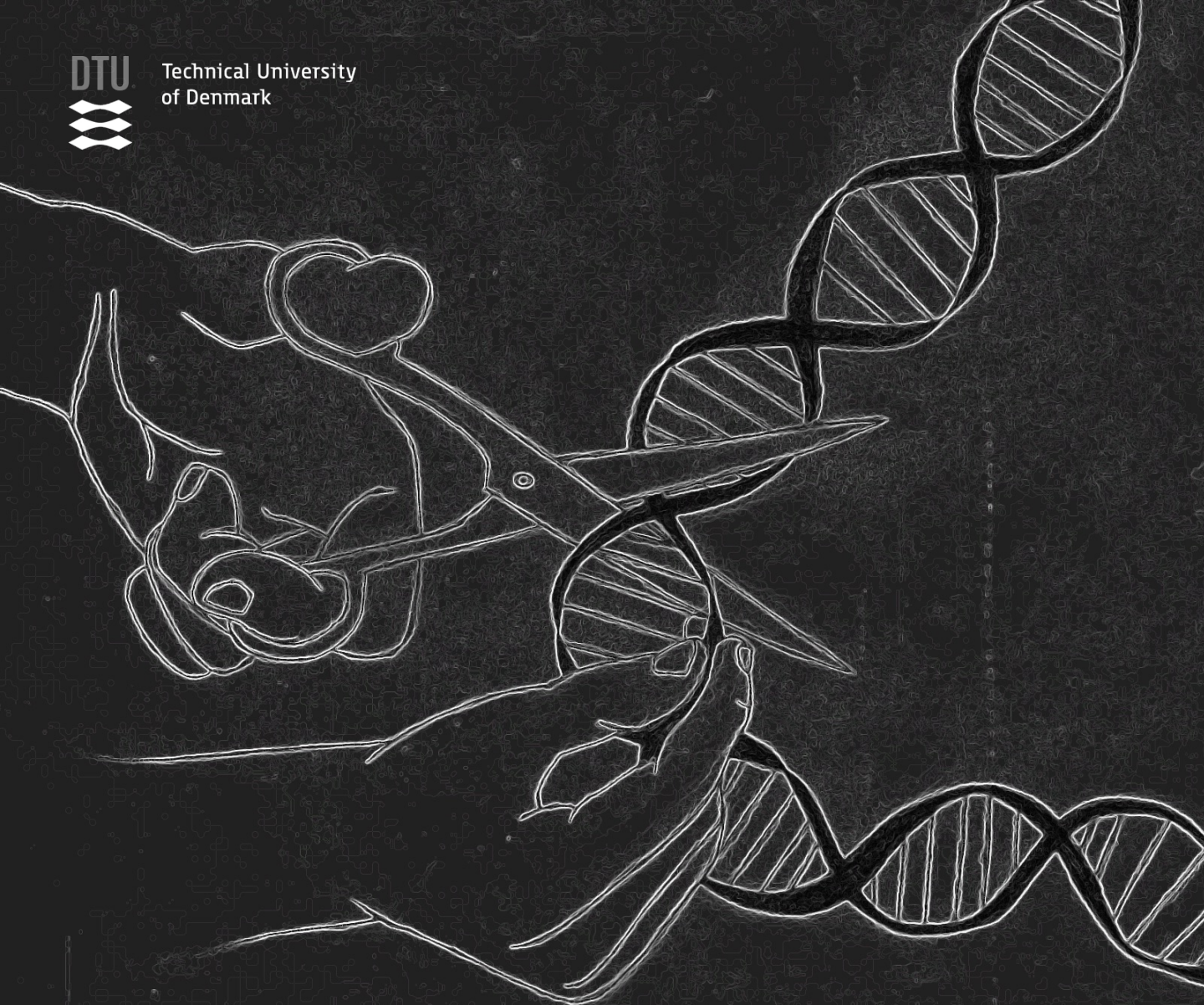DTU Library
Technical Information Center of Denmark

Accelerated and Rational Design of Improved CHO Cell Factories

Lise Marie Grav

Ph.D. Thesis
January 2018

Accelerated and Rational Design of Improved CHO Cell Factories

Ph.D. Thesis

Lise Marie Grav

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

January 2018

Supervisor: Senior Researcher and Co-PI Helene Faustrup Kildegaard
Co-Supervisors: Assistant Professor Jae Seong Lee & Professor MSO Mikael Rørdam Andersen

Cover art: Hand drawn illustration of precision genome editing. Lise Marie Grav.

Preface and acknowledgements

This PhD was carried out at the Novo Nordisk Foundation Center for Biosustainability at the Technical University of Denmark, in the research group of Senior Researcher and co-PI Helene Faustrup Kildegaard. The project was initiated December 1st, 2014, and was financed by the Novo Nordisk Foundation. The period January 2017 to May 2017 was spent learning valuable skills in big data handling and transcriptomic analysis as a visiting researcher at the University of Queensland under the supervision of Lars Keld Nielsen.

I feel extremely privileged and grateful to have been given the opportunity to complete my PhD in a research environment filled with utmost talented people, excellent collaborative efforts and guidance. It has truly been an adventure with many ups and downs, where I have developed both scientifically and personally.

Throughout this journey I have been fortunate to interact with many inspiring scientist, I couldn't have made it without you. I especially want to give a big thank you to my supervisors, Helene, Jae Seong and Mikael for their patience, excellent supervision, suggestions, ideas, advice and endless support. Helene, it has been a pleasure completing this PhD project in your group, you create an amazing work environment filled with trust and freedom and I am proud to be your first PhD student to graduate from your group. Daria, I want to thank you for much appreciated support and scientific input during my final year. Ankita, thank you for the endless conversations, patience and support concerning the topic of transcriptomics. Nusa, Thomas A. and Sara thank you for all the scientific and other fruitful discussions we have had in our time as office buddies. Julie, you have been essential for this work, you have provided both excellent professional and mental support. Thank you for getting me out of the laboratory and the office on occasions, and letting me unload all my frustrations on to you. I want to extend my gratitude to the CHO Core group, for their support in the laboratory. Special thanks go to Karen Kathrine, you have made our laboratory run smoothly during my entire PhD project. I am grateful for the FACS training and support you have provided me with. Lars, thank you for all your excellent ideas, suggestions and support. I was extremely pleased with my external stay in your research group, thank you for making it possible. Robin, thank you for teaching me valuable computer coding skills and how to set up a transcriptomic pipeline. You have been extremely patient, kind and helpful.

I am grateful to all my wonderful colleagues during my three years as a PhD, special thanks go to Thomas K. and Jae Seong for teaching me all your tricks and skills in the laboratory, for your patience and the excellent scientific education you have provided. You have both been exceptional mentors for me and I am eternally grateful for that.

Last and finally, I want to thank my family and friends for the support and encouragement I have received during this PhD project and life in general. I would not be where I am today, if it wasn't for you all.

Copenhagen, January 29th, 2018

Lise Marie Grav

Abstract

Recombinant production of therapeutic proteins provides huge benefits to human health and promises solutions to some of the most devastating and currently untreatable diseases in healthcare. Key to the development of new therapeutic proteins is to optimize and engineer living cells, namely cell factories, to produce therapeutic proteins by taking advantage of the cells protein production machinery. However, there is an insufficient understanding of the cellular processes and their regulatory mechanisms that hinder optimal therapeutic protein production such as cell growth, protein productivity and product quality. Our successes are dependent on getting a better understanding of the biological processes that comprise the cell factory, and the ability to directly engineer the cell factory for beneficial production of both new and existing therapeutic products.

In this thesis, the development of a number of novel tools is reported that aim to accelerate the construction of production cell lines for therapeutic proteins with optimal phenotypic attributes for industrial processes. Chinese hamster ovary (CHO) cells are the predominant production host for therapeutic proteins, and are the cell factory of interest in this thesis. The core of the thesis is revolved around the development and application of genome editing techniques that enable us to precisely engineer the genome of CHO cells by either rendering specific-targeted genes un-functional or inserting new genes in precise genomic locations. This allows us to i) facilitate improved studies of gene functions, ii) remove cellular traits that are disadvantageous to protein production, iii) control and predict the level and stability of protein expression, and iv) design and engineer new cellular functions.

Overall, the results of this thesis illustrate the benefits of advancing the toolbox for designing and engineering CHO cell lines. Each chapter demonstrates direct applications of strategies to improve the therapeutic production capabilities of CHO cell factories, with the goal to speed up the development process of new therapeutic proteins and reducing the costs in order to benefit a broader range of patients around the globe.

Dansk sammenfatning

Rekombinant produktion af terapeutiske proteiner giver enorme fordele for menneskers sundhed og præsenterer os for løsninger på nogle af de mest alvorlige og i øjeblikket ubehandlede sygdomme i sundhedsvæsenet. Nøglen til udviklingen af nye terapeutiske proteiner er at optimere og konstruere de levende celler, nemlig cellefabrikkerne, til at producerer terapeutiske proteiner ved at udnytte cellernes eget proteinproduktions maskineri. Vores forståelse af de cellulære processer og reguleringsmekanismer er imidlertid utilstrækkelig, og forhindrer optimal terapeutisk proteinproduktion, såsom øget cellevækst, protein produktivitet og protein kvalitet. Fremtidige succeshistorier er afhængige af at vi får en bedre forståelse af de biologiske processer, der omfatter cellefabrikken, og evnen til direkte at konstruere cellefabrikken til gavnlig produktion af både nye og eksisterende terapeutiske produkter.

I denne afhandling rapporteres udviklingen af en række nye værktøjer, der sigter mod at fremskynde konstruktionen af produktionscellelinier til fremstilling af terapeutiske proteiner med optimale fænotypiske egenskaber for industrielle processer. Kinesiske hamster ovarie (CHO) celler er den dominerende produktionsvært for terapeutiske proteiner, og er cellefabrikken af interesse i denne afhandling. Kernen i afhandlingen drejer sig om udvikling og anvendelse af genomredigeringsteknikker, der gør det muligt for os at præcis redigere genomet af CHO-celler ved at gøre udvalgte gener ufunktionelle eller indsætte nye gener på præcise genomiske steder. Dette tillader os at i) letter forbedre studier af genfunktioner, ii) fjerne cellulære træk, der er en ulempe for proteinproduktion, iii) kontrollere og forudsige niveauet og stabiliteten af proteinekspression, og iv) designe og konstruere nye cellulære funktioner.

Samlet set illustrerer resultaterne af denne afhandling fordelene ved at videreudvikle værktøjskassen til at designe og konstruere CHO-cellelinjer. Kapitlerne i afhandlingen demonstrerer direkte applikationer af strategier til at forbedre CHO-cellefabrikkernes terapeutiske produktionskapacitet, med det formål at fremskynde udviklingsprocessen af nye terapeutiske proteiner og reducere omkostningerne til gavn for en større andel patienter over hele verden.

List of publications

This thesis is compiled of the following articles and manuscripts (published and unpublished):

I. CRISPR/Cas9 mediated genome engineering of CHO cell factories: Application and perspectives
Jae Seong Lee, <u>Lise Marie Grav</u>, Nathan E. Lewis, Helene Faustrup Kildegaard. (2015). *Biotechnology Journal*, 10, 979-999. doi: 10.1002/biot.201500082

II. Application of CRISPR/Cas9 Genome Editing to Improve Recombinant Protein Production in CHO Cells
<u>Lise Marie Grav</u>, Karen Julie la Cour Karottki, Jae Seong Lee, Helene Faustrup Kildegaard. (2017). In: Meleady P. (eds) Heterologous Protein Production in CHO Cells. Methods in Molecular Biology, vol 1603. Humana Press, New York, NY. doi: 10.1007/978-1-4939-6972-2_7.

III. One-step generation of triple knockout CHO cell lines using CRISPR/Cas9 and fluorescent enrichment
<u>Lise Marie Grav</u>, Jae Seong Lee, Signe Gerling, Thomas Beuchert Kallehauge, Anders Holmgaard Hansen, Stefan Kol, Gyun Min Lee, Lasse Ebdrup Pedersen, Helene Faustrup Kildegaard. (2015). *Biotechnology Journal.* , 10, 1446-1456. doi: 10.1002/biot.201500027.

IV. Accelerated Homology-Directed Targeted Integration of Transgenes in Chinese Hamster Ovary Cells via CRISPR/Cas9 and Fluorescent Enrichment
Jae Seong Lee, <u>Lise Marie Grav</u>**,** Lasse Ebdrup Pedersen, Gyun Min Lee, Helene Faustrup Kildegaard. **(**2016). *Biotechnology and Bioengineering*, 113(11):2518-23. doi: 10.1002/bit.26002.

V. Minimized clonal variation for improved comparative systems biology studies
<u>Lise Marie Grav</u>, Jae Seong Lee, Daria Sergeeva, Mikael Rørdam Andersen, Lars Keld Nielsen, Gyun Min Lee, Helene Faustrup Kildegaard. (2018). (*Manuscript in preparation*)

In addition, the work during my PhD has contributed to the following manuscript, which have not been included in this thesis:

VI. Versatile microscale screening platform for improving recombinant protein productivity in Chinese hamster ovary cells
Henning Gram Hansen, Claes Nymand Nilsson, Anne Mathilde Lund, Stefan Kol, <u>Lise Marie Grav</u>, Magnus Lundqvist, Johan Rockberg, Gyun Min Lee, Mikael Rørdam Andersen, Helene Faustrup Kildegaard. (2015). *Scientific Reports*, 5, 18016. doi:10.1038/srep18016.

Table of contents

Introduction

Therapeutic proteins have enabled effective treatments for some of the most difficult medical conditions that threaten human health, as they are able to modulate physiological processes at a cellular level. The early therapeutic proteins were derived from animal- and subsequently human-sources. A groundbreaking achievement in the field of therapeutic proteins was the establishment of recombinant production of therapeutic proteins, where living cells are utilized as host systems for the protein production. It allows the production of large quantities of human proteins in a controlled environment. This has improved the access to these life-changing treatments, and simultaneously reduced the risk of pathogen transmission and immune reactions to potential foreign components in the products. Treatments are currently offered for a wide range of diseases such as cancer, autoimmune disorders, hormone and growth factor deficiencies, and benefits health of millions of people worldwide [1] [2].

The demand for new therapeutic products is continuously increasing in order to broaden the range of medical conditions that are being treated. To expand the access to these life-changing products, the products need to be made safer and more affordable. Increasing the production levels, and creating a more robust and reproducible production host system will accelerate the development and commercialization of new products and reduce manufacturing costs. The product itself should display maximal biochemical and biophysical uniformity in protein folding, monodispersity, and PTMs to increase product safety and keep up with the increasing regulatory demands [3].

Since the approval of the first recombinant therapeutic protein expressed in Chinese hamster ovary (CHO) cells over 30 years ago [2], CHO cell lines have become the primary production host for therapeutic proteins, producing 7 out of the top 10 best selling therapeutic proteins from 2016 [4]. CHO cells are favoured mainly due to their ability to produce proteins with post-translational modifications (PTMs) biologically similar to those found in humans. A requirement for optimal safety, quality and efficacy. Thus, therapeutic proteins produced in CHO cells are superior to those produced in microbial hosts. CHO cells have been successfully used in production of safe and high-quality therapeutic proteins for decades, generating a stronger regulatory track record than any other cell line used today [2]. They can grow in suspension culture making them ideal for large-scale bioprocesses and they achieve relatively high productivity levels for certain products [5].

During the development of new therapeutic proteins or biosimilars, a new production host cell line also needs to be developed. A typical cell line development pipeline involves the transfection of expression cassettes with a recombinant gene (encoding the protein of interest) and a selection marker, followed by selection and gene amplification to increase the number of recombinant gene copies. Due to random integration of the recombinant gene in the genome and the amplification process, the resulting clones display highly heterogeneous phenotypes, known as clonal variation. Subsequently, a large number of clones needs to be screened to identify a stable high-producing clone candidate, followed by scale-up and process optimization. The cell line development pipeline

typically requires a timeline of six to twelve months [6]. Successes currently achieved in CHO host cell development are impressive with substantial increase in titers since the introduction of the first recombinant therapeutic protein [5]. The successes are mostly driven by unpredictable, labour-intensive and time-consuming screening and bioprocess optimization [2,5]. The development processes are often product-specific and not easily reproduced, requiring the entire pipeline to be repeated for each new product. Additionally, the production processes still suffer from cellular limitations such as low productivity of certain proteins, requirement for advanced growth media, and limited growth capacity leading to longer and more expensive manufacturing processes as compared to bacterial- or yeast-based production host systems.

Advancements in cell line development technologies are crucial to support the effort of accelerating the development of safer, more affordable and more complex therapeutic products. With the availability of several CHO cell line draft genomes [7,8] and genome engineering technologies [9], the field is moving towards more rational engineering approaches. CHO cells have lagged behind microbial cell factories in terms of cell line engineering due to the late arrival of both annotated genomes and available genome editing tools. A combination of an increased understanding of cellular processes and the ability to precision engineer genomes will provide the means to start re-engineering CHO cells with new and improved capabilities - more attuned to highly efficient bioprocesses. The new precise genome editing tool - Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)/CRISPR-associated protein 9 (Cas9) - recently entered the field of CHO cell engineering [10]. This tool is easier to establish, less time-consuming, more efficient, and much more cost-effective than other gene editing methods such as transcription activator-like effector nucleases and zinc-finger nucleases. CRISPR/Cas9 technology has the potential to revolutionize current cell line generation strategies to enable rational design and engineering of CHO production hosts. With a rational engineering approach, the cell line development will be more predictable in terms of production levels and stability, speeding up development and bringing the product to the market in a shorter time frame.

The goal of this PhD project is to expand the available genome editing tools for CHO cell line engineering. It aims to contribute to the advancement of synthetic biology within the field of recombinant protein therapeutic production; with a vision of being able to one day build CHO cells from scratch with optimal functionalities for therapeutic protein production, in a fast, predictable and user-friendly manner.

Thesis structure

The research presented in this thesis was completed in a collaborative effort between me and my colleagues. This thesis is divided into five parts, centered around my main projects: the development and rational use of site-specific genome editing tools. In **Chapter 1**, a review serving as an introduction to the field of CRISPR/Cas9 genome editing, and its potential applications and future perspectives in the field of CHO cell engineering is presented. **Chapter 2** builds on chapter 1, and presents a protocol on how to step-by-step use the genome editing tool CRISPR/Cas9 to knockout target genes in CHO cells. **Chapter 3** present a study where we successfully utilized the CRISPR/Cas9 genome editing tool to knockout multiple genes simultaneously in CHO cells. In **Chapter 4**, the first case of targeted gene integration using CRISPR/Cas9 without the use of an antibiotic selection marker is presented. **Chapter 5**, shows how site-specific gene insertion tools can be applied to achieve more accurate comparative biological studies, using the field of transcriptomics as a test case. Finally, **Chapter 6** contains a few concluding remarks about the results and future perspectives for the research presented in this thesis.

Chapter 1

CRISPR/Cas9 as a genome editing tool

This chapter introduces the history and concept of CRISPR/Cas9, and its use as a genome editing tool for engineering purposes of CHO cell lines. CRISPR/Cas9 was introduced to mammalian cell engineering in the infancy of this PhD project and plays a major role throughout the studies presented in this thesis. We discuss the many different applications, future directions and potential of genome editing on the road to building next-generation CHO cell factories.

Review

CRISPR/Cas9-mediated genome engineering of CHO cell factories: Application and perspectives

Jae Seong Lee[1], Lise Marie Grav[1], Nathan E. Lewis[2,3] and Helene Faustrup Kildegaard[1]

[1] The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Hørsholm, Denmark
[2] Department of Pediatrics, University of California, San Diego, La Jolla, CA, USA
[3] The Novo Nordisk Foundation Center for Biosustainability at the University of California, San Diego School of Medicine, CA, USA

Chinese hamster ovary (CHO) cells are the most widely used production host for therapeutic proteins. With the recent emergence of CHO genome sequences, CHO cell line engineering has taken on a new aspect through targeted genome editing. The bacterial clustered regularly interspaced short palindromic repeat (CRISPR)/CRISPR-associated protein 9 (Cas9) system enables rapid, easy and efficient engineering of mammalian genomes. It has a wide range of applications from modification of individual genes to genome-wide screening or regulation of genes. Facile genome editing using CRISPR/Cas9 empowers researchers in the CHO community to elucidate the mechanistic basis behind high level production of proteins and product quality attributes of interest. In this review, we describe the basis of CRISPR/Cas9-mediated genome editing and its application for development of next generation CHO cell factories while highlighting both future perspectives and challenges. As one of the main drivers for the CHO systems biology era, genome engineering with CRISPR/Cas9 will pave the way for rational design of CHO cell factories.

Correspondence: Dr. Helene Faustrup Kildegaard, the Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Kogle Allé 6, 2970 Hørsholm, Denmark
E-mail: hef@biosustain.dtu.dk

Abbreviations: BER, base excision repair; **Cas**, CRISPR-associated; **Cas9**, Cas protein 9; **ChIP-seq**, chromatin immunoprecipitation-sequencing; **CHO**, Chinese hamster ovary**; CMV**, cytomegalovirus; **CRISPR**, clustered regularly interspaced short palindromic repeat; **CRISPRi**, CRISPR interference; **crRNA**, CRISPR RNA; **dCas9**, dead Cas9; **DSBs**, double strand breaks; **dsODN**, double-stranded oligodeoxynucleotide; **EF1-α**, elongation factor 1-α; **FACS**, fluorescence-activated cell sorting; **gRNA**, guide RNA; **HDR**, homology directed repair; **HR**, homologous recombination; **IDLV**, integrase-deficient lentivirus; **indel**, insertion/deletion; **MMEJ**, microhomology-mediated end joining; **NHEJ**, non-homologous endjoining; **NLS**, Nuclear localization signal; **PAM**, protospacer adjacent motif; **RFLP**, restriction fragment length polymorphism; **RGENs**, RNA-guided engineered nucleases; **RNAi**, RNA interference; **SpCas9**, Streptococcus pyogenes Cas9; **SSA**, single-strand annealing; **SSBs**, single strand breaks; **ssODNs**, single-stranded oligodeoxynucleotides; **TALENs**, transcription activator-like effector nucleases; **tracrRNA**, trans-activating crRNA; **ZFNs**, zinc-finger nucleases.

1 Introduction

Chinese hamster ovary (CHO) cells are the primary expression system for manufacturing biopharmaceuticals. CHO cells account for 35.5% of the total cumulative (1982–2014) biopharmaceutical product approvals since the first market approval of tissue plasminogen activator produced from recombinant CHO (rCHO) cells in 1987 [1, 2]. Despite extensive use of CHO cells as the workhorse of the biopharmaceutical industry, CHO cells were considered as a 'black box' due to lack of genomic information. This hindered efforts to understand the molecular basis of high level production of recombinant proteins in CHO cells. Thus, the impressive progress in CHO cell culture technology was achieved by empirical approaches such as screening and process optimization. Such approaches are effective in obtaining high level production of many recombinant proteins, including recombinant antibodies and Fc-fusion proteins, with yields up to 10 g/L and over [3]. However, such yields are often limited to only certain types of product. In addition, a high degree of variability

Biotechnology
Journal
www.biotechnology-journal.com

- Chapter 1 -

Biotechnol. J. 2015, 10, 979–994

BIO tec
visions
www.biotecvisions.com

of rCHO cells requires laborious and expensive processes to select the best clone for the production of each new therapeutic candidate [2].

The systems biology approach toward rational design of microbial production hosts has been driven by three main resources: genomic information, genome-scale pathway information/computational models, and genetic engineering tools [4]. The same drivers for CHO cells have recently arisen, thus permitting genome-scale science for CHO cells. The complete genomic sequence of CHO-K1 cells became publicly available in 2011 [5], and since then sequencing efforts have resulted in additional genomic sequence data being available for the Chinese hamster and various CHO cell lines [6–8]. Now an active effort is underway in the CHO community to refine the genome assembly and annotation for the Chinese hamster [9]. The emergence of CHO genomes enables in-depth interpretation of various 'omics data sets [10] and construction of CHO genome-scale models, such as computational models of the CHO metabolic network [4, 11]. Integration of 'omic resources with modeling approaches may unveil the underlying mechanism between genotype and phenotype in CHO cells. This could contribute to identification of either biomarkers for determining high producing and stable CHO cells or novel and non-intuitive targets for CHO cell line engineering.

The emerging targets will necessitate extensive and efficient validation of the genetic manipulations. Genome editing tools have been developed including zinc-finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENS) and the more recent RNA-guided engineered nucleases (RGENs) derived from the bacterial clustered regularly interspaced short palindromic repeat (CRISPR)/CRISPR-associated (Cas) system, which enables one to cut and modify specific target DNA sites in a precise manner [12]. Compared with conventional gene targeting strategies based on homologous recombination, these programmable nucleases innovatively enhance the induction of DNA damage repair pathways, triggered by site-specific DNA double strand breaks (DSBs). The induction of DNA damage repair pathways results in targeted genome modification including gene knockout, gene correction/insertion and chromosome rearrangements [12]. Armed with genome editing technologies, large-scale genetic manipulation and genome analysis of CHO cells is feasible. These engineering efforts now enable the development of next-generation recombinant protein production in CHO hosts, with their accompanying improved productivity and protein quality. Notably, of the current genome editing nucleases, RGENs, hereafter referred to as CRISPR/Cas9, have been evolving rapidly due to their simple composition based on the principles of Watson-Crick base pairing between small guide RNAs and the targeted genomic sites rather than the use of modular DNA recognition proteins. This feature confers ease of use, rapid and cost-effective design in addition to high targeting efficiency that has led to widespread implementation of CRISPR/Cas9 for genetic modification of various types of organism and cell [13–15]. A comparison of the advantages and disadvantages of the ZFNs, TALENS and RGENs has previously been reviewed [16, 17].

In conjunction with the advent of revolutionary CRISPR/Cas9-mediated genome engineering, tomes of research and review papers have been published since 2013, highlighting its short history and diverse applications [13–15]. The biotechnological potential of this system is of particular interest, and so here we highlight the role of CRISPR/Cas9 in CHO cell engineering. We first briefly describe the principle of the CRISPR/Cas9 system and genome editing in cells, followed by its potential with examples consisting of four applications. We then review experimental procedures for genome engineering with the CRISPR/Cas9 system in mammalian cells including CHO cells. Finally, we conclude by discussing the past, present and future of CHO cell factory engineering via CRISPR/Cas9-mediated genome editing.

2 CRISPR/Cas9 in adaptive immunity

CRISPR/Cas systems have evolved in prokaryotes to provide adaptive immunity against foreign genetic elements such as viruses and plasmids [18]. The CRISPR loci typically consist of a clustered set of Cas genes and a CRISPR array – a series of repeat sequences (direct repeats) interspaced by variable sequences (spacers). The spacers correspond to short sequences within foreign genetic elements (protospacers), such as phage sequences [19]. Three major CRISPR/Cas systems (I–III) are classified based on their genetic content and distinct mechanism to achieve CRISPR-mediated adaptive immunity [20]. The overall theme of type II CRISPR-mediated adaptive immunity can be divided into three stages after the invasion of foreign genetic elements. The first stage is spacer acquisition, where protospacers are cleaved by different Cas proteins and subsequently inserted into the CRISPR array as spacers [18]. Biogenesis and processing of CRISPR RNA (crRNA) form the second stage, where CRISPR arrays are transcribed and subsequently processed into smaller crRNAs, each composed of a direct repeat and a spacer [21]. Trans-activating crRNA (tracrRNA) is also transcribed from the CRISPR locus and contributes to the processing of crRNA, by hybridizing with the direct repeats and forming an RNA duplex that is cleaved by RNase III [22]. In the final step Cas9 and crRNAs (including tracrRNA) hybridize to form an active ribonucleoprotein complex. This complex scans invading DNA for a target complementary to the crRNA spacer sequence, and directs the cleavage of the target DNA by the Cas9 nuclease [23–25]. This targeted cleavage by CRISPR/Cas9 has been used to enable rational genome editing.

Biotechnology
Journal
www.biotechnology-journal.com

- Chapter 1 -

Biotechnol. J. 2015, 10, 979–994

BIO tec
visions
www.biotecvisions.com

3 CRISPR/Cas9-mediated genome editing

The foundation for utilizing the CRISPR/Cas9 as programmable nucleases for genome editing was laid by a few essential findings: DNA is the primary target of CRISPR/Cas systems [26], Cas9 is the only protein needed to generate DSBs in the target sequence [27], and crRNA hybridizes with trans-activating crRNA (tracrRNA) to facilitate RNA-guided targeting of the Cas9 protein [22]. In 2013, the CRISPR/Cas9 system was successfully engineered to accomplish genome editing in mammalian cells, by recombinant expression of crRNA-tracrRNA together with Cas9 to create DSBs [28, 29]. The RNA duplex tracrRNA-crRNA, which can be fused to a single guide RNA (gRNA), guides the Cas9 protein to target DNA where it introduces site-specific DSBs (Fig. 1A) [23]. Target recognition by Cas9 requires both a 20-nt target complementary sequence in the crRNA and a protospacer adjacent motif (PAM) sequence adjacent to the crRNA binding region in the target DNA [25, 30]. In the CRISPR/Cas9 system derived from *Streptococcus pyogenes*, the PAM is a 5′-NGG sequence [25]. For other Cas9 orthologs there may be different PAM requirements [23, 24, 27, 28, 31]. The Cas9-RNA complex rapidly searches for guide RNA complementarity by dissociating from non-PAM sites [32]. Only upon binding to a PAM sequence does the Cas9-RNA complex look for guide RNA complementary in the flanking DNA [32]. Base pairing between the target sequence and the crRNA leads to the activation of the two nuclease-domains of Cas9 (HNH and RuvC-like domains) [33], which also requires PAM recognition [32]. The Cas9 protein can be programmed to target and cleave any dsDNA sequence of interest if it primes the PAM sequence, simply by changing the DNA target binding sequence in the guide RNA. The Cas9 protein can also be converted into a nicking enzyme, which is a mutant version of Cas9 created by inactivating one of the two nuclease-domains, to facilitate genome editing with minimal mutagenic activity [28, 34, 35]. While ZFNs and TALENs have been used for a longer period of time, many
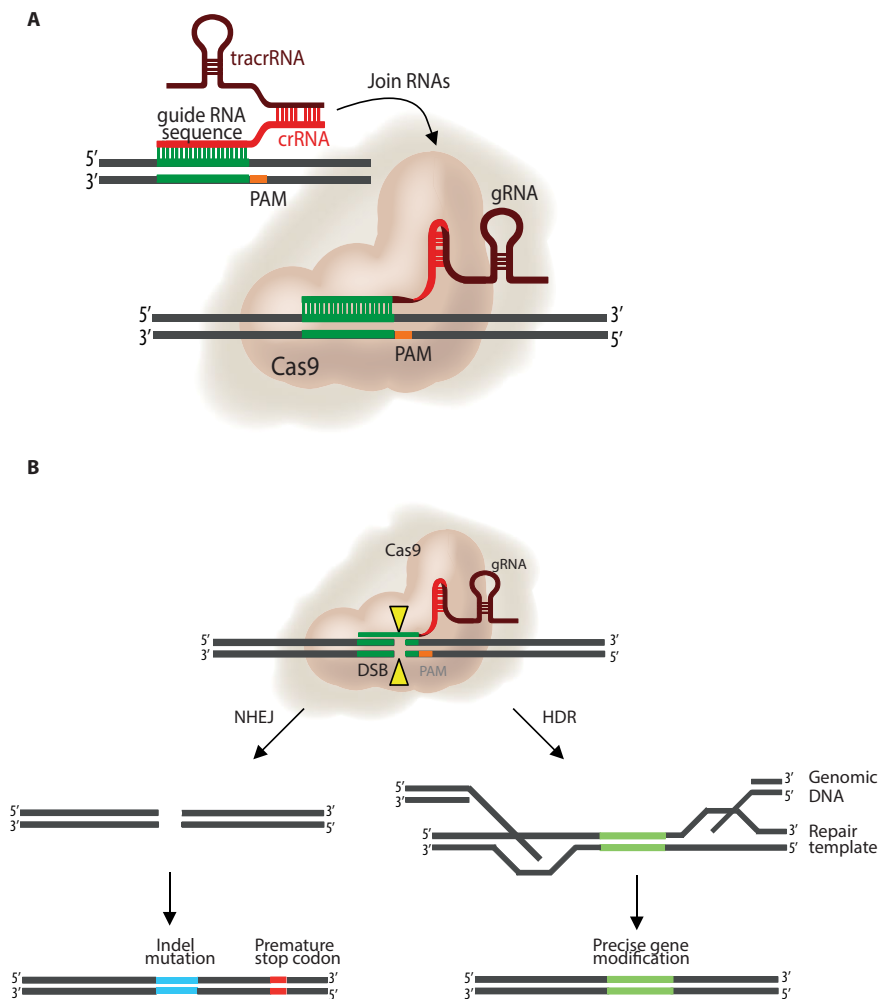
Figure 1. Genome editing with CRISPR/Cas9. (**A**) The trans-activating CRISPR RNA (tracrRNA) and the CRISPR (crRNA) are fused together as a single guide RNA (gRNA), and guide the Cas9 nuclease to the target DNA that is complementary to the guide sequence. (**B**) The Cas9 nuclease from the microbial CRISPR adaptive immune system is directed to specific DNA sequences via base-pairing of the guide sequence on its guide RNA (gRNA) (green) with the DNA target. Binding of a protospacer-adjacent motif (PAM, orange) downstream of the target site helps to direct Cas9-mediated DNA double strand breaks (DSBs). DNA DSBs are typically repaired by the error-prone non-homologous end joining (NHEJ) or error-free homology directed repair (HDR). In the NHEJ pathway, indels can be introduced during ligation of the dissected DNA ends and eventually lead to frameshift mutations and functional gene knock-out. In the HDR pathway, recruitment of accessory factors that direct genomic recombination with homology arms on an exogenous repair template facilitates the introduction of precise gene modification [13].

Table 1. Applications of the CRISPR Cas9 system for targeted genome editing in mammalian cells[a]

Cell type (Organism)	Application	Cas9 form	Delivery method	Target locus	Reference
HEK293FT (Human) N2A (Mouse)	Gene disruption	Cas9 nuclease[b] Cas9 nickase (D10A)	Plasmid/Chemical transfection	EMX1 PVALB Th (mouse)	[28]
HEK293T (Human) K562 (Human) PGP1 (Human)	Gene disruption Gene insertion	Cas9 nuclease	Plasmid (Cas9), RNA (gRNA), Plasmid or oligonucleotide (Donor)/ Chemical transfection or Nucleofection	AAVS1 DNMT3a/b	[30]
HEK293T (Human)	Gene disruption Gene insertion Gene activation	Cas9 nuclease Cas9 nickase (D10A) dCas9	Plasmid/Chemical transfection	ZFP42 (REX1) POU5F1 (OCT4) SOX2 NANOG AAVS1	[34]
HEK293FT (Human) HUES62 (Human)	Gene disruption Gene insertion	Cas9 nuclease Cas9 nickase (D10A, H840A)	Plasmid (Cas9) and PCR product (gRNA) or oligonucleotide (Donor)/ Chemical transfection or Nucleofection	EMX1 DYRK1A GRIN2B VEGFA	[35]
HEK293T (Human) K562 (Human)	Gene disruption	Cas9 nuclease	Plasmid (Cas9) and RNA (gRNA)/Nucleofection	CCR5 C4BPB	[44]
CHO (Chinese hamster)	Gene disruption	Cas9 nuclease	Plasmid/Nucleofection	C1GALT1C1 FUT8	[45]
HEK293T (Human) Dermal fibroblasts (Human)	Gene disruption Gene activation	Cas9 nuclease dCas9	Lentivirus[c]/Transduction	AAVS1 IL1RN HBG1	[46]
HEK293T (Human)	Gene disruption	Cas9 nuclease Cas9 nickase (D10A)	Plasmid[c]/Chemical transfection	HPRT1 ATM APC CDH1 AXIN2 CFTR	[47]
CHO (Chinese hamster)	Gene disruption	Cas9 nuclease	Plasmid/Chemical transfection	FUT8 BAK BAX	[48]
K562 (Human) HeLa (Human)	Gene disruption	Cas9 nuclease Cas9 nickase (D10A)	Plasmid (Cas9) and RNA (gRNA) or Plasmid for all components/ Chemical transfection or Nucleofection	CCR5 C4BPB VEGFA EMX1 AAVS1	[49]

Table 1. Applications of CRISPR Cas9 system for targeted genome editing in mammalian cells[a] (continue)

Cell type (Organism)	Application	Cas9 form	Delivery method	Target locus	Reference
Murine erythroleukemia (Mouse)	Genomic deletion	Cas9 nuclease	Plasmid/Electroporation	12 genomic loci	[50]
HAP1 (Human) KBM7 (Human)	Genomic deletion	Cas9 nuclease	Plasmid/Chemical transfection	Chr 15	[51]
HEK293T(Human) AALE (Human)	Chromosomal rearrangement	Cas9 nuclease	Plasmid/Chemical transfection or Nucleofection	CD74 ROS1 EML4 ALK KIF5B RET	[52]
CHO (Chinese hamster)	Gene insertion	Cas9 nuclease	Plasmid/Chemical transfection	C1GALT1C1 MGAT1 LDHA	[56]
HEK293T(Human)	Gene insertion	Cas9 nuclease	Plasmid/Chemical transfection	FBL	[57]
HEK293 (Human)	Gene repression	dCas9	Plasmid/Chemical transfection	EGFP reporter	[58]
HEK293 (Human)	Gene activation	dCas9	Plasmid/Chemical transfection	VEGFA NTF3	[59]
HEK293 (Human) HeLa (Human)	Gene activation Gene repression	dCas9	Lentivirus/Transduction or Plasmid/Chemical transfection	EGFP reporter CXCR4 CD71	[60]
KBM7 (Human) HL60 (Human)	Genome-scale screening	Cas9 nuclease	Lentivirus/Transduction	Coding exon for all RefSeq transcripts (7114 genes)	[64]
A375 (Human) HUES62 (Human)	Genome-scale screening	Cas9 nuclease	Lentivirus/Transduction	Constitutive exons for all coding genes (18 080 genes)	[65]
ESC (Mouse)	Genome-scale screening	Cas9 nuclease	Lentivirus/Transduction	Mouse proteincoding genes (19 150 genes)	[66]
K562 (Human)	Genome-scale gene regulation	dCas9	Lentivirus/Transduction	Entire set of protein coding genes (15 977 genes)	[68]
A375 (Human)	Genome-scale gene activation	dCas9	Lentivirus/Transduction	Human RefSeq coding isoforms (23 430 isoforms)	[69]
HeLa (Human)	Genome-scale screening	Cas9 nuclease	Lentivirus/Transduction	Library based on screening results (291 genes)	[70]

a) Examples are limited to the references in Section 4 due to space limitations, and are listed in the order in which they appear in the text.
b) Two Cas9 proteins from *Streptococcus pyogenes* (SpCas) and *Streptococcus thermophiles* (StCas) were used.
c) A single vector system containing multiple gRNA expression cassettes and a Cas9 expression cassette was incorporated.

Biotechnology Journal
www.biotechnology-journal.com

- Chapter 1 -

Biotechnol. J. 2015, 10, 979–994

BIOtec visions
www.biotecvisions.com

researchers are now also exploiting the CRISPR/Cas9 system as a genome editing tool, and employing it in a substantial number of species and cell types (see [15] and Table 1).

4 Applications of CRISPR/Cas9-mediated genome editing

The basis for performing targeted genome editing is the creation of DSBs in the genomic site of interest. Nuclease induced DSBs in mammalian genomes can be resolved by the endogenous DNA repair mechanism, which has evolved into two major pathways: nonhomologous end joining (NHEJ) and homology directed repair (HDR) (Fig. 1B) [16, 36, 37]. In the absence of repair templates, NHEJ ligates both dissected DNA ends directly without DNA-end resection, which can generate mutations at the DSB site. In contrast, HDR repair precisely restores DSBs on the basis of repair templates harboring a homologous DNA sequence that includes endogenous sister chromatids or exogenous donor repair templates. HDR includes homologous recombination (HR) and single-strand annealing (SSA), which are initiated by DNA-end resection. HR is the most common form of HDR and has sub-pathways using repair templates for accurate repair rather than using intra-homologous regions flanking the DSB in SSA, which results in the trimming of the intervening region [36, 37]. Given the inherent feature of the HR pathway, its precise manner of repair in order to preserve the genetic materials – HDR will be referred to as HR in this context. A recent alternative repair mechanism, called microhomology-mediated end joining (MMEJ), lies between NHEJ and HDR, and relies on microhomologous sequences for error-prone end joining following DNA-end resection [36, 38, 39]. Even though MMEJ is reported to be induced at considerably lower levels than HDR [38], it compensates the absence of NHEJ or HDR that is observed in cancer cells [40, 41].

In the following sections, we provide examples of CRISPR/Cas9-mediated genome editing applying the different DNA repair pathways for engineering purposes (Fig. 2).

4.1 CRISPR/Cas9-mediated gene disruption

In mammalian cells, DSBs generated by CRISPR/Cas9 are preferentially repaired by error-prone NHEJ, which causes insertion/deletion (indel) mutations at target sites (Fig. 2A). Indel formation can lead to frameshift mutations in the coding region of genes, which disrupts their proper translation and results in functional knockout of the target genes. HDR can also be applied to generate gene knockouts when Cas9 and gRNA are introduced with donor repair templates containing premature stop codon or various lengths of bases, which induces frameshift

mutations. Many mutations induced by CRISPR/Cas9 without exogenous donor repair templates are also likely to be associated with MMEJ [42]. In general, NHEJ has been applied due to its mechanistic flexibility, lack of requirement for DNA-end resection and repair templates, and because it is not confined to S and G2 phases during the cell cycle [36, 37]. HR-based gene targeting occurs infrequently in mammalian cells [43], and is typically observed at more variable frequency with programmable nuclease-induced DSBs [15]. Thus, NHEJ-driven approaches have been primarily exploited to achieve gene knockouts in mammalian cells via CRISPR/Cas9 [28, 30, 44]. Our group demonstrated high efficiency of CRISPR/Cas9-mediated disruptions of the C1GALT1-specific chaperone (COSMC) and fucosyltransferase 8 (FUT8) genes in CHO cells with indel frequencies up to 47.3% based on NHEJ [45]. Simple design and preparation of CRISPR/Cas9 enable simultaneous modification of multiple target sequences by introducing multiple gRNAs, referred to as multiplexing [28, 30, 46–48].

While Cas9 normally induces DSBs, single strand breaks (SSBs) can be created by Cas9 nickases, which only have one active nuclease domain. SSBs are repaired by the high-fidelity HDR or base excision repair (BER) pathway to decrease the frequency of unwanted indels generated at off-target sites. A double nicking strategy consisting of Cas9 nickases with a pair of offset gRNAs can produce SSBs on different DNA strands that result in production of composite DSBs and efficient indel mutations at on-target sites (Fig. 2B). Because off-target nicking is repaired precisely, the double nicking approach increases the targeting specificity of Cas9-mediated knockout significantly [28, 34, 35, 49].

Simultaneous introduction of DSBs using paired CRISPR/Cas9 cleavage in different genomic loci can give rise to deletions of the intervening chromosomal segment from kilobase-size [49, 50] to megabase-size (Fig. 2C) [50, 51] or chromosomal rearrangements including interchromosomal translocations and intrachromosomal inversions depending on target loci (Fig. 2C and 2D) [52]. The feasibility of chromosome-scale genome engineering facilitates not only inactivation of entire gene clusters but also investigation of a structural change of chromosomes and its effect. In particular, large deletions of genomic regions allow characterization of genes and genetic elements such as promoters, enhancers and other regulatory elements.

4.2 CRISPR/Cas9-mediated site-specific gene integration

Conventional gene targeting for defined sequence insertion or replacement by HR requires several kilobase (kb) of homologous sequences and positive/negative selection markers on targeting vectors [53]. However, the success rate is markedly low at a frequency of one per 10^5 to 10^7
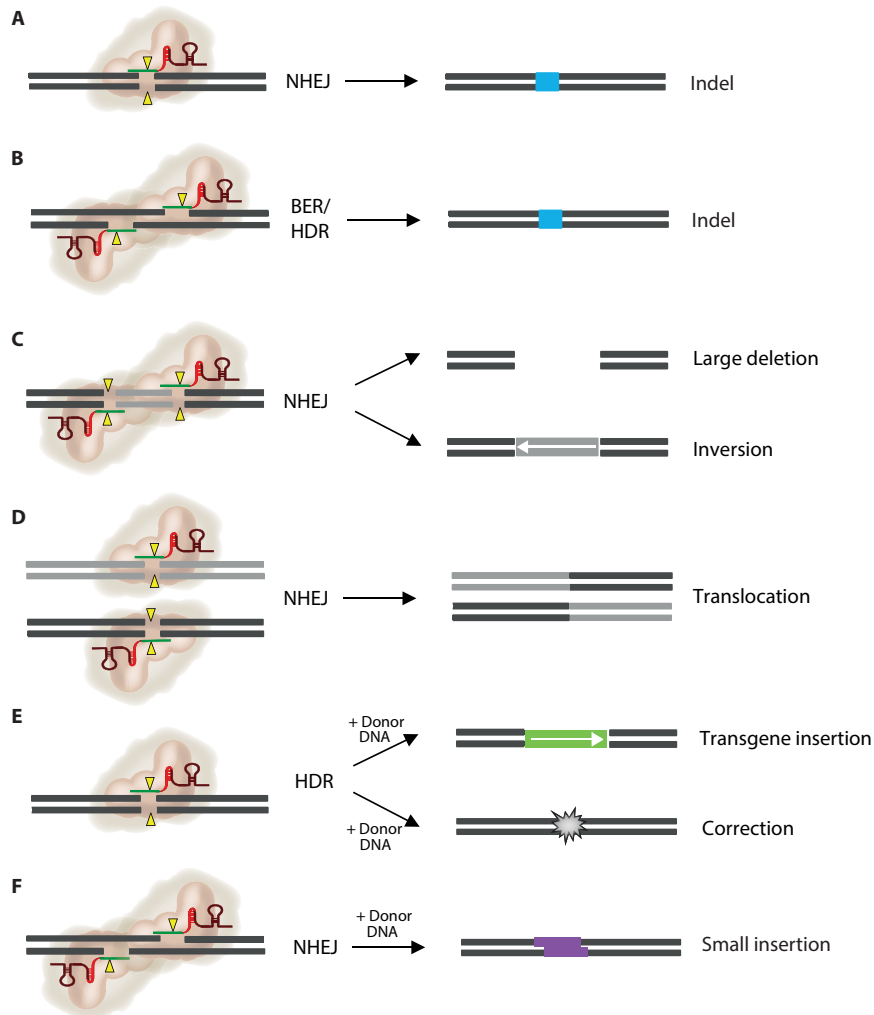
- Chapter 1 -

Biotechnol. J. 2015, 10, 979–994

Biotechnology
Journal
www.biotechnology-journal.com

BIO tec
visions
www.biotecvisions.com

Figure 2. Types of genome editing application using CRISPR/Cas9. (**A**) CRISPR/Cas9-mediated generation of a single DNA DSB can result in insertion and deletion (indel) mutations if the break is repaired by the non-homologous end joining (NHEJ). (**B**) Repair of two simultaneously introduced DNA single strand breaks (SSBs) by nicking Cas9 can result in indel mutations if repaired by base excision repair (BER) or homology directed repair (HDR). (**C**) Repair of two simultaneously introduced DNA DSBs can result in large deletion or inversion of the entire intervening stretch of DNA if repaired by NHEJ. (**D**) Repair of DNA DSBs on two different chromosomes can result in translocation. (**E**) Repair of a single DNA DSB using donor DNA with a transgene can result in transgene insertion or gene correction if the donor DNA contains base-pair changes. (**F**) Repair of two simultaneously introduced DNA SSBs by nicking Cas9 using donor DNA can result in small insertion.

treated cells, which hampers the use of HR [54]. The introduction of DSBs can overcome these limitations while increasing the frequency of HR and reducing constraints on the vector design. Donor repair templates in the form of either plasmid DNA or single-stranded oligodeoxynucleotides (ssODNs) can be used to insert a gene of interest into target sites coupled with CRISPR/Cas9 [55]. Homologous sequences, referred to as homology arms, flank the insertion sites, and their size can be less than 1 kb each on plasmid-based donor repair templates. Co-delivery of plasmid donors with CRISPR/Cas9 components has achieved site-specific integration of transgenes encoding reporter proteins and/or antibiotic resistance markers [30, 56] or introduction of restriction sites into a target locus in mammalian cells (Fig. 2E) [28, 55]. Using ssODNs provides an alternative method for short modifications at target sites with high efficiencies and simple synthesis. And 40–50 bp of short flanking homologous sequences on each side of ssODNs can successfully lead

to HDR-based gene replacements [28, 30, 35, 55]. A double nicking-induced HDR can also be leveraged to introduce a restriction site at a target locus with comparable efficiencies obtained with wild-type Cas9 nuclease [35].

HDR-independent knock-in strategies have expanded options for CRISPR/Cas9-mediated gene insertion [35, 57]. Double nicking-mediated generation of defined overhangs at a target genomic site could insert a double-stranded oligodeoxynucleotide (dsODN), containing compatible overhangs, via non-HDR mediated ligation (Fig. 2F) [35]. MMEJ-mediated gene knock-ins can also be applied in human cells, resulting in cassette integration [57].

4.3 CRISPR/Cas9-mediated gene activation and repression

Inactivation of both nuclease domains of Cas9 nuclease results in a nuclease-null Cas9, often termed 'dead' Cas9 (dCas9). Despite loss of nuclease activity, dCas9 retains

target specific DNA binding ability when guided by gRNAs. Targeting of dCas9-gRNA complex to specific sites was shown to interfere with transcriptional elongation, RNA polymerase binding, or transcriptional factor binding. This results in reversible repression of targeted genes, referred to as CRISPR interference (CRISPRi) [58]. dCas9 can also be utilized to regulate endogenous gene expression particularly when dCas9 fuses with effector domains such as a transcriptional activation domain VP64 or repression domain KRAB [34, 59, 60]. Multiplexed recruitment of dCas9-activators results in a synergistic effect of higher levels of gene activation [34, 59], and dCas9-repressors enhance the repressive function of CRISPRi [60]. dCas9 fused to epigenetic enzymes may enable targeted epigenetic changes such as histone modifications or DNA methylation, as described with TALE DNA-binding domains [61].

4.4 Genome scale knockout screens for target identification

For decades, desirable traits in CHO cells have been identified with screens and selection systems, implemented after mutagenesis, drug treatment, or media optimization. This process has enabled the identification of improved cell lines and in a few cases, mutations underlying these traits [62]. The CRISPR/Cas9 system now provides a way to systematize the identification of mutations conferring a desirable trait. Specifically, while most CRISPR/Cas9 applications have targeted individual genes, advances in array-based oligonucleotide library synthesis [63] have recently enabled the development of genome-wide loss-of-function and gene overexpression screens [64–69]. For these screens, thousands of gRNAs are synthesized in parallel on microarrays and introduced into lentiviral transfer plasmids along with Cas9 (Fig. 3A–C). Viral infections can be titrated so that most infected cells receive a single gRNA, thus allowing one to conduct negative and positive screens on a population of cells (Fig. 3D and 3E). Mutants exhibiting the desired traits can be identified by quantifying gRNAs in the population of cells before and after the screen via next-generation sequencing (Fig. 3F and 3G). While these screens are conceptually similar to RNA interference (RNAi)-mediated screening, CRISPR/Cas9 genome-wide loss-of-function screens usually completely inactivate the genes and there are fewer ambiguities arising from residual activity. Furthermore, different gRNAs against the same gene tend to give more consistent phenotypes than different short-hairpin RNAs (shRNAs) in RNAi screens [65]. Thus CRISPR/Cas9 screens have recently been applied to identify genes responsible for cancer drug resistance [65, 69] and mechanisms underlying toxin activity [68, 70]. In theory, these approaches could be employed to identify mutants for any desired trait for which a high-throughput selection system can be implemented.
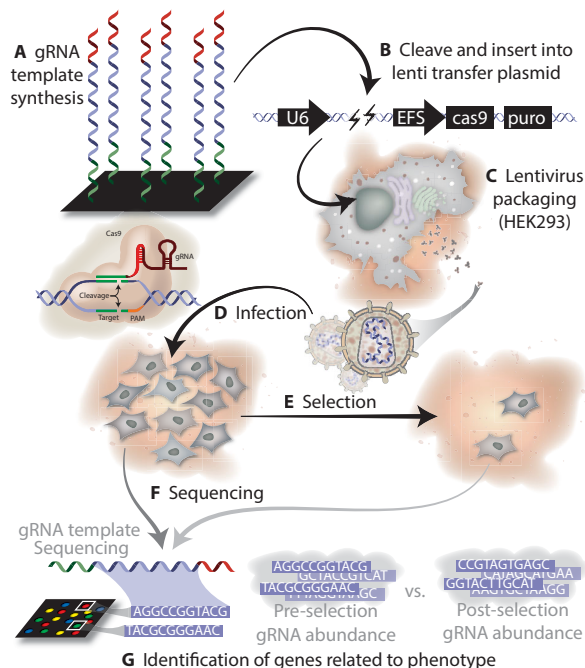
Figure 3. Genome-wide screening using CRISPR/Cas9. Genome-wide knockout, activation, and repression screens can be conducted. (**A**) To do this, gRNA constructs for all targets are synthesized in parallel on microarrays. (**B**) RNA constructs are loaded into a lentiviral transfer plasmid, and (**C**) packaged into lentiviruses. (**D**) The lentiviruses, harboring all possible gRNAs, are used to infect a population of cells. (**E**) The cells are then subjected to a selection pressure. (**F**) gRNAs are sequenced before and after the screen. (**G**) Positive and negative genetic changes can be identified by searching for gRNA constructs that increase or decrease, respectively, in the population over the course of the screen.

5 Experimental design

Genome editing of cells requires several steps including target site selection, primer design, cloning of CRISPR/Cas9 reagents, transfection, cell line generation and analysis. An overview of the experimental steps is given in Fig. 4 and will be described in detail below.

5.1 Target site selection and primer design

For selection of a 20 bp target sequence there are two main considerations: the presence of a 5'-NGG-3' PAM sequence directly downstream of the target sequence for *Streptococcus pyogenes* Cas9 (SpCas9) and the minimization of potential off-target activity. For this purpose, several online target selection tools are available. For target selection in mammalian cells, one must consider the various exons and splice variants for the gene of interest (Fig. 4A). It is recommended to design at least two gRNAs for each target locus and test their target modification efficiency, as gRNAs occasionally may not work for

A Target site selection

B Primer design

C Construction of CRISPR Cas9 reagents

DNA

RNA

Protein

D Transfection/infection

E Single cell isolation and expansion

F Analysis of modifications
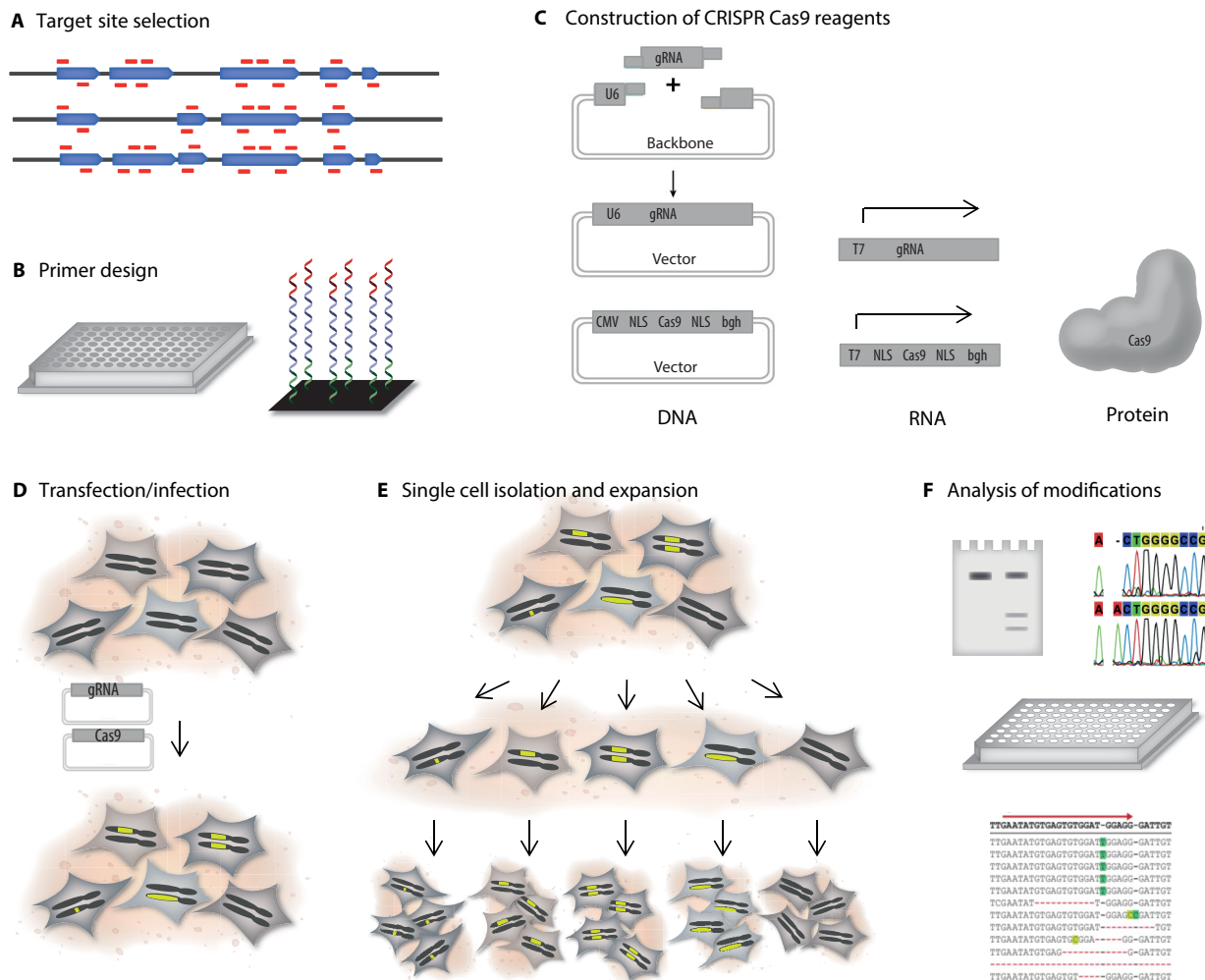
Figure 4. Experimental overview of genome engineering. Genome engineering contains several experimental steps including: (**A**) Selection of the target site that could be in the protein coding regions of a gene of interest. It is important to consider the potential splice variants of the gene of interest. (**B**) Upon target site selection, oligonucleotides for gRNA construction can be designed and ordered in microtiter plates or spotted on arrays for high-throughput cloning. (**C**) The gRNA can be cloned in a simple step with annealed oligonucleotides using ligation-free cloning and be delivered together with Cas9 as DNA. For RNA delivery, expression vectors can be generated to facilitate in vitro transcription of Cas9 and gRNA. For protein delivery, the Cas9 protein can be purified and delivered together with in vitro transcribed gRNA. The U6 polymerase III promoter is often used to drive expression of the gRNA. For Cas9 expression, the cytomegalovirus (CMV) vector is often applied. The nuclear localization signal (NLS) facilitates targeting of Cas9 to the nucleus. Bgh encodes the bovine growth hormone poly adenylation signal. The T7 promoter can be applied for in vitro transcription.(**D**) The Cas9 and gRNA reagents are introduced into cells during transfection or infection to modify the genomes. (**E**) Modified cells are isolated using FACS or limiting dilution, and isolated cells are expanded to create genome engineered monoclonal cell lines.(**F**) Different analytical assays apply preparation and digestion of amplicons followed by gel separation that includes the Surveyor nuclease assay or restriction fragment length polymorphism (RFLP). Sequencing can be applied for high-throughput screening in microtiter plates to obtain information about introduced DNA changes in cell pools or clonal cell lines.

unknown reasons [55]. Online target selection tools assist in identifying potential target sites in a genomic sequence of interest and provide computationally predicted off-target sites for each identified target. Examples of these tools are the CRISPR Design (http://tools.genome-engineering.org) [55], the ZiFiTtargeter (http://zifit.partners.org) [71] and the E-CRISP (http://www.e-crisp.org) [72]. At the time of writing this manuscript, these tools were unable to search for similarities in the CHO genome.

However, for genome engineering of CHO cells, the CRISPy tool (http://staff.biosustain.dtu.dk/laeb/crispy) is available [45]. This online bioinformatics tool provides a precompiled database with identified target sites in the CHO-K1 genome. In CHO-K1 genome there are 2 379 237 Cas9 targets with the format GN19NGG in CDS regions of 23 750 genes (ref-seq assembly GCF_000223135.1). For increased targeting specificity, one may use the D10A nickase mutant of Cas9 (Cas9n) together with a pair of

987

gRNAs, and the CRISPy tool also provides potential partner gRNAs for nicking Cas9 compatible with CHO cells. The selection of gRNAs regarding orientation and spacing is critical and offsets between -4 and 100 bp are recommended [35]. In addition, CRISPy aids in primer design for verifying cleavage efficiency at on-target sites and computationally predicted off-target sites. Upon selection of target sites, oligonucleotides for gRNA synthesis can be designed. It is important to note that the PAM site is required in the target locus but is not part of the 20 bp target sequence of the gRNA. Also, the U6 RNA polymerase III promoter, often used to express the gRNA, prefers a guanine (G) nucleotide as the first base of its transcript. In cases where the target sequence does not contain a G as the 20 bp nucleotide upstream of the PAM, a G can be added to introduce G as the first base of the gRNA transcript [55]. For high-throughput cloning, oligonucleotides can be delivered in microtiter plates or spotted on arrays (Fig. 4B).

5.2 Construction of CRISPR/Cas9 reagents

Depending on the genome engineering task, gRNAs can be delivered as gRNA-expressing plasmids or as PCR amplicons containing an expression cassette. The cloning of gRNA expression plasmids can be done rapidly by applying a single cloning step with a pair of partially complementary oligonucleotides (Fig. 4C). The generation of PCR amplicons is even faster and is well suited for large-scale testing of gRNAs or generation of large knockout libraries. Depending on the purpose of the genome editing event, the gRNA and Cas9 can be delivered as RNA upon in vitro transcription or even as Cas9 purified protein together with in vitro transcribed gRNA (Fig. 4C) [73]. An overview of delivery methods applied in mammalian genome editing including references is given in Table 1. The most used promoter to drive expression of gRNAs in mammalian cells is the RNA polymerase III U6 promoter. Common promoters for driving the expression of Cas9 in mammalian cells are the cytomegalovirus (CMV) promoter and the elongation factor 1-α (EF1-α) promoter. For RNA delivery of Cas9 and gRNA, the T7 promoter is often applied for in vitro transcription. Fusing 2A-GFP or antibiotic resistance genes to Cas9 can facilitate screening or selection of transfected cells with Cas9 expression. Isolation of cell populations with increasingly higher Cas9 expression levels can facilitate isolation of cells with increased genome editing rates [74]. Increased efficiency of genome editing may also result in a trade-off effect of increasing non-targeted cleavage and mutagenesis effects.

5.3 DNA repair templates

Generation of DNA repair templates is often based on plasmid-based vectors with homology arms of varying length. Often, the homology arms are longer than 500 bp. The ssODNs require short homology arms of minimum 40 bp and can therefore be prepared without cloning [75]. They can be oriented in either the sense or the antisense direction relative to the target locus.

5.4 Engineered cell line generation

The CRISPR/Cas9 reagents aimed at modifying the genome can be introduced by traditional transfection methods including chemical transfection, electroporation, nucleofection and by viral delivery depending on the cell type and application (Fig. 4D, Table 1). The transfection process generally results in a heterogeneous mixture of genomic modifications in the targeted cell pool, and efficiencies are highly dependent on transfection efficiency. Several assays are available to quantify and characterize indels and modifications introduced. Isolation of modified cells from the mixed population of wild type and modified cells are often desired and can be performed by fluorescence-activated cell sorting (FACS) or limiting dilutions (Fig. 4E). It is important to note that different cell types may respond differently to single cell isolation. The isolated cells can then be expanded to establish new clonal cell lines. The assays described below can also be applied to characterize the generated clones to identify clones with desired genomic modifications.

5.5 Analysis of modifications

Different assays are available to monitor the heterogeneous mixture of mutations introduced by NHEJ or HDR upon CRISPR/Cas9-mediated genome editing (Fig. 4F). For detection of indels, two mutation screening assays use the Surveyor nuclease the T7 endonuclease I, which are two mismatch-recognizing nucleases. The enzymes recognize heteroduplex DNA in an amplicon pool after reannealing with wild type sequences and facilitate selective digestion of mismatched duplex amplicons. This facilitates estimation of the mutated fraction of an amplicon pool [76, 77]. Some mutations can be determined using the restriction fragment length polymorphism (RFLP) method if a restriction site is introduced during the genome editing process. With this method, the amplicons are digested by selected restriction enzymes and separated by size before analysis. The Surveyor nuclease and RFLP assay has been applied in clone verification in several genome editing studies including CRISPR/Cas9-mediated generation of mutated mice [78]. Genomic changes in a modified pool of cells or clones can also be analyzed by Sanger sequencing. When sequencing amplicons with a heterogenous mixture of mutations, subcloning can be used to isolate the different amplicons before sequencing. For high-throughput identification of genome modifications, deep sequencing can be applied for direct sequencing of amplicons without subcloning.

Biotechnology Journal
www.biotechnology-journal.com

- Chapter 1 -

Biotechnol. J. 2015, 10, 979–994

BIO tec visions
www.biotecvisions.com

Combined with barcoding, amplicons from hundreds of different cell pools or clones can be sequenced simultaneously using a benchtop sequencer like the Miseq (Illumina Inc.).

6 Off-target analysis

Since Cas9 recognizes a target by simple Watson-Crick base pairing rules, it is presumed that the off-target activity results as the nuclease binds a sequence with some level of homology to the target sequence and subsequently induces a DSB. Off-target activity can lead to alterations that have more or less serious consequences for the cells. It can disrupt the normal cell function [79], and cause unwanted chromosomal rearrangements such as deletions, inversions and translocations [80–83].

A wide variety of assays exist to measure the extent to which a given guide sequence exhibits off-target activity. The most straightforward approach to analyze specific off-target activity is based on an initial prediction of sites in the genome that might be prone to off-target activity. Available prediction tools incorporate combinations of biological principles and factors that drive the Cas9-DNA interactions in their prediction algorithms. These could include information such as that the PAM proximal region is crucial for target recognition and cleavage by Cas9 [25, 28] and tolerates a single base mismatch, that three to five mismatches in the PAM distal region can be tolerated [34, 84, 85], and that a 5 nt seed region next to PAM is required for Cas9 binding [86]. The off-target prediction is usually followed by either gel-based assays such as T7 endonuclease I or Sanger- or deep-sequencing to look for off-target effects at the predicted sites, of which the sequencing-based approaches are more sensitive [49]. However, contrasting results from this approach have been reported [44, 49, 87].

Methods for genome-wide identification of off-target activity exist. These methods are both based on Cas9 binding sites and Cas9 cleavage. Chromatin immunoprecipitation-sequencing (ChIP-seq) can assay Cas9 binding sites. The nuclease is pulled down and the bound DNA fragments are sequenced and mapped to their native location in the genome [86, 88]. The Cas9 binding site methods often lead to identification of false-positives, as Cas9 does not always cleave the sites bound by dCas9. Additional methods to assay off-target Cas9 cleavage include integrase-deficient lentivirus (IDLV) at the DSBs [89] and in vitro selection based on a DNA substrate library [90]. The IDLV capture is much lower than the actual mutation frequency; thus, many genuine off-target sites are not detected [91]. For the in vitro selection, many of the identified off-target sites do not exist in the genome [90]. Newer, more sensitive genome-wide methods are high-throughput, genome-wide translocation sequencing (HTGTS) [92], genome-wide, unbiased identification

of DSBs enabled by sequencing (GUIDE-seq) [93] and in vitro Cas9-digested whole-genome sequencing (Digenome-seq) [91]. Digenome-seq is the only method not limited by chromatin accessibility. Even with sensitive genome-wide identification of off-target activity, it is still demanding to determine with certainty which mutations are spontaneous and which are due to off-target activity of the nuclease.

A good solution would be to minimize the effects of off-target activity, and several strategies have been tested. Adjusting the concentration as well as the ratio of Cas9 to gRNA, changes the amount of off-target cleavage [85]. Using paired nickases, as described earlier, does not promote unwanted translocations or double the specificity of Cas9-mediated mutagenesis [34, 35]. The use of 3´ truncated gRNAs or gRNAs with two extra guanine nucleotides at the 5´ end yields better on-/off-target ratios [49, 90, 91], but at lower absolute on-target efficiencies. Using 5´ truncated gRNAs shows decreased mutagenic effects and enhanced sensitivity to single or double mismatches at the DNA-RNA interface [94]. Fusion of catalytically dCas9 to FokI nuclease improves the specificity of genome editing [95, 96], and thereby decreases the off-target activity. Selection of unique target sites that lack any homologous sequence elsewhere in the genome can also minimize off-target effects [49].

7 Past, present and future of CHO cell factory engineering via CRISPR/Cas9-mediated genome editing

7.1 Past of CHO cell engineering

CHO cell engineering efforts have been forced to rely on traditional methods because of the lack of genomic sequence data. These methods included overexpression of endogenous CHO or exogenous genes. RNA interference mediated regulation of gene expression has also been used to increase the time integral of viable cell concentration and/or specific productivity (q) [97, 98]. Targeted genome engineering for CHO cells has primarily focused on knockout, which has been conducted by HR and ZFNs [99–102]. Both alleles of the a1,6-fucosyltransferase (*FUT8*) were disrupted by HR in CHO-DG44 cells following screening of approximately 120 000 transfectants [99]. Application of the ZFNs significantly increased the targeting efficiencies of gene disruptions up to >1%, which allowed isolation of knockout clones from one 96-well plate of single-cell derived clones [100–102]. Sequential rounds of ZFN-mediated gene disruptions achieved accumulation of knockouts on the background of previously generated knockout clones, enabling the generation of double and triple knockout CHO cell lines [101].

Site-specific integration of transgenes encoding recombinant proteins has also been performed to circumvent random integration followed by extensive clone screening for high producers. Site-specific recombinase-based systems, including Cre/loxP and Flp/FRT, enabled insertion of a gene of interest into the predefined specific target sites where a high-level of transcription and amplifiable capacity were warranted [103, 104]. The PhiC31/R4 integrase system was also successfully applied to place a gene of interest into transcriptionally active pseudo attP sites [105]. Moreover, homology-independent target-specific integration of short (<100 bp) oligonucleotides [106] or linear transgenes [107] has been driven by ZFNs in CHO cells. Of note, Cristea et al. achieved targeted integration and expression of large transgenes encoding antibody at *FUT8* locus of CHO cells with the use of ZFNs or TALENs following physiological selection for *FUT8* knockout [107].

7.2 State of the art in CHO cell engineering – toward facile genome engineering

In 2014 and 2015, we demonstrated successful application of CRISPR/Cas9 for generating highly efficient gene disruptions and gene insertions in CHO cells [45, 56]. Using Cas9 proteins from *Streptococcus pyogenes*, 18 gRNAs were used to target four genes. These resulted in relatively high indel frequencies in the range of 7.6 to 47.3%, regardless of CHO host cell line (CHO-K1 or CHO-S), transfection method (nucleofection or chemical transfection), culture mode (adherent or suspension), or locus (*C1GALT1C1*, *FUT8*, *MGAT1* or *LDHA*) [45, 56]. Most indels were single base pair insertions that resulted in frame shift mutations. Current protocols are now enabling the simultaneous disruption of up to three genes with CRISPR/Cas9-based multiplexing. Simultaneous introduction of multiple gRNAs and enrichment for transfectants is yielding 58.8% of all clones harboring multiple knockouts [48]. HDR-mediated integration of large gene expression cassettes into desired loci was achieved in CHO cells by combining the CRISPR/Cas9 system with donor plasmids harboring short homology arms [56]. A targeting efficiency of between 7.4 and 27.8% was obtained depending on the target locus and CRISPR/Cas9 activity. Since HDR-based transgenesis is considered to be a rare event in CHO cells due to their intrinsically low levels of HDR [107], this result makes it worthwhile to revisit the HDR-based targeted integration. Recent genome engineering efforts via CRISPR/Cas9 conducted in most common nonmammalian systems including *Escherichia coli* [58, 108, 109] and *Saccharomyces cerevisiae* [110, 111] reiterate the applicability and necessity of this state of the art technology toward accelerated cell factory construction.

7.3 Future of CHO genome engineering – toward rational design of mammalian cell factories

The CRISPR/Cas9 technology will be a valuable tool as we aim to develop the next generation of CHO cell factories. Advances in three areas will help this goal come to fruition, namely, further refinements of the CRISPR/Cas9 technology, increased understanding of the CHO genome, and detailed characterization of cellular pathways.

First, future efforts will continue to improve the precision and targeting efficiency of CRISPR/Cas9. The present CRISPR/Cas9 has a high genome editing success rate, and nearly up to 100% of multigene mutation frequencies were achieved in *Escherichia coli* [109] and *Saccharomyces cerevisiae* [111]. However, it still has a broad range of genome editing activities (2.3 to 79%) obtained in mammalian cells depending on the cell type, delivery method and target locus [16]. Given the current status of absence of reliable prediction rule for mutation rates, increasing the mutation rates must be pursued to accelerate genome editing in CHO cells. For example, rational design of highly active gRNAs [112] and engineered Cas9 [113] will enhance the functionality of the CRISPR/Cas9 system. The improved accessibility of Cas9 and gRNAs by efficient delivery approaches through expression vector engineering or mRNA/protein delivery of CRISPR/Cas9 components can also be easily implemented. Instead, an efficient enrichment of cells with high Cas9 expression levels [74] or genome editing events [114] could provide another simple option for acquisition of modified cells. In contrast with an acceptable frequency of gene disruptions, the rate of HDR-mediated integration of large transgenes remained relatively low, as a drug or phenotypic selection process was required for the generation of targeted integrants [56]. In addition to the aforementioned general approaches, methods to improve the HDR:NHEJ ratio are needed because high nuclease activity does not necessarily ensure high rates of HDR. Lin et al. recently reported interesting results showing that chemically induced cell cycle synchronization and subsequent delivery of pre-assembled Cas9-guide RNA ribonucleoprotein complexes achieved HDR-mediated genome editing at levels up to 38% in human cells [115]. The reapplication of dominant NHEJ-driven knock-in approaches [107] together with CRISPR/Cas9 will also broaden the scope of knock-in strategies for rCHO cell construction.

While improvements in the targeting efficiency are being achieved, undesired off-target mutations can still arise. In the area of cell culture engineering, unexpected mutations may not be critical compared with therapeutic applications, particularly when well-characterized genes or sites are targeted. However, the potentially confounding effects of off-target mutations may generate false-positive effects, thus obstructing evaluation of novel sites. We anticipate that better understanding of the CRISPR/Cas9 system at structural and biochemical levels could over-

come this issue. This will require improved approaches to reduce off-target mutations of Cas9 (Section 6) and refinement of bioinformatics tools for design of gRNA target sites with minimized off-target effects (Section 5.1).

Second, the establishment of improved Chinese hamster and CHO cell line reference genomes will improve genome editing efforts and subsequent in-depth analyses of outcomes. The well-defined reference genomes will allow us to analyze genome-wide off-target effects of Cas9 nucleases as well as structural variation analysis of CHO cells (e.g., indels and chromosomal translocations) [116] upon exposure to CRISPR/Cas9. Improvements in reference genome sequences will also decrease false-negative rates in genome-scale knockout, activation, or repression screening, thus helping to identify more candidate genes that are essential for desired phenotypes.

Third, computational models of metabolism [117], glycosylation [118], signaling [119], DNA damage response [120] and transcriptional regulation [121] will aid in the design of novel expression hosts. These models, when combined with 'omics data (e.g. RNA-Seq or epigenome) [122], will help to identify pathways that when engineered will yield desired traits ranging from enhanced protein secretion to the identification of ideal transgene insertions sites ensuring high and stable expression of genes. Given the homogeneity of transgene expression levels observed in targeted integrants [56], site-specific integration of a gene of interest into any desired target sites could lead to construction of production CHO cell lines assuring sustainable and high expression levels in a short time.

8 Concluding remarks

The advent of genomic sequences and genome editing technologies makes genome-scale science for CHO cells inevitable. Apart from the pros and cons of individual programmed nucleases, the cost-effective and simple CRISPR/Cas9 system definitely offers several advantages that industry would welcome. Together with improved reference genomes for Chinese hamster and CHO cells, the highly efficient CRISPR/Cas9 system is now ready to address long-held questions inherent to CHO cells. Multiplexed genome engineering including gene knockouts, gene knock-ins and controlled expression of target genes will accelerate understanding of the molecular basis driving traits that make CHO cells a predominant mammalian cell factory. This could ultimately lead to engineering of the next generation of CHO cell factories.
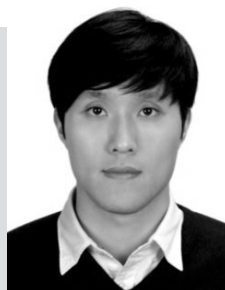
Jae Seong Lee is currently a postdoctoral researcher in the Novo Nordisk Foundation Center for Biosustainability (CFB) at the Technical University of Denmark (2013-present). Dr. Lee earned his B.S and Ph.D. in Biological Sciences from Korea Advanced Institute of Science and Technology (KAIST) under the mentorship of Prof. Gyun Min Lee. His research has focused on the genetic and metabolic engineering of CHO cells for high level production of recombinant proteins. His current research interests are development of efficient CHO genome editing technologies and rational design of recombinant CHO cells.

Helene Faustrup Kildegaard is currently co-Principal Investigator at the Novo Nordisk Foundation Center for Biosustainability (CFB) at the Technical University of Denmark. Dr. Kildegaard earned her Ph.D. in molecular cancer biology from the Danish Cancer Society/Technical University of Denmark. She was then awarded a postdoctoral research grant on therapeutic protein production in CHO cells from the Lundbeck Foundation and the Technical University of Denmark. Her current research focuses on applying system biology and developing efficient genome editing tools to engineer CHO cell factories for increased production and improved quality of therapeutic proteins.

9 References

[1] Walsh, G., Biopharmaceutical benchmarks 2014. *Nat. Biotechnol.* 2014, *32*, 992–1000.

[2] Jayapal, K. P., Wlaschin, K. F., Hu, W. S., Yap, M. G., Recombinant protein therapeutics from CHO cells – 20 years and counting. *Chem. Eng. Prog.* 2007, *103*, 40–47.

[3] Huang, Y. M., Hu, W., Rustandi, E., Chang, K. et al., Maximizing productivity of CHO cell-based fed-batch culture using chemically defined media conditions and typical manufacturing equipment. *Biotechnol. Prog.* 2010, *26*, 1400–1410.

[4] Hefzi, H., Lewis, N. E., From random mutagenesis to systems biology in metabolic engineering of mammalian cells. *Pharm. Bioprocess.* 2014, *2*, 355–358.

[5] Xu, X., Nagarajan, H., Lewis, N. E., Pan, S. et al., The genomic sequence of the Chinese hamster ovary (CHO)-K1 cell line. *Nat. Biotechnol.* 2011, *29*, 735–741.

[6] Lewis, N. E., Liu, X., Li, Y., Nagarajan, H. et al., Genomic landscapes of Chinese hamster ovary cell lines as revealed by the *Cricetulus griseus* draft genome. *Nat. Biotechnol.* 2013, *31*, 759–765.

[7] Brinkrolf, K., Rupp, O., Laux, H., Kollin, F. et al., Chinese hamster genome sequenced from sorted chromosomes. *Nat. Biotechnol.* 2013, *31*, 694–695.

Biotechnology Journal
www.biotechnology-journal.com

- Chapter 1 -

Biotechnol. J. 2015, 10, 979–994

BIOtec visions
www.biotecvisions.com

[8] Hammond, S., Swanberg, J. C., Kaplarevic, M., Lee, K. H., Genomic sequencing and analysis of a Chinese hamster ovary cell line using Illumina sequencing technology. *BMC Genomics* 2011, *12*, 67.

[9] Hammond, S., Kaplarevic, M., Borth, N., Betenbaugh, M. J., Lee, K. H., Chinese hamster genome database: An online resource for the CHO community at www.CHOgenome.org. *Biotechnol. Bioeng.* 2012, *109*, 1353–1356.

[10] Kildegaard, H. F., Baycin-Hizal, D., Lewis, N. E., Betenbaugh, M. J., The emerging CHO systems biology era: Harnessing the 'omics revolution for biotechnology. *Curr. Opin. Biotechnol.* 2013, *24*, 1102–1107.

[11] Kaas, C. S., Fan, Y., Weilguny, D., Kristensen, C., Kildegaard, H. F., Andersen, M. R., Toward genome-scale models of the Chinese hamster ovary cells: Incentives, status and perspectives. *Pharm. Bioprocess.* 2014, *2*, 437–448.

[12] Gaj, T., Gersbach, C. A., Barbas III, C. F., ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends Biotechnol.* 2013, *31*, 397–405.

[13] Hsu, P. D., Lander, E. S., Zhang, F., Development and applications of CRISPR-Cas9 for genome engineering. *Cell* 2014, *157*, 1262–1278.

[14] Doudna, J. A., Charpentier, E., Genome editing. The new frontier of genome engineering with CRISPR-Cas9. *Science* 2014, *346*, 1258096.

[15] Sander, J. D., Joung, J. K., CRISPR-Cas systems for editing, regulating and targeting genomes. *Nat. Biotechnol.* 2014, *32*, 347–355.

[16] Kim, H., Kim, J. S., A guide to genome engineering with programmable nucleases. *Nat. Rev. Genet.* 2014, *15*, 321–334.

[17] Ain, Q. U., Chung, J. Y., Kim, Y. Current and future delivery systems for engineered nucleases: ZFN, TALEN and RGEN, *J. Controlled Release* 2015. In Print. DOI: 10.1016/j.jconrel.2014.12.036.

[18] Barrangou, R., Fremaux, C., Deveau, H., Richards, M. et al., CRISPR provides acquired resistance against viruses in prokaryotes. *Science* 2007, *315*, 1709–1712.

[19] Mojica, F. J., Díez-Villaseñor, C., García-Martínez, J., Soria, E., Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J. Mol. Evol.* 2005, *60*, 174–182.

[20] Makarova, K. S., Aravind, L., Wolf, Y. I., Koonin, E. V., Unification of Cas proteins families and a simple scenario for the origin and evolution of CRISPR Cas systems, *Biology Direct,* 2011, *6*, 38.

[21] Brouns, S. J., Jore, M. M., Lundgren, M., Westra, E. R. et al., Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* 2008, *321*, 960–964.

[22] Deltcheva, E., Chylinski, K., Sharma, C. M., Gonzales, K. et al., CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature* 2011, *471*, 602–607.

[23] Garneau, J. E., Dupuis, M. È., Villion, M., Romero, D. A. et al., The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* 2010, *468*, 67–71.

[24] Gasiunas, G., Barrangou, R., Horvath, P., Siksnys, V., Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc. Natl. Acad. Sci. USA* 2012, *109*, E2579–E2586.

[25] Jinek, M., Chylinski, K., Fonfara, I., Hauer, M. et al., A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 2012, *337*, 816–821.

[26] Marraffini, L. A., Sontheimer, E. J., CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. *Science* 2008, *322*, 1843–1845.

[27] Sapranauskas, R., Gasiunas, G., Fremaux, C., Barrangou, R. et al., The *Streptococcus thermophilus* CRISPR/Cas system provides immunity in *Escherichia coli. Nucleic Acids Res.* 2011, *39*, 9275–9282.

[28] Cong, L., Ran, F. A., Cox, D., Lin, S. et al., Multiplex genome engineering using CRISPR/Cas systems. *Science* 2013, *339*, 819–823.

[29] Jinek, M., East, A., Cheng, A., Lin, S. et al., RNA-programmed genome editing in human cells. *Elife* 2013, *2*, e00471.

[30] Mali, P., Yang, L., Esvelt, K. M., Aach, J. et al., RNA-guided human genome engineering via Cas9. *Science* 2013, *339*, 823–826.

[31] Zhang, Y., Heidrich, N., Ampattu, B. J., Gunderson, C. W. et al., Processing-independent CRISPR RNAs limit natural transformation in *Neisseria meningitidis. Mol. Cell* 2013, *50*, 488–503.

[32] Sternberg, S. H., Redding, S., Jinek, M., Greene, E. C., Doudna, J. A., DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature* 2014, *507*, 62–67.

[33] Jinek, M., Jiang, F., Taylor, D. W., Sternberg, S. H. et al., Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. *Science* 2014, *343*, 1247997.

[34] Mali, P., Aach, J., Stranges, P. B., Esvelt, K. M. et al., CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nat. Biotechnol.* 2013, *31*, 833–838.

[35] Ran, F. A., Hsu, P. D., Lin, C. Y., Gootenberg, J. S. et al., Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell* 2013, *154*, 1380–1389.

[36] Huertas, P., DNA resection in eukaryotes: Deciding how to fix the break. *Nat. Struct. Mol. Biol.* 2010, *17*, 11–16.

[37] Lieber, M. R., The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. *Annu. Rev. Biochem.* 2010, *79*, 181–211.

[38] Truong, L. N., Li, Y., Shi, L. Z., Hwang, P. Y. et al., Microhomology-mediated end joining and homologous recombination share the initial end resection step to repair DNA double-strand breaks in mammalian cells. *Proc. Natl. Acad. Sci. U.S.A.* 2013, *110*, 7720–7725.

[39] Kent, T., Chandramouly, G., McDevitt, S. M., Ozdemir, A. Y., Pomerantz, R. T., Mechanism of microhomology-mediated end-joining promoted by human DNA polymerase θ. *Nat. Struct. Mol. Biol.* 2015, *22*, 230–237.

[40] Ceccaldi, R., Liu, J. C., Amunugama, R., Hajdu, I. et al., Homologous-recombination-deficient tumours are dependent on Polθ-mediated repair. *Nature* 2015, *518*, 258–262.

[41] Mateos-Gomez, P. A., Gong, F., Nair, N., Miller, K. M. et al., Mammalian polymerase θ promotes alternative NHEJ and suppresses recombination. *Nature* 2015, *518*, 254–257.

[42] Bae, S., Kweon, J., Kim, H. S., Kim, J. S., Microhomology-based choice of Cas9 nuclease target sites. *Nat. Methods* 2014, *11*, 705–706.

[43] Sedivy, J. M., Sharp, P. A., Positive genetic selection for gene disruption in mammalian cells by homologous recombination. *Proc. Natl. Acad. Sci. U.S.A.* 1989, *86*, 227–231.

[44] Cho, S. W., Kim, S., Kim, J. M., Kim, J. S., Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. *Nat. Biotechnol.* 2013, *31*, 230–232.

[45] Ronda, C., Pedersen, L. E., Hansen, H. G., Kallehauge, T. B. et al., Accelerating genome editing in CHO cells using CRISPR Cas9 and CRISPy, a web-based target finding tool. *Biotechnol. Bioeng.* 2014, *111*, 1604–1616.

[46] Kabadi, A. M., Ousterout, D. G., Hilton, I. B., Gersbach, C. A., Multiplex CRISPR/Cas9-based genome engineering from a single lentiviral vector. *Nucleic Acids Res.* 2014, *42*, e147.

[47] Sakuma, T., Nishikawa, A., Kume, S., Chayama, K., Yamamoto, T., Multiplex genome engineering in human cells using all-in-one CRISPR/Cas9 vector system. *Sci. Rep.* 2014, *4*, 5400.

[48] Grav, L. M., Lee, J. S., Gerling, S., Kallehauge, T. B. et al., One-step generation of triple knockout CHO cell lines using CRISPR/Cas9 and fluorescent enrichment. *Biotechnol. J.* 2015. DOI: 10.1002/biot.201500027.

- Chapter 1 -

Biotechnol. J. 2015, 10, 979–994

Biotechnology Journal
www.biotechnology-journal.com

BIO tec visions
www.biotecvisions.com

[49] Cho, S. W., Kim, S., Kim, Y., Kweon, J. et al., Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. *Genome Res.* 2014, *24*, 132–141.

[50] Canver, M. C., Bauer, D. E., Dass, A., Yien, Y. Y. et al., Characterization of genomic deletion efficiency mediated by clustered regularly interspaced palindromic repeats (CRISPR)/Cas9 nuclease system in mammalian cells. *J. Biol. Chem.* 2014, *289*, 21312–21324.

[51] Essletzbichler, P., Konopka, T., Santoro, F., Chen, D. et al., Megabase-scale deletion using CRISPR/Cas9 to generate a fully haploid human cell line. *Genome Res.* 2014, *24*, 2059–2065.

[52] Choi, P. S., Meyerson, M., Targeted genomic rearrangements using CRISPR/Cas technology. *Nat. Commun.* 2014, *5*, 3728.

[53] Sorrell, D. A., Kolb, A. F., Targeted modification of mammalian genomes. *Biotechnol. Adv.* 2005, *23*, 431–469.

[54] Vasquez, K. M., Marburger, K., Intody, Z., Wilson, J. H., Manipulating the mammalian genome by homologous recombination. *Proc. Natl. Acad. Sci. U.S.A.* 2001, *98*, 8403–8410.

[55] Ran, F. A., Hsu, P. D., Wright, J., Agarwala, V. et al., Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.* 2013, *8*, 2281–2308.

[56] 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*, 8572.

[57] Nakade, S., Tsubota, T., Sakane, Y., Kume, S. et al., Microhomology-mediated end-joining-dependent integration of donor DNA in cells and animals using TALENs and CRISPR/Cas9. *Nat. Commun.* 2014, *5*, 5560.

[58] Qi, L. S., Larson, M. H., Gilbert, L. A., Doudna, J. A. et al., Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* 2013, *152*, 1173–1183.

[59] Maeder, M. L., Linder, S. J., Cascio, V. M., Fu, Y. et al., CRISPR RNA-guided activation of endogenous human genes. *Nat. Methods* 2013, *10*, 977–979.

[60] Gilbert, L. A., Larson, M. H., Morsut, L., Liu, Z. et al., CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell* 2013, *154*, 442–451.

[61] Konermann, S., Brigham, M. D., Trevino, A. E., Hsu, P. D. et al., Optical control of mammalian endogenous transcription and epigenetic states. *Nature* 2013, *500*, 472–476.

[62] Urlaub, G., Käs, E., Carothers, A. M., Chasin, L. A., Deletion of the diploid dihydrofolate reductase locus from cultured mammalian cells. *Cell* 1983, *33*, 405–412.

[63] Kosuri, S., Church, G. M., Large-scale de novo DNA synthesis: Technologies and applications. *Nat. Methods* 2014, *11*, 499–507.

[64] Wang, T., Wei, J. J., Sabatini, D. M., Lander, E. S., Genetic screens in human cells using the CRISPR-Cas9 system. *Science* 2014, *343*, 80–84.

[65] Shalem, O., Sanjana, N. E., Hartenian, E., Shi, X. et al., Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science* 2014, *343*, 84–87.

[66] Koike-Yusa, H., Li, Y., Tan, E. P., Velasco-Herrera, M. D. C., Yusa, K., Genome-wide recessive genetic screening in mammalian cells with a lentiviral CRISPR-guide RNA library. *Nat. Biotechnol.* 2014, *32*, 267–273.

[67] Sanjana, N. E., Shalem, O., Zhang, F., Improved vectors and genome-wide libraries for CRISPR screening. *Nat. Methods* 2014, *11*, 783–784.

[68] Gilbert, L. A., Horlbeck, M. A., Adamson, B., Villalta, J. E. et al., Genome-scale CRISPR-mediated control of gene repression and activation. *Cell* 2014, *159*, 647–661.

[69] Konermann, S., Brigham, M. D., Trevino, A. E., Joung, J. et al., Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature* 2015, *517*, 583–588.

[70] Zhou, Y., Zhu, S., Cai, C., Yuan, P. et al., High-throughput screening of a CRISPR/Cas9 library for functional genomics in human cells. *Nature* 2014, *509*, 487–491.

[71] Hwang, W. Y., Fu, Y., Reyon, D., Maeder, M. L. et al., Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nat. Biotechnol.* 2013, *31*, 227–229.

[72] Heigwer, F., Kerr, G., Boutros, M., E-CRISP: fast CRISPR target site identification. *Nat. Methods* 2014, *11*, 122–123.

[73] Zuris, J. A., Thompson, D. B., Shu, Y., Guilinger, J. P. et al., Cationic lipid-mediated delivery of proteins enables efficient protein-based genome editing in vitro and in vivo. *Nat. Biotechnol.* 2015, *33*, 73–80.

[74] Duda, K., Lonowski, L. A., Kofoed-Nielsen, M., Ibarra, A. et al., High-efficiency genome editing via 2A-coupled co-expression of fluorescent proteins and zinc finger nucleases or CRISPR/Cas9 nickase pairs. *Nucleic Acids Res.* 2014, *42*, e84.

[75] Chen, F., Pruett-Miller, S. M., Huang, Y., Gjoka, M. et al., High-frequency genome editing using ssDNA oligonucleotides with zinc-finger nucleases. *Nat. Methods* 2011, *8*, 753–755.

[76] Guschin, D. Y., Waite, A. J., Katibah, G. E., Miller, J. C. et al., A rapid and general assay for monitoring endogenous gene modification. *Methods Mol. Biol.* 2010, *649*, 247–256.

[77] Mashal, R. D., Koontz, J., Sklar, J., Detection of mutations by cleavage of DNA heteroduplexes with bacteriophage resolvases. *Nat. Genet.* 1995, *9*, 177–183.

[78] Wang, H., Yang, H., Shivalila, C. S., Dawlaty, M. M. et al., One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. *Cell* 2013, *153*, 910–918.

[79] Kim, H. J., Lee, H. J., Kim, H., Cho, S. W., Kim, J. S., Targeted genome editing in human cells with zinc finger nucleases constructed via modular assembly. *Genome Res.* 2009, *19*, 1279–1288.

[80] Brunet, E., Simsek, D., Tomishima, M., DeKelver, R. et al., Chromosomal translocations induced at specified loci in human stem cells. *Proc. Natl. Acad. Sci. U.S.A.* 2009, *106*, 10602–10625.

[81] Lee, H. J., Kim, E., Kim, J. S., Targeted chromosomal deletions in human cells using zinc finger nucleases. *Genome Res.* 2010, *20*, 81–89.

[82] Lee, H. J., Kweon, J., Kim, E., Kim, S., Kim, J. S., Targeted chromosomal duplications and inversions in the human genome using zinc finger nucleases. *Genome Res.* 2012, *22*, 539–548.

[83] Kim, Y., Kweon, J., Kim, A., Chon, J. K., et al., A library of TAL effector nucleases spanning the human genome. *Nat. Biotechnol.* 2013, *31*, 251–258.

[84] Fu, Y., Foden, J. A., Khayter, C., Maeder, M. L. et al., High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat. Biotechnol.* 2013, *31*, 822–826.

[85] Hsu, P. D., Scott, D. A., Weinstein, J. A., Ran, F. A. et al., DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat. Biotechnol.* 2013, *31*, 827–832.

[86] Wu, X., Scott, D. A., Kriz, A. J., Chiu, A. C. et al., Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. *Nat. Biotechnol.* 2014, *32*, 670–676.

[87] Lin, Y., Cradick, T. J., Brown, M. T., Deshmukh, H. et al., CRISPR/Cas9 systems have off-target activity with insertions or deletions between target DNA and guide RNA sequences. *Nucleic Acids Res.* 2014, *42*, 7473–7485.

[88] Kuscu, C., Arslan, S., Singh, R., Thorpe, J., Adli, M. Genome-wide analysis reveals characteristics of off-target sites bound by the Cas9 endonuclease. *Nat. Biotechnol.* 2014, *32*, 677–683.

[89] Wang, X., Wang, Y., Wu, X., Wang, J. et al., Unbiased detection of off-target cleavage by CRISPR-Cas9 and TALENs using integrase-defective lentiviral vectors. *Nat. Biotechnol.* 2015, *33*, 175–178.

[90] Pattanayak, V., Lin, S., Guilinger, J. P., Ma, E., Doudna, J. A. et al., High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nat. Biotechnol.* 2013, *31*, 839–843.

Biotechnology Journal
www.biotechnology-journal.com

- Chapter 1 -

Biotechnol. J. 2015, 10, 979–994

BIO tec visions
www.biotecvisions.com

[91] Kim, D., Bae, S., Park, J., Kim, E. et al., Digenome-seq: Genome-wide profiling of CRISPR-Cas9 off-target effects in human cells. *Nat. Methods* 2015, *12*, 237–243.

[92] Frock, R. L., Hu, J., Meyers, R. M., Ho, Y. J. et al., Genome-wide detection of DNA double-stranded breaks induced by engineered nucleases. *Nat. Biotechnol.* 2015, *33*, 179–186.

[93] Tsai, S. Q., Zheng, Z., Nguyen, N. T., Liebers, M. et al., GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nat. Biotechnol.* 2015, *33*, 187–197.

[94] Fu, Y., Sander, J. D., Reyon, D., Cascio, V. M., Joung, J. K., Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. *Nat. Biotechnol.* 2014, *32*, 279–284.

[95] Guilinger, J. P., Thompson, D. B., Liu, D. R., Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification. *Nat. Biotechnol.* 2014, *32*, 577–582.

[96] Tsai, S. Q., Wyvekens, N., Khayter, C., Foden, J. A. et al., Dimeric CRISPR RNA-guided FokI nucleases for highly specific genome editing. *Nat. Biotechnol.* 2014, *32*, 569–576.

[97] Krämer, O., Klausing, S., Noll, T., Methods in mammalian cell line engineering: From random mutagenesis to sequence-specific approaches. *Appl. Microbiol. Biotechnol.* 2010, *88*, 425–436.

[98] 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.

[99] Yamane-Ohnuki, N., Kinoshita, S., Inoue-Urakubo, M., Kusunoki, M. et al., Establishment of FUT8 knockout Chinese hamster ovary cells: An ideal host cell line for producing completely defucosylated antibodies with enhanced antibody-dependent cellular cytotoxicity. *Biotechnol. Bioeng.* 2004, *87*, 614–622.

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

[101] Liu, P. Q., Chan, E. M., Cost, G. J., Zhang, L. et al., Generation of a triple-gene knockout mammalian cell line using engineered zinc-finger nucleases. *Biotechnol. Bioeng.* 2010, *106*, 97–105.

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

[103] Huang, Y., Li, Y., Wang, Y. G., Gu, X. et al., An efficient and targeted gene integration system for high-level antibody expression. *J. Immunol. Methods* 2007, *322*, 28–39.

[104] Kito, M., Itami, S., Fukano, Y., Yamana, K., Shibui, T., Construction of engineered CHO strains for high-level production of recombinant proteins. *Appl. Microbiol. Biotechnol.* 2002, *60*, 442–448.

[105] Lieu, P. T., Machleidt, T., Thyagarajan, B., Fontes, A. et al., Generation of site-specific retargeting platform cell lines for drug discovery using phiC31 and R4 integrases. *J. Biomol. Screen.* 2009, *14*, 1207–1215.

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

[107] Cristea, S., Freyvert, Y., Santiago, Y., Holmes, M. C. et al., In vivo cleavage of transgene donors promotes nuclease-mediated targeted integration. *Biotechnol. Bioeng.* 2013, *110*, 871–880.

[108] Jiang, W., Bikard, D., Cox, D., Zhang, F., Marraffini, L. A. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nat. Biotechnol.* 2013, *31*, 233–239.

[109] Jiang, Y., Chen, B., Duan, C., Sun, B. et al., Multigene editing in the *Escherichia coli* genome using the CRISPR-Cas9 system. *Appl. Environ. Microbiol.* 2015, *81*, 2506–2514.

[110] DiCarlo, J. E., Norville, J. E., Mali, P., Rios, X. et al., Genome engineering in *Saccharomyces cerevisiae* using CRISPR-Cas systems. *Nucleic Acids Res.* 2013, *41*, 4336–4343.

[111] Jakociunas, T., Bonde, I., Herrgård, M., Harrison, S. J. et al., Multiplex metabolic pathway engineering using CRISPR/Cas9 in *Saccharomyces cerevisiae*. *Metab. Eng.* 2015, *28*, 213–222.

[112] Doench, J. G., Hartenian, E., Graham, D. B., Tothova, Z. et al., Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation. *Nat. Biotechnol.* 2014, *32*, 1262–1267.

[113] Oakes, B. L., Nadler, D. C., Savage, D. F., Protein engineering of Cas9 for enhanced function. *Methods Enzymol.* 2014, *546*, 491–511.

[114] Ramakrishna, S., Cho, S. W., Kim, S., Song, M. et al., Surrogate reporter-based enrichment of cells containing RNA-guided Cas9 nuclease-induced mutations. *Nat. Commun.* 2014, *5*, 3378.

[115] Lin, S., Staahl, B. T., Alla, R. K., Doudna, J. A., Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery. *Elife* 2014, *3*, e04766.

[116] Baik, J. Y., Lee, K. H. Toward product attribute control: Developments from genome sequencing. *Curr. Opin. Biotechnol.* 2014, *30*, 40–44.

[117] Hyduke, D. R., Lewis, N. E., Palsson, B. Ø., Analysis of omics data with genome-scale models of metabolism. *Mol. Biosyst.* 2013, *9*, 167–174.

[118] Spahn, P. N., Lewis, N. E., Systems glycobiology for glycoengineering. *Curr. Opin. Biotechnol.* 2014, *30*, 218–224.

[119] Hyduke, D. R., Palsson, B. Ø., Towards genome-scale signalling network reconstructions. *Nat. Rev. Genet.* 2010, *11*, 297–307.

[120] Workman, C. T., Mak, H. C., McCuine, S., Tagne, J. B. et al., A systems approach to mapping DNA damage response pathways. *Science* 2006, *312*, 1054–1059.

[121] Le Novère, N., Quantitative and logic modelling of molecular and gene networks. *Nat. Rev. Genet.* 2015, *16*, 146–158.

[122] Lewis, N. E., Abdel-Haleem, A. M., The evolution of genome-scale models of cancer metabolism. *Front. Physiol.* 2013, *4*, 237.

20

Chapter 2

Knockout of single target genes

The CRISPR/Cas9 tool is a major player in the process of transitioning from the traditional random approach of cell line development for therapeutic protein production towards a more rational engineering approach. Precise genome editing has long been the missing tool in mammalian cell line engineering, and with its tidy arrival we hope to catch up to the many existing tools for microbial cells to build synthetic gene networks [11]. New methods and technologies, including gene editing, will help to characterize and modulate gene expression in biological context that is necessary to build synthetic gene networks with predictable activity. Gene editing with CRISPR/cas9 can be used for targeted mutagenesis of both coding and noncoding genetic regions, and for the insertion of transgenes at precise genomic locations.

In this chapter and the following chapters, several studies demonstrating successful application of CRISPR/Cas9 genome editing for CHO cell engineering for various purposes is introduced, in an effort to contribute to the toolbox for CHO cell line development. This chapter is comprised of a book chapter describing a step-by-step method on how to generate a single gene knockout in CHO cells using the CRISPR/Cas9 genome editing tool. With the availability of CHO genome sequences and genome editing tools came the opportunity to knockout genes, making it possible to remove unwanted activities associated with these genes. Knockout of genes have since played an important role in improving culture performance, product quality and enabling gene amplification methods, among other beneficial attributes [12]. With the emergence of the CRISPR/Cas9 genome editing tool [13,14], the process of knocking out genes in CHO cells has significantly improved [10]. The tool is cheaper, more efficient, and relatively simple and flexible (allowing targeting of almost any gene) in comparison to other genome editing tools. With this in mind, it provides the opportunity to target and knockout a broader range of genes than in the pre-CRISPR/Cas9 era. This significantly improves our ability to study more genes and their biological function, allowing us to identify further disadvantageous traits and reroute biological pathways to gain phenotypes more optimal for production of therapeutic proteins.

Chapter 7

Application of CRISPR/Cas9 Genome Editing to Improve Recombinant Protein Production in CHO Cells

Lise Marie Grav, Karen Julie la Cour Karottki, Jae Seong Lee, and Helene Faustrup Kildegaard

Abstract

Genome editing has become an increasingly important aspect of Chinese Hamster Ovary (CHO) cell line engineering for improving production of recombinant protein therapeutics. Currently, the focus is directed toward expanding the product diversity, controlling and improving product quality and yields. In this chapter, we present our protocol on how to use the genome editing tool Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR)/CRISPR-associated protein 9 (Cas9) to knockout engineering target genes in CHO cells. As an example, we refer to the glutamine synthetase (GS)-encoding gene as the knockout target gene, a knockout that increases the selection efficiency of the GS-mediated gene amplification system.

Key words Chinese Hamster Ovary Cells, CRISPR/Cas9, Genome editing, Glutamine synthetase, Knockout, Recombinant protein production

1 Introduction

Chinese Hamster Ovary (CHO) cells are extensively used as a host cell system for the production of recombinant protein therapeutics. The progress and success of CHO cell culture technology has long been depending on large-scale screening of highly productive cell lines and process optimization. Despite the established success of CHO cells, there are increasing demands to expand product diversity, control and improve product quality, and improve cellular production capacities. Advances in this area have been made by genetic engineering approaches including the repression or knockout of disadvantageous genes.

Knockout of genes has improved production of recombinant proteins in CHO cells and the performance of CHO production cells on several levels. Prolonging cell cultures by targeting pro-apoptotic genes, increasing product quality by targeting genes involved in glycosylation of proteins, and enabling amplification of

101

genes by targeting the dihydrofolate reductase (DHFR) gene are among the achievements using gene knockouts. An overview of genes targeted using gene knockout for improvement of CHO production cells was recently presented by Fischer et al. [1].

The bacterial clustered regularly interspaced short palindromic repeat (CRISPR)/CRISPR-associated protein 9 (Cas9) gene editing tool has drastically improved the knockout of genes—making it easier, cheaper, and more efficient. CRISPR/Cas9 is a relatively simple genome editing system comprised only of the nuclease Cas9 and a single guide RNA (sgRNA), which has shown to be highly applicable for genome editing in CHO cells [2]. In this system target recognition is enabled by a 20-nt target complementary sequence in the sgRNA, and a protospacer adjacent motif (PAM) sequence directly downstream of the target sequence [3]. The PAM sequence may differ between Cas9 orthologs [2]. The sgRNA guides the Cas9 nuclease to the target DNA, where it introduces a double strand break (DSB) [4]. The DSB can be repaired by the cells' own DNA repair system. In CHO cells the most frequently used DNA repair pathway is the error-prone non-homologous end joining (NHEJ). Through the NHEJ pathway both dissected DNA ends are ligated directly without DNA-end resection, which may cause insertion or deletion of one or more base pairs—also known as indel formation. Indel formation can cause a frameshift in the coding region of genes, disrupting their translation and leaving the genes dysfunctional. The Cas9 protein can be programmed to target any DNA sequence of interest that is followed by a PAM sequence, simply by changing the 20-nt target complementary sequence in the sgRNA.

In this chapter, we describe a general and robust platform for generating a single-gene knockout in CHO cells using the CRISPR/Cas9 genome editing tool. As an example, we refer to the gene glul, encoding glutamine synthetase (GS), as the target gene. However, the methods are general and can be used to knockout other genes or even to multiplex knockouts [5]. GS is the enzyme that converts glutamate and ammonia to glutamine, and without glutamine in the growth medium GS is essential for cell survival [6]. The widely used GS expression system ™ (Lonza) exploits the glutamine metabolism in mammalian cells, by using a transfected GS-encoding gene as a selectable marker as it permits growth in medium lacking glutamine. The system works well in cell lines that do not express sufficient GS to survive. In cell lines such as CHO cells, which express sufficient endogenous GS, the addition of the GS inhibitor methionine sulfoximine (MSX) is required to inhibit excess GS activity. Increasing the MSX concentration can result in gene amplification and increased productivity [7]. It is desirable to eliminate the endogenous GS-encoding gene from CHO cell lines to improve the selection stringency and efficiency of the GS system in CHO cells [8].

2 Materials

2.1 sgRNA
Expression Plasmid
Construction

1. Target sequence analysis software, e.g., CRISPy tool available for free online.

2. Glycerol stock of *E. coli* transformed with sgRNA expression plasmid (from Ronda et al. [9]).

3. 2× YT medium.

4. Kanamycin.

5. 500 mL baffled Erlenmeyer shake flask.

6. Sterile pipette tips.

7. Incubator with shaker, 37 °C, 250 rpm.

8. Plasmid midi- or maxiprep kit (Machery-Nagel).

9. Sterile Milli-Q water.

10. NanoDrop 2000 (Thermo Scientific).

11. PCR primers for amplification of sgRNA backbone (for design instructions *see* Subheading 3.2.2).

12. Primers containing the sgRNA sequence (for design instructions, *see* Subheading 3.2.3).

13. Phusion U polymerase (Thermo Scientific).

14. 5× HF Buffer (Thermo Scientific).

15. dNTPs.

16. PCR tubes.

17. Thermocycler.

18. Fast Digest DpnI enzyme (Thermo Scientific).

19. 10× Green Buffer (Thermo Scientific).

20. 1 kb DNA ladder.

21. 1% agarose gel: 1 g agarose powder (Bio-Rad) dissolved in 100 mL 1× TAE buffer (Sigma).

22. Gel chamber and power source.

23. PCR and gel purification kit (Machery-Nagel).

24. 10× NEBuffer 4 (New England Biolabs).

25. Heat block.

26. USER enzyme (New England Biolabs).

27. 10× BSA (New England Biolabs).

28. 1.5 mL eppendorf tubes.

29. Mach1 competent *E. coli* cells (Thermo Scientific).

30. Heat block, 37 °C, 300 rpm.

31. Table top centrifuge.

32. Sterile spatula.

33. LB-kanamycin plates: 15 g/L Agar, 10 g/L Tryptone, 10 g/L NaCl, 5 g/L Yeast Extract, 50 μg/mL kanamycin.

34. 10 mL bacterial culture tubes.

35. Plasmid miniprep kit (Machery-Nagel).

36. Access to Sanger sequencing facility.

37. Sequencing primers.

38. Sequence analysis software (e.g., CLC Main Workbench).

2.2 Prepare GFP 2A Peptide-Linked Cas9 Expression Plasmid

1. Glycerol stock with *E. coli* transformed with GFP 2A peptide-linked Cas9 expression plasmid (from Grav et al. [5]).

2. Ampicillin.

3. LB-ampicillin agar plates: 15 g/L Agar, 10 g/L Tryptone, 10 g/L NaCl, 5 g/L Yeast Extract, 60 μg/mL ampicillin.

2.3 Transfection of CHO-S Cells

1. CHO-S cells (Life Technologies).

2. NucleoCounter NC-200 Cell counter (ChemoMetec).

3. Growth medium: CD CHO medium (Life technologies) supplemented with 8 mM L-glutamine (Lonza).

4. 15 or 50 mL centrifuge tubes.

5. 6-well plate, flat bottom (Corning #351146).

6. Humidified incubator, 37 °C, 5% CO_2, 120 rpm.

7. OptiPro™ SFM reduced serum medium (Life Technologies).

8. FreeStyle™ MAX reagent (Life Technologies).

9. sgRNA expression plasmid generated in Subheading 3.2.

10. GFP 2A peptide-linked Cas9 expression plasmid prepared in Subheading 3.3.

2.4 Analysis of Indel Generation: T7 Endonuclease Assay

1. Quick extract (Epicentre).

2. 2× Phusion Master Mix (Thermo Scientific).

3. Primers (for design *see* Subheading 3.5).

4. 10× NEBuffer 2 (New England Biolabs).

5. T7 endonuclease (New England Biolabs).

6. 4% E-Gel (Invitrogen #G501804) or equivalent.

7. Mother E-base (Invitrogen).

2.5 Generation of Clonal Cell Lines: Fluorescence-Activated Cell Sorting (FACS)

1. Fluorescence-activated cell sorter.

2. 384-well plates, flat bottom (Corning #3542).

3. FACS sorting medium: CD CHO medium (Life Technologies) supplemented with 8 mM L-glutamine (Lonza), 1% Antibiotic-Antimycotic 100× (Gibco), and 1.5% HEPES (Life Technologies).

4. FACS tubes.

5. 30 μm cell strainer.

6. Celigo cytometer or microscope.

7. Humidified incubator, 37 °C, 5% CO_2, no shake.

8. 96-well plates, flat bottom (Corning #351172).

9. Clone expansion medium: CD CHO medium (Life Technologies) supplemented with 8 mM L-glutamine (Lonza), 1% Antibiotic-Antimycotic 100× (Gibco), and 1 μL/mL Anti-clumping agent (Life Technologies #0010057AE).

10. 96-well plates, V-Shaped (Greiner bio-one #651161).

11. Breathable plastic bag.

2.6   Analysis of Gene Modifications: Sanger Sequencing

1. Primers from Subheading 3.5.

2. 2× Phusion Master Mix (Thermo Scientific).

3. DNA quick extract from Subheading 3.6.

2.7   Expansion of Clones

1. 12-well plates, flat bottom (Corning #351143).

2. 125 mL shake flask (Corning #431143).

3. DMSO (Sigma-Aldrich #472301).

4. Cryotubes.

3   Methods

The following section is a general protocol that we use for single knockouts in our lab. To make the protocol easier to follow, we refer to the GS-encoding gene as an example of a specific target. The protocol can be used to target other genes, and can easily be adapted to multiple knockouts [5]. An overview of the protocol is shown in Fig. 1.

3.1   Identification of Target Site and sgRNA Primer Design

1. Go to http://staff.biosustain.dtu.dk/laeb/crispy/.

2. Search for your target gene by name, id or symbol, e.g., the GS-encoding gene glul. Be aware of that there can be multiple genes annotated to encode one protein (*see* **Note 1**).

3. A selection of different target sequences will be displayed. You can sort them according to the number of exact matches (preferably one) and see where the different target sequences are located in relation to exons and introns (*see* **Note 2**). The target sequence is followed by a PAM sequence (5′-NGG-3′) in the target gene, as shown in Fig. 2a.

5′-GNNNNNNNNNNNNNNNNNNNNNGG-3′
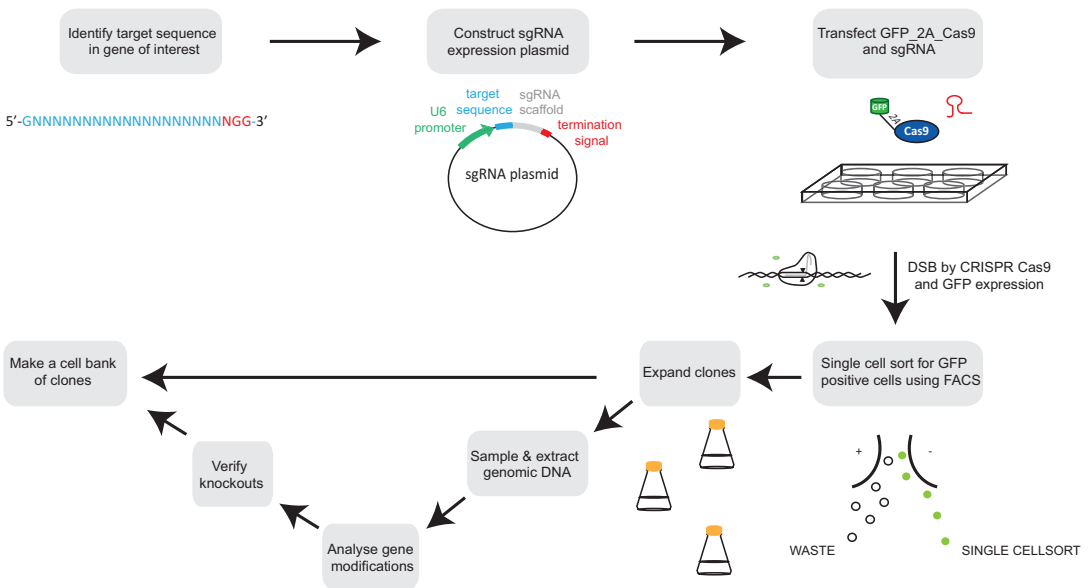
Fig. 1 Schematic outline of the experimental setup for the method described in this chapter

4. Design sgRNA primers using your selected target sequence from Subheading 3.1. The target sequence minus the PAM sequence is the only varying region when designing primers and constructing the sgRNA plasmids for different target sequences, as described in Fig. 2b.

3.2 sgRNA Plasmid Construction

3.2.1 Prepare sgRNA Backbone Plasmid

1. Request the sgRNA plasmid from Ronda et al. [9] and generate a bacterial glycerol stock (or prepare it from scratch following the method described in the publication).

2. Use the tip of a sterile pipette tip and scrape the bacterial stock, add the pipette tip to 100–200 mL of 2× YT medium supplemented with 50 µg/mL kanamycin, and incubate overnight at 37 °C with shaking at 250 rpm.

3. Isolate plasmid DNA using a plasmid midi- or maxiprep kit, resuspend at 1 µg/µL in Milli-Q water. Measure the concentration using Nanodrop 2000.

3.2.2 Amplify sgRNA Backbone

1. Use the sgRNA plasmid map, and design and order uracil-containing primers to amplify the sgRNA backbone. The primers should amplify the sgRNA backbone so that it acquires overhangs after Uracil-Excision Specific Reagent (USER) treatment that is compatible for USER fusion with the overhangs of the annealed primers from Subheading 3.2.3, as described in Fig. 2b. Alternatively, apply primers previously published [9].
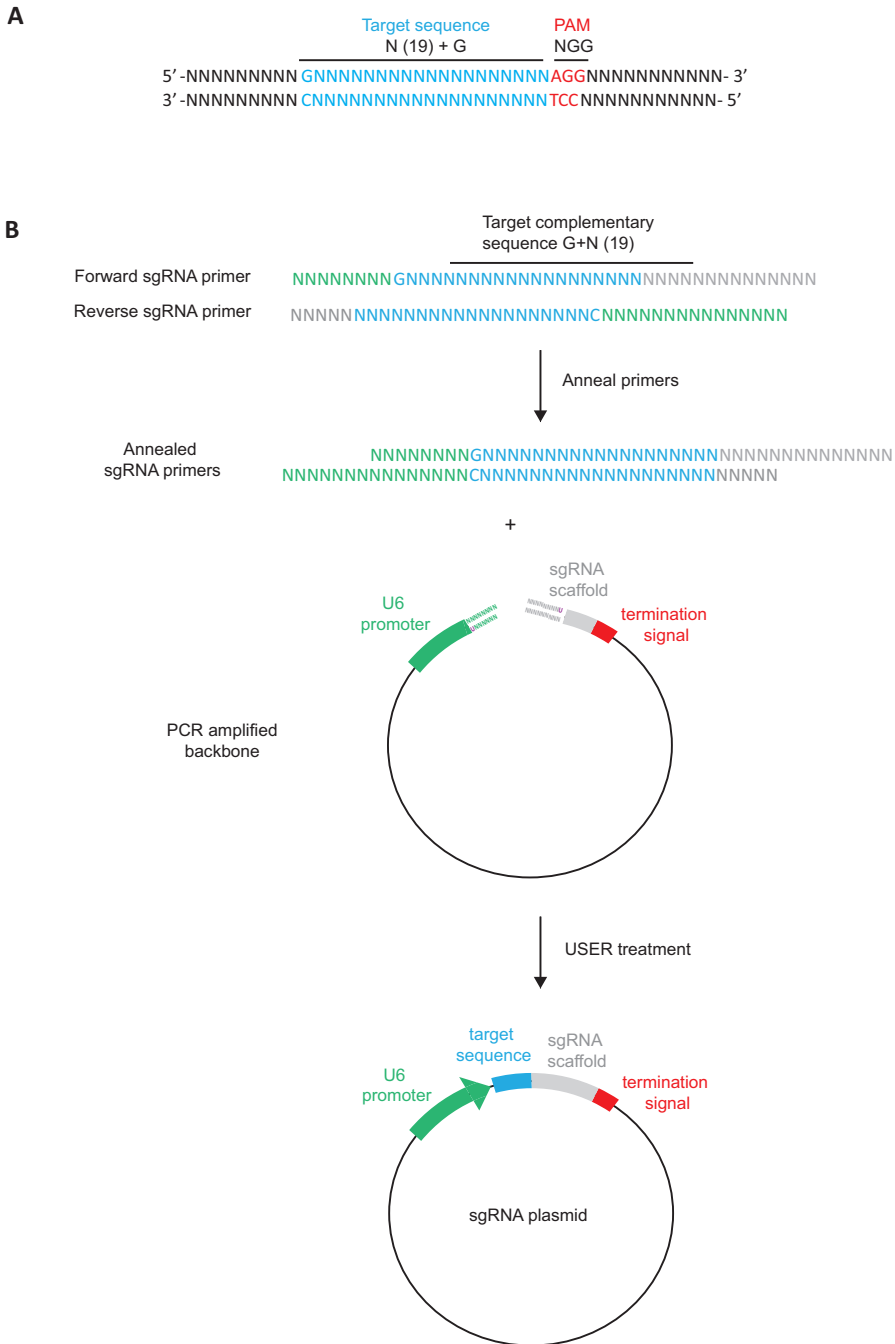
Fig. 2 An outline of the target gene and target sequence, and a schematic overview of how to construct the sgRNA expression plasmid. (**a**) An outline of the target gene showing the target sequence (5′-G+N(19)-3′) and the PAM sequence (5′-NGG-3′) in relation to each other. (**b**) Schematic overview of the sgRNA plasmid construction described in Subheading 3.2. A simple way to construct your sgRNA plasmid is by ordering your target complementary sequence as primers. You can keep your sgRNA "constant" and just exchange the target complementary sequence. The primers for the target sequence are designed to anneal and give rise to overhangs that match overhangs generated after USER treatment of the amplified sgRNA backbone. In this case, the following uracil containing primers would be used for amplifying the backbone: Fwd: AGCTAGAAA UAGCAAGTTAAAATAAGGC and Rev.: ACAAGATAUATAAAGCCAAGAAATCGA. After assembly of the annealed primers and the amplified sgRNA backbone upon USER enzyme treatment, you will attain the complete sgRNA expression construct

Table 1
PCR program to amplify the sgRNA backbone

Temperature (°C)	Time (min)	Number of cycles
98	00:30	1
98	00:10	35
57	00:30	
72	01:15	
72	10:00	1
4	∞	1

2. Mix the following components in a PCR tube:
   – 10 μL 5× HF buffer.
   – 1 μL dNTPs (10 mM).
   – 2.5 μL Primer forward (10 μM).
   – 2.5 μL Primer reverse (10 μM).
   – 0.5 μL Phusion U polymerase.
   – 1 μL sgRNA plasmid template (1.7 ng/uL) (prepared in Subheading 3.2.1).
   – 32 μL Milli-Q water.

3. Place the PCR tube in thermocycler, and run the following program (as shown in Table 1).

4. Treat the sgRNA backbone amplicon with DpnI enzyme to remove methylated DNA by mixing the following components:
   – 44 μL of the sgRNA backbone PCR reaction mixture.
   – 5 μL 10× Green buffer.
   – 1 μL Fast Digest DpnI enzyme.

5. Incubate the mixture at 37 °C for 1 h, this will degrade methylated DNA.

6. Run the PCR product on a 1% agarose gel alongside a 1 kb DNA ladder.

7. Cut out the band at approximately 4.2 kb and purify the PCR product using a PCR and gel purification kit.

3.2.3 Annealing Primers for sgRNA Construct

1. Mix the following components in an Eppendorf tube:
   – 10 μL 10× NEBuffer 4.
   – 10 μL sgRNA Forward primer (100 μM).
   – 10 μL sgRNA Reverse primer (100 μM).
   – 70 μL Milli-Q water.

Table 2
Components of reaction for assembly of backbone and sgRNA insert

Component	Negative control, µL	sgRNA reaction, µL
Backbone	1	1
Annealed sgRNA primers	–	7
10× BSA	0.5	0.5
NEBuffer 4	0.5	0.5
USER enzyme	1	1
Milli-Q water	7	–

2. Incubate the mixture at 95 °C for 5 min on a heat block, turn off the heat block, and leave the mixture on the heat block overnight for gradual cooling. Store at −20 °C.

3.2.4 Assembly of Backbone and sgRNA Insert

1. Mix components in a PCR tube according to Table 2, including a negative control (backbone only).

2. Incubate the mixed reactions at 37 °C for 40 min, and 25 °C for 30 min. Store at −20 °C.

3.2.5 Transformation of sgRNA Plasmid in E. coli

1. Add 1.5 µL of USER reaction to 15 µL competent *E. coli* cells in an Eppendorf tube and incubate on ice for 30 min.

2. Heat shock at 42 °C for 30 s.

3. Return to ice and keep it there for 1 min.

4. Add 1 mL 2× YT medium to the Eppendorf tube and incubate the mixture at 37 °C for 1 h at 300 rpm shake.

5. Pellet the cells at $2000 \times g$ for 3 min.

6. Remove the supernatant, resuspend the pellet in 100 µL 2× YT medium, and plate it using a sterile spatula on a pre-warmed (37 °C) kanamycin agar plate.

7. Incubate the plates upside down at 37 °C overnight.

3.2.6 Analyze and Prepare sgRNA Plasmid

1. Pick a colony using a pipette tip and transfer it to 4 mL 2× YT medium supplemented with 50 µg/mL kanamycin in a 10 mL bacterial culture tube.

2. Shake cells at 250 rpm at 37 °C for 5 h.

3. Isolate plasmid DNA using a plasmid miniprep kit and use this for sanger sequencing. Use a primer for sequencing that anneals before the U6 promoter and covers the target sequence, to make sure your sgRNA expression cassette does not contain any mutations using a sequence analysis software, e.g., CLC Main Workbench.

4. Grow a midi- or maxiprep culture of the correct transformant by inoculating 100–200 mL of 2× YT medium supplemented with 50 μg/mL kanamycin with 1 mL culture from **step 2**. Incubate overnight at 37 °C with shaking at 250 rpm.

5. Isolate plasmid DNA using a midi- or maxiprep kit. Resuspend plasmid DNA at approximately 1 μg/μL in sterile Milli-Q water. Measure the concentration using Nanodrop 2000. Use this product for transfection.

3.3 Prepare the GFP 2A Peptide-Linked Cas9 Expression Plasmid

1. Request the GFP 2A peptide-linked Cas9 expression plasmid (GFP_2A_Cas9) plasmid from Grav et al. [5], and generate a bacterial glycerol stock (or prepare it from scratch following the method described in the publication).

2. Use the tip of a sterile pipette tip and scrape the bacterial stock, add the pipette tip to 100–200 mL of 2× YT medium supplemented with 60 μg/mL ampicillin, and incubate overnight at 37 °C with shaking at 250 rpm.

3. Isolate plasmid DNA using a plasmid midi- or maxiprep kit, resuspend 1 μg/μL in sterile Milli-Q water. Measure the concentration using Nanodrop 2000. Use this product for transfection.

3.4 Transfection of CHO-S Cells

3.4.1 Day 0: Washing and Seeding Cells for Transfection

1. Use healthy (above 95% viability) CHO-S cells at a low passage.

2. Count cells using a NucleoCounter.

3. Harvest $1.5–2 \times 10^6$ cells (for a single transfection in 1×6 well), spin down cells at $200 \times g$ for 5 min, and remove the supernatant.

4. Wash cells with preheated CD CHO medium supplemented with 8 mM glutamine (no anti-clumping agent), spin down at $200 \times g$ for 5 min, and remove the supernatant.

5. Inoculate cells at $5–6 \times 10^5$ cells/mL in 1×6 well with preheated growth medium.

6. Incubate cells at 37 °C, 5% CO_2, and shake at 120 rpm for 1 day.

3.4.2 Day 1: Transfection

1. Count cells and inoculate cells at 1×10^6 cells/mL in 3 mL preheated growth medium in a 6-well plate.

2. Use a total amount of 3.75 μg of plasmid (1:1 (w/w) of sgRNA plasmid and Cas9 expression plasmid). If you want to use multiple sgRNAs, *see* **Note 3**.

3. Gently mix plasmids with 60 μL OptiPRO™ SFM reduced serum medium.

4. Dilute 3.75 μL FreeStyle™ MAX reagent in 60 μL OptiPRO™ SFM reduced serum medium. Mix gently and add to plasmid premix (**step 3**).

5. Incubate for 5 min.

6. Gently add transfection mix to the cells from **step 1**.

7. Incubate cells at 37 °C, 5% CO_2 and shake at 120 rpm for 2 days.

3.5 Optional: T7 Endonuclease Assay to Check if Your sgRNA Works

1. Harvest 50 µL cells 2 days after transfection, spin down at $1000 \times g$, remove the supernatant, and add 20 µL quick extract and incubate at 65 °C for 15 min followed by 95 °C for 5 min. Store at −20 °C.

2. Design primers using, e.g., NCBI Primer-BLAST tool (http://www.ncbi.nlm.nih.gov/tools/primer-blast/) so that a product between 600 and 1000 bp will be amplified. This product should span the selected target sequence in your target gene. The T7 endonuclease will cleave the product where an indel is present when hybridized to wild-type sequence. Design your primers so that the PCR product after cleavage will give bands of different sizes that are separable on an agarose gel.

3. Mix the following components in a PCR tube (prepare one for quick extract of the transfected pool of cells and one for quick extract from CHO-S cells):

   – 10 µL 2× Phusion Master Mix.

   – 1 µL Primer Forward (10 µM).

   – 1 µL Primer Reverse (10 µM).

   – 1 µL DNA template (quick extract from **step 1**).

   – 7 µL Milli-Q water.

4. Place the PCR tube in a thermocycler, and run the following program as outlined in Table 3:

5. Run 5 µL of the PCR product next to a 1 kb DNA ladder on a 1% agarose gel. There should be only one clear band (if not redo and/or troubleshoot your PCR).

6. Transfer 10 µL to a new PCR tube.

7. Place the PCR tube in a thermocycler and run the following T7 endonuclease annealing program as outlined in Table 4.

8. Divide the PCR product into two PCR tubes (5 µL in each tube).

9. Mix the components for the following two reactions (T7+ and T7−) as outlined in Table 5.

10. Incubate the PCR tubes for 30 min at 37 °C.

11. Load the products on a 4% E-Gel or equivalent, and run for approximately 30 min. Your results should be similar to what is shown in Fig. 3a.

Table 3
PCR program for T7 endonuclease assay

Temperature (°C)	Time (min)	Number of cycles
98	00:30	1
98	00:10	10
68	00:30	
−1 per cycle		
72	00:30	
98	00:10	20
58	00:30	
72	00:30	
72	05:00	1
4	∞	1

Table 4
T7 endonuclease annealing program

Temperature (°C)	Time (min)	Number of cycles
95	05:00	1
95	00:01	5
-2 s^{-1}		
85	00:01	300
-0.2 s^{-1}		
4	∞	

Table 5
Components of T7+ and T7− reactions

Component	T7+, μL	T7− (negative control), μL
10× NEBuffer 2	1	1
T7 endonuclease	0.5	–
PCR product	5	5
Milli-Q water	3.5	4
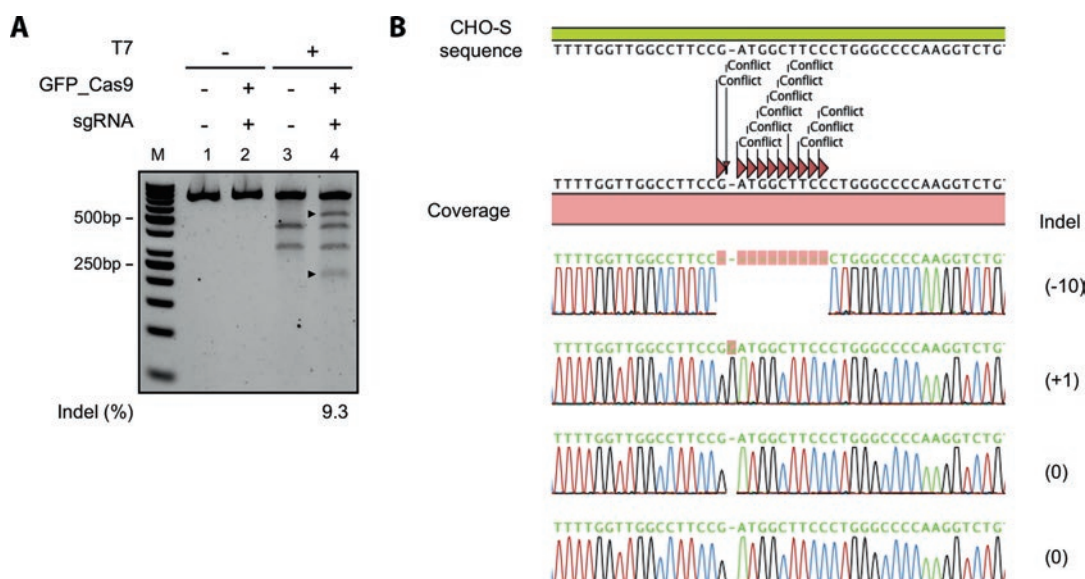
Fig. 3 Analysis of genome modifications. (**a**) T7 endonuclease assay of a pool of cells 2 days after transfection, showing that a sgRNA complementary sequence to a selected target site is capable of generating indels. The samples are analyzed on a 4% E-Gel, where the two first lanes (*1* and *2*) are not treated with T7 endonuclease, and the two last lanes (*3* and *4*) are treated with T7 endonuclease. Samples in *lane 2* and *4* are transfected with GFP_2A_Cas9 and a sgRNA, while the other two samples are un-transfected. After T7 endonuclease treatment, the un-transfected sample shows two bands, while the transfected sample shows two additional bands at expected sizes, designated by *arrows*. A result like this shows that a selected sgRNA is capable of generating indels at the target site. The percentage of indels generated in this case is estimated to 9.3, by using ImageJ software. (**b**) An example of sanger sequencing analysis of a target sequence region. The alignment shows that out of four sequences, in this case, there are one sequence with an indel of −10, one with an indel of +1, and two that still have the wild-type sequence (no indel)

3.6 Generation of Clonal Cell Lines Using FACS	1. Two to three days after transfection, prepare wanted number of 384-well flat-bottom plates (*see* **Note 4**) with 30 μL FACS sorting medium.

1. Two to three days after transfection, prepare wanted number of 384-well flat-bottom plates (*see* **Note 4**) with 30 μL FACS sorting medium.

2. Strain cells through a 30 μm cell strainer into a FACS tube.

3. Using a FACS, sort for GFP-positive single cells into one or more pre-warmed (37 °C) 384-well plates. If a FACS is not available, an alternative method can be used (*see* **Note 5**).

4. Spin plates at $200 \times g$ for 5 min to make sure cells reach the medium.

5. Place cells in a breathable plastic bag (to limit evaporation), and incubate cells at 37 °C, 5% CO_2, no shake for 10 days.

6. Check for surviving cells using a microscope or Celigo cytometer. Cell count should preferably be around >1000 in a well or confluency >50%.

7. Carefully pipette up and down three times and transfer cells to a 96-well flat-bottom plate with 180 μL of clone expansion medium.

8. After 4 days, check plates in microscope. When the clones have a confluency >50%, carefully pipette up and down three times and transfer 50 μL cell suspension to a 96-well V-shaped plate.

9. Spin down the V-shaped 96-well plate at $1000 \times g$ for 5 min, remove the supernatant, add 20 μL quick extract, resuspend the pellets, and move them to PCR tubes or plates. Incubate at 65 °C for 15 min and 95 °C for 5 min. Store at −20 °C.

3.7 Analysis of Gene Modifications: Sanger Sequencing

1. Mix the following components in a PCR tube (per clone you have generated):
   – 10 μL 2× Phusion Master Mix.
   – 1 μL Primer Forward (10 μM).
   – 1 μL Primer Reverse (10 μM).
   – 1 μL DNA template (quick extract from Subheading 3.6).
   – 7 μL Milli-Q water.

2. Place the PCR tube in thermocycler, and run the following program as shown in Table 6.

3. Run the PCR product on a 1% agarose gel, cut out the band with the expected amplicon size, and purify it using a gel and PCR purification kit. Measure the concentration using Nanodrop 2000.

Table 6
PCR program for analysis of gene modifications

Temperature (°C)	Time (min)	Number of cycles
98	00:30	1
98	00:10	10
68	00:30	
−1 per cycle		
72	00:30	
98	00:10	20
58	00:30	
72	00:30	
72	05:00	1
4	∞	1

4. Sequence the product using the forward primer designed for the T7 endonuclease assay (*see* **Note 6**). Mix primer and purified PCR product according to the instruction provided by the sequencing service you use. We discuss application of alternative methods to analyze gene modifications in Subheading 4 (*see* **Note 7**).

5. Analyze your sequencing results using a sequence analysis software, e.g., CLC Main Workbench and align the results to the wild-type target sequence (e.g., the GS sequence). The results should be similar to what is shown in Fig. 3b.

3.8   Expansion of Clone Candidates

1. Select clones with indels that lead to a frameshift, which indicates that you have rendered the gene dysfunctional. Even if the analysis shows there is a frameshift, it is important to verify that it is a real knockout, e.g., by western blotting (*see* **Note 8**) and/or a functional assay (*see* **Note 9**).

2. Move the selected clones from the 96-well plate when >90% confluent to a 12-well flat-bottom plate, maintain in the 12-well plate until confluent, and then move to a 6-well flat-bottom plate, when confluent seed in a 125 mL shake flask at 3×10^5 cells/mL.

3. Take out the samples you need and bank the clones, using 10^7 cells in 1 mL conditioned medium with 5–10% DMSO. Freeze in a Styrofoam box at −80 °C the first 24 h before moving to permanent storage at −180 °C.

4   Notes

1. When selecting a sgRNA sequence to target the gene of interest, it is important to be aware that there can be several genes annotated for the same protein that may be isoforms or pseudogenes. When this is the case you could either select a sgRNA specific for each of the genes, select a sgRNA that matches all of the genes, or select sgRNAs based on expression levels of the gene variants.

2. Examples of sgRNAs in CRISPy with exact match in each of the two glul gene variants in the CHO-K1 genome (GCF_000223135.1) are "GGCCCAGGGAAGCCATCGGAAGG" (GeneID:100, 689,337) and "GGCCTCCTCGATGTGCCTGGTGG" (GeneID:100,764,163), or that matches both genes (in the 3′-end) are "GAGAAGGCATGTGCGGACGATGG." When selecting your target sequence, you should consider the various exons and splice variants for the gene of interest, and which target position will most likely make a break that leaves the gene dysfunctional. If opportunity allows it, it is a good idea

to test a minimum of two different sgRNAs. Always remember that the annotation of some genomes can be incomplete at the time of analysis, and it is always important to validate your knockout, either by a Western Blot (*see* **Note 8**) and/or a functional assay (*see* **Note 9**). The CRISPy tool takes the presence of a 5′-NGG-3′ PAM sequence directly downstream of the target sequence for *Streptococcus pyogenes* Cas9 (spCas9) into consideration. If you are using a different Cas9 ortholog, there might be different requirements to your PAM sequence that you need to consider.

3. When using multiple sgRNAs, the weight ratio between the Cas9 expression plasmid and sgRNA plasmids is still (1:1). The weight ratio between the different sgRNAs is also (1:1). For instance, if you use three different sgRNAs the weight ratio should be (1:1:1). Another recommendation when multiplexing gene knockouts is to use deep sequencing when analyzing the gene modifications, as explained in **Note 6**.

4. The number of 384-well plates necessary for sorting depends on your FACS sorting efficiency. It is important to use stringent settings so that you only sort one cell per well. Depending on your sorting efficiency (how many wells get a single cell sorted into it) and survival (how many single cells survive), you can adjust the number of plates you sort. If your sorting efficiency is around 30%, two 384-well plates should give a sufficient number of knockout cells to choose among.

5. Limiting dilution is an alternative way to single cell sort your cells, if you do not have access to a FACS [10]. In this case you will need to screen more clones, as you cannot enrich for GFP_2A_Cas9 expressing cells. When applying limiting dilution it is sufficient to use a plasmid expressing Cas9 that is not linked to GFP with a 2A peptide.

6. If your coverage is not good when only using the forward primer, you can also use the reverse primer for sequencing, or design new sequencing primers. The Sanger sequencing method is very simple, but not optimal. You might experience some unclear sequencing result, due to, e.g., mixed indels on different alleles. Select only the clones that show clear sequencing results indicating a frameshift; otherwise, it is necessary to use one of the alternative methods described in **Note 7**. When potential functional knockout clones are selected, we recommend applying an additional primer set for verification.

7. TOPO™ cloning-based sequencing and deep sequencing are alternative methods that can be used to analyze gene modifications in greater detail. A description of how to perform these analyses can be found in Ronda et al. [9]. Both methods can reveal if you have an indel on one or more alleles of your targeted gene. However, TOPO cloning-based sequencing
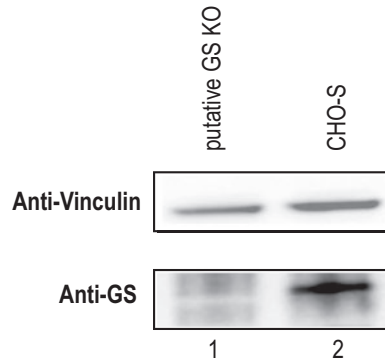
Fig. 4 Knockout validation. An example of a western blot of cell lysates from a putative GS knockout (KO) clone in *lane 1* and wild-type CHO-S in *lane 2*, which were analyzed with anti-GS antibody (Abcam ab 49873) and anti-vinculin antibody (Sigma-Aldrich V9131). Anti-vinculin is used as a loading control. The putative GS knockout in *lane 1* shows no band at the expected size of GS (45 kDa), while the wild-type shows a band at the expected size of GS, indicating that GS is not expressed in this putative GS knockout

requires a much higher extent of screening than deep sequencing. It is therefore recommended to use deep sequencing analysis when analyzing gene modifications from multiplexed knockouts, as it minimizes the work load and gives a higher coverage. When potential functional knockout clones are selected, we recommend applying an additional primer set for verification.

8. Even though you see clear gene modifications in your analysis, the protein could still be expressed in an intact or truncated version. A simple way to ensure that you have disrupted the protein expression is to perform a Western Blot of your putative clones, using an antibody against target gene. In the case of a GS knockout, you can use an antibody against GS as shown in Fig. 4.

9. In addition to ensuring that you have lost the protein expression, it is important to ensure that you have acquired the wanted phenotypic change. As in the case of a GS knockout, a simple test is to grow the cells in the presence and absence of glutamine. If the cells die in the absence of glutamine (and not in the presence), you likely have a functional knockout.

Acknowledgments

References

1. Fischer S, Handrick R, Otte K (2015) The art of CHO cell engineering: a comprehensive retrospect and future perspectives. Biotechnol Adv 33(8):1878–1896

2. Lee JS, Grav LM, Lewis NE, Faustrup Kildegaard H (2015) CRISPR/Cas9-mediated genome engineering of CHO cell factories: application and perspectives. Biotechnol J 10:979–994

3. Jinek M, Chylinski K, Fonfara I, Hauer M et al (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science 337:816–821

4. Garneau JE, Dupuis MÈ, Villion M, Romero DA et al (2010) The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. Nature 468:67–71

5. Grav LM, Lee JS, Gerling S, Kallehauge TB, Hansen AH, Kol S, Lee GM, Pedersen LE, Kildegaard HF (2015) One-step generation of triple knockout CHO cell lines using CRISPR/Cas9 and fluorescent enrichment. Biotechnol J 10:1446–1456

6. Wurm FM (2004) Production of recombinant protein therapeutics in cultivated mammalian cells. Nat Biotechnol 22:1393–1398

7. Jun SC, Kim MS, Hong HJ, Lee GM (2006) Limitations to the development of humanized antibody producing Chinese hamster ovary cells using glutamine synthetase-mediated gene amplification. Biotechnol Prog 22:770–780

8. Fan L, Kadura I, Krebs LE, Hatfield CC, Shaw MM, Frye CC (2012) Improving the efficiency of CHO cell line generation using glutamine synthetase gene knockout cells. Biotechnol Bioeng 109:1007–1015

9. Ronda C, Pedersen LE, Hansen HG, Kallehauge TB, Betenbaugh MJ, Nielsen AT, Kildegaard HF (2014) Accelerating genome editing in CHO cells using CRISPR Cas9 and CRISPy, a web-based target finding tool. Biotechnol Bioeng 111:1604–1616

10. Freshney RI (2010) Cloning and selection in: culture of animal cells, 6th edn. Wiley, New York, NY

Chapter 3

Knockout of multiple genes simultaneously

Here we present an efficient way of removing several genes simultaneously in order to speed up rational cell line development processes in a more cost-effective manner, as generation of a single knockout clone is rather time-consuming. In the past decades, genetic engineering of CHO production host cells has focused on the manipulation of single target genes [12]. However, changes in cellular phenotypes are most likely not the result of altering the expression of a single gene but rather of a plethora of genes involved in the same or different pathways. It is expected that engineering multiple genes and eventually entire signalling pathways will have a higher probability to significantly improve phenotypic outcomes.

Research Article

One-step generation of triple knockout CHO cell lines using CRISPR/Cas9 and fluorescent enrichment

Lise Marie Grav[1],*, Jae Seong Lee[1],*, Signe Gerling[1], Thomas Beuchert Kallehauge[1], Anders Holmgaard Hansen[1], Stefan Kol[1], Gyun Min Lee[2], Lasse Ebdrup Pedersen[1] and Helene Faustrup Kildegaard[1]

[1] The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Hørsholm, Denmark
[2] Department of Biological Sciences, KAIST, Daejeon, Republic of Korea

The CRISPR/Cas9 genome editing technology has previously been shown to be a highly efficient tool for generating gene disruptions in CHO cells. In this study we further demonstrate the applicability and efficiency of CRISPR/Cas9 genome editing by disrupting *FUT8*, *BAK* and *BAX* simultaneously in a multiplexing setup in CHO cells. To isolate Cas9-expressing cells from transfected cell pools, GFP was linked to the Cas9 nuclease via a 2A peptide. With this method, the average indel frequencies generated at the three genomic loci were increased from 11% before enrichment to 68% after enrichment. Despite the high number of genome editing events in the enriched cell pools, no significant off-target effects were observed from off-target prediction followed by deep sequencing. Single cell sorting of enriched multiplexed cells and deep sequencing of 97 clones revealed the presence of four single, 23 double and 34 triple gene-disrupted cell lines. Further characterization of selected potential triple knockout clones confirmed the removal of Bak and Bax protein and disrupted fucosylation activity as expected. The knockout cell lines showed improved resistance to apoptosis compared to wild-type CHO-S cells. Taken together, multiplexing with CRISPR/Cas9 can accelerate genome engineering efforts in CHO cells even further.

Supporting information available online

1 Introduction

Chinese hamster ovary (CHO) cells are widely used in the biopharmaceutical industry as a host for the production of complex pharmaceutical proteins. Thus, genome engineering of CHO cells for improved productivity and product quality is of great interest.

In biotechnology and medical research, generating desirable genomic traits in mammalian cell lines is often highly valuable. Traditionally, genome editing has been performed mainly by conventional gene targeting strategies based on homologous recombination [1]. However, the low frequency of spontaneous breakage of chromosomal DNA increases the need for nucleases to generate site-specific DNA double strand breaks (DSBs). Genome editing nucleases including meganucleases, zinc-finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENS), and the more recent clustered regularly interspaced short palindromic repeats (CRISPR)/ CRISPR-associated (Cas) system have paved the way for rapid and efficient induction of DSBs at defined genomic sites, leading to site-specific gene manipulations by non-

Abbreviations: *BAK*, Bcl-2 homologous antagonist killer; ***BAX*,** Bcl-2-associated X; **Cas**, CRISPR-associated; **Cas9**, Cas protein 9; **CHO**, Chinese hamster ovary; **CRISPR,** clustered regularly interspaced short palindromic repeats; **DSBs**, double stranded breaks; **FACS**, fluorescence-activated cell sorting; ***FUT8*,** fucosyltransferase 8; **GFP**, green fluorescent protein; **HDR**, homology-directed repair; **LCA**, *Lens culinaris* agglutinin; **NHEJ**, non-homologous end joining; **PAM**, protospacer adjacent motif; **sgRNA**, single-guide RNA; **TALENs**, transcription activated-like effector nucleases; **USER**, uracil specific excision reagent; **ZFNs**, zinc finger nucleases

* These authors contributed equally to this work.

Biotechnology Journal
www.biotechnology-journal.com

- Chapter 3 -

Biotechnol. J. 2015, 10, 1446–1456

BIO tec visions
www.biotecvisions.com

homologous end joining (NHEJ) repair or homology-directed repair (HDR) [2]. We have previously shown that the CRISPR/Cas protein 9 (Cas9) genome editing tool is very efficient in CHO cells by generating NHEJ-mediated gene disruptions in two genes involved in glycosylation: fucosyltransferase 8 (*FUT8*) and *COSMC* encoding the C1GALT1-specific chaperone 1 [3]. In this system, a single-guide RNA (sgRNA) directs the Cas9 nuclease to cause DSBs in matching target DNA sequences if they are adjacent to short sequences known as protospacer adjacent motifs (PAMs) [4, 5]. The target site specificity is determined by 20-base pair (bp) sequences at the 5' end of the sgRNA that form base pairs with desired target DNA sequences [4, 5]. This technology can edit multiple target sequences simultaneously by simply expressing multiple sgRNAs together with a single Cas9 nuclease, which confers multiplex genome engineering [5–9]. Additionally, a recent study provides a simple method to increase genome editing frequencies by ZFNs and CRISPR/Cas9 via 2A peptide-coupled co-expression of nuclease and fluorescent protein followed by FACS enrichment of cell populations with high nuclease expression levels [10]. Up to now, multiple knockout of genes was shown to be feasible in CHO cells through the sequential application of different pairs of ZFNs [11, 12]. Successive transfection and screening of knockout cell lines achieved generation of triple gene knockout cell lines at frequencies of >1% [12].

In this study, we explore the potential of the CRISPR/Cas9 system to generate several gene disruptions simultaneously by multiplexing. Our goal was to accelerate the genome editing process during cell line generation and this study focused on the development of an efficient multiplexing process for CHO cells. To facilitate enrichment of cells with generated indels, a vector with green fluorescent protein (GFP) reporter linked via a 2A peptide to Cas9 was constructed and transfected cells were sorted with fluorescence-activated cell sorting (FACS). Deep sequencing was applied to analyze the genome editing efficiencies at both on-target and off-target sites for three target genes, *FUT8*, Bcl-2 homologous antagonist killer (*BAK*) and Bcl-2-associated X (*BAX*). Selected engineered cell lines were shown to be unable to fucosylate and to be resistant to apoptotic inducers, thus confirming a functional knockout of *FUT8*, *BAK* and *BAX*. These results support the conclusion that application of this multiplexing process leads to the successful generation of genetic and functional triple knockout CHO cell lines.

2 Materials and methods

2.1 Plasmid construction and sgRNA target design

Cloning of the GFP_2A_Cas9 was performed with seamless uracil-specific excision reagent (USER) cloning to insert GFP_2A directly upstream of the Cas9 reading frame in pJ607-Cas9 [3]. The uracil-containing primers are listed in Supporting information, Table S1 and PCR (98°C for 2 min; 35 times: 98°C for 10 s, 60°C for 45 s, 72°C for 40 s/5 min; 72°C for 5 min) was performed with the X7 DNA polymerase [13]. Subsequent to DpnI (Thermo Fisher Scientific, Waltham, MA) treatment of the backbone fragment (pJ607-Cas9), amplicons were purified from 1% agarose TAE gel using NucleoSpin® Gel and PCR Clean-up kit (Macherey-Nagel, Düren, Germany). The two PCR products were mixed and treated with USER enzyme (New England Biolabs, Ipswich, MA) according to the manufacturer's recommendations. Subsequent to USER enzyme treatment, the reaction mixture was transformed into *Escherichia coli* One Shot® Mach1™ competent cells (Life technologies, Thermo Scientific, Rockford, IL) according to standard procedures. Transformant cells were selected on 100 μg/mL ampicillin LB plates. The sequence of the construct was verified by sequencing and purified plasmid was obtained using the Nucleo-Bond® Xtra Midi kit (Macherey-Nagel) according to the manufacturer's recommendations. The online bioinformatics tool "CRISPy" was applied for sgRNA target selection for *BAK* and *BAX* [3]. For *FUT8*, the sgRNA3_F was selected based on a previous study [3]. The *BAK*, *BAX* and *FUT8* target sites are presented in Supporting information, Table S2. The sgRNA expression vectors were constructed as previously described [3]. The oligos for sgRNA construction are listed in Supporting information, Table S3.

2.2 Cell cultivation and transfections

CHO-S suspension cells obtained from Life Technologies were grown in CD CHO medium supplemented with 8 mM L-glutamine and 1 μL/mL anti-clumping agent (Life Technologies). Cells were expanded in Corning vent cap shake flasks (Sigma-Aldrich, St. Louis, MO) in a humidified incubator at 120 rpm, 37°C and 5% CO_2. Cell growth was monitored using the NucleoCounter NC-200 Cell Counter (ChemoMetec, Allerod, Denmark) and cells were passed into fresh medium every two to three days with seeding densities at 3 to 4×10^5 cells/mL. Cells were washed and seeded at 5 to 6×10^5 cells/mL without anti-clumping agents one day prior to transfection. Cells were transfected with expression vectors encoding Cas9 or GFP_2A_Cas9 and sgRNA targeting *FUT8*, *BAK* and/or *BAX*. For each sample, a density of 1×10^6 cells/mL in a six multi-well plate (BD Biosciences, San Jose, CA) or in 125 mL flasks were transfected with a total of 3.8 μg or 17.7 μg of DNA respectively using FreeStyle™ MAX reagent together with OptiPRO SFM medium (Life Technologies) according to the manufacturer's recommendations. Transfections with pmaxGFP® vector (Lonza, Basel, Switzerland) were applied as control for transfection efficiencies. Anti-clumping agent (1.5 μL/mL) was added

Biotechnology
Journal
www.biotechnology-journal.com

- Chapter 3 -

Biotechnol. J. 2015, 10, 1446–1456

BIO tec
visions
www.biotecvisions.com

one day after transfection. For comparison of Cas9 and GFP_2A_Cas9 efficiencies, transfected cells were harvested on day 3 after transfection. For multiplexing, cells were harvested on day 2 after transfection, before or after FACS. For each sample 500 000 cells were harvested.

2.3 Isolation of fluorescent-positive cells with FACS

A FACSJazz (BD Biosciences) was used to bulk sort cells from selected transfections. Before cell sorting, cells were prepared by centrifugation at 100 g for 5 min and resuspended in 1 mL fresh medium. The cells were filtered through a 40 µm cell strainer to achieve single cell suspension and transferred to a FACS-compatible tube. Fluorescent-positive cell populations were gated based on non-transfected wild-type CHO-S cells. From the multiplexed samples, 500 000 fluorescent-positive cells were isolated followed by single cell sorting into 96-well U-bottom plates (BD Biosciences). For single cell sorting, one cell was seeded per well in 200 µL 80% CD CHO medium supplemented with 8 mM L-glutamine, and 1% penicillin-streptomycin and 20% 20 µm filtered spent medium. After 10 days, anti-clumping agent was added to each well with colonies at a final concentration of 2 µL/mL. At day 14, the colonies were moved to 96-well flat-bottom plates (BD Biosciences) for further expansion. Confluent colonies were split and replicated plates were seeded and harvested for deep sequencing analysis when close to confluent. Genomic DNA was extracted from cell pellets (cell pools and clonal cells) using QuickExtract DNA extraction solution (Epicentre, Illumina, Madison, WI) according to the manufacturer's instructions.

2.4 Deep sequencing analysis

Deep sequencing was performed on a MiSeq Benchtop Sequencer (Illumina, San Diego, CA). Libraries were prepared based on Illumina 16S Metagenomic Sequencing Library Preparation with some modifications. Genomic regions flanking target genomic sites were amplified using Phusion Hot Start II HF Pfu polymerase (Thermo Fisher Scientific) by touchdown PCR (95°C for 5 min; 10 times: 95°C for 45 s, 69–59°C (–1°C/cycle) for 30 s, 72°C for 45 s; 20 times: 95°C for 45 s , 59°C for 30 s, 72°C for 45 s; 72°C for 10 min), with primers containing overhang sequences compatible with Illumina Nextera XT indexing (forward primer overhang: TCGTCGGCAGCGTCAG ATGTGTATAAGAGACAG, reverse primer overhang: GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG). PCR primers are listed in Supporting information, Table S4. After PCR amplification, amplicons were purified using AMPure XP beads (Beckman Coulter, Brea, CA). Illumina Nextera XT Index (Illumina #FC-131-1001) sequencing adapters were integrated into the amplicons by PCR (98°C for 3 min; 8 times: 95°C for 30 s, 55°C for 30 s, 72°C for 30 s; 72°C for 5 min). The final libraries were purified

once again before quantification. The purified libraries were quantified with the Qubit® dsDNA BR Assay kit (Life Technologies; #Q32850), and the product size within the library was verified by 2% agarose gel. Sequencing was performed as a 151 paired-end run. Sequence data were analyzed using custom python code that compared the sequence data with the expected wild-type sequence from the CHO genome [3]. In short, the sequence data were searched for high similarity to the expected wild-type genomic sequences and then analyzed for the potential presence and size of an indel. The final result is a spreadsheet with a row for each target or off-target, which states how often we see the wild-type, how often we see an indel and what size that indel is.

2.5 Off-target site prediction

Off-target sites were predicted using custom python code that compared the targets with all other potential targets in the CHO genome (ref-seq assembly GCF_000223135.1). Off-targets were then scored based on the following criteria: (i) the first 5 bp immediately adjacent to the PAM sequence: cutoff value of 0; (ii) the number of mismatches in the first 13 bp: cutoff value of up to 1; and (iii) the total number of mismatches including number of mismatches in first 13 bp: cutoff value of up to 5. From the scoring, the 15 highest potential off-target sites for each of the *FUT8*, *BAK* and *BAX* sgRNA were selected (Supporting information, Table S5).

2.6 Re-sequencing of genomic regions flanking sgRNA target sites

Genomic regions flanking the sgRNA target site for *FUT8*, *BAK* and *BAX* were amplified from genomic DNA extracts using DreamTaq DNA polymerase (Thermo Fischer Scientific) by touchdown PCR (95°C for 5 min; 10 times: 95°C for 45 s, 69–59°C (–1°C/cycle) for 30 s, 72°C for 45 s; 20 times: 95°C for 45 s, 59°C for 30 s, 72°C for 45 s; 72°C for 10 min) using PCR primers listed in Supporting information, Table S6. The PCR products were purified from 1% agarose TAE gel with the NucleoSpin® Gel and PCR Clean-up kit (Macherey-Nagel). Each purified PCR product was Sanger sequenced with the corresponding forward primer using EuroFins sequencing service (Eurofins Scientific, Luxembourg).

2.7 Analysis of Bak and Bax protein level

Five million cells were harvested and re-suspended in a fresh lysis buffer (10 mM Tris-HCl, pH 7.4, 100 mM NaCl, 2.5 mM MgCl$_2$, 0.5% Triton X-100, supplemented with 1× Roche protease inhibitor cocktail). The lysate was centrifuged at 12 000 g at 4°C for 10 min, and the soluble lysate was reduced using DTT. 15 µL of the reduced soluble lysates were separated on 4–12% Bis-Tris NuPAGE®

Biotechnology Journal
www.biotechnology-journal.com

- Chapter 3 -

Biotechnol. J. 2015, 10, 1446–1456

BIO tec
visions
www.biotecvisions.com

gel (Life technologies) in 1× MES buffer, followed by transfer to nitrocellulose membranes using iBlot® dry blotting system (Life Technologies). Membranes were blocked with 5% dry skimmed milk diluted in 1× PBS and 0.1% Tween20 for 1 h, and further probed with anti-Bak (Sigma Aldrich, B5897) or anti-Bax (Abcam, ab32503) antibodies at 1:2000 dilutions overnight. The blots were then probed with a HRP-conjugated donkey anti-rabbit IgG secondary antibody (1:5000) for 50 min and detected using Amersham ECL detection system (GE Healthcare Life Sciences, Buckinghamshire, UK). As a loading control, the membranes were re-probed with an anti-Vinculin antibody (Sigma Aldrich, V9131) at 1:1000 dilution and a mouse IgG secondary antibody (1:5000).

2.8 Phenotypic analysis of *FUT8* knockout cells

A lectin stain was performed to analyze the phenotypic change of *FUT8* knockout cells. Cells were seeded in a 96-well tissue culture treated black/clear® microplate (Greiner Bio-one, Frickenhausen, Germany) and incubated for 45 min at room temperature in medium containing 20 µg/mL fluorescein-*Lens culinaris* agglutinin (F-LCA, Vector Laboratories, Peterborough, UK) and NucBlue® Live ReadyProbes® Reagent (Life Technologies). Cells were then washed three times with fresh culture medium followed by complete removal. Celigo Imaging Cell Cytometer (Nexcelom Bioscience, Lawrence, MA) was used to measure fluorescence using the mask + target1 application. The blue and green fluorescence channels were used as mask and target1 respectively.

2.9 Glycoprofiling

Exponentially growing cells were seeded at 1×10^6 cells/mL in 250 mL Corning vent cap shake flasks (Sigma Aldrich). Four days after seeding, 40 mL supernatant was harvested and centrifuged at 2000 g for 15 min. The supernatant was filtered using sterile Nalgene rapid flow 0.2 µm SFCA filter (Thermo Scientific). Proteins contained in the supernatant were concentrated by three to four cycles of centrifugation at 4000 g at room temperature for 10 min in Amicon Ultra column with 3000 Da cutoff, until 1 mL was left in column. N-glycans from retained proteins were released and fluorescently labeled with GlykoPrep Rapid N-Glycan kit (ProZyme Inc., Hayward, CA) according to the manufacturer's instructions using 2-aminobenzamide (2-AB) as a fluorescent label. Labeled N-glycans were analyzed on LC-MS. LC-MS System was a Thermo Ultimate 3000 HPLC with fluorescence detector coupled on-line to a Thermo Velos Pro Iontrap MS. Separation was performed on a BEH Glycan column 100 mm × 2.1, 1.7 µm (Waters, Milford, MA) and solvents: A: 100% acetonitrile, B: 50 mM ammonium formate, pH 4.4 adjusted with formic acid and filtered 0.2 µm. Separation gradient from 39% buffer A to 47% buffer A over 16 min at 0.5 mL/min

flow-rate. Fluorescence detector set to high power lamp and 360 nm excitation, 428 nm emission. MS settings: fullscan: 700–2000 m/z, source fragmentation 60 V, polarity negative. All annotated sugar structures are peaks with correct mass and at least a signal to noise value of 20:1 calculated with the Chromeleon software (Thermo Scientific). Almost all annotated sugar structures were in the plus 300 S/N intensities.

2.10 Apoptosis analysis

Exponentially growing cells were seeded at 3 to 4×10^5 cells/mL into polystyrene transparent square 96-half-deepwell microplates with sandwich cover (Enzyscreen, Leiden, Netherlands) containing 250 µL culture medium and the culture plates were incubated on the Multitron cell (INFORS HT, Bottmingen, Switzerland) at 300 rpm in humidified 5% CO_2 at 37°C. To induce apoptosis, 10 µg/mL actinomycin D (Sigma Aldrich) was added to the culture on day 3. After 24 h cultivation in the presence of actinomycin D, 15 000 cells were transferred to 96-well tissue culture-treated black/clear® microplates (Greiner Bio-one), and caspase activity was measured using the Homogeneous Caspases Assay kit (Roche, Basel, Switzerland) according to the manufacturer's instructions. Briefly, caspase substrate labeled with Rhodamine 110 was added and incubated for 1 h at 37°C. Free Rhodamine 110 resulting from cleavage by active caspases was measured by microplate reader (Synergy Mx, BioTek Instruments, Inc., Winooski, VT).

2.11 Transient expression of antibody and quantification

For transient expression of antibody, cells were transfected with expression vector encoding rituximab and cell growth was measured daily as described in Section 2.2. Cell suspension was harvested daily and centrifuged at 4000 g for 5 min and supernatant was kept for rituximab quantification. Rituximab quantification was performed using an Octet RED96 (Pall, Menlo Park, CA, USA). Protein A biosensors were equilibrated in sample diluent buffer (PBS containing 0.1% BSA and 0.02% Tween-20) and rituximab was measured for 120 s at 30°C. Absolute concentrations of rituximab were calculated by comparison with a calibration curve generated from a dilution series of human IgG (GENSA01006, VWR, Dublin, Ireland). Regeneration of biosensor tips between measurements was performed in 10 mM glycine pH 1.7.

3 Results

3.1 Facilitation of enrichment for genome-edited cells via GFP 2A peptide-linked Cas9 expression vector combined with FACS

The overall aim of this study was to: (i) accelerate the generation of engineered CHO cell lines with multiple gene disruptions; and (ii) lower number of clones screened for isolation of triple knockout cell lines. To facilitate this, an expression vector with GFP 2A peptide-linked to CHO codon-optimized Cas9 [3] was constructed for identification and enrichment of the Cas9 expressing cells when combined with FACS (Supporting information, Fig. S1). The 2A peptide leads to coupled co-expression of GFP and Cas9 nuclease, which are driven by a single promoter. Therefore cells that are fluorescent after transfection indicate concurrent expression of Cas9 nuclease [10]. To test the method, the glycosylation enzyme Fut8 and the two pro-apoptotic proteins, Bak and Bax, were selected for disruption based on industrial relevance and previous characterization [11, 12].

Three sgRNA expression vectors were constructed to target *FUT8*, *BAK* and *BAX*. To compare the indel generating efficiency of GFP_2A_Cas9 to the previous characterized Cas9 expression vector, CHO-S cells were transfected with the vectors encoding Cas9 or GFP_2A_Cas9 together with sgRNAs targeting either *FUT8*, *BAK* or *BAX*. The indels generated at the target sites were investigated by deep sequencing (Fig. 1A). The GFP_2A_Cas9 expression vector was shown to be functional and generated close to the same amount of indels at the sgRNA target sites in *FUT8*, *BAK* and *BAX* as Cas9.

To test the functionality of the GFP_2A_Cas9 expression vector in multiplexing and in FACS enrichment of cells with genome editing events, CHO-S cells were trans-
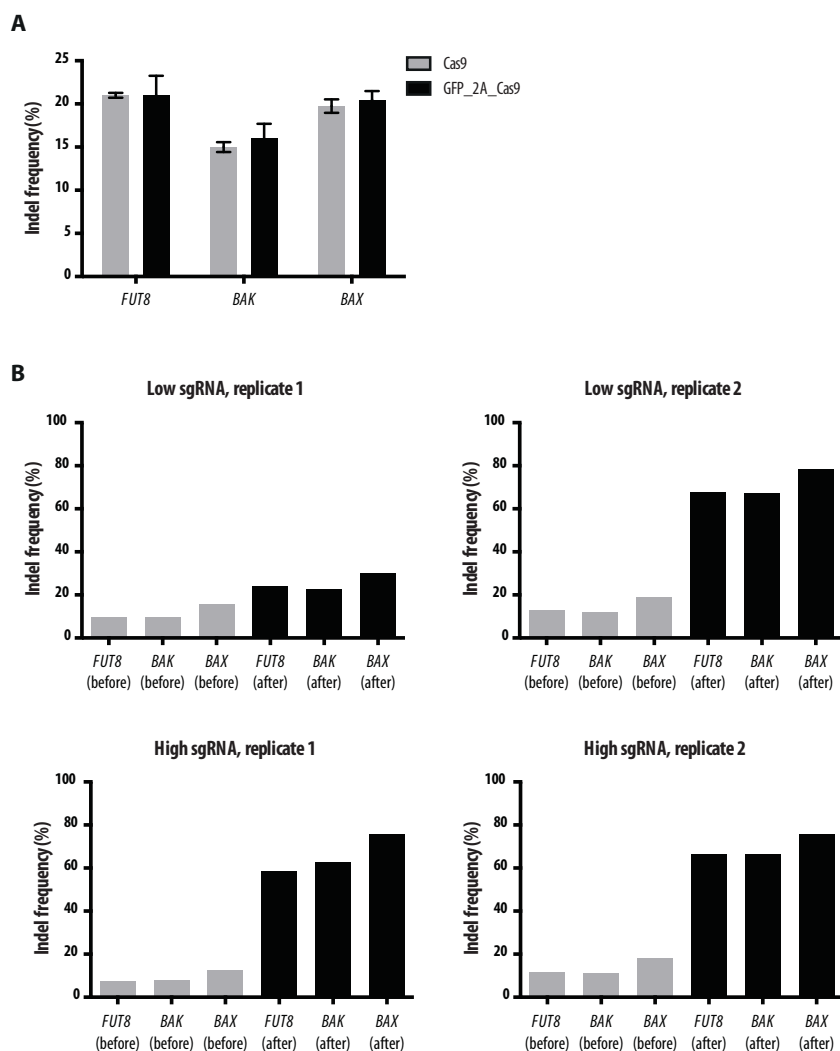
Figure 1. Validation of the GFP 2A peptide-linked Cas9 expression vector. (**A**) The graph shows the frequency of indels generated at *FUT8*, *BAK* and *BAX* sgRNA target sites upon transient transfection of Cas9 and GFP_2A_Cas9, together with the respective sgRNAs. The percentage of indels is measured from deep sequencing of amplicons covering the target sites in the genome of transfected CHO-S cells harvested three days after transfection. The data are generated from two biological replicates. (**B**) The graph shows the percentage of indels generated at *FUT8*, *BAK* and *BAX* sgRNA target sites upon transient transfection of GFP_2A_Cas9 and the three respective sgRNAs simultaneously. The indels are measured from deep sequencing of amplicons covering the target sites in the genome of transfected cells before and after FACS two days after transfection. Low sgRNA shows indels generated from transfection of GFP_2A_Cas9 to each of the sgRNAs in the molar ratio of 1:(0.8/0.8/0.8). High sgRNA shows indels generated from transfection of GFP_2A_Cas9 to each of the sgRNAs in the molar ratio of 1:(2.4/2.4/2.4). The results from two biological replicates are shown.

Table 1. Off-target analysis of the 15 most potential off-target sites for each of the sgRNAs in multiplexed and FACS-enriched cell populations

*FUT8*_on target	66.3%	*BAK*_on target	66.3%	*BAX*_on target	75.8%
*FUT8*_off1	0.2%	*BAK*_off1	0.1%	*BAX*_off1	0.1%
*FUT8*_off2	0.1%	*BAK*_off2	0.2%	*BAX*_off2	0.2%
*FUT8*_off3	0.2%	*BAK*_off3	0.1%	*BAX*_off3	N/A[a]
*FUT8*_off4	0.2%	*BAK*_off4	0.2%	*BAX*_off4	0.6%
*FUT8*_off5	0.2%	*BAK*_off5	0.2%	*BAX*_off5	46.3%[b]
*FUT8*_off6	0.0%	*BAK*_off6	N/A[a]	*BAX*_off6	0.2%
*FUT8*_off7	0.1%	*BAK*_off7	0.2%	*BAX*_off7	0.7%
*FUT8*_off8	0.1%	*BAK*_off8	0.4%	*BAX*_off8	0.2%
*FUT8*_off9	0.1%	*BAK*_off9	0.4%	*BAX*_off9	0.1%
*FUT8*_off10	N/A[a]	*BAK*_off10	N/A[a]	*BAX*_off10	0.1%
*FUT8*_off11	0.1%	*BAK*_off11	0.1%	*BAX*_off11	0.1%
*FUT8*_off12	0.2%	*BAK*_off12	0.2%	*BAX*_off12	0.1%
*FUT8*_off13	0.7%	*BAK*_off13	0.2%	*BAX*_off13	0.1%
*FUT8*_off14	0.2%	*BAK*_off14	0.1%	*BAX*_off14	0.1%
*FUT8*_off15	0.1%	*BAK*_off15	0.2%	*BAX*_off15	0.7%

a) N/A indicates that no PCR products were obtained from samples.
b) The high indel frequency observed in this sample is also observed in the GFP_2A_Cas9 only sample.

fected with the vectors encoding GFP_2A_Cas9 and sgRNAs against *FUT8*, *BAK* and *BAX* simultaneously. Two different molar ratios between GFP_2A_Cas9 and sgRNA expression vectors were selected to test the effect of relative molar ratios of GFP_2A_Cas9 and sgRNA on efficiency during multiplexing. The selected ratios were 1:(0.8/0.8/0.8) as used previously [3] and 1:(2.4/2.4/2.4) to increase sgRNA. Two days after transfection, the cell pools were FACS sorted for fluorescent cells. From each sample, both bulk sorting and single cell sorting was performed to acquire both a GFP-enriched population and GFP-positive single cells. The indels generated at *FUT8*, *BAK* and *BAX* sgRNA target sites in the enriched bulk-sorted cell populations were analyzed by deep sequencing (Fig. 1B). Despite the discrepancy observed between the two low sgRNA replicates, the amount of sgRNA is not likely to limit the frequency of genome editing events.

In particular, FACS enrichment of the CHO-S cells transfected with a high sgRNA to GFP_2A_Cas9 plasmid ratio resulted in a large enrichment of cells with generated indels at the three genomic sites with an average of 11% before enrichment and 68% after enrichment. These data demonstrate the feasibility of GFP_2A_Cas9 expression vector and FACS enrichment for CRISPR/Cas9-mediated multiplex genome engineering.

3.2 Analysis of off-target sites for *FUT8*, *BAK* and *BAX* sgRNAs

The high number of genome editing events in the enriched cell pool led us to investigate whether potential off-target sites were being modified as well. Bioinformatics analysis comparing the *FUT8*, *BAK* and *BAX* sgRNA target sites to other sequences in the genome allowed us

to predict the highest potential off-target sites from scoring the number of mismatches (Supporting information, Table S5). To analyze the genome editing efficiencies at these potential off-target sites, deep sequencing of 15 different off-target sites for each sgRNA was performed (primers given in Supporting information, Table S4). Comparisons of genome editing efficiencies of the predicted off-target sites were performed from pools of cells treated with GFP_2A_Cas9 only and GFP_2A_Cas9 together with the three sgRNAs before and after FACS enrichment (Supporting information, Table S7). From the deep sequencing analysis, no significant off-target modifications were observed at the investigated sites in the enriched cell pool despite on-target genome editing efficiencies at 66, 66 and 76% for *FUT8*, *BAK* and *BAX*, respectively (Table 1). The observed low mutation levels at potential off-target sites support the reliability of the current multiplexing process.

3.3 Generation and characterization of single, double and triple gene-disrupted CHO-S cell lines

Single cells from each of the four indel-enriched cell populations (analyzed in Fig. 1B) were expanded to generate clonal cell lines and resulted in 187 colonies. To analyze the indels generated at the *FUT8*, *BAK* and *BAX* loci of these clones, deep sequencing was performed (Supporting information, Table S8 and S9). From the 187 clones, 90 clones were discarded based on failed or unclear deep sequencing data yielding 97 clones with clear genotypes (Supporting information, Table S8 and S9). Of the 97 analyzed clones, 36 clones showed no indels in *FUT8*, *BAK* or *BAX* while four, 23 and 34 clones showed indels in one, two or three genes, respectively (Fig. 2A).
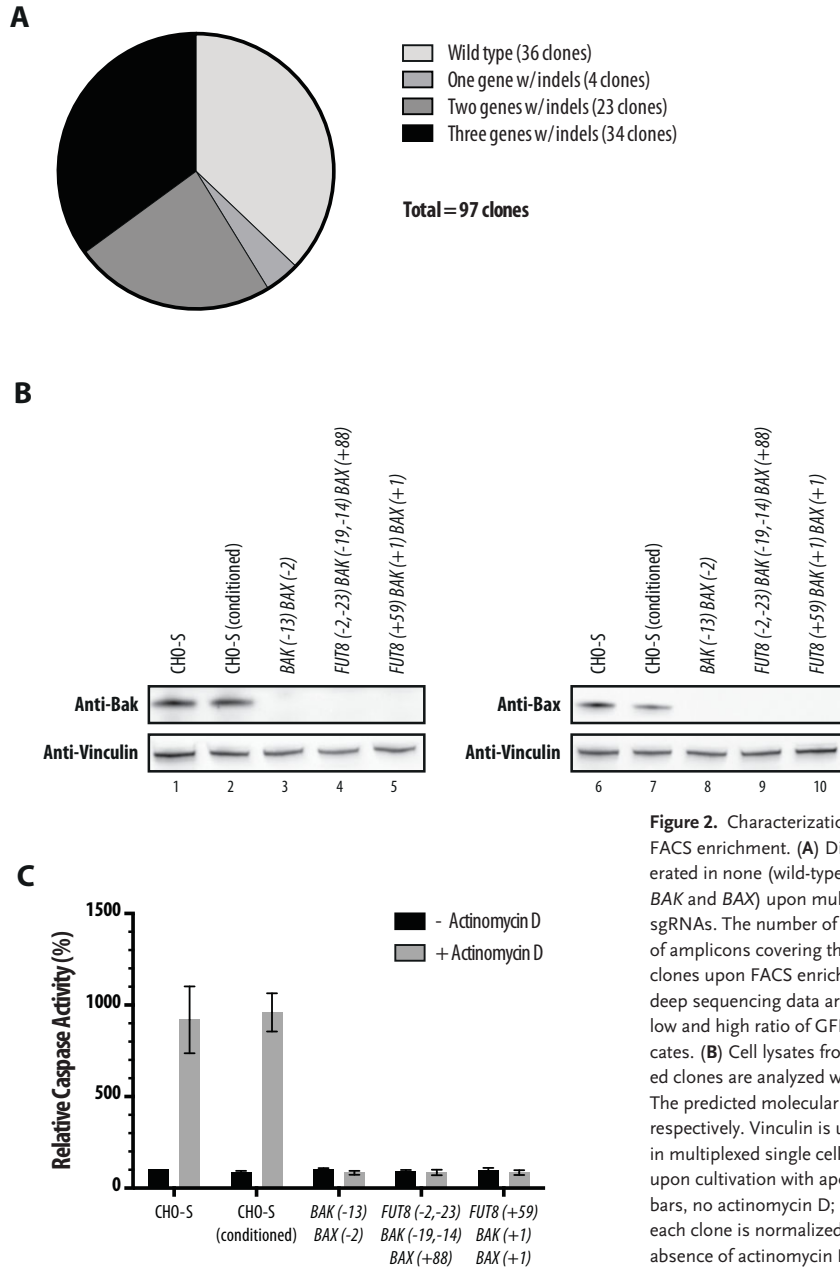
Figure 2. Characterization of multiplexed single cell sorted clones upon FACS enrichment. (**A**) Distribution of the number of clones with indels generated in none (wild-type), one, two or three of the targeted genes (*FUT8*, *BAK* and *BAX*) upon multiplexing with GFP_2A_Cas9 and the respective sgRNAs. The number of genes with indels is measured by deep sequencing of amplicons covering the three sgRNA target sites in 187 single cell sorted clones upon FACS enrichment. Only data for 97 of the clones that had clear deep sequencing data are shown. The data include clones generated upon low and high ratio of GFP_2A_Cas9 to sgRNAs and both biological replicates. (**B**) Cell lysates from CHO-S, CHO-S (conditioned) and three selected clones are analyzed with anti-Bak, anti-Bax, or anti-Vinculin antibodies. The predicted molecular weight of Bak and Bax is 23 kDa and 20 kDa, respectively. Vinculin is used as a loading control. (**C**) Analysis of apoptosis in multiplexed single cell sorted clones. Caspase activities are measured upon cultivation with apoptosis inducer, actinomycin D, for 24 h. Black bars, no actinomycin D; gray bars, 10 µg/mL actinomycin D. The activity of each clone is normalized by that of the wild-type CHO-S cells in the absence of actinomycin D. Average value of triplicates ± SD is shown.

One potential *BAK* and *BAX* double knockout cell line and two potential *FUT8*, *BAK* and *BAX* triple knockout cell lines were selected for further characterization. Primers were designed to verify the indels generated at the three sgRNA target sites, and to facilitate analysis of a larger region surrounding the target site compared to the deep sequencing primers. The flanking regions surrounding the *FUT8*, *BAK* and *BAX* target site were amplified, gel purified and Sanger sequenced (Supporting information, Table S10). The purified bands are presented in Supporting information, Fig. S2. Alignment of the

sequencing data was performed (Supporting information, Fig. S3, S4 and S5) which verified the deep sequencing data obtained for the three clones (Supporting information, Table S11).

To investigate whether the genome modifications introduced by CRISPR/Cas9 genome editing technology resulted in a functional knockout, three knockout cell lines with validated genotypes were further analyzed for protein levels and their functions. Western blot analysis revealed no detectable Bak and Bax protein levels in one *BAK* and *BAX* double knockout cell line and two *FUT8*,
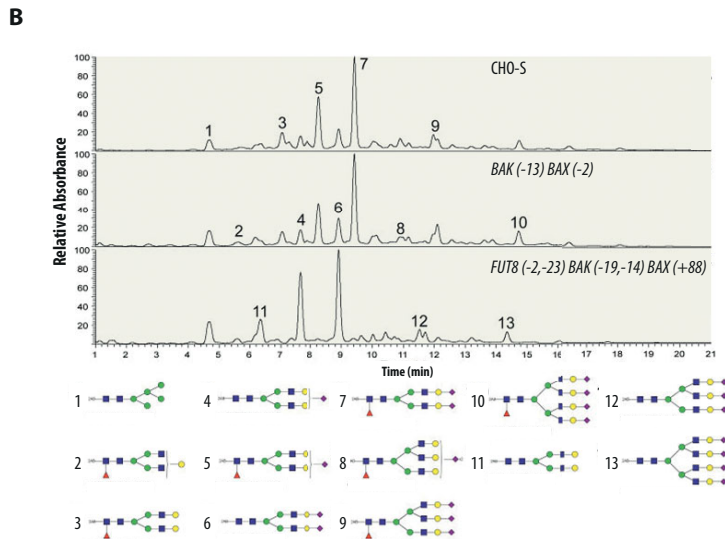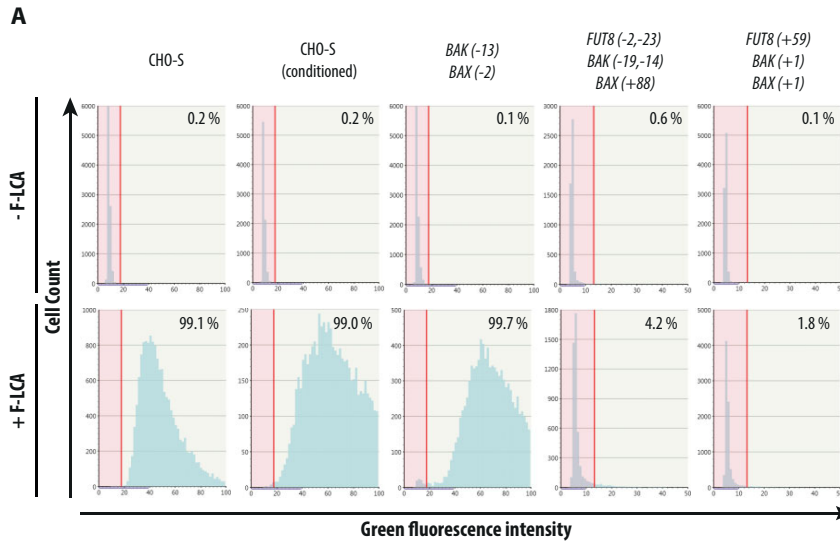
Figure 3. Characterization of glycosylation of multiplexed single cell sorted clones. (**A**) Lectin staining characterized by alteration of green fluorescence intensity upon incubation with F-LCA. The figure shows the representative images of triplicate experiments and the percentages of F-LCA positive cells are shown as the average value of the triplicates (n = 3). (**B**) Glycoprofiling of all proteins in pooled supernatants from duplicate cultivations of CHO-S, the double knockout cell line *BAK (-13) BAX (-2)* and the triple knockout cell line *FUT8 (-2,-23) BAK (-19,-14) BAX (+88)*. The upper panel shows the chromatograms of the released and fluorescent labeled glycoforms of the secreted proteins from the three cell lines respectively. The numbers indicate the peaks of the different glycoforms, which are presented in the lower panel. Blue square with black outline, *N*-acetylglucosamine; green circle with black outline, mannose; yellow circle with black outline, galactose; red triangle with black outline, fucose; purple diamond with black outline, *N*-acetylneuraminic acid.

BAK and *BAX* triple knockout cell lines (Fig. 2B, Lane 3–5 for Bak; Lane 8–10 for Bax). Truncated Bak and Bax proteins were not observed either, which confirmed successful knockout of *BAK* and *BAX* via frameshift mutations in the exons yielding no full-length Bak and Bax proteins.

Since Bak and Bax regulate the intrinsic apoptotic pathway through mitochondrial permeabilization leading to cytochrome *c* release and subsequent activation of caspase cascade [11, 14], double knockout of *BAK* and *BAX* is expected to render cells resistant to apoptosis. When cells lacking Bak and Bax were treated for 24 h with pro-apoptotic transcription inhibitor, actinomycin D, which is known to induce cytochrome *c* release from mitochondria and cell death [15], caspase activation hardly occurred. In contrast, wild-type CHO-S cells and one clonal cell line without indels in *FUT8*, *BAK* and *BAX* (conditioned wild-type CHO-S) showed significant induction of caspases in the presence of actinomycin D (Fig. 2C).

The phenotypic change caused by disruption of *Fut8* was assessed by a lectin stain and glycoprofiling of secreted proteins from *FUT8*, *BAK* and *BAX* triple knockout cell lines. Since the enzyme activity of Fut8 is responsible for transferring fucose to core oligosaccharides, cells lacking the Fut8 enzyme activity do not bind LCA that selectively recognizes fucosylated plasma membrane proteins [3, 12]. The two knockout cell lines harboring the knockout of *Fut8* showed no F-LCA binding, whereas both wild-type cells and one *BAK* and *BAX* double knockout cell line were completely stained with F-LCA (Fig. 3A). Comparison of the glycoprofiles of all secreted proteins from wild-type CHO-S cells, one *BAK* and *BAX* double knockout cell line and one *FUT8*, *BAK* and *BAX* triple knockout cell line, showed clear differences in the

1453

Biotechnology Journal
www.biotechnology-journal.com

- Chapter 3 -

Biotechnol. J. 2015, 10, 1446–1456

BIO tec visions
www.biotecvisions.com
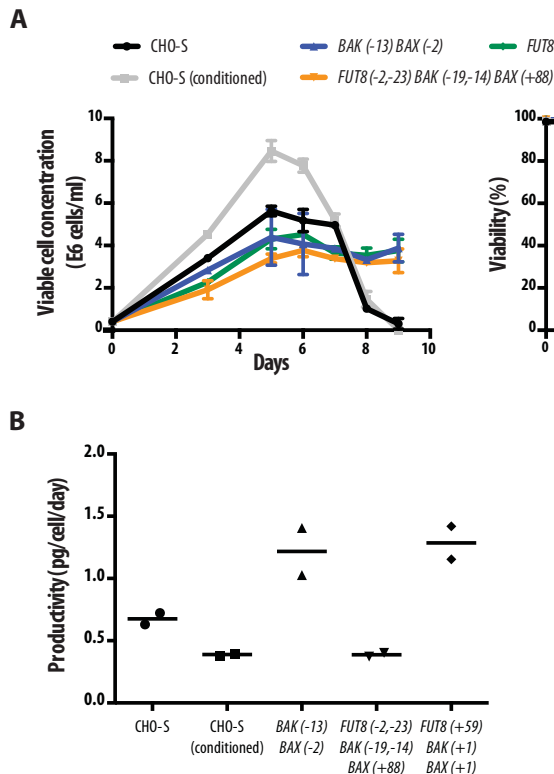
Figure 4. Growth profiles and productivity of multiplexed single cell sorted clones from batch cultures. (**A**) Viable cell concentration and viability. CHO-S, CHO-S (conditioned) and three selected clones are cultivated in shake flasks with seeding densities at 4×10^5 cells/mL. The error bars show the deviation between data points from each biological replicate (n = 2). (**B**) Productivity upon transient expression with antibody expression plasmid. The productivity is calculated from the growth of the cell lines and the amount of secreted rituximab protein during cultivation for three days after transient transfection. The center line represents the average of the biological replicates (n = 2).

amount of fucosylated and non-fucosylated glycoforms (Fig. 3B and Supporting information, Table S12). In the *FUT8*, *BAK* and *BAX* triple knockout cell line, the main peaks of the fucosylated glycoforms (peaks 5 and 7) were not present and the main peaks of non-fucosylated glycoforms (peaks 4 and 6) were increased. The analysis confirmed the loss of Fut8 enzyme activity in the triple knockout cell line as expected. Taken together these data demonstrate a highly efficient method for generation of genetic and functional multiple knockout CHO cell lines using multiplexing with CRISPR/Cas9.

3.4 Cell culture performance of double and triple knockout CHO-S cell lines

To evaluate the effect of multiple knockouts on culture performance, the generated double and triple knockout cell lines were cultivated together with wild-type and conditioned wild-type CHO-S cells. In accordance with the apoptosis inhibition (Fig. 2C), the double and triple knockout cell lines harboring knockout *BAK* and *BAX* were found to delay viability drop, resulting in extended culture longevity compared to CHO-S cells (Fig. 4A). Subsequently, the capability to produce recombinant therapeutic proteins was assessed by transient expression of antibody constructs encoding rituximab for three days (two biological replicates). The cell growth and product titer was measured daily and the raw data are presented

in Supporting information, Table S13. From the viable cell concentration (Supporting information, Fig. S6) and the production of rituximab (Supporting information, Fig. S7), the productivity of the five cell lines was calculated (Fig. 4B). The double and two triple knockout cell lines were capable of producing rituximab comparable to those from wild-type CHO-S cells, and even better for two of the clones.

4 Discussion

With the availability of CHO cell lines and Chinese hamster genome sequences, extensive genetic manipulation and genomic analysis of CHO cells has become possible, which has lead us to engineer CHO cells via sequence-specific approaches [16, 17]. Here, we have developed an efficient multiplexing platform for CHO cells via simultaneous transient expression of sgRNAs together with co-expression of GFP and Cas9 nuclease. To avoid a potential decrease in Cas9 activity by being fused to GFP, a linker peptide harboring the 2A peptide was applied to couple GFP to Cas9. The use of a self-cleaving 2A peptide allows stoichiometric co-expression of 2A peptide-linked proteins from the same transcript following post-translational cleavage [10, 18]. Thus, the cleavage of GFP from Cas9 is expected to yield wild-type Cas9 and free GFP at comparable levels, which can be used for sorting trans-

Biotechnology Journal
www.biotechnology-journal.com

BIOtec visions
www.biotecvisions.com

fected cells. FACS enrichment for fluorescence-positive cells achieved isolation of cell populations with high genome editing events, which allowed for the simultaneous generation of single, double, and triple gene knockout clones in a single round of transfection and clone generation. Notably, 57 out of 97 single cell sorted clones (efficiency = 58.8%) harbored multiple knockouts, in which 59.6% of clones were triple knockout clones, demonstrating the efficiency of CRISPR/Cas9 genome editing in multiplexing. The resultant multiple knockout CHO cell lines showed the expected phenotypes with the practical industrial relevance including enhanced culture lifespan and capability to produce non-fucosylated proteins (Figs. 3 and 4A). In particular, stable production of highly active non-fucosylated antibodies [19] can be achieved from knockout CHO cell lines lacking Fut8, which will be executed in future studies.

By virtue of its simple, fast, and highly efficient multi-gene knockout capability, this platform can offer distinct advantages over current targeted methods as well as traditional methods for CHO cell line engineering including overexpression or introduction of new genes, RNA interference-mediated knockdown or regulation of gene expression [20, 21]. Permanent gene modifications through transient expression of sgRNA and Cas9 nuclease abolish the need for a lengthy selection process with a limited number of available selection markers when compared to traditional methods. This feature enables us to exclude the addition of expensive selection markers during cell cultivation for sustainable expression of exogenous transgenes while reducing metabolic burdens caused by expression of selection marker genes. The high frequency of gene disruption can reduce screening time and effort dramatically by generating independent knockout clones from less than one 96-well plate of single cell sorted clones. In comparison to previous triple knockout CHO cell line generation using ZFNs [12], the present platform is fast and efficient in regard to the one-step generation of multiple gene knockout CHO cell lines. The application of CRISPR/Cas9 technology also benefits from the straightforward, rapid, and low cost of sgRNA design compared to other genome editing nucleases.

Multiplex genome engineering with transfection of separate sgRNAs against each target site in a pool has advantages and disadvantages compared to all-in-one vector systems containing multiple sgRNA and Cas9 nuclease expression cassettes [8, 9]. Co-transfection of multiple plasmids can cause a lower transfection efficiency and variable targeting efficiency at individual target sites due to dilution of each construct and uncoupled expression levels from separate entities. Nevertheless, this was not the case in this study as shown by the high and comparable genome editing efficiency at the three genomic sites after FACS enrichment (Fig. 1B). It can rather be exploited for various combinatorial multiplexing through co-transfection of several sgRNAs without con-

struction of single vectors every time for particular combinations. Introduction of more sgRNAs could expand this system further, thus it will be valuable to investigate how many target sites can be dealt with at once.

Although CRISPR/Cas9 technology generates on-target modifications specifically, off-target mutations can also occur as reported in previous studies in mammalian cells [22–26]. Based on the recent progress on off-target analysis, the potential off-target sites for each sgRNA were predicted by extending our previous bioinformatics tool, CRISPy (Section 2.5), followed by deep sequencing analysis aiming at the top 15 candidate off-target sites (Table 1). Interestingly, no off-target mutations were observed even after FACS enrichment, which supports the high fidelity of the present system. The application of a CHO genome-oriented bioinformatics tool for design of sgRNA target sites with the minimal number of possible off-target sites could facilitate multiplex genome engineering while minimizing the risk of off-target mutations. Furthermore, multiplexing with the use of Cas9 nickase can also be utilized to improve on-target specificity of Cas9 [26, 27].

Our study provides a simple and efficient platform for multiplex genome engineering of CHO cells. With the advent of improved CHO genome characterization and global gene expression profile, it will not be confined to knockout of well characterized genes, but may be harnessed for multiple gene integration [10] and transcriptional control of endogenous genes [28, 29] in CHO cells through slight modifications. The multiple modifications of industrially-relevant genes will improve the production of biopharmaceuticals and enhance the understanding of the gene network underlying protein production.

5 References

[1] Sorrell, D. A., Kolb, A. F., Targeted modification of mammalian genomes. *Biotechnol. Adv.* 2005, *23*, 431–469.

[2] Kim, H., Kim, J. S., A guide to genome engineering with programmable nucleases. *Nat. Rev. Genet.* 2014, *15*, 321–334.

[3] Ronda, C., Pedersen, L. E., Hansen, H. G., Kallehauge, T. B. et al., Accelerating genome editing in CHO cells using CRISPR/Cas9 and CRISPy, a web-based target finding tool. *Biotechnol. Bioeng.* 2014, *111*, 1604–1616.

[4] Jinek, M., Chylinski, K., Fonfara, I., Hauer, M. et al., A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 2012, *337*, 816–821.

[5] Cong, L., Ran, F. A., Cox, D., Lin, S. et al., Multiplex genome engineering using CRISPR/Cas systems. *Science* 2013, *339*, 819–823.

[6] Mali, P., Yang, L., Esvelt, K. M., Aach, J. et al., RNA-guided human genome engineering via Cas9. *Science* 2013, *339*, 823–826.

[7] Wang, H., Yang, H., Shivalila, C. S., Dawlaty, M. M. et al., One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. *Cell* 2013, *153*, 910–918.

[8] Kabadi, A. M., Ousterout, D. G., Hilton, I. B., Gersbach, C. A., Multiplex CRISPR/Cas9-based genome engineering from a single lentiviral vector. *Nucleic Acids Res.* 2014, *42*, e147.

[9] Sakuma, T., Nishikawa, A., Kume, S., Chayama, K., Yamamoto, T., Multiplex genome engineering in human cells using all-in-one CRISPR/Cas9 vector system. *Sci. Rep.* 2014, *4*, 5400.

[10] Duda, K., Lonowski, L. A., Kofoed-Nielsen, M., Ibarra, A. et al., High-efficiency genome editing via 2A-coupled co-expression of fluorescent proteins and zinc finger nucleases or CRISPR/Cas9 nickase pairs. *Nucleic Acids Res.* 2014, *42*, e84.

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

[12] Liu, P. Q., Chan, E. M., Cost, G. J., Zhang, L. et al., Generation of a triple-gene knockout mammalian cell line using engineered zinc-finger nucleases. *Biotechnol. Bioeng.* 2010, *106*, 97–105.

[13] Nørholm, M. H., A mutant Pfu DNA polymerase designed for advanced uracil-excision DNA engineering. *BMC Biotechnol.* 2010, *10*, 21.

[14] Wei, M. C., Zong, W. X., Cheng, E. H., Lindsten, T. et al., Proapoptotic BAX and BAK: A requisite gateway to mitochondrial dysfunction and death. *Science* 2001, *292*, 727–730.

[15] Arnoult, D., Parone, P., Martinou, J. C., Antonsson, B. et al., Mitochondrial release of apoptosis-inducing factor occurs downstream of cytochrome c release in response to several proapoptotic stimuli. *J. Cell Biol.* 2002, *159*, 923–929.

[16] Xu, X., Nagarajan, H., Lewis, N. E., Pan, S. et al., The genomic sequence of the Chinese hamster ovary (CHO)-K1 cell line. *Nat. Biotechnol.* 2011, *29*, 735–741.

[17] Lewis, N. E., Liu, X., Li, Y., Nagarajan, H. et al., Genomic landscapes of Chinese hamster ovary cell lines as revealed by the *Cricetulus griseus* draft genome. *Nat. Biotechnol.* 2013, *31*, 759–765.

[18] Szymczak, A. L., Workman, C. J., Wang, Y., Vignali, K. M. et al., Correction of multi-gene deficiency in vivo using a single 'self-cleaving' 2A peptide-based retroviral vector. *Nat. Biotechnol.* 2004, *22*, 589–594.

[19] Shields, R. L., Lai, J., Keck, R., O'Connell, L. Y. et al., Lack of fucose on human IgG1 N-linked oligosaccharide improves binding to human Fcgamma RIII and antibody-dependent cellular toxicity. *J. Biol. Chem.* 2002, *277*, 26733–26740.

[20] Krämer, O., Klausing, S., Noll, T., Methods in mammalian cell line engineering: From random mutagenesis to sequence-specific approaches. *Appl. Microbiol. Biotechnol.* 2010, *88*, 425–436.

[21] Jadhav, V., Hackl, M., Druz, A., Shridhar, S. et al., CHO microRNA engineering is growing up: Recent successes and future challenges. *Biotechnol. Adv.* 2013, *31*, 1501–1513.

[22] Fu, Y., Foden, J. A., Khayter, C., Maeder, M. L. et al., High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat. Biotechnol.* 2013, *31*, 822–826.

[23] Pattanayak, V., Lin, S., Guilinger, J. P., Ma, E. et al., High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nat. Biotechnol.* 2013, *31*, 839–843.

[24] Hsu, P. D., Scott, D. A., Weinstein, J. A., Ran, F. A. et al., DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat. Biotechnol.* 2013, *31*, 827–832.

[25] Wu, X., Scott, D. A., Kriz, A. J., Chiu, A. C. et al., Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. *Nat. Biotechnol.* 2014, *32*, 670–676.

[26] Cho, S. W., Kim, S., Kim, Y., Kweon, J. et al., Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. *Genome Res.* 2014, *24*, 132–141.

[27] Ran, F. A., Hsu, P. D., Lin, C. Y., Gootenberg, J. S. et al., Double nicking by RNA-guided CRISPR/Cas9 for enhanced genome editing specificity. *Cell* 2013, *154*, 1380–1389.

[28] Konermann, S., Brigham, M. D., Trevino, A. E., Joung, J. et al., Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature* 2014, doi: 10.1038/nature14136. [Epub ahead of print].

[29] Cheng, A. W., Wang, H., Yang, H., Shi, L. et al., Multiplexed activation of endogenous genes by CRISPR-on, an RNA-guided transcriptional activator system. *Cell Res.* 2013, *23*, 1163–1171.

3.2 Applicability and future perspectives

A major issue of eukaryotic expression systems is the continuous acidification of culture medium due to accumulation of lactate, which is generated by the conversion of pyruvate by the lactate dehydrogenase (LDH). The complete knockout of the LDH-A gene has been shown to be lethal for CHO cells [15], which could the case for several other genes as well. However, a recent study by Hefzi et. al [16] successfully managed to knockout LDH-A by simultaneously knocking one or more of the four genes PDK1, PDK2, PDK3 and PDK4, using the multiplexing technique presented in this chapter. The resulting CHO cells were devoid of lactate dehydrogenase activity, an achievement previously believed to be impossible. With this testimony that the impossible can be made possible, it is clear that this multiplexed genome editing technique will completely transform our ability to study and engineer more complex biological problems.

Chapter 4

Site-specific knockin of transgenes

This chapter presents a platform for site-specific gene integration in CHO cells mediated by CRISPR/Cas9 and homology-directed DNA repair pathway, where the use of antibiotics is avoided. Site-specific integration of genes provides a point of control in the development of recombinant protein producing CHO cells by avoiding position effect variation and phenotypic heterogeneity associated with random gene integration. The technique allows the use of predetermined loci in the genome where, for example, the gene expression levels and stability have been assessed over time. Resulting in predictable levels of protein production and accelerated cell line generation. The need to screen thousands of clones will no longer be necessary if sufficient production levels can be achieved in this controllable manner. Site-specific integration into precise genomic locations is also necessary in order to characterize the function of regulatory elements in mammalian cells in their natural chromosomal position.

We opt to avoid the use of antibiotics in this study, as it has been shown that the expression of antibiotic selection markers can occupy a substantial amount of the protein production machinery [17]. Additionally, by avoiding the use of antibiotic selection we do not limit the use of the platform to the number of antibiotic selection markers that are available. This opens up for the possibility to consequently target several additional loci in the genome. Avoiding the use of antibiotics will possibly increase the control and predictability of the platform, as any changes to the environment cells are grown in can increase the genetic plasticity [18].

Accelerated Homology-Directed Targeted Integration of Transgenes in Chinese Hamster Ovary Cells Via CRISPR/Cas9 and Fluorescent Enrichment

Jae Seong Lee,[1] Lise Marie Grav,[1] Lasse Ebdrup Pedersen,[1] Gyun Min Lee,[1,2] Helene Faustrup Kildegaard[1]

[1]The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Hørsholm 2970, Denmark; telephone: +45 20124629; fax: +45 45258001; e-mail: hef@biosustain.dtu.dk

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

ABSTRACT: Targeted gene integration into site-specific loci can be achieved in Chinese hamster ovary (CHO) cells via CRISPR/Cas9 genome editing technology and the homology-directed repair (HDR) pathway. The low efficiency of HDR often requires antibiotic selection, which limits targeted integration of multiple genes at multiple sites. To improve HDR-mediated targeted integration, while avoiding the use of selection markers, chemical treatment for increased HDR, and fluorescent enrichment of genome-edited cells was assessed in CHO cells. Chemical treatment did not improve HDR-mediated targeted integration. In contrast, fluorescent markers in Cas9 and donor constructs enable FACS enrichment, resulting in a threefold increase in the number of cells with HDR-mediated genome editing. Combined with this enrichment method, large transgenes encoding model proteins (including an antibody) were successfully targeted integrated. This approach provides a simple and fast strategy for targeted generation of stable CHO production cell lines in a rational way.

Biotechnol. Bioeng. 2016;113: 2518–2523.

© 2016 Wiley Periodicals, Inc.

KEYWORDS: Chinese hamster ovary cells; CRISPR/Cas9; fluorescent enrichment; homology-directed repair; targeted integration

Introduction

The extensive use of Chinese hamster ovary (CHO) cells as a production host for complex therapeutic proteins, has long been

based on random integration, large-scale screening for highly productive cell lines and process optimization (Kim et al., 2012). In order to meet industrial requirements, such as high and stable transgene expression, and increasing demands for new therapeutic proteins, CHO cell line engineering is transitioning into a more rational and targeted approach (Nielsen and Borth, 2015). The ability to make targeted modifications of the CHO genome drastically improved with the bacterial clustered regularly interspaced short palindromic repeat (CRISPR)/CRISPR-associated protein 9 (Cas9) gene editing system, which is based on the generation of double-strand breaks (DSBs) at desired locations in the genome and consequent gene modification using the DNA repair pathways, non-homologous end joining (NHEJ), or homology-directed repair (HDR) (Lee et al., 2015a). It allows for the direct manipulation of existing genes as well as insertion of new genes into desired genomic locations.

NHEJ is the predominant form of DNA repair in CHO cells (Mansour et al., 2008), causing a low efficiency of CRISPR/Cas9-mediated targeted gene insertions by HDR. This makes isolation of clones with targeted gene insertions time-consuming and labor intensive, creating a desire for more efficient approaches to accelerate genome engineering. We recently demonstrated efficient and precise targeted integration of a gene cassette (3.7 kb in size) into site-specific loci in CHO cells, using CRISPR/Cas9 genome editing system and donor plasmid harboring homology sequences (Lee et al., 2015b). It is a simple method, providing a high percentage of clones with targeted gene insertions based on HDR and antibiotic selection. There are a few drawbacks with our previous method: antibiotic selection is not desired in an industrial setting due to high costs, selection genes are available in a limited number—making it difficult to employ this method for multiple integrations, and selection genes can potentially impose a further metabolic burden on the cells (Veraitch and Al-Rubeai, 2005). It is thus attractive to have efficient marker-free systems.

In order to avoid using antibiotic selection, we tested two approaches: chemical treatment and FACS enrichment; aiming to

enhance HDR efficiency or to enrich targeted integrants resulting from HDR events (Fig. 1A). To investigate which approaches are most effective with regard to improving targeted integration efficiency, we used short DNA fragments (a BamHI restriction site with linker sequences, 15 bp in total) as gene of interest (GOI). This construct allows for deep sequencing analysis to determine the relative ratio of Cas9-induced DSBs being repaired by the NHEJ versus HDR pathway (Fig. 1B).

The efficiency of HDR-mediated genome editing in human cells has previously been increased several fold by chemical treatment for inhibition of NHEJ (Chu et al., 2015; Maruyama et al., 2015) or cell cycle synchronization to select for HDR (Lin et al., 2014). We therefore chose the DNA ligase IV inhibitor Scr7 (Chu et al., 2015; Maruyama et al., 2015) and the G2/M cell cycle inhibitor lithium chloride (LiCl) (Ha et al., 2014) as chemical additives to test their effects on CRISPR/Cas9-mediated genome editing in CHO cells. Based on previous studies (Chu et al., 2015; Ha et al., 2014; Maruyama et al., 2015), we chose five concentrations of Scr7 (0.1, 1, 5, 10, and 20 μM) and 10 mM LiCl. Scr7 inhibited cell growth of CHO cells in a dose-dependent manner, while cell viability was maintained (Supplementary Fig. S1). Highest insertion efficiency was reported upon treatment with 1 μM Scr7 in human cells (Chu et al., 2015; Maruyama et al., 2015). The use of 10 mM LiCl was validated in CHO cell cultures to induce cell cycle arrest at the G2/M phase (Supplementary Fig. S2) while maintaining cell viability (Ha et al., 2014).

CHO cells were transfected with GFP_Cas9 and sgRNA expression vectors together with donor plasmids targeting two endogenous target sites, COSMC (C1GALT1C1) and FUT8 (Supplementary Table SI). Chemical additives were applied before and/or after transfection, and deep sequencing of target sites was performed (Methods). Transfection efficiency was estimated to be approximately 90% for replicates of both target sites in controls without chemical treatment (Supplementary Fig. S3). The use of Scr7 concentrations above 5 μM decreased transfection efficiency, in particular by 25–50% with 20 μM Scr7 regardless of the timing of the treatment. Severe cell clumping was observed in the samples with 20 μM Scr7, and likely led to the large difference in transfection efficiency between the two target sites and replicates. LiCl slightly reduced transfection efficiency (∼20%), while it was unaffected by pre-treatment (condition 3). We observed an average of 35.7/32.0% indel (NHEJ) frequencies (light gray) and 1.8/0.9% targeted integration (HDR) frequencies (black) at COSMC and FUT8 target sites, respectively in the absence of chemical additives (Fig. 1C and D). It corresponds to NHEJ versus HDR ratio of 20 and 35 at the two target sites, indicating a preference toward NHEJ. The HDR frequency was only enhanced at FUT8 in the presence of 1 μM Scr7 after transfection, but it was ineffective at COSMC. Overall, we observed no significant increase in HDR efficiency upon Scr7 and LiCl treatment (Fig. 1C and D, HDR fold), but rather negative effects on total editing including both NHEJ and HDR. The negative effects could be due to decreased transfection efficiency. However, 1 μM Scr7 and 10 mM LiCl before transfection did not lead to increase in HDR frequencies despite similar levels of transfection efficiency. This limited effect may arise from the refractory nature of CHO cells against chemical manipulation of DNA repair pathways caused by different kinetics of DSB repair. Recent studies also reported

minimal effects of chemical treatment in mammalian systems, suggesting context or cell line dependent responses to these treatments (Pinder et al., 2015; Song et al., 2016). Genetic manipulation of key molecules of NHEJ/HDR (Ismail et al., 2015; Kim et al., 2001; Lambert and Lopez, 2000; Mansour et al., 2008) and investigation of alternative NHEJ pathways (Ceccaldi et al., 2015; Mateos-Gomez et al., 2015) could perhaps unveil the underlying mechanism.

Our second approach is based on a previous study where we show that CRISPR/Cas9 targeted cells are enriched via a GFP-labelled Cas9 expression vector combined with FACS, resulting in cell populations with high genome editing events (Grav et al., 2015). By applying two fluorescent proteins, GFP 2A peptide-linked Cas9 and donor DNA expressing mCherry, we can enrich cells transfected with both Cas9 and donor DNA that are likely to be exposed to genome editing events. Since mCherry is placed outside the homology regions of the donor DNA, it is not integrated upon correct targeting of the donor plasmids to the desired locus via HDR (Fig. 1B).

Fluorescent enrichment resulted in a threefold increase in the number of cells with HDR-mediated genome editing at the two target sites relative to non-enriched samples, yielding HDR frequencies of ∼7% (Fig. 1E). NHEJ frequencies up to 85% were also achieved upon enrichment, but no effects were observed in control samples without sgRNAs. Time course analysis from day 0 to 5 after enrichment showed no significant changes in HDR and NHEJ frequencies, revealing that genome editing events occurred within 48 h post-transfection. Overall, these data demonstrate the feasibility of efficient isolation of genome-edited cells with targeted integration.

To assess the effect of fluorescent enrichment on targeted integration of longer DNA fragments, we inserted large transgenes encoding model proteins, erythropoietin (EPO, 2.1 kb in size excluding homology arms), or antibody (Rituximab, 5.1 kb in size excluding homology arms) into COSMC (Fig. 2A). Successful targeted integration of transgenes in enriched pool of cells was verified by 5′/3′ junction PCR targeting genome-donor boundaries (Fig. 2B, Lanes 4 and 8). We then isolated non-fluorescent clonal cells from each enriched pool of cells for EPO or Rituximab, in which 62% or 80% of cells were both mCherry and GFP negative. Selection of non-fluorescent clonal cells excludes many cells with randomly integrated donor DNA and facilitates a screening process to select proper targeted integrants, as only the transgene expression cassettes inside homology regions can be inserted through HDR. 81 EPO and 76 Rituximab clonal cells were acquired from six 96-well plates of single cell sorted clones (i.e., Sorting survival rate of 10.5% or 9.9%). One out of 81 EPO clonal cells and one out of 76 Rituximab clonal cells were verified as correct targeted integrants (Supplementary Table SIII) by out–out PCR using a primer set targeting outside of the homology arms (Fig. 2C, Lane 2: EPO, amplicon size of 3.7 kb, Lane 4: Rituximab, amplicon size of 6.8 kb) compared with non-targeted cells (Fig. 2C, Lanes 1 and 3: amplicon size of 1.6 kb). The generated targeted integrants harbored the same copy number of transgenes compared to a mCherry expressing targeted integrant at the same locus that was derived from the previous selection-dependent method (Clone #1, Lee et al., 2015b) (Fig. 2D). The Rituximab construct consists of two
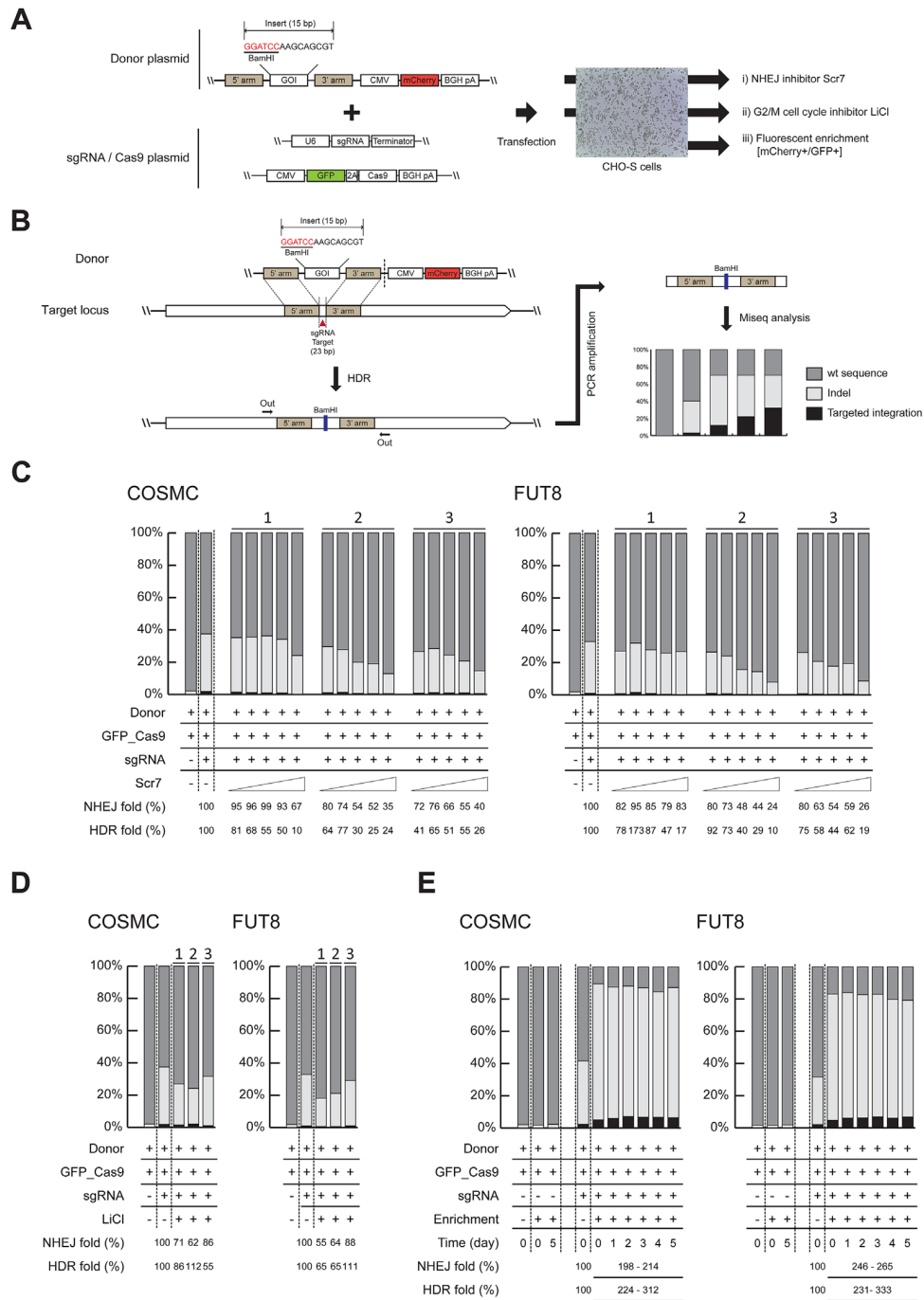
Figure 1. The effect of chemical treatment or fluorescent enrichment on NHEJ-directed indel generation and HDR-directed targeted integration in CHO cells. (**A**) Schematic illustration of two approaches for increasing targeting efficiency including chemical treatment (i and ii) and fluorescent enrichment (iii). The donor plasmid contains target GOI, BamHI restriction site with additional linker sequence flanked by homology arms and a mCherry expression cassette outside homology arms. In GFP_Cas9 expression plasmid, GFP is linked via a 2A peptide to Cas9. (**B**) Experimental schematic of detection of genome editing. Upon DSB generation and subsequent HDR-mediated targeted integration of GOI, targeted genomic regions were PCR-amplified with primers that anneal outside of the homology arms. Targeted deep sequencing of the PCR products was performed using MiSeq. The percentage of wild type sequences (dark gray), NHEJ-directed indel sequences (light gray), and sequences corresponding to precise insertion of target GOI (black) are illustrated in the bar plot. (**C** and **D**) Analysis of genome editing events upon chemical treatment. Cells were treated with five different concentrations (0.1, 1, 5, 10, and 20 μM) of Scr7 (**C**) or 10 mM concentration of LiCl (**D**) immediately after transfection (denoted by 1), both before and after transfection (denoted by 2), or before transfection only (denoted by 3) (See Methods). Genomic DNA harvested 2 days after transfection was subjected to deep sequencing analysis as shown in panel B. (**E**) Analysis of genome editing events upon fluorescent enrichment. After 2 days of transfection, cells were FACS sorted for mCherry and GFP fluorescent positive populations (day 0), followed by 5 incubation days. Time course analysis of genome editing events was performed on genomic DNA as described earlier. NHEJ and HDR fold change indicates the relative percentage values of indel and targeted integration rate, compared to control samples transfected with GFP_Cas9, sgRNA, and donor plasmid without chemical treatment or fluorescent enrichment—set at 100%. Results of approximately 50,000–230,000 sequencing reads from two biological replicates are shown.
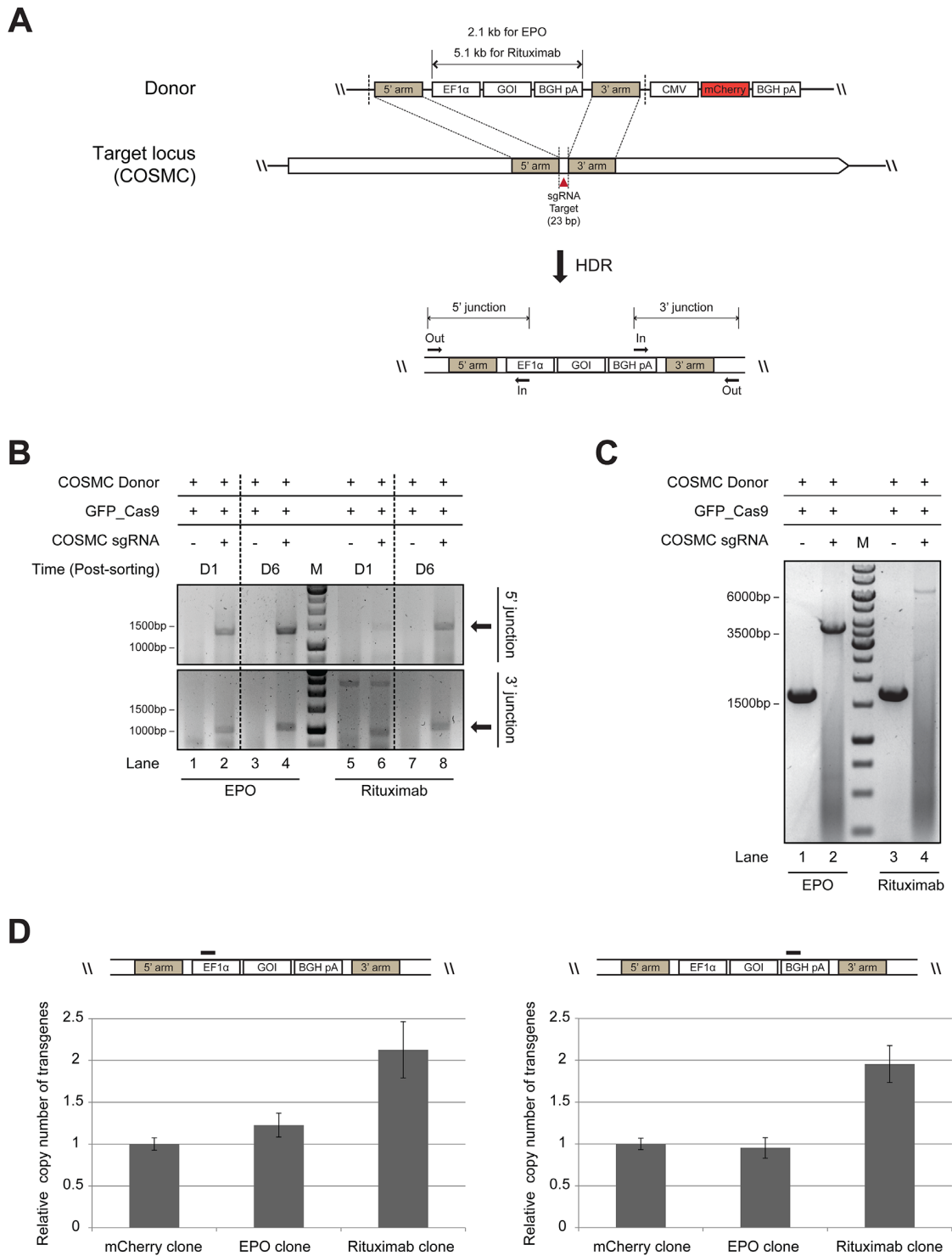
Figure 2. Application of fluorescent enrichment for targeted integration of large transgenes encoding model proteins. (**A**) Schematic illustration of the targeting strategy. The donor plasmids contain expression cassettes of two model proteins, EPO or antibody (Rituximab), as described in Figure 1A. Integration events were analyzed using primer sets that anneal either inside and outside (5′/3′ junction regions) or outside of the homology arms. (**B** and **C**) Detection of integration events upon fluorescent enrichment using 5′/3′ junction PCR (**B**) or out–out PCR (**C**). Genomic DNA harvested from enriched pool of cells was analyzed for 5′/3′ junction regions. After single cell sorting, genomic DNA harvested from clonal cells was analyzed for regions outside of the homology arms (Lanes 2 and 4 in **C**). Lanes 1 and 3 in **C** serve as a negative control, using genomic DNA of a pool of cells transfected with only GFP_Cas9 and donor plasmid to provide PCR products of wild type COSMC locus. M, 1 kb DNA ladder (**D**) Relative copy number of transgenes in clonal cells with targeted integration. Target amplicon regions on genomic DNA are depicted. Genomic DNA of COSMC targeted integrant expressing mCherry (Lee et al., 2015b) was used as reference. The error bars represent the standard deviations ($n = 3$).

expression cassettes encoding either heavy or light chains, thus twofold higher copy number was observed. Targeting results comparable to those obtained from the previous integration platform supports the reliability of the current integration platform.

In this way, we not only shorten the clone generation time, but also generate clonal CHO cell lines with site-specific, marker-free (no antibiotic selection required), and clean (no unwanted DNA present) targeted integration of transgenes from one 96-well plate of single-cell derived clones, offering several advantages over the previous method. Improvement in HDR efficiency through small molecule screening (Yu et al., 2015) or transient modulation of DNA repair proteins (Chu et al., 2015) could fully optimize targeted gene insertion in CHO cells. Taken together, the present platform has the huge potential to accelerate targeted generation of stable CHO production cell lines. It could ultimately lead to a rational construction of CHO production cell lines assuring stable high expression levels in a short time when combined with the selection of ideal transgene insertion sites.

Methods

Plasmid Construction

GFP_Cas9 and sgRNA expression vectors used in this study were constructed as described previously (Supplementary Table SI). Donor plasmids harboring BamHI or model protein-encoding genes were constructed via uracil-specific excision reagent (USER) cloning method (Lund et al., 2014) using primers listed in Supplementary Table SII. For construction of donor plasmids harboring BamHI, four PCR amplified DNA parts including 5′ homology arm with BamHI sequence at 3′ end, 3′ homology arm, mCherry expression cassette, and backbone were assembled with USER enzyme (New England Biolabs, Ipswich, MA). The EPO donor plasmid was made of six DNA parts that include 5′ and 3′ homology arms, EF-1α promoter, EPO-BGH pA, mCherry expression cassette, and backbone. The antibody (Rituximab) construct was assembled using seven DNA bricks: 5′ and 3′ homology arms, EF-1α promoter, heavy chain-BGH pA, light chain expression cassette (EF-1α—light chain—BGH pA), mCherry expression cassette, and backbone. The length of each 5′ and 3′ homology arm was set at 750 bp. Plasmids used as PCR templates for EPO and antibody parts have previously been described (Kol et al., 2015). Purified plasmid was obtained using NucleoBond Xtra Midi EF (Macherey-Nagel, Düren, Germany) according to manufacturer's instructions following sequence verification.

Cell Culture and Transfection

CHO-S cells were grown in CD CHO medium supplemented with 8 mM L-glutamine (Life technologies, Thermo Scientific, Rockford, IL) and cultivated in Corning vent cap shake flasks (Sigma–Aldrich, St. Louis, MO) in a humidified incubator at 120 rpm, 37°C, and 5% CO_2. Cell growth and viability was monitored using the NucleoCounter NC-200 Cell Counter (ChemoMetec, Allerod, Denmark) based on two fluorescent dyes, acridine orange and DAPI for the total and dead cell populations, respectively. Cells at a concentration of 1×10^6 cells/mL were transfected with donor

plasmid, expression vectors encoding GFP_Cas9, and/or sgRNA targeting the integration site at a ratio of 1:1:1 (w:w:w). Transfection was performed in 6- or 12-well plates (BD Biosciences, San Jose, CA) with a total of 3×10^6 cells and 3.75 μg of DNA or 1×10^6 cells and 1.25 μg of DNA, respectively using FreeStyle™ MAX reagent together with OptiPRO SFM medium (Life Technologies) according to the manufacturer's recommendations. Transfection efficiency was measured based on % population of mCherry-positive cells using the mask + target 1 application on Celigo Cell Imaging Cytometer (Nexcelom Bioscience, Lawrence, MA) as described previously (Lee et al., 2015b).

Cell Cycle Analysis

The cell cycle was analyzed using a Celigo Cell Imaging Cytometer. Cells were incubated with culture media containing various concentrations (1, 5, 10, 15, and 20 mM) of LiCl (Sigma–Aldrich) for 24 h, followed by Vybrant® DyeCycle™ Green stain (final stain concentration of 10 μM, Thermo Fisher Scientific) for 60 min in the dark according to manufacturer's instructions. Approximate 10,000–15,000 cells were analyzed using the Expression Analysis application on Celigo.

Chemical Treatment

Three treatment conditions were tested in this study: (i) cells were incubated with Scr7 (0.1, 1, 5, 10, and 20 μM, ApexBio, Houston, TX) or LiCl (10 mM) for 48 h after transfection; (ii) cells were treated with Scr7 or LiCl for both 24 h before transfection and 48 h immediately after transfection; and (iii) cells were treated with Scr7 or LiCl for 24 h before transfection, washed to remove the chemical upon transfection, and then incubated without any chemical for an additional 48 h until analysis.

Fluorescent Enrichment

A 48 h post-transfection, cells were bulk sorted for both mCherry and GFP positive cell population using a BD FACSJazz cell sorter (BD Biosciences, San Jose, CA) according to a previous study (Grav et al., 2015). Bulk sorted cells were maintained in 80% CD CHO medium supplemented with 8 mM L-glutamine, and 1% Antibiotic-Antimycotic (Life Technologies) and 20% 20 μm filtered spent medium for 1–2 weeks in order to make them restore and further expand before single cell sorting. Single cell sorting was performed as described previously (Grav et al., 2015) upon gating for mCherry and GFP negative cell population to exclude cells with randomly integrated donor DNA.

Genomic DNA Extraction and PCR Amplification of Target Region

Genomic DNA was extracted from cell pellets (cell pools and clonal cells) using QuickExtract DNA extraction solution (Epicentre, llumina, Madison, WI) according to the manufacturer's instructions. GeneJET Genomic DNA Purification Kit was used for genomic DNA preparation for copy number analysis, as described below. 5′/3′ junction PCR was carried out using DreamTaq DNA

polymerase (Thermo Fisher Scientific) by touchdown PCR (95°C for 2 min; 10×: 95°C for 30 s, 65–55°C [−1°C/cycle] for 30 s, 72°C for 2 min; 30×: 95°C for 30 s, 55°C for 30 s, 72°C for 2 min; 72°C for 5 min). Out–Out PCR was performed using Phusion High-Fidelity PCR Master Mix (Thermo Fisher Scientific) by touchdown PCR (98°C for 30 s; 10×: 98°C for 10 s, 68–58°C [−1°C/cycle] for 30 s, 72°C for 1 min [BamHI], 2 min [EPO], or 3 min 30 s [Rituximab]; 30×: 98°C for 10 s, 58°C for 30 s, 72°C for 1 min [BamHI], 2 min [EPO], or 3 min 30 s [Rituximab]; 72°C for 10 min). PCR primers are listed in Supplementary Table SII.

Deep Sequencing Analysis

Deep sequencing analysis of PCR products was performed on a MiSeq Benchtop Sequencer (Illumina, San Diego, CA) using the protocol described previously (Grav et al., 2015). As described in Figure 1B, PCR amplicons for deep sequencing were generated from the out–out PCR products to avoid amplifying the donor plasmids.

Quantitative Real-Time PCR (qRT-PCR) for Copy Number Analysis

Relative copy number of transgenes was determined using Brilliant III Ultra-Fast SYBR® Green QPCR Master Mix (Agilent Technologies, Santa Clara, CA) on Mx3005P qPCR System (Agilent Technologies) as described previously (Lee et al., 2015b). Target genomic regions including EF-1α, BGHpA, and Vinculin were amplified using primers listed in Supplementary Table SII that were validated by melting curve analysis and agarose gel electrophoresis. A delta–delta threshold cycle ($\Delta\Delta C_T$) method was applied using a COSMC targeted integrant expressing mCherry (Clone #1) in which mCherry coding sequence is flanked by EF-1α promoter and BGHpA sequences (Lee et al., 2015b), as calibrator. As unknown samples harbored coding sequences flanked by the same EF-1α promoter and BGHpA sequences, EF-1α promoter and BGHpA amplicons were used as target GOI for comparative quantification. Vinculin was used as an internal control gene for normalization.

References

Ceccaldi R, Liu JC, Amunugama R, Hajdu I, Primack B, Petalcorin MI, O'Connor KW, Konstantinopoulos PA, Elledge SJ, Boulton SJ, Yusufzai T, D'Andrea AD. 2015. Homologous-recombination-deficient tumours are dependent on Polθ-mediated repair. Nature 518(7538):258–262.

Chu VT, Weber T, Wefers B, Wurst W, Sander S, Rajewsky K, Kühn R. 2015. Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells. Nat Biotechnol 33(5):543–548.

Grav LM, Lee JS, Gerling S, Kallehauge TB, Hansen AH, Kol S, Lee GM, Pedersen LE, Kildegaard HF. 2015. One-step generation of triple knockout CHO cell lines using CRISPR/Cas9 and fluorescent enrichment. Biotechnol J 10(9):1446–1456.

Ha TK, Kim YG, Lee GM. 2014. Effect of lithium chloride on the production and sialylation of Fc-fusion protein in Chinese hamster ovary cell culture. Appl Microbiol Biotechnol 98(22):9239–9248.

Ismail IH, Gagné JP, Genois MM, Strickfaden H, McDonald D, Xu Z, Poirier GG, Masson JY, Hendzel MJ. 2015. The RNF138 E3 ligase displaces Ku to promote DNA end resection and regulate DNA repair pathway choice. Nat Cell Biol 17(11):1446–1457.

Kim JY, Kim YG, Lee GM. 2012. CHO cells in biotechnology for production of recombinant proteins: Current state and further potential. Appl Microbiol Biotechnol 93(3):917–930.

Kim PM, Allen C, Wagener BM, Shen Z, Nickoloff JA. 2001. Overexpression of human RAD51 and RAD52 reduces double-strand break-induced homologous recombination in mammalian cells. Nucleic Acids Res 29(21):4352–4360.

Kol S, Kallehauge TB, Adema S, Hermans P. 2015. Development of a VHH-based erythropoietin quantification assay. Mol Biotechnol 57(8):692–700.

Lambert S, Lopez BS. 2000. Characterization of mammalian RAD51 double strand break repair using non-lethal dominant-negative forms. EMBO J 19(12):3090–3099.

Lee JS, Grav LM, Lewis NE, Faustrup Kildegaard H. 2015a. CRISPR/Cas9-mediated genome engineering of CHO cell factories: Application and perspectives. Biotechnol J 10(7):979–994.

Lee JS, Kallehauge TB, Pedersen LE, Kildegaard HF. 2015b. Site-specific integration in CHO cells mediated by CRISPR/Cas9 and homology-directed DNA repair pathway. Sci Rep 5:8572.

Lin S, Staahl BT, Alla RK, Doudna JA. 2014. Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery. Elife 3:e04766.

Lund AM, Kildegaard HF, Petersen MB, Rank J, Hansen BG, Andersen MR, Mortensen UH. 2014. A versatile system for USER cloning-based assembly of expression vectors for mammalian cell engineering. PLoS ONE 9:e96693.

Mansour WY, Schumacher S, Rosskopf R, Rhein T, Schmidt-Petersen F, Gatzemeier F, Haag F, Borgmann K, Willers H, Dahm-Daphi J. 2008. Hierarchy of nonhomologous end-joining, single-strand annealing and gene conversion at site-directed DNA double-strand breaks. Nucleic Acids Res 36(12):4088–4098.

Maruyama T, Dougan SK, Truttmann MC, Bilate AM, Ingram JR, Ploegh HL. 2015. Increasing the efficiency of precise genome editing with CRISPR-Cas9 by inhibition of nonhomologous end joining. Nat Biotechnol 33(5):538–542.

Mateos-Gomez PA, Gong F, Nair N, Miller KM, Lazzerini-Denchi E, Sfeir A. 2015. Mammalian polymerase θ promotes alternative NHEJ and suppresses recombination. Nature 518(7538):254–257.

Nielsen L, Borth N. 2015. Editorial: On the cusp of rational CHO cell engineering. Biotechnol J 10(7):929–930.

Pinder J, Salsman J, Dellaire G. 2015. Nuclear domain 'knock-in' screen for the evaluation and identification of small molecule enhancers of CRISPR-based genome editing. Nucleic Acids Res 43(19):9379–9392.

Song J, Yang D, Xu J, Zhu T, Chen YE, Zhang J. 2016. RS-1 enhances CRISPR/Cas9- and TALEN-mediated knock-in efficiency. Nat Commun 7:10548.

Veraitch FS, Al-Rubeai M. 2005. Enhanced growth in NS0 cells expressing aminoglycoside phosphotransferase is associated with changes in metabolism, productivity, and apoptosis. Biotechnol Bioeng 92(5):589–599.

Yu C, Liu Y, Ma T, Liu K, Xu S, Zhang Y, Liu H, La Russa M, Xie M, Ding S, Qi LS. 2015. Small molecules enhance CRISPR genome editing in pluripotent stem cells. Cell Stem Cell 16(2):142–147.

Supporting Information

Additional supporting information may be found in the online version of this article at the publisher's web-site.

4.3 Applicability and future perspectives

Using CRISPR/Cas9 for site-specific integration of genes provides a completely new and extremely flexible way to integrate genes in the genome. It allows targeting of several more annotated loci than previous genome-editing tools, which lets us test many more genomic locations for potential suitability for stable and high transgene expression. Not being limited by the number of available selection markers also opens the possibility to insert genes in many different loci in the same system. This provides a huge advantage if there is a need to express different genes at different levels, which often is the case for genes working in consort with each other. The efficiency that we have managed to reach in this study, is still quite low and requires screening of a significant number of clones. However, if coupled with recombinase-mediated cassette exchange (RMCE) landing pads, as demonstrated in Chapter 5, the targeted integration site becomes re-usable, leaving the current efficiency more than sufficient.

Chapter 5

Minimized clonal variation for improved comparative systems biology studies

In this chapter, we present a study where we have combined site-specific gene integration with a recombinase-mediated cassette exchange (RMCE) method to create isogenic clones. In order to study and understand the effect of genes or elucidate why it is difficult to produce certain proteins, it is necessary to utilize systems that are comparable. Immortalized cell lines, like CHO, are known for their inherent genomic instability, making it difficult to assess the effect different transgenes impose on the cells. The necessity to use isogenic cells for comparative studies is due to the difficulty in distinguishing if an observed change in phenotype is related to the actual introduction of a transgene to the CHO cells, or if it simply transpires from clonal variation.

A transcriptomics analysis is performed on the isogenic clones generated in the study, to illustrate the applicability of the technique for comparative studies like differential gene expression. Transcriptomics is the study of the complete set of transcripts generated by the genome, and is one of the most sensitive omics tools due to the fundamentally stochastic process of gene expression, and is highly influenced by phenotypic variation. Presenting an ideal case to test the applicability of isogenic cell lines in this setting.

Minimized clonal variation for improved comparative systems biology studies

Lise Marie Grav[1], Jae Seong Lee[2], Daria Sergeeva[1], Mikael Rørdam Andersen[3], Lars Keld Nielsen[1,4], Gyun Min Lee[1,5], 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 Biotechnology and Biomedicine, Technical University of Denmark, Kgs. Lyngby, Denmark
[4] Australian Institute for Bioengineering and Nanotechnology, The University of Queensland, Brisbane, Australia
[5] Department of Biological Sciences, KAIST, Daejeon, Republic of Korea

Key Words: Chinese hamster ovary cells; CRISPR/Cas9; targeted integration; recombinase-mediated cassette exchange; cell line engineering; transcriptome; stable gene expression; clonal variation

Abstract

Mammalian cell lines, such as Chinese hamster ovary (CHO) cells, are widely used to stably express transgenes for the production of complex biopharmaceuticals. Nevertheless, the effect of recombinant protein expression on the cell is difficult to investigate using current cell line development platforms, possessing high genomic and phenotypic diversity. It hampers comparative transcriptomic approach for discovery of potential engineering targets for the generation of CHO cell lines with improved protein production performance. Here, we present a cell line development platform where the genetic background that may influence large transcriptional variation is minimized, allowing us to study the impact of any transgenes on the global transcriptome of mammalian cells. The platform cell line was developed by clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated protein 9 (Cas9)-mediated site-specific integration of a landing pad for recombinase-mediated cassette exchange (RMCE). Upon RMCE we created derivative subclones stably expressing four commercially relevant recombinant proteins, which show an extraordinary comparability in growth and transcript levels of transgenes across all subclones. We further use these isogenic cell lines to examine their exact transcriptional response to recombinant proteins and observe low variation in global transcriptomic levels based on RNA-sequencing data. We also observe little variation in phenotypic behaviour and global gene expression within producer clones that are grown in unchanged environmental conditions for three months. Thus, this cell line development platform enables robust comparative studies of transgene effect on the cell. The platform has potential to accelerate the discovery and verification process of CHO cell line engineering targets when combined with omics techniques.

Introduction

Stable expression of transgenes in mammalian cells is crucial for both basic biological studies and biotechnology, especially for the production of biopharmaceuticals. Nowadays, the preferred choice for the production of biopharmaceuticals is CHO cell lines. Current mammalian cell line development 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 a clonal variation. This clonal heterogeneity obliges researchers to do time-consuming and labour-intensive screening to find cell lines with desired performance [1].

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 [2,3]. 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 emphasize the importance of genomic environment surrounding the transgene [4]. Moreover, upon random integration transgene cassette can be rearranged, affecting the original vector elements, which may confer unpredictable expression of the transgene (Lee et al., in preparation). Another source of heterogeneity is non-genetic variation due to epigenetics, stochastic gene expression and changing environmental conditions [5]. Currently, clonal variation is a barrier that hinders our ability to understand the biology of CHO cell lines.

One of the fundamental questions, which remains to be answered in the CHO field, is what impact a specific transgene has on the cell. Elucidation of the cell's response to a transgene is crucial for the discovery of potential engineering targets to 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 [6]. Thus, there is a need for a robust cell line development 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. Systems for site-specific integration of transgenes allows the generation of isogenic cell lines with low genomic variation and predictable and stable transgene expression [7]. The first generation of targeted integration approaches is based on site-specific recombination and RMCE, and have been widely used for genetic modification [8]. 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 27 kb [9]). As second-generation targeted integration tools, programmable endonucleases, such as CRISPR/Cas9, have been employed. CRISPR/Cas9 has been successfully used for site-specific integration of transgenes in mammalian cells, particularly in CHO [7,10] 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 [11].

Several site-specific recombinases have been described and used for engineering of mammalian genomes. The most commonly used recombinase 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 [8,12] By using a pair of heterotypic sites for a particular recombinase, site-directional RMCE can occur. In RMCE systems, 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 apply recombinase-mediated targeted integration in CHO cell lines. The Flp/*FRT* system has been used to generate cell lines expressing either monoclonal antibodies (mAbs) [13–15] and antibody fragments [16], or other recombinant proteins, such as erythropoietin (EPO) [17], tissue plasminogen activator [18], secreted alkaline phosphatase [18], bone morphogenic protein 2 [11] and G-protein coupled receptors [19]. Analogous to Flp/*FRT* system, Cre/*lox* recombinase system has been used to generate mAb-expressing CHO cell lines [20]. 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 [17], surface marker [13] or antibody [14]), in order to preselect transcriptionally active integration site. The site-specific integration of RMCE landing pad into predefined loci is an evident improvement of the RMCE platform. Targeted integration ensures that the landing pad will not disrupt any coding genes or essential genetic elements, which can have phenotypic consequences. By using CRISPR/Cas9-mediated homology-directed integration and fluorescent enrichment Flp/*FRT*- and Bxb1/*att*-landing pad master cell lines were created and used for efficient generation of mAb-expressing cell lines [21].

The use of RMCE-based site-specific integration platform allows faster cell line generation compared to traditional random integration approach, and 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 [14,22]. Interestingly, it has been shown that RMCE-derived clones, expressing similar monoclonal antibodies or antibody fragments, can have different productivity [22,16]. However, only one study went beyond descriptive analysis of clone performance, and tried to reveal the cause of different productivity between clones expressing two similar antibody fragments by proteomics analysis [23].

In this study, we developed a platform for the generation of isogenic CHO-S cell lines (isoCHO) to reveal changes in the transcriptome caused by recombinant protein expression. By CRISPR/Cas9-mediated targeted integration of a landing pad for Cre/*lox* RMCE, we created master cell lines (MCLs) with a stable and homogeneous expression of reporter fluorescent protein. Upon RMCE and fluorescence-activated cell sorting we generated derivative stable producer cell lines, expressing four recombinant proteins: etanercept (ETC), EPO, growth/differentiation factor 5 (GDF5) and C1 esterase inhibitor (C1INH). A set of isogenic producer cell lines, which were genetically identical aside from the expressed gene of interest (GOI), showed highly comparable growth and transcript levels of transgenes. This cell line development platform allowed us to evaluate the transcriptomic response to transgenes by comparative transcriptome analysis, while overcoming effects accrued by clonal variation in CHO cells.

Results

Stable master cell lines generated using CRISPR/Cas9 and a pre-selected 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 genomic region located close to essential genes, to potentially avoid subsequent loss of transgenes and transcriptional silencing. Based on our previously published method for CRISPR/Cas9-mediated homology-directed targeted integration [7], we inserted the RMCE "landing pad" into the preselected site, creating a MCL for the second stage of cell line generation (RMCE with GOIs). To ensure that the GOI only will be expressed from our target locus upon RMCE, we introduced *loxP/lox2272* sites [24] flanking the mCherry reporter gene (Fig. 1a). The EF-1α promoter (EP) driving the mCherry expression and the polyA signal are consequently located outside *lox* sites, creating a promoter/polyA trap for the incoming promoterless GOI upon co-transfection with Cre recombinase (Fig. 1b). This ensures that expression of the GOI only occurs upon exchange with mCherry. After CRISPR/Cas9-mediated targeted integration of the landing pad, selection, fluorescence-activated single cell sorting and clone expansion, 13 clones having high confluency in 96-well plates were identified as junction PCR-positive. Among the targeted integrants, seven clones contained one copy of the Cherry expression cassette (Fig. 1c). Out of two further characterized
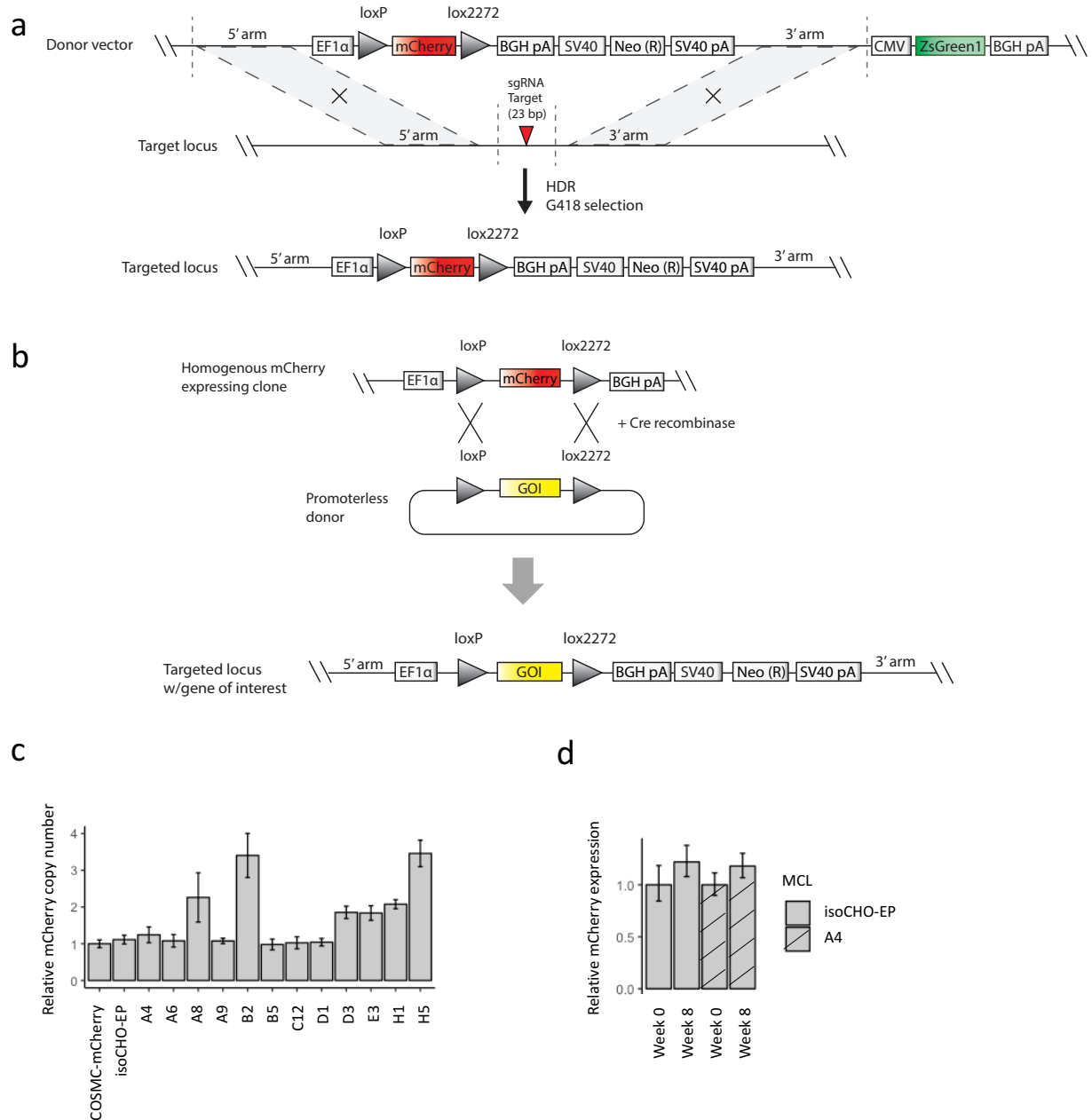
Fig. 1 | Targeted integration of landing pad into preselected locus using CRISPR/Cas9 and RMCE strategy. (a) Schematic representation of the targeting strategy of the landing pad into the preselected target locus. The donor plasmid contains the landing pad with *loxP* (34bp) and *lox2272* (34bp) sequences flanking mCherry gene. EF-1α promoter and BGHpA polyA is located outside the lox sequences allowing for both promoter and polyA trapping upon RMCE. **(b)** Schematic of RMCE strategy: co-transfection 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. **(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 [7]. 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).

clones, showing highly stable mCherry expression after an eight-week long cultivation period (Fig. 1d), one clone (isoCHO-EP) was selected as the MCL based on homogenous and high levels of mCherry expression (Supplementary Fig. S1 and Table S1).

Isogenic subclones expressing therapeutic proteins show similar growth and low transcriptional variation.
To generate a set of isogenic subclones from the master cell line, we cloned four promoterless vectors for RMCE with GOIs encoding recombinant proteins ETC, EPO, GDF5, or C1INH. After co-transfections 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 from the RMCE donor vector, using isoCHO-EP as gating control. For each GOI exchange, we verified exchange with GOI in generated subclones using insert PCR with primers aligning outside the *lox* sequences and sequenced the products. We selected 12 clones (three clones for each GOIs) for further analysis. To study the growth behaviour of these subclones, batch cultures in shake flasks were performed. All 12 subclones derived from isoCHO-EP displayed highly similar growth phenotypes (Fig. 2a). On day four in late exponential phase, cells were harvested for RNA-sequencing to capture the transcriptome for further analysis of the variations between all subclones expressing different recombinant proteins (dataset 1). Based on the transcriptomics data, highly comparable relative transgene expression levels were observed within biological replicates (three clones expressing the same GOI) (Fig. 2b), which was confirmed by qRT-PCR (Supplementary Fig. S2). The reproducibility among the biological replicates were evaluated following guidelines for technical replicates [25], with Spearman R^2>0.9 (Supplementary Fig. S3). To assess the variation in global gene expression between subclone sets expressing different GOIs, principal component analysis (PCA) of the transcriptomics data, excluding transgene expression, was performed. Little variation in the gene expression profiles of the isoCHO-EP subclones is 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 (Fig. 2c). To further assess the variation in all genes expressed in the subclones, differential gene expression (DGE) analysis was performed on the dataset 1 (Supplementary Fig. S4a). Only 160 out of a total of 12646 tested genes (1.27%) were significant differentially expressed, indicating very similar gene expression profiles of all 12 clones. From these data, we can conclude that our two-stage isogenic cell lines generation platform facilitated efficient generation of highly reproducible subclones with low variations 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 ETC and C1INH were cultivated up to three months 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), mid-culture (1.5 months) and end of culture (3 months) were thawed and cultivated in batch cultures simultaneously. Unfortunately, one biological replicate of both ETC and C1INH was lost after 1.5 months and were therefore not included in the batch. The viable cell densities were analysed revealing almost no variation in growth irrespective of product or time of harvest indicating very stable and similar growth of the subclones (Fig. 3a). From the batch cultures, cells were harvested in late exponential phase for RNA-sequencing analysis (dataset 2). RNA-sequencing revealed low variation in
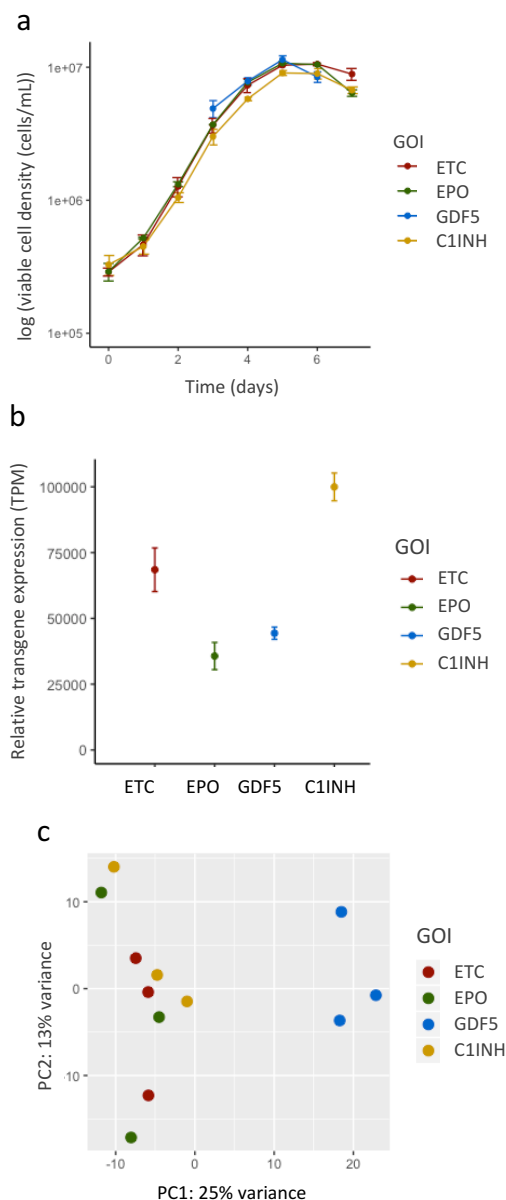
Fig.2 | Isogenic subclones generated by RMCE show low variation in phenotypes and transcriptomes. (a) Viable cell densities of isoCHO-EP subclones expressing ETC, EPO, GDF5 or C1INH. The error bars of each line represent the standard deviations of three isogenic subclones expressing the same GOI (n=3). **(b)** Relative levels of transgene expression, as measured in transcripts per kilobase million (TPM) of ETC, 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)** PCA of whole transcriptome data from isoCHO-EP subclones expressing ETC, EPO, GDF5 or C1INH sampled at day 4 in late exponential phase.

relative transgene expression levels over the time in culture, indicating stable expression of GOIs from the integration site (Fig. 3b). PCA of the transcriptomics data (not including transgenes) reveal 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
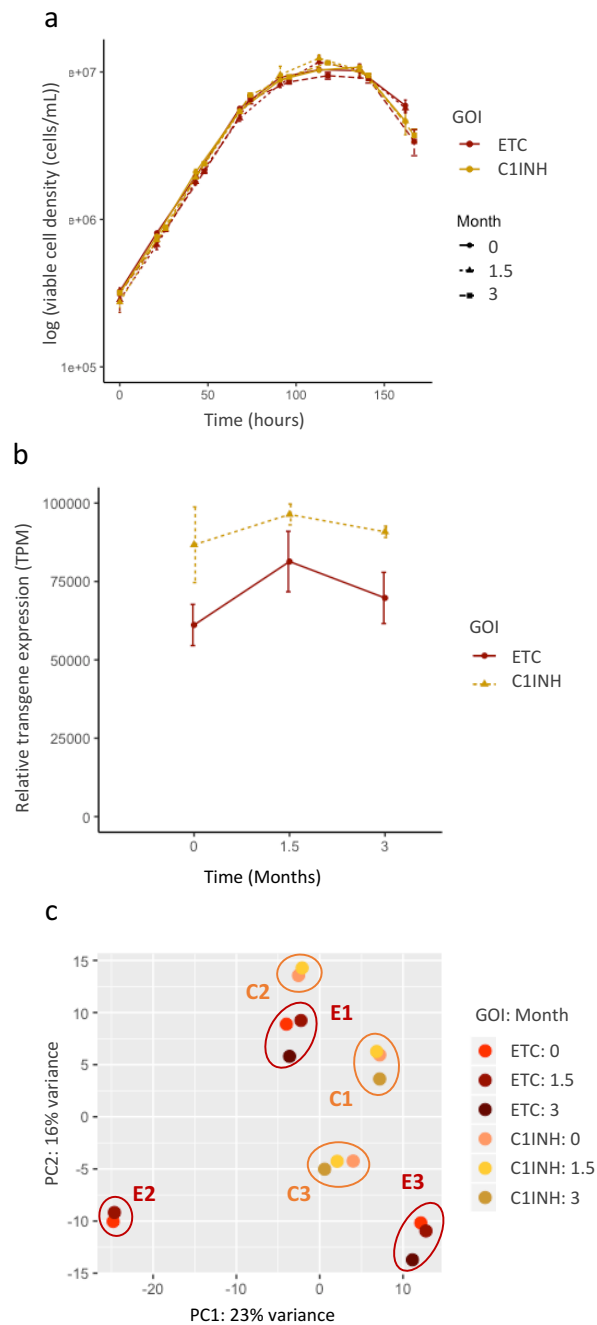
Fig.3 | Long-term stability of gene expression in isogenic subclones. (a) Viable cell densities of isoCHO-EP subclones expressing ETC and C1INH cultivated upon thawing of cells harvested after 0, 1.5 and 3 months in culture. The error bars represent the standard deviations of three (0 and 1.5 months) or two (3 months) subclones expressing the same GOI (n=2-3). **(b)** Levels of transgene expression over time of isoCHO-EP ETC 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)** PCA of whole transcriptome data from isoCHO-EP subclones expressing ETC 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 ETC producing subclones (E1, E2, E3) and C1INH producing subclones (C1, C2, C3), and no clustering observed due to time in culture.

individual subclones (Fig. 3c). The gene expression of biological replicates was deemed as reproducible, with Spearman R^2>0.90 (Supplementary Fig. S5). From the DGE analysis of subclones producing the same recombinant protein (Supplementary Fig. S4b), only 10 out of 12932 tested genes (0.08%) were differentially expressed over the three months period of passaging. This indicates very low changes in gene expression levels over long-term cultivation. Moreover, only 4 out of 12935 tested genes (0.03%), including transgenes, were detected as differentially expressed when comparing the ETC cultures at 0, 1.5 and 3 months versus the C1INH cultures at the same time points (Supplementary Fig. S4c), indicating a low or similar effect of ETC and C1INH expression on the transcriptome in these cells. Together, the data indicate that the generated clones behave very similarly and show low transcriptomic changes over time.

Different responses to transgene expression levels from same promoters in same target site. 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 therapeutic proteins. We created a new donor vector for targeted integration of the landing pad, where the full length human EF-1α promoter is replaced with a commercial composite promoter (CP) consisting of mouse CMV enhancer, minimal human EF-1α and HTLV RU5' 5'UTR region. After CRISPR/Cas9-mediated targeted integration, selection, single cell sorting and clone expansion, 11 targeted integrants with high confluency in 96-well plates were verified by junction PCR. Among the verified targeted integrants, seven clones contained one copy of the mCherry expression cassette (Fig. 4a). Of two selected clones, both showed stable mCherry expression after an eight-week long cultivation period (Fig. 4b). One clone (isoCHO-CP) was selected as a new master cell line based on homogenous mCherry expression (Supplementary Table S2) and that it showed about 1.8 times higher mCherry expression level than isoCHO-EP (Supplementary Fig. S6).

Similar to isoCHO-EP, isoCHO-CP was used to generate subclones expressing four recombinant proteins by RMCE. In addition, non-producing 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. 15 clones (three clones expressing each GOIs and three non-producer clones) derived from isoCHO-CP were selected for further analysis. To study the growth behaviour of these subclones, batch cultures were performed. Of the isoCHO-CP subclones, the three EPO and three GDF5 subclones grow similar as non-producing isoCHO-CP subclones. Interestingly, C1INH and ETC expressing clones grow to slightly lower maximum VCDs (Supplementary Fig. S7a). On day four in late exponential phase, cells were harvested for RNA-sequencing to analyse the variation in transcriptome of all 15 subclones (dataset 3). From the RNA-sequencing data, highly comparable relative expression levels were observed within each set of ETC and GDF5, while larger error bars were observed within the EPO and C1INH expression clone sets (Supplementary Fig. S7b). However, qPCR showed lower variation in gene expression among the EPO and C1INH clones (Supplementary Fig. S8). Unexpectedly, only EPO expression levels were increased by 1.7 fold in isoCHO-CP clones compared to isoCHO-EP (Fig. 4c). The expression level of ETC, GDF5 and C1INH were similar in both isoCHO-EP and isoCHO-CP clones.

Large transcriptional variation observed in relation to increased protein expression and clonal variation between MCLs. To study the variation in global gene expression between subclones expressing different

GOIs at different expression levels, PCA of the transcriptomics data from the 15 subclones was performed (dataset 3). PCA of gene expression data from isoCHO-CP subclones, not including transgenes, display slightly higher variation between subclones than dataset 1, resulting in PC1 explaining 33% of the variation and PC2 explaining 12% of the variation (Fig. 5a). The gene expression of biological replicates was compared and
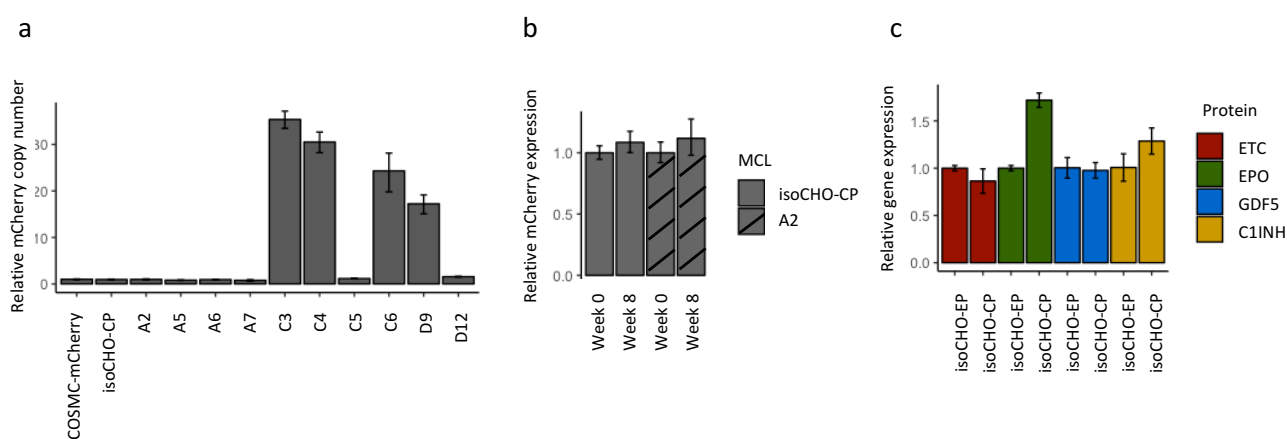
Fig. 4| Generation of clones with composite promoter. (a) Relative copy number of mCherry transgene in clonal cells. The plot shows the relative copy number of mCherry in of all targeted cell lines verified by junction PCR, in comparison to reference sample COSMC-mCherry from Lee et al [7]. The error 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 average value of isoCHO-EP subclones. The error bars represent the standard deviations of three isogenic clones expressing the same GOI (n=3).

deemed to be reproducible with Spearman $R^2>0.90$ (Supplementary Fig. S9). A large amount of differentially expressed genes (DEGs) were observed when comparing all five groups of subclones (Supplementary Fig. S4d), 4622 DEGs out of a total of 12967 tested genes (35.6%). With a log fold change cut-off at 1.5, the number of DEGs is reduced to 229 (5%). To check if the increase in DEG between isoCHO-CP subclones could be due to the increase in EPO productivity, we compared the gene expression of all subclones excluding EPO producers (Supplementary Fig. S4e). This yielded a much lower number of DEGs: 137 out of a total of 12967 tested genes (1.1%). Interestingly, this smaller set of DEGs completely overlap in terms of gene identity with the DEGs identified in dataset 1.

We did some further analysis of the DEGs identified when including EPO producers in the comparison, to see if the DEGs were related to the increased transgene expression. DEGs with human homologs were subjected to enrichment analysis of canonical pathways showing that DEGs are most enriched in mitochondrial dysfunction, oxidative phosphorylation, and EIF2 signalling canonical pathways (Supplementary Fig. S10). Analysis of non-redundant gene interaction networks shows that the most represented cellular functions in our set of DEGs are molecular transport, post-translational modification, protein synthesis, lipid metabolism and cell signalling.

To assess clonal variation, we compared subclones expressing ETC and C1INH derived from both isoCHO-EP and isoCHO-CP together with their respective non-producing clones (dataset 4). PCA shows 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 (Fig. 5b). The PC2 only accounts for 4% and the distance between subclones derived from the same MCL. When comparing subclones from different MCLs (Supplementary Fig. S4g), 6884 out of a total of 13031 genes (52,8%), including transgenes, were identified as differentially expressed. With a log fold change cut-off at 1.5, the DEG are reduced to 1100 (8,4%). The DEG 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.
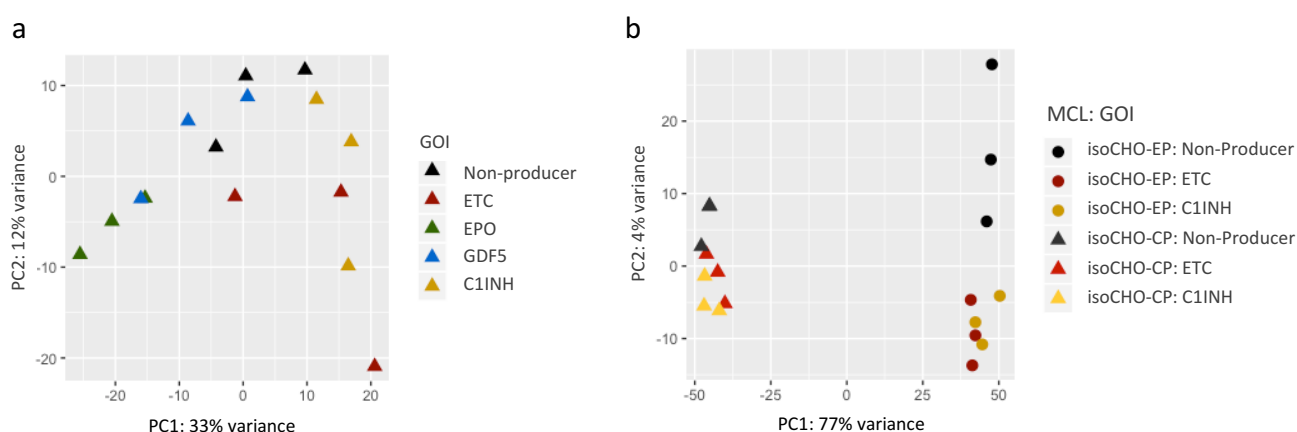
Fig. 5 | Variation within subclone populations. (a) PCA of whole transcriptome data from isoCHO-CP subclones expressing ETC, EPO, GDF5 or C1INH and non-producer subclones sampled at day 4 in late exponential phase. **(b)** PCA of whole transcriptome data from isoCHO-EP and isoCHO-CP subclones expressing ETC or C1INH and non-producer subclones sampled at day 4 in late exponential phase.

Discussion

Traditionally, mammalian cell lines are generated by random integration of transgenes leading to large uncontrollable variations in growth and transgene expression levels. Random integration creates a high diversity of clones, which can be beneficial in the search for high producing clones [3]. However, the large clonal variation complicates transcriptomics data analysis due to difficulties in deciphering which transcriptional changes are originating from transgene expression and which are originating from clonal variation [26]. 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 the generation of 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 over 1.5 to 3 months of cultivation. A similar strategy for the selection of target sites between essential genetic   elements has been applied in yeast [27].   Our preliminary finding might suggest that this strategy will be a successful selection criteria for target sites in CHO cells as well. The   second step of our approach allows fast and reproducible generation of cell lines originating from the same   master cell line that only differ in the transgene coding sequence, without the use of antibiotic selection 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 [21]. In their approach, 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 was required upon RMCE. Selection pressure applied after RMCE exchange could lead to some perturbations in the cell [3], which could explain the variable levels of expression in RMCE derived clones reported in their study. The design of RMCE   promoter/polyA 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 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 polyA tail. Removal of the NeoR marker could potentially release move transcriptional and translational capacity for growth and protein production in the host cell line as observed previously [28]. In addition, this could facilitate re-use of neoR selection marker for efficient future targeted integration events.

With our two-staged approach, the generated subclones expressing ETC, EPO, GDF5 and C1INH show very similar growth profiles between biological   replicates as well   as between different transgene   expressing clones (Fig. 2a and Supplementary Fig. S6). 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. The similar growth might also be due to a mild impact of RMCE on the MCLs phenotype or the subcloning of CHO-S cells   during the first CRISPR-mediated targeted integration step. 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 [7]. This could indicate that the CHO-S cells used as a starting point in both studies contain 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. Furthermore, Ko et al. shows that subclones of a mAb producing clone displays a wide variety in growth, titer and product quality [29]. This might be due to random integration of transgene in unstable regions in the parental clone population.

Large challenges in transcriptomics studies arises from the plasticity of CHO cell genomes and the cell's easy adaptation to environmental changes. It was therefore surprising to us that the isoCHO-EP clones expressing ETC and C1INH showed high stability in their transcriptome over 3 months in culture. This could be explained by the constant culture conditions and stability of the integration site. 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 [30].

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, DEG 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). Especially the isoCHO-EP -derived clones showed few DEGs (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 analyse if the clones indeed show similar uptake of nutrients and excretion of bi-products.

The transcriptomic analysis of C1INH and ETC producing clones together with non-producing clones originating from isoCHO-EP and isoCHO-CP showed low variation among the subclones originating from either isoCHO-EP and isoCHO-CP, indicating a low or similar effect of ETC and C1INH expression on the transcriptome in these cells. On the contrary, large variation was observed between the two groups of isogenic cell lines originated from different MCLs (Fig. 5b). This indicates that even the two MCLs with similar growth profiles could have quite different transcriptomic profiles, as similar subclones derived from the different MCLs display a large amount of DEGs (52,8%). A similar observation was reported in Orellana et. al, showing that two cell lines originated from the same cell pool have 58% of DEGs despite only modest differences in growth rate [26]. With such high degree of variation, it is almost impossible to use transcriptomics data to identify potential engineering targets without minimizing clonal variation, as was performed in our study.

The higher level of EPO expression in isoCHO-CP-derived clones allowed us to further investigate the transcriptional impact of EPO expression in the cell lines. Increase in EPO productivity 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 down-regulated, indicating that the "metabolic effort" from these pathways has been rerouted to protein production. Down-regulation of oxidative phosphorylation genes was also reported in producer HEK239 cultures [31]. EIF2 signaling is known to be involved in global translational control and has previously been shown to regulate translation and protein production in CHO cells [32]. The most represented cellular functions in our set of DEGs were molecular transport, post-translational modification, protein synthesis, lipid metabolism and cell signaling. These are all cellular functions we would expect to change when protein production is increased, indicating that the increase in DEG 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. It can be applied in numerous studies ranging from multi-omics studies, studying the effect of media supplementations and the impact of specific genes or genetic elements. The platform can be further advanced by inserting the landing pad into multiple sites,

combining it with inducible promoters or promoters of different strengths, in order to study multiple genes in coherence to each other.

Methods

Plasmids and plasmid construction. GFP_2A_Cas9 and sgRNA vectors were constructed as described previously [33]. sgRNA target sequence was designed manually after selection of target loci. 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. Plasmids used as PCR templates for EF-1α sequence, NeoR and ZsGreen expression cassettes have been described previously [7]. 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 [7]. The composite promoter (CP) was amplified from pDRIVE5-GFP-2 plasmid included in Invivogen's PromTest™ kit (version # 13F06-MM). Promoterless RMCE vectors was made of two parts, the backbone and GOI (ETC, EPO, GDF5, C1INH) with *loxP* sequence at 5' end and *lox2272* sequence at 3' end. Plasmids used as PCR templates for ETC, EPO and C1INH have previously been described [34-36]. The plasmid with the coding sequence for human GDF5 was ordered via GeneArt and codon optimized for CHO, sequence listed in Supplementary Table S3. Primers are listed in Supplementary Table S4. 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).

Generation of RMCE master cell lines 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 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 was 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_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 mg/mL; Sigma-Aldrich). During selection with G418, 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 non-tissue treated 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 1x Antibiotic-Antimycotic (Gibco). 14 days after single-cell sorting, the entire volume of sub-confluent clones was transferred to 180 µl of CD CHO medium supplemented with 1x 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^{TM} DNA extraction solution (Epicentre, 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 2x 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 2 min; 30 x: 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 insert PCR are listed in Supplementary Table S4, primers for junction PCR are not disclosed due to proprietary reasons. 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 relative copy number of transgene. 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; 40X: 95°C for 20 s, 60°C 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) [7], using GAPDH as internal control gene for normalization. Primers are listed in Supplementary Table S6, and were validated by melting curve analysis and primer efficiency test. Each experiment included no template controls in every PCR running, 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^{TM} 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). To determine relative mCherry expression level Brilliant III Ultra-Fast SYBR® Green QPCR Master Mix (Agilent Technologies) was used, as described earlier. Relative recombinant protein expression levels were determined using TaqMan^{TM} Gene Expression Master Mix and custom-made TaqMan^{TM} probes for transgenes and pre-made TaqMan assay for GAPDH (Cg04424038_gh) (Thermo Fisher Scientific). Amplification was executed with the following conditions: 50°C for 2 min, 95°C for 10 min; 40X: 95°C for 15 s, 60°C for 1 min. Using $\Delta\Delta C_T$ method, relative expression level was calculated by normalization to expression

level of GAPDH. Primers and probes are listed in Supplementary Table S5, and were validated by melting curve analysis and primer efficiency test. Each experiment included no template controls in every PCR running, and had 3 replicates with 2 times repetition.

RMCE subclone generation. MCLs at concentration 1 x 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^{TM} MAX transfection reagent. For generation of non-producer clones only Cre-recombinase vector was transfected. Cells were passaged two times after transfection. After 7 days cell pool was 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 non-producer clones).

Batch cultivation. Cells were seeded at 3 x 10^5 cells/mL in 60 mL CD CHO medium, supplemented with 8 mM L-Glutamine and 1 μL/mL anti-clumping agent (Life Technologies), in Corning vent cap 250 mL shake flask (Sigma-Aldrich). 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 seven days in culture.

Long-term cultivation. 1 x 10^7 cells were thawed in 30 mL CD CHO medium, supplemented with 8 mM L-Glutamine and 1 μL/mL anti-clumping agent (Life Technologies), in Corning vent cap 125-mL shake flask (Sigma-Aldrich). They were passaged to 3 x 10^5 cells/mL three times a week for three months. Cells were cryopreserved at every passage: 1 x 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 x 10^6 cells were harvested on day four in late-exponential phase during batch cultivation. Cells were centrifuged at 1000g for 4 min and the supernatant was discarded. The pellet was completely re-suspended 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 was used for RNA sequencing (RIN or RIQ > 9).

Library preparation and RNA-seq. The total RNA samples were processed by the NGS lab at Novo Nordisk Foundation Center for Biosustainability (Technical University of Denmark), and prepared and depleted for ribosomal RNA 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 x 150 bp paired-end reads for the dataset 1 with about 10 million reads per sample, and NextSeq 500/550 High Output v2 kit and 300 bp paired-end reads for remaining subclone datasets analyzed (datasets 2, 3 and 4) with about 20 million reads per sample (Supplementary Table S6).

RNA-Seq Analysis and Differential Gene Expression Analysis. The transcript levels in all the experimental datasets, were quantified following the pipeline depicted in Supplementary Fig S11. 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 illumina 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 [37] 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 reference. The transgenes (ETC, EPO, GDF5, C1INH, and NeoR) coding sequences used in this study were all added to the transcriptome reference, and included in the DEG analysis. The salmon output was imported with tximport [38] into R [40] and analysed. Principal component analysis was performed by applying the prcomp R function to the complete dataset (excluding transgenes) and visualized in conjunction with ggplot2[41]. Differential gene expression was determined with limma-voom [39] using ANOVA-style F-test on the comparisons of interest (Supplementary Fig. S11). Genes were considered differentially expressed when the adjusted p value of the F-test was < 0.05. Functional enrichment analysis of canonical pathways and network analysis was carried out using QIAGEN's Ingenuity pathway analysis (IPA) tool. All statistically significant DEGs with a human homolog 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 DEGs while reducing the noise. These genes and their corresponding log fold change (between EPO and non-producer subclones) were integrated into a core analysis using the IPA tool.

Acknowledgements

References

1. Noh, S. M., Sathyamurthy, M. & Lee, G. M. Development of recombinant Chinese hamster ovary cell lines for therapeutic protein production. *Curr. Opin. Chem. Eng.* **2,** 391–397 (2013).

2. Vcelar, S. *et al.* Changes in chromosome counts and patterns in CHO cell lines upon generation of recombinant cell lines and subcloning. *Biotechnol. J.* (2018). doi:10.1002/biot.201700495

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

4. Wilson, C., Bellen, H. J. & Gehring, W. J. Position Effects on Eukaryotic Gene Expression. *Annu. Rev. Cell Biol.* **6,** 679–714 (1990).

5. Pilbrough, W., Munro, T. P. & Gray, P. Intraclonal protein expression heterogeneity in recombinant CHO cells.

PLoS One **4,** e8432 (2009).

6. Hansen, H. G., Pristovšek, N., Kildegaard, H. F. & Lee, G. M. 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 (2017).

7. 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.* **5,** 8572 (2015).

8. Wirth, D. *et al.* Road to precision: recombinase-based targeting technologies for genome engineering. *Curr. Opin. Biotechnol.* **18,** 411–419 (2007).

9. Duportet, X. *et al.* A platform for rapid prototyping of synthetic gene networks in mammalian cells. *Nucleic Acids Res.* **42,** 13440–13451 (2014).

10. 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.* **113,** 2518–2523 (2016).

11. Phan, Q. V., Contzen, J., Seemann, P. & Gossen, M. Site-specific chromosomal gene insertion: Flp recombinase versus Cas9 nuclease. *Sci. Rep.* **7,** 17771 (2017).

12. Merrick, C., Zhao, J. & Rosser, S. Serine Integrases: Advancing Synthetic Biology. *ACS Synth. Biol.* (2018). doi:10.1021/acssynbio.7b00308

13. Baumann, M. *et al.* Preselection of recombinant gene integration sites enabling high transcription rates in CHO cells using alternate start codons and recombinase mediated cassette exchange. *Biotechnol. Bioeng.* **114,** 2616–2627 (2017).

14. Zhang, L. *et al.* Recombinase-mediated cassette exchange (RMCE) for monoclonal antibody expression in the commercially relevant CHOK1SV cell line. *Biotechnol. Prog.* **31,** 1645–1656 (2015).

15. Nehlsen, K. *et al.* Recombinant protein expression by targeting pre-selected chromosomal loci. *BMC Biotechnol.* **9,** 100 (2009).

16. Mayrhofer, P. *et al.* Accurate comparison of antibody expression levels by reproducible transgene targeting in engineered recombination-competent CHO cells. *Appl. Microbiol. Biotechnol.* **98,** 9723–9733 (2014).

17. Kim, M. S. & Lee, G. M. Use of Flp-mediated cassette exchange in the development of a CHO cell line stably producing erythropoietin. *J. Microbiol. Biotechnol.* **18,** 1342–1351 (2008).

18.   Zhou, H., Liu, Z.-G., Sun, Z.-W., Huang, Y. & Yu, W.-Y. 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,** 122–129 (2010).

19.   Schucht, R. *et al.* Rapid establishment of G-protein-coupled receptor-expressing cell lines by site-specific integration. *J. Biomol. Screen.* **16,** 323–331 (2011).

20.   Kito, M., Itami, S., Fukano, Y., Yamana, K. & Shibui, T. Construction of engineered CHO strains for high-level production of recombinant proteins. *Appl. Microbiol. Biotechnol.* **60,** 442–448 (2002).

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

22.   Scarcelli, J. J., Shang, T. Q., Iskra, T., Allen, M. J. & Zhang, L. Strategic deployment of CHO expression platforms to deliver Pfizer's Monoclonal Antibody Portfolio. *Biotechnol. Prog.* **33,** 1463–1467 (2017).

23.   Sommeregger, W. *et al.* Proteomic differences in recombinant CHO cells producing two similar antibody fragments. *Biotechnol. Bioeng.* **113,** 1902–1912 (2016).

24.   Araki, K., Araki, M. & Yamamura, K.-I. Site-directed integration of the cre gene mediated by Cre recombinase using a combination of mutant lox sites. *Nucleic Acids Res.* **30,** e103 (2002).

25.   Conesa, A. *et al.* A survey of best practices for RNA-seq data analysis. *Genome Biol.* **17,** 13 (2016).

26.   Orellana, C. A. *et al.* RNA-Seq highlights high clonal variation in monoclonal antibody producing CHO cells. *Biotechnol. J.* (2018). doi:10.1002/biot.201700231

27.   Mikkelsen, M. D. *et al.* Microbial production of indolylglucosinolate through engineering of a multi-gene pathway in a versatile yeast expression platform. *Metab. Eng.* **14,** 104–111 (2012).

28.   Kallehauge, T. B. *et al.* Ribosome profiling-guided depletion of an mRNA increases cell growth rate and protein secretion. *Sci. Rep.* **7,** 40388 (2017).

29.   Ko, P. *et al.* Probing the importance of clonality: Single cell subcloning of clonally derived CHO cell lines yields widely diverse clones differing in growth, productivity, and product quality. *Biotechnol. Prog.* (2017). doi:10.1002/btpr.2594

30.   Vcelar, S. *et al.* Karyotype variation of CHO host cell lines over time in culture characterized by chromosome counting and chromosome painting. *Biotechnol. Bioeng.* **115,** 165–173 (2018).

31.   Dietmair, S. *et al.* A multi-omics analysis of recombinant protein production in Hek293 cells. *PLoS One* **7,** e43394

(2012).

32. Underhill, M. F., Birch, J. R., Smales, C. M. & Naylor, L. H. eIF2alpha phosphorylation, stress perception, and the shutdown of global protein synthesis in cultured CHO cells. *Biotechnol. Bioeng.* **89,** 805–814 (2005).

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

34. Pristovšek, N. *et al.* 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.* (2018). doi:10.1002/biot.201700216

35. Kol, S., Kallehauge, T. B., Adema, S. & Hermans, P. Development of a VHH-Based Erythropoietin Quantification Assay. *Mol. Biotechnol.* **57,** 692–700 (2015).

36. Hansen, H. G. *et al.* Versatile microscale screening platform for improving recombinant protein productivity in Chinese hamster ovary cells. *Sci. Rep.* **5,** 18016 (2015).

37. Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. & Kingsford, C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods* **14,** 417–419 (2017).

38. Soneson, C., Love, M. I. & Robinson, M. D. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Res.* **4,** 1521 (2015).

39. Law, C. W., Chen, Y., Shi, W. & Smyth, G. K. voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol.* **15,** R29 (2014).

40: R Core Team 2014 R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria 2014.

41: H. Wickham. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2009.

5.2 Applicability and future perspectives

In the quest to better understand existing biological processes and engineer new biological pathways, it is crucial to have an optimal system where we can study the effect of expressing transgenes. The system can be used for studying the variation in amounts of product being produced and the reasons behind the phenomenon. This platform offers an unprecedented solution for testing and generating new strategies to specifically tailor the expression and secretory needs of recombinant therapeutic proteins. The technique can be further adapted and utilized to introduce several genes in different locations in a controllable and comparable manner, beginning the journey toward repairing or inserting entire pathways. The ability to study transgenes in combination with each other will allow us to better evaluate possible engineering strategies e.g. the optimal ratios for expression of different transgenes. Furthermore, the utilization of RMCE landing pads, allows these cell lines to be readily re-used in an endless amount of studies. The application of the platform can be extended beyond comparative analysis of transgenes; it would e.g. be optimal for the difficult but indispensable task of testing and developing existing and new genetic elements for optimal, coherent and regulated gene expression.

Chapter 6

Concluding remarks

The aim of this PhD thesis has been to develop new genome editing tools to advance the CHO cell engineering toolbox and to demonstrate their relevance to the field of recombinant production of therapeutic proteins by direct applications. The primary contribution of the thesis is thus both the techniques presented and their demonstrated applicability.

To continue in the paths that started with this thesis, several projects in our laboratory have been initiated that build on using the tools and results presented here. CRISPR has enabled us to remove, insert and study a plethora of different genes. It has proved to be extremely efficient, and our biggest bottleneck is now the characterization of the clones we generate. A major obstacle currently facing the CHO cell engineering community, is how to ensure that changes observed in phenotypes are indeed due to our engineering efforts and not simply due to clonal variation. Chapter 5 clearly indicates that clonal variation needs to be taken into account when studying gene effects and phenotypic changes.

With the more precise and controllable tools presented in this thesis we have advanced the engineering toolbox for future CHO cell line development for the production of therapeutic proteins. The tools will help us gain an increasing understanding of biological processes to guide our engineering efforts towards our end goal of generating superior CHO cell lines for optimal production of a wide range of therapeutic proteins.

References

1. Walsh, G. Biopharmaceutical benchmarks—2003. *Nat. Biotechnol.* **21,** 865–870 (2003).

2. Walsh, G. Biopharmaceutical benchmarks 2014. *Nat. Biotechnol.* **32,** 992–1000 (2014).

3. Frye, C. *et al.* Industry view on the relative importance of 'clonality' of biopharmaceutical-producing cell lines. *Biologicals* **44,** 117–122 (2016).

4. Kesik-Brodacka, M. Progress in biopharmaceutical development. *Biotechnol. Appl. Biochem.* (2017). doi:10.1002/bab.1617

5. Kuo, C.-C. *et al.* The emerging role of systems biology for engineering protein production in CHO cells. *Curr. Opin. Biotechnol.* **51,** 64–69 (2017).

6. Browne, S. M. & Al-Rubeai, M. Selection methods for high-producing mammalian cell lines. *Trends Biotechnol.* **25,** 425–432 (2007).

7. Xu, X. *et al.* The genomic sequence of the Chinese hamster ovary (CHO)-K1 cell line. *Nat. Biotechnol.* **29,** 735–741 (2011).

8. Lewis, N. E. *et al.* Genomic landscapes of Chinese hamster ovary cell lines as revealed by the Cricetulus griseus draft genome. *Nat. Biotechnol.* **31,** 759–765 (2013).

9. Carroll, D. Genome engineering with targetable nucleases. *Annu. Rev. Biochem.* **83,** 409–439 (2014).

10. Ronda, C. *et al.* Accelerating genome editing in CHO cells using CRISPR Cas9 and CRISPy, a web-based target finding tool. *Biotechnol. Bioeng.* **111,** 1604–1616 (2014).

11. Cameron, D. E., Bashor, C. J. & Collins, J. J. A brief history of synthetic biology. *Nat. Rev. Microbiol.* **12,** 381–390 (2014).

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

13. Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* **339,** 819–823 (2013).

14. Jinek, M. *et al.* RNA-programmed genome editing in human cells. *Elife* **2,** e00471 (2013).

15. Yip, S. S. M. *et al.* Complete knockout of the lactate dehydrogenase A gene is lethal in pyruvate

dehydrogenase kinase 1, 2, 3 down-regulated CHO cells. *Mol. Biotechnol.* **56,** 833–838 (2014).

16. Hefzi, H. *et al.* Ribosome profiling-guided depletion of an mRNA increases cell growth rate and protein secretion. *Sci. Rep.* **7,** 40388 (2017).

17. Kallehauge, T. B. *et al.* Mammalian cells devoid of lactate dehydrogenase activity. U.S. Patent WO2017192437, published November 9 (2017).

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

Supplementary Material for

One-step generation of triple knockout CHO cell lines using CRISPR Cas9 and fluorescent enrichment

Lise Marie Grav[1,*], Jae Seong Lee[1,*], Signe Gerling[1], Thomas Beuchert Kallehauge[1], Anders Holmgaard Hansen[1], Stefan Kol[1], Gyun Min Lee[2], Lasse Ebdrup Pedersen[1], Helene Faustrup Kildegaard[1,#]

[1]The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, 2970 Hørsholm, Denmark

[2]Department of Biological Sciences, KAIST, 291 Daehak-ro, Yuseong-gu, Daejeon 305-701, Republic of Korea

Correspondence: Helene Faustrup Kildegaard, The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Kogle Allé 6, 2970 Hørsholm, Denmark. **E-mail**: hef@biosustain.dtu.dk

Supplementary Figure S1: Vector map for GFP_2A_Cas9. The GFP_2A_Cas9 expression vector is prepared with USER cloning to insert GFP and 2A in front of Cas9 in the Cas9 expression vector constructed previously [3].
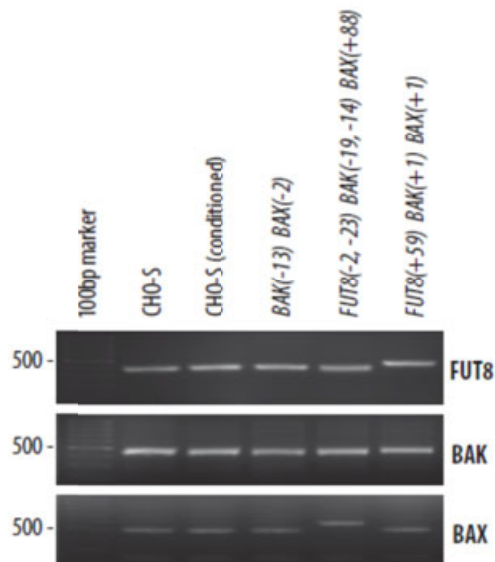
Supplementary Figure S2: Separation of amplicons covering the respective sgRNA target sites in *FUT8*, *BAK* and *BAX*. The seperated bands are cut out of the gel and Sanger sequenced to verify the indels generated at the *FUT8*, *BAK* and *BAX* loci of the multiplexed and FACS enriched single cell sorted clones.

2
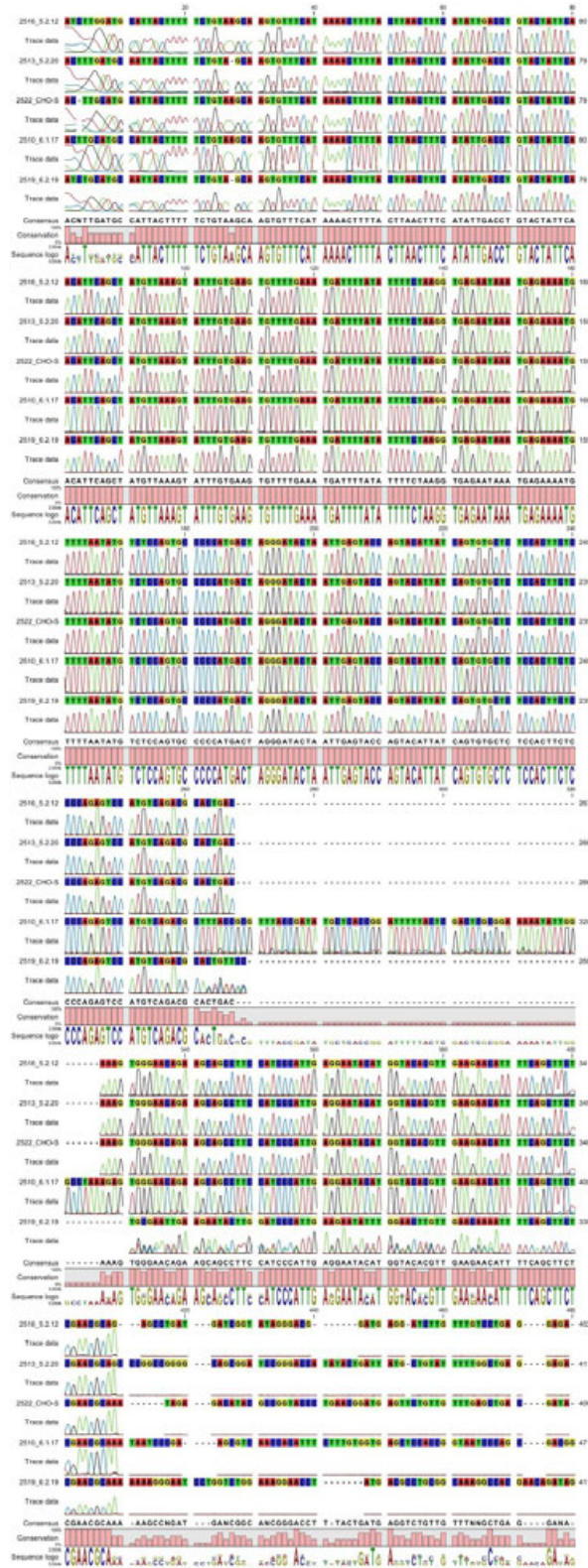
Supplementary Figure S3: Verification of the indels generated in the multiplexed and FACS enriched clones. Alignment of Sanger sequences of amplicons covering the *FUT8* sgRNA target site seperated on agarose gels using the CHO-S, CHO-S conditioned (5.2.20), *BAK(-13) BAX(-2)* double knockout cell line (5.2.12), *FUT8(-2,-23) BAK(-19,-14) BAX(+88)* triple knockout cell line (6.2.19) and *FUT8(+59) BAK(+1) BAX(+1)* triple knockout cell line (6.1.17) as templates.

3
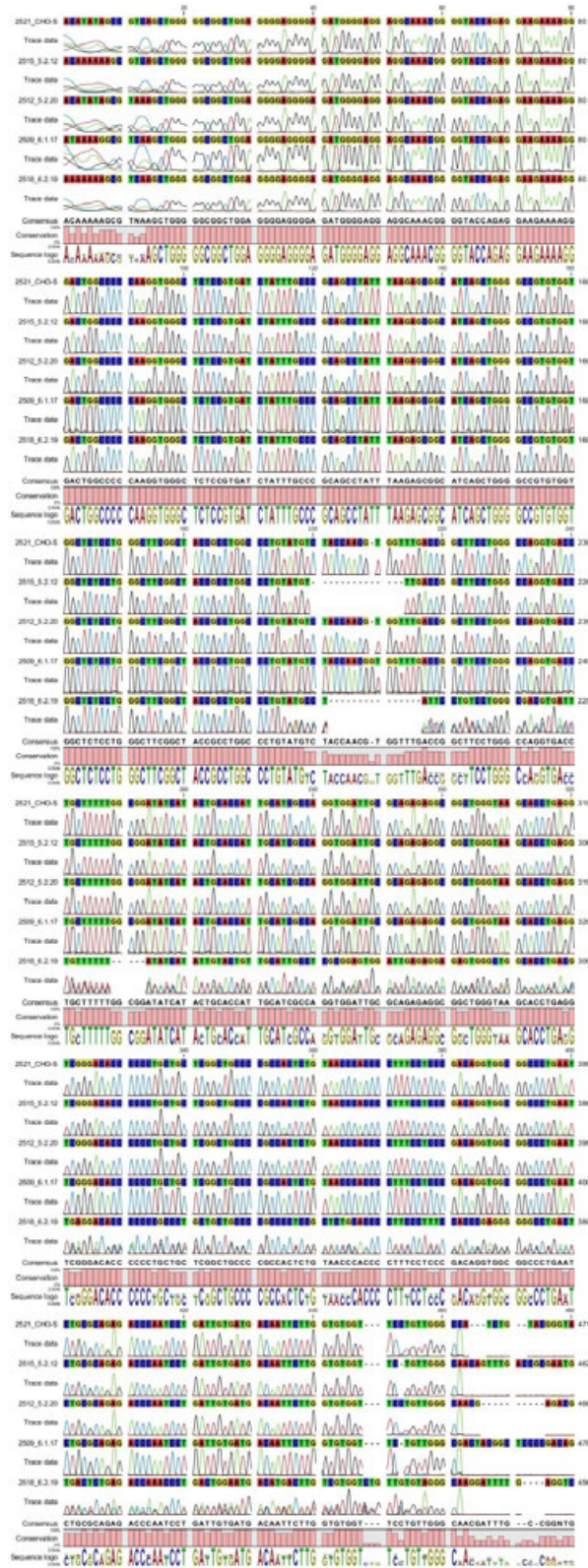
Supplementary Figure S4: Verification of the indels generated in the multiplexed and FACS enriched clones. Alignment of Sanger sequences of amplicons covering the *BAK* sgRNA target site seperated on agarose gels using the CHO-S, CHO-S conditioned (5.2.20), *BAK(-13) BAX(-2)* double knockout cell line (5.2.12), *FUT8(-2,-23) BAK(-19,-14) BAX(+88)* triple knockout cell line (6.2.19) and *FUT8(+59) BAK(+1) BAX(+1)* triple knockout cell line (6.1.17) as templates.

4
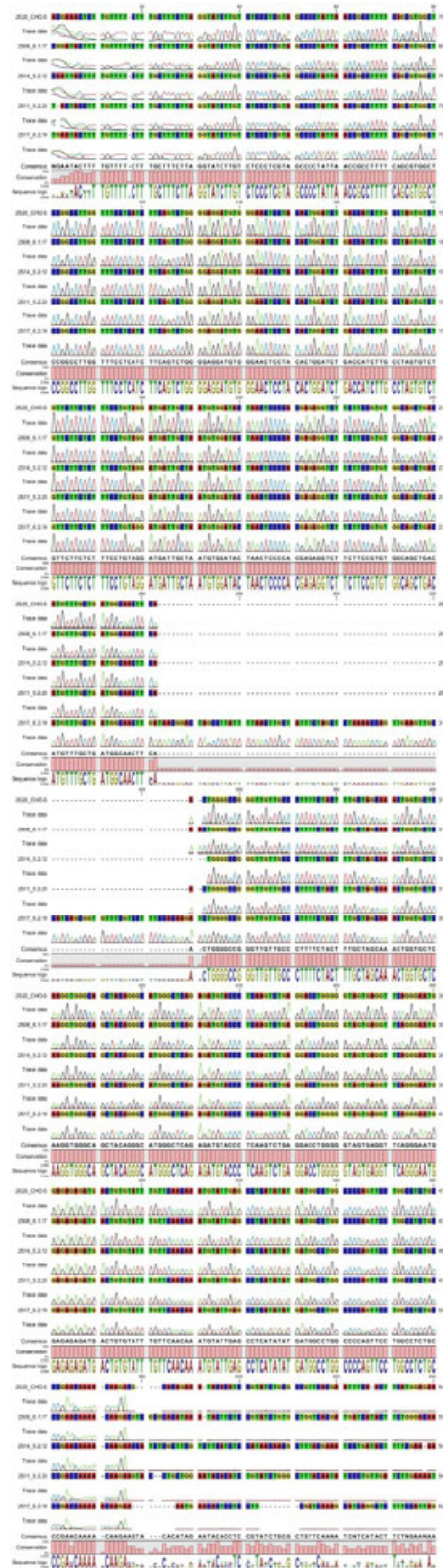
Supplementary Figure S5: Verification of the indels generated in the multiplexed and FACS enriched clones. Alignment of Sanger sequences of amplicons covering the *BAX* sgRNA target site seperated on agarose gels using the CHO-S, CHO-S conditioned (5.2.20), *BAK(-13) BAX(-2)* double knockout cell line (5.2.12), *FUT8(-2,-23) BAK(-19,-14) BAX(+88)* triple knockout cell line (6.2.19) and *FUT8(+59) BAK(+1) BAX(+1)* triple knockout cell line (6.1.17) as templates.

5
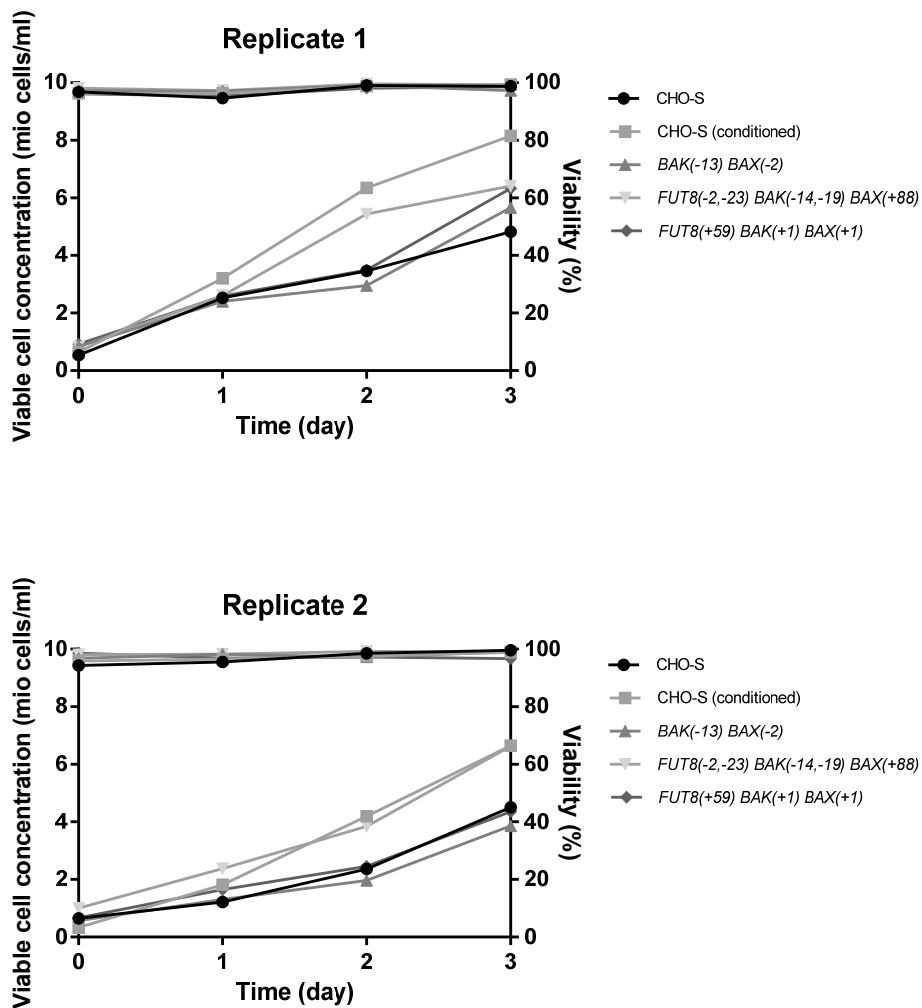
Supplementary Figure S6: Growth and viability of the multiplexed and FACS enriched clones. The graph shows the viable cell concentration (left Y-axis) and viability (right Y-axis) of CHO-S, CHO-S conditioned, *BAK(-13) BAX(-2)* double knockout cell line, *FUT8(-2,-23) BAK(-19,-14) BAX(+88)* triple knockout cell line and *FUT8(+59) BAK(+1) BAX(+1)* triple knockout cell line for three days upon transient transfection with antibody (Rituximab) expression vector.

6

Replicate 1

Replicate 2
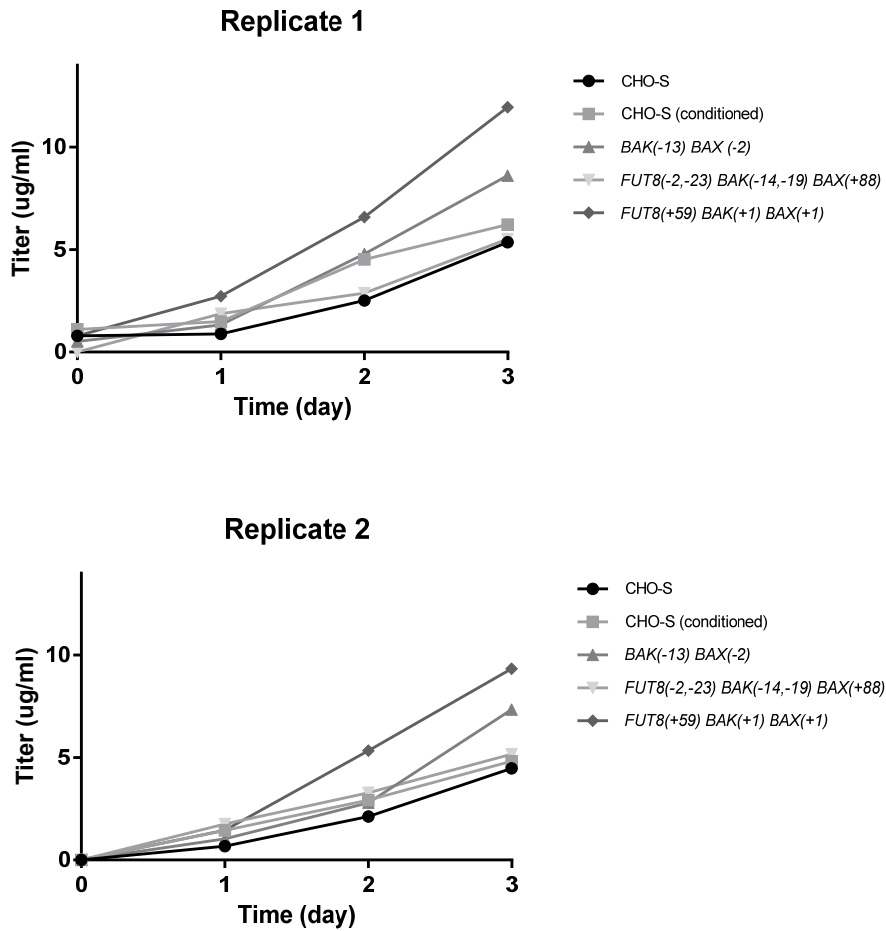
Supplementary Figure S7: Antibody production with the multiplexed and FACS enriched clones. The graph shows the amount of antibody (Rituximab) obtained upon transient transfection of CHO-S, CHO-S conditioned, *BAK(-13) BAX(-2)* double knockout cell line, *FUT8(-2,-23) BAK(-19,-14) BAX (+88)* triple knockout cell line and *FUT8(+59) BAK(+1) BAX(+1)* triple knockout cell line with antibody (Rituximab) expression vector and growth for three days.

Supplementary Table S1: Primer list for GFP_2A_Cas9 cloning. The bases in red mark the USER overhangs applied.

Primer number	Primer name	Sequence (5' – 3')
PR0239	BB_Cas9_2A-link_fw	AGAAGCAUGGACAAGAAATACTCCATCG
PR0240	BB_Cas9_2A-link_rv	ATGCATGGUGGCGGCGCTAGCCAGCTTC
PR0241	eGFP_2A_2A-link_fw	ACCATGCAUGTGAGCAAGGGCGAGGAGCTG
PR0242	eGFP_2A_2A-link_rv	ATGCTTCUCGATCGTGGGCCAGGATTCTCCTCGACG

Supplementary Table S2: *FUT8*, *BAK* and *BAX* sgRNA target sequences. The bases in red mark the PAM site.

Target gene	Target sequence (5' – 3')
FUT8	GTCAGACGCACTGACAAAGTGGG
BAK	GGAAGCCGGTCAAACCACGTTGG
BAX	GCTGATGGCAACTTCAACTGGGG

Supplementary Table S3: Oligos for *FUT8*, *BAK* and *BAX* sgRNA expression vector cloning

Oligo number	Oligo name	Oligo sequence (5' – 3')
7823	Oligo sgRNA3_F_fw/ sgRNA_FUT8_681494_fwd	GGAAAGGACGAAACACCGTCAGACGCACTGACAAAGTGTTTTAGAGCTAGAAAT
7919	Oligo sgRNA3_F_rv/ sgRNA_FUT8_681494_rev	CTAAAACACTTTGTCAGTGCGTCTGACGGTGTTTCGTCCTTTCCACAAGATAT
7747	sgRNA_BAK1_1544257_fwd	GGAAAGGACGAAACACCGGAAGCCGGTCAAACCACGTGTTTTAGAGCTAGAAAT
7843	sgRNA_BAK1_1544257_rev	CTAAAACACGTGGTTTGACCGGCTTCCGGTGTTTCGTCCTTTCCACAAGATAT
7760	sgRNA_BAX_1345650_fwd	GGAAAGGACGAAACACCGCTGATGGCAACTTCAACTGGTTTTAGAGCTAGAAAT
7856	sgRNA_BAX_1345650_rev	CTAAAACCAGTTGAAGTTGCCATCAGCGGTGTTTCGTCCTTTCCACAAGATAT

8

Supplementary Table S4: Primer list for deep sequencing (MiSeq). The primers contain overhang sequences compatible with Illumina Nextera XT indexing (forward primer overhang: TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG, reverse primer overhang: GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG).

Primer number	Primer name	Primer sequence (5' – 3')
PR0222	FUT8 gRNA3_F_Nex	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTGCCCCCATGACTAGGGATA
PR0223	FUT8 gRNA3_R_Nex	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTCTGCGTTCGAGAAGCTGAAA
PR0461	Miseq_BAK1_1544257_fw	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCAAGGTGGGCTCTCCGTGAT
PR0462	Miseq_BAK1_1544257_rv	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCGATGCAATGGTGCAGTATGAT
PR0463	Miseq_BAX_1345650_fw	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTGTGGATACTAACTCCCCACG
PR0464	Miseq_BAX_1345650_rv	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTCCCTGAACCTCACTACCCC
A1	MiSeq_Fut8_off1_fwd	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTCTGTCTGCCATCTCAACTGT
B1	MiSeq_Fut8_off2_fwd	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCTCTTGTCCCATGCTCCCTC
C1	MiSeq_Fut8_off3_fwd	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGAGTCTCTATGTGCACGGGAG
D1	MiSeq_Fut8_off4_fwd	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTCTCTGTGCTCTCATGCTCC
E1	MiSeq_Fut8_off5_fwd	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCATTGGGTGCATAGCTGGA
F1	MiSeq_Fut8_off6_fwd	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGGGGGGTCTTGTGGCTCTTTA
G1	MiSeq_Fut8_off7_fwd	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGATATCTGTGTGACCCCACGG
H1	MiSeq_Fut8_off8_fwd	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTTACCACCCAAGGTCCCCAA
A2	MiSeq_Fut8_off9_fwd	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTGCTGCGGAATCGATAGCAA
B2	MiSeq_Fut8_off10_fwd	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCACACTGCAGCTTACAACAG
C2	MiSeq_Fut8_off11_fwd	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCTGGGAAGTGGACAGCAGAC
D2	MiSeq_Fut8_off12_fwd	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGACTGACCAGCGAGACCCTAT
E2	MiSeq_Fut8_off13_fwd	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGAGAGCAGGGAGAGCTACCAA
F2	MiSeq_Fut8_off14_fwd	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCCTCCTTCCTTAGGCCTCT
G2	MiSeq_Fut8_off15_fwd	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCCATTCCTGCAGTGATGGT
A1	MiSeq_Fut8_off1_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGAGAGGCAAGATGAAGGTGGACG
B1	MiSeq_Fut8_off2_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCAAGAGGCTGTCACTGCAGT
C1	MiSeq_Fut8_off3_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGACCAGGCTCAGAGCTAGGAT
D1	MiSeq_Fut8_off4_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGCTGTTCCAGGCTCTGCTTA
E1	MiSeq_Fut8_off5_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGCATGCTAGCGGCTTGTTTT
F1	MiSeq_Fut8_off6_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCTGGTCATCACCACTGCTCA
G1	MiSeq_Fut8_off7_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCTGGCCAGACCCAAAGGATC
H1	MiSeq_Fut8_off8_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGAACCTCTTCCTGGGGGCTAT
A2	MiSeq_Fut8_off9_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGATCTACCCAAGCACCACCT
B2	MiSeq_Fut8_off10_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCATCTCTCCAGCCCCTTCTT
C2	MiSeq_Fut8_off11_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGACCCTCCTCTCTCCCAAAGA
D2	MiSeq_Fut8_off12_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGGAGAGCAGGCTGAGTGAAG
E2	MiSeq_Fut8_off13_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCCAGTGCCAGTTTCAATGGC
F2	MiSeq_Fut8_off14_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGAGGATTGAGGGGGTCACTGA
G2	MiSeq_Fut8_off15_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGACACCCTGGGTCTGAGAAAC
A4	MiSeq_Bak1_off1_fwd	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGAGCAGGGTTCATCTGAGTGC

9

B4	MiSeq_Bak1_off2_fwd	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTAGCTTGCTGCCTTTGAGCA
C4	MiSeq_Bak1_off3_fwd	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCAGGCCGAGAAGCACATAA
D4	MiSeq_Bak1_off4_fwd	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGAGGAGGGCTTCTGGATATGGT
E4	MiSeq_Bak1_off5_fwd	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGGCCAAACTCCAGGAGTCC
F4	MiSeq_Bak1_off6_fwd	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGCACAGTTAACAGGTTGGTTCC
G4	MiSeq_Bak1_off7_fwd	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCGTGGGGATATCTTGCCTC
H4	MiSeq_Bak1_off8_fwd	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTCAAACCCAATAAGTGAGACCT
A5	MiSeq_Bak1_off9_fwd	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCCTGCTCACTAACACCCAT
B5	MiSeq_Bak1_off10_fwd	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGAAGCCACATAATTGCCCCCT
C5	MiSeq_Bak1_off11_fwd	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCATTACCCAGTGCTCCCCG
D5	MiSeq_Bak1_off12_fwd	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGCCTTTCTTGACATGGCCAA
E5	MiSeq_Bak1_off13_fwd	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCACAGTGAATTTGGCCACGG
F5	MiSeq_Bak1_off14_fwd	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTCAGGTTCGGCACTTAGGTT
G5	MiSeq_Bak1_off15_fwd	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCCTTTTCTTCTGCTGTCACA
A4	MiSeq_Bak1_off1_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTGCCTGCCTTACTGACTTTGA
B4	MiSeq_Bak1_off2_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCTACCAACGCCCCACAGTTA
C4	MiSeq_Bak1_off3_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGACTCCTGCTACTCCCTCACC
D4	MiSeq_Bak1_off4_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTCACGACCTGGACCTCACTA
E4	MiSeq_Bak1_off5_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTAGACTTCCCAAGCCCTCGA
F4	MiSeq_Bak1_off6_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCTGGGTTTGAGGCTACCCTG
G4	MiSeq_Bak1_off7_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCCAGATATGTCACAACCCCGT
H4	MiSeq_Bak1_off8_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGAGGCCCCATAGACTACCCA
A5	MiSeq_Bak1_off9_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGCTGATGGAAGCAAAGCAGAC
B5	MiSeq_Bak1_off10_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGACCCCACAGACTGACCAAG
C5	MiSeq_Bak1_off11_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTAGCGCCCTCTTGGACTTTG
D5	MiSeq_Bak1_off12_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGAGCTCCTCCTGCACGTTTTT
E5	MiSeq_Bak1_off13_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGCCCCTTCATCCCAGCATAG
F5	MiSeq_Bak1_off14_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGCATAGATGGCCGCTTGC
G5	MiSeq_Bak1_off15_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGGGGCCAAGCACTGTGAGAAT
A7	MiSeq_Bax_off1_fwd	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCAGTGACACCATCCCTGTCA
B7	MiSeq_Bax_off2_fwd	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGAGCCACATTGACTCTCAGGC
C7	MiSeq_Bax_off3_fwd	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTCAATGGGGTTGGTCTTCCG
D7	MiSeq_Bax_off4_fwd	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCTCCTGGCTGCTCATTCCAT
E7	MiSeq_Bax_off5_fwd	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTGAGCTTCAGAGGTTGGCTG
F7	MiSeq_Bax_off6_fwd	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCGTGACCCCCAAATGAGAA
G7	MiSeq_Bax_off7_fwd	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGCCTTTAAATGGCTGGTTCCA
H7	MiSeq_Bax_off8_fwd	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTAACAGGCAGGTCCACCTTC
A8	MiSeq_Bax_off9_fwd	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGCCTGAGATGCCCAGTATCC
B8	MiSeq_Bax_off10_fwd	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTGTGTACAGTGGTGTTGCCA
C8	MiSeq_Bax_off11_fwd	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCTGCTGCCACCTAAGAGGAC
D8	MiSeq_Bax_off12_fwd	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTGGCAGAAGATCCCCTAGGT

10

E8	MiSeq_Bax_off13_fwd	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTGTGTCCATGGTAGTTTCTTTGTG
F8	MiSeq_Bax_off14_fwd	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGAGCAAACCAACACAGAATGAACA
G8	MiSeq_Bax_off15_fwd	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGACAAGAAAGAAGGAAAGGAAGAAAGG
A7	MiSeq_Bax_off1_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTGTGTGGCCTGCTTTCTAGT
B7	MiSeq_Bax_off2_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCCTGTCTAAGGGTGGCACTG
C7	MiSeq_Bax_off3_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGAGGTGGGAGGGACACAAAC
D7	MiSeq_Bax_off4_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGTCCCCAGTGCCCACATTAA
E7	MiSeq_Bax_off5_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGAGCTTGATTTCCACGGCTCA
F7	MiSeq_Bax_off6_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCACCCCAAACCCCACATTCA
G7	MiSeq_Bax_off7_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGACTCTGTGTCGTGGTAATCCA
H7	MiSeq_Bax_off8_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGTTAGTGTGCCACCTACCCC
A8	MiSeq_Bax_off9_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGCATCAACCACGAGCTGTTG
B8	MiSeq_Bax_off10_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTCCCTCTTCACTCCAGCACA
C8	MiSeq_Bax_off11_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTGAGTACACACGTGCACACT
D8	MiSeq_Bax_off12_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTTGAGGGCCCAAGAATGCAT
E8	MiSeq_Bax_off13_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGATGACTGCTACCTGGCGTT
F8	MiSeq_Bax_off14_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCCCCCACTTGGTTTTGGTCT
G8	MiSeq_Bax_off15_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGAGACCCAGAAGCCCAAACAG

11

Supplementary Table S5: Off-target sites identification. The top 15 sites for each sgRNA targeting *FUT8*, *BAK* and *BAX* respectively are shown and have been selected based on the Ham score. ==HAM score refers to Hamming distance indicating the number of positions in a sequence that differs from a reference sequence [1]. In this case, it is the number of positions in which the off-target sequence differs from the on-target sequence in either the 13 bp closest to the PAM or in total.== The bases in red mark the sequence difference from the on-target site.

	Ham score in first 13 bp	Ham score in total	Contig	Strand	Position	Sequence
Fut8_on target	0	0	NW_003613860	+	710987	GTCAGACGCACTGACAAAGT
Fut8_off1	0	4	NW_003613648	-	1396948	GTACTATGCACTGACAAAGT
Fut8_off2	0	4	NW_003614846	-	151698	GACGTAGGCACTGACAAAGT
Fut8_off3	0	6	NW_003613896	+	39301	TAATGTGGCACTGACAAAGT
Fut8_off4	1	3	NW_003613649	-	2963243	GTCAAAAGCACTGACAAGGT
Fut8_off5	1	3	NW_003613951	-	740130	GCCCGACGCACTGAGAAAGT
Fut8_off6	1	4	NW_003614027	+	545477	GACAGCATCACTGACAAAGT
Fut8_off7	1	5	NW_003613583	-	4130466	AGAAGAGGCACTGACAAACT
Fut8_off8	1	5	NW_003613591	+	1717331	CCCGGAGGCACTTACAAAGT
Fut8_off9	1	5	NW_003613596	+	4119735	AATAAACGCACTTACAAAGT
Fut8_off10	1	5	NW_003613689	+	2115570	GGCACGGGCACTTACAAAGT
Fut8_off11	1	5	NW_003613750	-	1719535	TGAAGAAGCACTGAGAAAGT
Fut8_off12	1	5	NW_003613777	+	239782	ACCGGAAGCACCGACAAAGT
Fut8_off13	1	5	NW_003613780	+	483228	GGTAGGTGCACTGCCAAAGT
Fut8_off14	1	5	NW_003613842	-	344431	GCAACAGGCAATGACAAAGT
Fut8_off15	1	5	NW_003613844	-	1138412	TTCTATCTCACTGACAAAGT
Bak1_on target	0	0	NW_003614972	-	227662	GGAAGCCGGTCAAACCACGT
Bak1_off1	1	4	NW_003613599	+	195202	GTAATCAGGTCAAACCATGT
Bak1_off2	1	4	NW_003615993	-	85713	GGAAAGGGGTCAAACCAGGT
Bak1_off3	1	5	NW_003613681	-	1683603	GTGACCAGGTCAAACCACTT
Bak1_off4	1	5	NW_003614007	+	38086	GAAAAGAGGTCAAACCAGGT
Bak1_off5	1	5	NW_003614165	+	227394	GCAAATAGGTCAAACCATGT
Bak1_off6	1	5	NW_003614340	+	338518	AATAGCAGGTCAAACCAAGT
Bak1_off7	1	5	NW_003615215	+	28553	TAAATCAGGTCAAACCACAT
Bak1_off8	1	5	NW_003615261	+	144994	GCAAAGAGGTCAAACCATGT
Bak1_off9	1	5	NW_003615939	-	147940	GCAAAGTGGTCAAACCAGGT
BAK_off10	1	5	NW_003616605	+	61822	AGAAAAGGTCAAACCAAGT
Bak1_off11	1	6	NW_003613725	+	1131426	CCAGCCAGGTCAAACCAGGT
Bak1_off12	1	6	NW_003613753	-	1078815	TGTGGTTGGTCAAACAACGT
Bak1_off13	1	6	NW_003613831	+	460720	CAGGGCTGGTCAAGCCACGT
Bak1_off14	1	6	NW_003613831	+	722520	GGCCAGAGGACAAACCACGT
Bak1_off15	1	6	NW_003613839	+	1346280	CTGAGTTGGTCAAACCACTT
Bax_on target	0	0	NW_003614570	+	281815	GCTGATGGCAACTTCAACTG
Bax_off1	0	5	NW_003613691	+	514917	GGAAGAGGCAACTTCAACTG
Bax_off2	0	5	NW_003617041	+	69462	GAGAGTAGCAACTTCAACTG
Bax_off3	1	3	NW_003613598	+	671039	TCTGATTGCAACTTCAACTC
Bax_off4	1	3	NW_003614075	+	114499	GCTAAGGGCAACTTCAAGTG
Bax_off5	1	4	NW_003613602	+	412112	GCTAGAGGGAACTTCAACTG
Bax_off6	1	4	NW_003613618	-	1971061	TTTTATGTCAACTTCAACTG
Bax_off7	1	4	NW_003613646	+	321775	TCTTAGGGCAACTTCCACTG
Bax_off8	1	4	NW_003613686	-	1252493	GAAGAGGGGAACTTCAACTG
Bax_off9	1	4	NW_003613866	-	672417	GATGTAGGCAACATCAACTG
Bax_off10	1	4	NW_003614684	-	92790	AGTGATTGCAACTTGAACTG
Bax_off11	1	5	NW_003613585	+	4201621	GTGAGTGGCAAGTTCAACTG
Bax_off12	1	5	NW_003613608	+	1389389	TGTGAAAGCACCTTCAACTG
Bax_off13	1	5	NW_003613649	-	3145890	GGGGAAAGCAACTTCATCTG
Bax_off14	1	5	NW_003613754	+	1713055	TGCCATGGCAACTCCAACTG
Bax_off15	1	5	NW_003613775	-	36278	AAAGAAGGCAACCTCAACTG

12

Supplementary Table S6: Primers for Sanger sequencing of genomic regions flanking the sgRNA target sites.

Primer number	Primer name	Sequence (5' – 3')
PR0750	BAK1_topo_fwd	ACTTTCTGACCTGGACTGGC
PR0751	BAK1_topo_rev	GCCCAACAGAACCACACCAA
PR0752	BAX_topo_fwd	CTGACTCCCCAGGACCTTGA
PR0753	BAX_topo_rev	CTTGTTTTGTTCGGGCAGAGG
PR0754	FUT8_topo_fwd	AATCTGTTGATTCCAGGTTCC
PR0755	FUT8_topo_rev	TGCGTTCGAGAAGCTGAAAA

13

Supplementary Table S7: Deep sequencing (MiSeq) of the most potential off-target sites predicted for each of the sgRNAs targeting *FUT8*, *BAK* and *BAX* respectively. The analysis is performed on the GFP_2A_Cas9 only and the two multiplexed samples with high sgRNA to GFP_2A_Cas9 ratio before and after FACS enrichment, which contain the cell population with most generated indels (biological replicate number 2). Samples marked with an asterisk are high in all samples. N/A indicates that no PCR products are obtained from samples.

	GFP_2A_Cas9 no sgRNA	GFP_2A_Cas9 *FUT8*, *BAK and BAX* Before enrichment	GFP_2A_Cas9 *FUT8*, *BAK and BAX* After enrichment
*FUT8*_on target	0.4%	11.5%	66.3%
*FUT8*_off1	0.2%	0.2%	0.2%
*FUT8*_off2	0.1%	0.1%	0.1%
*FUT8*_off3	0.2%	0.2%	0.2%
*FUT8*_off4	0.2%	0.2%	0.2%
*FUT8*_off5	0.2%	0.2%	0.2%
*FUT8*_off6	0.1%	0.1%	0.0%
*FUT8*_off7	0.1%	0.2%	0.1%
*FUT8*_off8	0.1%	0.1%	0.1%
*FUT8*_off9	0.2%	0.1%	0.1%
*FUT8*_off10	0.2%	0.3%	N/A
*FUT8*_off11	0.1%	0.1%	0.1%
*FUT8*_off12	0.2%	0.2%	0.2%
*FUT8*_off13	3.7%	1.6%	0.7%
*FUT8*_off14	0.2%	0.2%	0.2%
*FUT8*_off15	0.1%	0.1%	0.1%
*BAK*_on target	0.1%	11.0%	66.3%
*BAK*_off1	0.2%	0.2%	0.1%
*BAK*_off2	0.1%	0.2%	0.2%
*BAK*_off3	0.1%	0.1%	0.1%
*BAK*_off4	0.3%	0.3%	0.2%
*BAK*_off5	0.1%	0.2%	0.2%
*BAK*_off6	0.4%	0.4%	N/A
*BAK*_off7	0.1%	0.2%	0.2%
*BAK*_off8	0.5%	0.5%	0.4%
*BAK*_off9	1.8%	0.6%	0.4%
*BAK*_off10	N/A	N/A	N/A
*BAK*_off11	0.1%	0.1%	0.1%
*BAK*_off12	0.2%	0.2%	0.2%
*BAK*_off13	0.2%	0.2%	0.2%
*BAK*_off14	0.1%	0.1%	0.1%
*BAK*_off15	0.2%	0.2%	0.2%
*BAX*_on target	0.2%	17.9%	75.8%
*BAX*_off1	0.1%	0.1%	0.1%
*BAX*_off2	0.2%	0.2%	0.2%
*BAX*_off3	N/A	N/A	N/A
*BAX*_off4	0.1%	0.1%	0.6%
*BAX*_off5	48.4%*	48.4%*	46.3%*
*BAX*_off6	0.2%	0.2%	0.2%
*BAX*_off7	0.8%	0.8%	0.7%
*BAX*_off8	0.2%	0.2%	0.2%
*BAX*_off9	0.2%	0.2%	0.1%
*BAX*_off10	0.1%	0.1%	0.1%
*BAX*_off11	0.1%	0.1%	0.1%
*BAX*_off12	0.1%	0.1%	0.1%
*BAX*_off13	0.1%	0.1%	0.1%
*BAX*_off14	0.1%	0.1%	0.1%
*BAX*_off15	0.7%	0.6%	0.7%

Supplementary Table S8: Deep sequencing results of clones (run #1). Each clone has its own index (column 1). Column 3, 4 and 5 show the number of reads obtained in total, of wild-type sequences and sequences with indels respectively. Clones with unclear data are indicated as fail in column 7 and are also marked as dark gray in column 12. Column 9, 10 and 11 show the type of indels identified in the clones (i.e. basepair difference from wild-type), where peak n = 0 means wild-type alleles. The clones marked with light gray indicate absences of genes with indels (wild-type). The clones marked with yellow indicate presence of one gene with indels (potential single knockout cell line). The clones marked with orange indicate presence of two genes with indels (potential double knockout cell line). The clones marked with red indicate presence of three genes with indels (potential triple knockout cell line). Clones with mixed alleles (presence of both WT allele and allele with indel) are also marked as dark gray and are also not included in Figure 2A. The clones marked with an asterisk are selected for further characterization that is referred to in the manuscript.

Index	Product	Counts	WT Count	WT	Indels	Knockout	Peaks	Peak 1	Peak 2	Peak 3	CLONE	
1	FUT8_681494	18682	18660	99,88%	0,12%	interesting		1	0			
1	BAK1_1544257	79188	79009	99,77%	0,23%			1	0			
1	BAX_1345650	80493	80224	99,67%	0,33%		0	1	0			5.1.1
2	FUT8_681494	9979	9965	99,86%	0,14%	interesting		1	0			
2	BAK1_1544257	45222	45121	99,78%	0,22%			1	0			
2	BAX_1345650	45355	45231	99,73%	0,27%		0	1	0			5.1.2
3	FUT8_681494	18085	18065	99,89%	0,11%	interesting		1	0			
3	BAK1_1544257	77702	77540	99,79%	0,21%			1	0			
3	BAX_1345650	79928	79679	99,69%	0,31%		0	1	0			5.1.3
4	FUT8_681494	7764	7755	99,88%	0,12%	Fail		1	0			
4	BAK1_1544257	77527	77404	99,84%	0,16%			1	0			
4	BAX_1345650	89856	80294	89,36%	10,64%		0	2	0	-9		5.1.4
5	FUT8_681494	14595	14585	99,93%	0,07%	interesting		1	0			
5	BAK1_1544257	70637	70484	99,78%	0,22%			1	0			
5	BAX_1345650	69541	69361	99,74%	0,26%		0	1	0			5.1.5
6	FUT8_681494	38477	38443	99,91%	0,09%	Fail		1	0			
6	BAK1_1544257	366	305	83,33%	16,67%			1	0			
6	BAX_1345650	159566	4039	2,53%	97,47%		1	1	-16			5.1.6
7	FUT8_681494	15332	15314	99,88%	0,12%	interesting		1	0			
7	BAK1_1544257	75063	74923	99,81%	0,19%			1	0			
7	BAX_1345650	92755	92506	99,73%	0,27%		0	1	0			5.1.7
8	FUT8_681494	60	57	95,00%	5,00%	Fail		1	0			
8	BAK1_1544257	54654	30001	54,89%	45,11%	1/2 allelle hit/2 clones		2	0	-13		
8	BAX_1345650	89131	88859	99,69%	0,31%		0	1	0			5.1.8
9	FUT8_681494	6381	6375	99,91%	0,09%	interesting		1	0			
9	BAK1_1544257	32314	32266	99,85%	0,15%			1	0			
9	BAX_1345650	26243	26152	99,65%	0,35%		0	1	0			5.1.9
10	FUT8_681494	11938	51	0,43%	99,57%	interesting		1	116			
10	BAK1_1544257	23030	411	1,78%	98,22%			1	-3			
10	BAX_1345650	49234	3678	7,47%	92,53%		3	1	77			5.1.10
11	FUT8_681494	17004	16986	99,89%	0,11%	interesting		1	0			
11	BAK1_1544257	84651	84467	99,78%	0,22%			1	0			
11	BAX_1345650	84467	84206	99,69%	0,31%		0	1	0			5.1.11
12	FUT8_681494	11851	11832	99,84%	0,16%	interesting		1	0			
12	BAK1_1544257	94059	93850	99,78%	0,22%			1	0			
12	BAX_1345650	98495	98205	99,71%	0,29%		0	1	0			5.1.12
13	FUT8_681494	22836	22812	99,89%	0,11%	interesting		1	0			
13	BAK1_1544257	103364	102816	99,47%	0,53%			1	0			
13	BAX_1345650	61267	60901	99,40%	0,60%		0	1	0			5.2.1
14	FUT8_681494	11380	14	0,12%	99,88%	interesting		2	35	-4		
14	BAK1_1544257	47424	159	0,34%	99,66%			2	62	2		
14	BAX_1345650	11550	11409	98,78%	1,22%		2	1	0			5.2.2
15	FUT8_681494	5938	2639	44,44%	55,56%	1/2 allelle hit/2 clones		2	-2	0		
15	BAK1_1544257	47295	267	0,56%	99,44%	Fail		1	-2			
15	BAX_1345650	116374	108415	93,16%	6,84%		1	1	0			5.2.3
16	FUT8_681494	10160	5273	51,90%	48,10%	1/2 allelle hit/2 clones		2	0	-11		
16	BAK1_1544257	52469	637	1,21%	98,79%	Fail		2	-3	-81		
16	BAX_1345650	69377	69049	99,53%	0,47%		1	1	0			5.2.4
17	FUT8_681494	20225	20190	99,83%	0,17%	interesting		1	0			
17	BAK1_1544257	77800	220	0,28%	99,72%			1	-14			
17	BAX_1345650	51430	51142	99,44%	0,56%		1	1	0			5.2.5
18	FUT8_681494	8515	34	0,40%	99,60%	interesting		1	-54			
18	BAK1_1544257	93177	166	0,18%	99,82%			2	-7	-4		
18	BAX_1345650	88923	2903	3,26%	96,74%		3	1	-3			5.2.6
19	FUT8_681494	1883	50	2,66%	97,34%	interesting		1	-2			
19	BAK1_1544257	45823	2997	6,54%	93,46%			1	1			
19	BAX_1345650	72622	67760	93,31%	6,69%		2	1	0			5.2.7
20	FUT8_681494	60	42	70,00%	30,00%	Fail		1	0			
20	BAK1_1544257	550	308	56,00%	44,00%			1	0			
20	BAX_1345650	107083	100796	94,13%	5,87%		0	1	0			5.2.8
21	FUT8_681494	6458	6448	99,85%	0,15%	interesting		1	0			
21	BAK1_1544257	29700	29522	99,40%	0,60%			1	0			
21	BAX_1345650	19043	813	4,27%	95,73%		1	1	1			5.2.9
22	FUT8_681494	10600	26	0,25%	99,75%	interesting		1	-15			
22	BAK1_1544257	59178	254	0,43%	99,57%			2	-14	-8		
22	BAX_1345650	67517	2186	3,24%	96,76%		3	1	-2			5.2.10
23	FUT8_681494	12816	78	0,61%	99,39%	interesting		1	1			
23	BAK1_1544257	49419	428	0,87%	99,13%			1	-4			
23	BAX_1345650	42223	41855	99,13%	0,87%		2	1	0			5.2.11
24	FUT8_681494	24594	24565	99,88%	0,12%	interesting		1	0			
24	BAK1_1544257	91694	276	0,30%	99,70%			1	-13			
24	BAX_1345650	92741	3012	3,25%	96,75%		2	1	-2		5.2.12	*

15

#	Name					Note						Section
25	FUT8_681494	67	31	46,27%	53,73%	Fail (1/2 hit)		2	0	1		
25	BAK1_1544257	653	316	48,39%	51,61%			1	0			
25	BAX_1345650	732	540	73,77%	26,23%		0	2	0	-13		5.2.13
26	FUT8_681494	17467	27	0,15%	99,85%	interesting		1	102			
26	BAK1_1544257	53617	259	0,48%	99,52%			1	72			
26	BAX_1345650	6227	6095	97,88%	2,12%		2	1	0			5.2.14
27	FUT8_681494	68	26	38,24%	61,76%	Fail		3	0		-8	
27	BAK1_1544257	148261	339	0,23%	99,77%			2	-10	-31		
27	BAX_1345650	84223	3482	4,13%	95,87%		2	1	-46			5.2.15
28	FUT8_681494	21173	31	0,15%	99,85%	Fail		1	1			
28	BAK1_1544257	43415	217	0,50%	99,50%			1	-57			
28	BAX_1345650	96570	27825	28,81%	71,19%		2	2	-15	0		5.2.16
29	FUT8_681494	18518	9691	52,33%	47,67%	Fail		2	0	-8		
29	BAK1_1544257	77868	34076	43,76%	56,24%			2	-10	0		
29	BAX_1345650	80508	80214	99,63%	0,37%		0	1	0			5.2.17
30	FUT8_681494	26551	23	0,09%	99,91%	interesting		2	-1	-3		
30	BAK1_1544257	104468	275	0,26%	99,74%			1	-11			
30	BAX_1345650	91614	91125	99,47%	0,53%		2	1	0			5.2.18
31	FUT8_681494	23495	23415	99,66%	0,34%	interesting		1	0			
31	BAK1_1544257	109343	108881	99,58%	0,42%			1	0			
31	BAX_1345650	80827	3408	4,22%	95,78%		1	1	-37			5.2.19
32	FUT8_681494	14481	14417	99,56%	0,44%	interesting		1	0			
32	BAK1_1544257	72212	71829	99,47%	0,53%			1	0			
32	BAX_1345650	62236	62007	99,63%	0,37%		0	1	0			5.2.20
33	FUT8_681494	5109	5085	99,53%	0,47%	interesting		1	0			
33	BAK1_1544257	27653	27509	99,48%	0,52%			1	0			
33	BAX_1345650	23638	23507	99,45%	0,55%		0	1	0			5.2.21
34	FUT8_681494	13617	13563	99,60%	0,40%	interesting		1	0			
34	BAK1_1544257	63610	63240	99,42%	0,58%			1	0			
34	BAX_1345650	62646	62355	99,54%	0,46%		0	1	0			5.2.22
35	FUT8_681494	6675	41	0,61%	99,39%	interesting		1	-1			
35	BAK1_1544257	149591	5355	3,58%	96,42%			2	-3	71		
35	BAX_1345650	108304	107053	98,84%	1,16%		2	1	0			5.2.23
36	FUT8_681494	30028	52	0,17%	99,83%	interesting		2	1	-8		
36	BAK1_1544257	97591	487	0,50%	99,50%			2	1	-10		
36	BAX_1345650	95504	5425	5,68%	94,32%		3	1	-13			5.2.24
37	FUT8_681494	18747	12961	69,14%	30,86%	Fail		2	0	-8		
37	BAK1_1544257	90913	32164	35,38%	64,62%			3	0		-2	
37	BAX_1345650	59705	59257	99,25%	0,75%		0	1	0			5.2.25
38	FUT8_681494	7656	7619	99,52%	0,48%	interesting		1	0			
38	BAK1_1544257	38359	38209	99,61%	0,39%			1	0			
38	BAX_1345650	32167	31569	98,14%	1,86%		0	1	0			5.2.26
39	FUT8_681494	13547	30	0,22%	99,78%	interesting		2	-2	-15		
39	BAK1_1544257	65952	510	0,77%	99,23%			2	1	-11		
39	BAX_1345650	63463	2399	3,78%	96,22%		3	1	-3			5.2.27
40	FUT8_681494	10737	19	0,18%	99,82%	Fail		2	1	-6		
40	BAK1_1544257	57817	28243	48,85%	51,15%	1/2 allelle hit/2 clones		2	1	0		
40	BAX_1345650	51229	2469	4,82%	95,18%		2	1	-2			5.2.28
41	FUT8_681494	8933	8906	99,70%	0,30%	interesting		1	0			
41	BAK1_1544257	48698	48160	98,90%	1,10%			1	0			
41	BAX_1345650	52138	51928	99,60%	0,40%		0	1	0			5.2.29
42	FUT8_681494	11287	5348	47,38%	52,62%	1/2 allelle hit/2 clones		2	1	0		
42	BAK1_1544257	41232	332	0,81%	99,19%	Fail		1	2			
42	BAX_1345650	44456	3135	7,05%	92,95%		2	1	1			5.2.30
43	FUT8_681494	7216	31	0,43%	99,57%	interesting		1	-2			
43	BAK1_1544257	32049	469	1,46%	98,54%			1	-14			
43	BAX_1345650	10100	9894	97,96%	2,04%		2	1	0			5.2.31
44	FUT8_681494	14408	29	0,20%	99,80%	interesting		2	-1	-4		
44	BAK1_1544257	60962	417	0,68%	99,32%			1	-15			
44	BAX_1345650	25494	25319	99,31%	0,69%		2	1	0			5.2.32
45	FUT8_681494	4855	17	0,35%	99,65%	Fail		1	-12			
45	BAK1_1544257	168	117	69,64%	30,36%			1	0			
45	BAX_1345650	24415	697	2,85%	97,15%		2	1	1			5.2.33
46	FUT8_681494	5300	13	0,25%	99,75%	interesting		1	-11			
46	BAK1_1544257	26902	298	1,11%	98,89%			1	1			
46	BAX_1345650	12573	11581	92,11%	7,89%		2	1	0			5.2.34
47	FUT8_681494	40394	40326	99,83%	0,17%	interesting		1	0			
47	BAK1_1544257	192086	191536	99,71%	0,29%			1	0			
47	BAX_1345650	174340	173707	99,64%	0,36%		0	1	0			5.2.35
48	FUT8_681494	12745	12684	99,52%	0,48%	interesting		1	0			
48	BAK1_1544257	99759	439	0,44%	99,56%			2	-14	-8		
48	BAX_1345650	49987	1362	2,72%	97,28%		2	1	1			5.2.36

*

16

#	Name					Comment						Label
49	FUT8_681494	24035	24	0,10%	99,90%	Fail		1	-3			
49	BAK1_1544257	132821	470	0,35%	99,65%			2	18	-13		
49	BAX_1345650	101749	53945	53,02%	46,98%	1/2 allele hit/2 clones		2	0	4		6.1.1
50	FUT8_681494	15660	12	0,08%	99,92%	interesting		1	-11			
50	BAK1_1544257	67602	202	0,30%	99,70%			1	-15			
50	BAX_1345650	57438	57208	99,60%	0,40%		2	1	0			6.1.2
51	FUT8_681494	37107	48	0,13%	99,87%	interesting		1	1			
51	BAK1_1544257	117005	263	0,22%	99,78%			1	-28			
51	BAX_1345650	94985	94485	99,47%	0,53%		2	1	0			6.1.3
52	FUT8_681494	14023	13966	99,59%	0,41%	interesting		1	0			
52	BAK1_1544257	61089	60707	99,37%	0,63%			1	0			
52	BAX_1345650	74698	74330	99,51%	0,49%		0	1	0			6.1.4
53	FUT8_681494	67	26	38,81%	61,19%	Fail		4	0			
53	BAK1_1544257	106415	298	0,28%	99,72%			1	-13			
53	BAX_1345650	101088	86391	85,46%	14,54%		1	1	0			6.1.5
54	FUT8_681494	12876	15	0,12%	99,88%	Interesting		2	-1	-5		
54	BAK1_1544257	80300	174	0,22%	99,78%			2	87	-14		
54	BAX_1345650	108677	4760	4,38%	95,62%		3	1	71			6.1.6
55	FUT8_681494	64	21	32,81%	67,19%	Fail		3	0		1	
55	BAK1_1544257	79546	477	0,60%	99,40%			2	1	6		
55	BAX_1345650	141046	131312	93,10%	6,90%		1	1	0			6.1.7
56	FUT8_681494	33112	23	0,07%	99,93%	Fail		2	117	-1		
56	BAK1_1544257	73097	394	0,54%	99,46%			1	-5			
56	BAX_1345650	64062	6273	9,79%	90,21%		3	1	-2			6.1.8
57	FUT8_681494	10655	6	0,06%	99,94%	interesting		1	-1			
57	BAK1_1544257	40452	109	0,27%	99,73%			1	-13			
57	BAX_1345650	40207	611	1,52%	98,48%		3	1	-1			6.1.9
58	FUT8_681494	36698	22	0,06%	99,94%	Fail		2	1	-4		
58	BAK1_1544257	30698	30307	98,73%	1,27%			1	0			
58	BAX_1345650	66373	51144	77,06%	22,94%		1	2	0	-2		6.1.10
59	FUT8_681494	182	103	56,59%	43,41%	Fail		3	0		1	
59	BAK1_1544257	832	335	40,26%	59,74%			2	0	-13		
59	BAX_1345650	169551	3866	2,28%	97,72%		1	1	-9			6.1.11
60	FUT8_681494	39426	19930	50,55%	49,45%	1/2 allele hit/2 clones		2	0	3		
60	BAK1_1544257	103954	103300	99,37%	0,63%			1	0			
60	BAX_1345650	103467	102960	99,51%	0,49%		0	1	0			6.1.12
61	FUT8_681494	13118	13082	99,73%	0,27%	Fail		1	0			
61	BAK1_1544257	96546	77843	80,63%	19,37%			2	0	1		
61	BAX_1345650	110058	109658	99,64%	0,36%		0	1	0			6.1.13
62	FUT8_681494	9206	9186	99,78%	0,22%	Fail		1	0			
62	BAK1_1544257	52844	52621	99,58%	0,42%			1	0			
62	BAX_1345650	54665	50570	92,51%	7,49%		0	1	0			6.1.14
63	FUT8_681494	10241	10207	99,67%	0,33%	Fail		1	0			
63	BAK1_1544257	87369	86914	99,48%	0,52%			1	0			
63	BAX_1345650	118299	101373	85,69%	14,31%		0	2	0	-9		6.1.15
64	FUT8_681494	15600	28	0,18%	99,82%	interesting		2	-20	-1		
64	BAK1_1544257	51660	51399	99,49%	0,51%			1	0			
64	BAX_1345650	95285	94929	99,63%	0,37%		1	1	0			6.1.16
65	FUT8_681494	40218	32	0,08%	99,92%	Interesting		1	59			
65	BAK1_1544257	85525	677	0,79%	99,21%			1	1			
65	BAX_1345650	77133	5397	7,00%	93,00%		3	1	1			6.1.17 *
66	FUT8_681494	68	31	45,59%	54,41%	Fail		4	0			
66	BAK1_1544257	108290	492	0,45%	99,55%			2	-3	-25		
66	BAX_1345650	112739	3767	3,34%	96,66%		2	1	1			6.1.18
67	FUT8_681494	19155	19100	99,71%	0,29%	interesting		1	0			
67	BAK1_1544257	99142	98736	99,59%	0,41%			1	0			
67	BAX_1345650	110305	109900	99,63%	0,37%		0	1	0			6.1.19
68	FUT8_681494	15867	15802	99,59%	0,41%	interesting		1	0			
68	BAK1_1544257	78876	78342	99,32%	0,68%			1	0			
68	BAX_1345650	63627	63274	99,45%	0,55%		0	1	0			6.1.20
69	FUT8_681494	7408	7385	99,69%	0,31%	Interesting		1	0			
69	BAK1_1544257	46556	46427	99,72%	0,28%			1	0			
69	BAX_1345650	19289	19175	99,41%	0,59%		0	1	0			6.1.21
70	FUT8_681494	16190	19	0,12%	99,88%	Interesting		1	-1			
70	BAK1_1544257	84119	497	0,59%	99,41%			2	1	-2		
70	BAX_1345650	61238	1673	2,73%	97,27%		3	1	-8			6.1.22
71	FUT8_681494	18796	18364	97,70%	2,30%	Interesting		1	0			
71	BAK1_1544257	116174	115719	99,61%	0,39%			1	0			
71	BAX_1345650	103058	102592	99,55%	0,45%		0	1	0			6.1.23
72	FUT8_681494	116	75	64,66%	35,34%	Fail		3	0		59	
72	BAK1_1544257	68344	930	1,36%	98,64%			1	1			
72	BAX_1345650	52091	51789	99,42%	0,58%		1	1	0			6.1.24

17

73 FUT8_681494	18158	18112	99,75%	0,25%	interesting		1	0				
73 BAK1_1544257	89109	88831	99,69%	0,31%			1	0				
73 BAX_1345650	82668	82392	99,67%	0,33%		0	1	0			6.2.1	
74 FUT8_681494	12311	12278	99,73%	0,27%	interesting		1	0				
74 BAK1_1544257	58106	57958	99,75%	0,25%			1	0				
74 BAX_1345650	51038	50875	99,68%	0,32%		0	1	0			6.2.2	
75 FUT8_681494	24751	24723	99,89%	0,11%	interesting		1	0				
75 BAK1_1544257	111244	110932	99,72%	0,28%			1	0				
75 BAX_1345650	103628	102727	99,13%	0,87%		0	1	0			6.2.3	
76 FUT8_681494	448	34	7,59%	92,41%	Fail		1	-1				
76 BAK1_1544257	16595	470	2,83%	97,17%			2	-10	-14			
76 BAX_1345650	72563	67314	92,77%	7,23%		2	1	0			6.2.4	
77 FUT8_681494	17911	17883	99,84%	0,16%	interesting		1	0				
77 BAK1_1544257	87181	86904	99,68%	0,32%			1	0				
77 BAX_1345650	84167	83925	99,71%	0,29%		0	1	0			6.2.5	
78 FUT8_681494	33488	66	0,20%	99,80%	Fail		1	1				
78 BAK1_1544257	626	403	64,38%	35,62%			1	0				
78 BAX_1345650	117101	6665	5,69%	94,31%		2	1	-16			6.2.6	
79 FUT8_681494	17456	17423	99,81%	0,19%	interesting		1	0				
79 BAK1_1544257	81727	81470	99,69%	0,31%			1	0				
79 BAX_1345650	102928	102634	99,71%	0,29%		0	1	0			6.2.7	
80 FUT8_681494	21365	21325	99,81%	0,19%	interesting		1	0				
80 BAK1_1544257	99746	99450	99,70%	0,30%			1	0				
80 BAX_1345650	89009	88730	99,69%	0,31%		0	1	0			6.2.8	
81 FUT8_681494	10777	10753	99,78%	0,22%	interesting		1	0				
81 BAK1_1544257	39288	39146	99,64%	0,36%			1	0				
81 BAX_1345650	26199	26090	99,58%	0,42%		0	1	0			6.2.9	
82 FUT8_681494	13943	46	0,33%	99,67%	interesting		2	-1	-2			
82 BAK1_1544257	107010	512	0,48%	99,52%			2	40	3			
82 BAX_1345650	51396	1266	2,46%	97,54%		3	1	-6			6.2.10	
83 FUT8_681494	6501	3885	59,76%	40,24%	Fail		2	0	-10			
83 BAK1_1544257	1918	1773	92,44%	7,56%			1	0				
83 BAX_1345650	203190	202660	99,74%	0,26%		0	1	0			6.2.11	
84 FUT8_681494	15196	15166	99,80%	0,20%	Fail		1	0				
84 BAK1_1544257	118970	118657	99,74%	0,26%			1	0				
84 BAX_1345650	117472	108005	91,94%	8,06%		0	1	0			6.2.12	
85 FUT8_681494	12093	20	0,17%	99,83%	Interesting		1	-5				
85 BAK1_1544257	81864	465	0,57%	99,43%			2	-2	98			
85 BAX_1345650	85296	4578	5,37%	94,63%		3	1	-1			6.2.13	
86 FUT8_681494	57	11	19,30%	80,70%	Fail		4	0				
86 BAK1_1544257	568	228	40,14%	59,86%			1	0				
86 BAX_1345650	112079	111508	99,49%	0,51%		0	1	0			6.2.14	
87 FUT8_681494	24663	21	0,09%	99,91%	Interesting		2	-4	-10			
87 BAK1_1544257	109915	352	0,32%	99,68%			2	-10	-13			
87 BAX_1345650	111986	6664	5,95%	94,05%		3	1	-4			6.2.15	
88 FUT8_681494	11829	6	0,05%	99,95%	interesting		1	-10				
88 BAK1_1544257	114771	274	0,24%	99,76%			2	47	-1			
88 BAX_1345650	60776	60418	99,41%	0,59%		2	1	0			6.2.16	
89 FUT8_681494	21087	19	0,09%	99,91%	interesting		1	-8				
89 BAK1_1544257	123857	623	0,50%	99,50%			1	1				
89 BAX_1345650	88189	87830	99,59%	0,41%		2	1	0			6.2.17	
90 FUT8_681494	11016	47	0,43%	99,57%	Interesting		1	-15				
90 BAK1_1544257	55887	363	0,65%	99,35%			1	3				
90 BAX_1345650	72389	5509	7,61%	92,39%		3	1	1			6.2.18	
91 FUT8_681494	12221	20	0,16%	99,84%	Interesting		2	-2	-23			
91 BAK1_1544257	56027	370	0,66%	99,34%			2	-19	-14			*
91 BAX_1345650	137793	7295	5,29%	94,71%		3	1	88			6.2.19	
92 FUT8_681494	18613	18544	99,63%	0,37%	interesting		1	0				
92 BAK1_1544257	95209	94622	99,38%	0,62%			1	0				
92 BAX_1345650	78199	77813	99,51%	0,49%		0	1	0			6.2.20	
93 FUT8_681494	25742	20	0,08%	99,92%	Fail		2	39	1			
93 BAK1_1544257	26419	201	0,76%	99,24%			1	-14				
93 BAX_1345650	31876	6375	20,00%	80,00%		2	2	-13	0		6.2.21	
94 FUT8_681494	7770	1338	17,22%	82,78%	Fail		3	-8		0		
94 BAK1_1544257	81415	36319	44,61%	55,39%	1/2 allelle hit/2 clones		2	-11	0			
94 BAX_1345650	102215	100212	98,04%	1,96%		0	1	0			6.2.22	
95 FUT8_681494	3693	36	0,97%	99,03%	Fail		2	-3	-23			
95 BAK1_1544257	109274	64280	58,82%	41,18%	1/2 allelle hit/2 clones		2	0	-5			
95 BAX_1345650	164246	159004	96,81%	3,19%		1	1	0			6.2.23	
96 FUT8_681494	4276	1485	34,73%	65,27%	Fail		3	-22		-3		
96 BAK1_1544257	141070	563	0,40%	99,60%			2	2	106			
96 BAX_1345650	100062	99498	99,44%	0,56%		1	1	0			6.2.24	

18

Supplementary Table S9: Deep sequencing results of clones (run #2). Each clone has its own index (column 1). Column 3, 4 and 5 show the number of reads obtained in total, of wild-type sequences and sequences with indels respectively. Clones with unclear data are indicated as fail in column 7 and are also marked as dark gray in column 12. Column 9, 10 and 11 show the type of indels identified in the clones (i.e. basepair difference from wild-type), where peak n = 0 means wild-type alleles. The clones marked with light gray indicate absences of genes with indels (wild-type). The clones marked with yellow indicate presence of one gene with indels (potential single knockout cell line). The clones marked with orange indicate presence of two genes with indels (potential double knockout cell line). The clones marked with red indicate presence of three genes with indels (potential triple knockout cell line). Clones with mixed alleles (presence of both WT allele and allele with indel) are also marked as dark gray and are also not included in Figure 2A.

Index	Product	Counts	WT Count	WT	Indels		Knockout	Peaks	Peak 1	Peak 2	Peak 3	CLONE
1	FUT8_681494	3828	16	0,42%	99,58%		Fail	1	1			5.2.37
1	BAK1_1544257	258	48	18,60%	81,40%			2	-2	0		
1	BAX_1345650	62804	681	1,08%	98,92%	2		1	-15			
2	FUT8_681494	4597	9	0,20%	99,80%		Interesting	1	-30			5.2.38
2	BAK1_1544257	40536	91	0,22%	99,78%			1	1			
2	BAX_1345650	35743	620	1,73%	98,27%	3		1	-2			
3	FUT8_681494	16038	15843	98,78%	1,22%		Fail	1	0			5.2.39
3	BAK1_1544257	51636	51370	99,48%	0,52%			1	0			
3	BAX_1345650	51356	37784	73,57%	26,43%	0		2	0	1		
4	FUT8_681494	8189	9	0,11%	99,89%		Interesting	1	-4			5.2.40
4	BAK1_1544257	45281	40	0,09%	99,91%			1	-2			
4	BAX_1345650	47724	529	1,11%	98,89%	3		1	-7			
5	FUT8_681494	10374	10241	98,72%	1,28%		Interesting	1	0			5.2.41
5	BAK1_1544257	48475	32	0,07%	99,93%			2	8	-3		
5	BAX_1345650	42953	1359	3,16%	96,84%	2		1	-19			
6	FUT8_681494	9198	6	0,07%	99,93%		Interesting	2	-5	-23		5.2.42
6	BAK1_1544257	64131	60	0,09%	99,91%			2	-8	39		
6	BAX_1345650	38936	776	1,99%	98,01%	3		1	-13			
7	FUT8_681494	8933	8476	94,88%	5,12%		Fail	1	0			5.2.43
7	BAK1_1544257	40555	16061	39,60%	60,40%			2	-2	0		
7	BAX_1345650	55214	54609	98,90%	1,10%	0		1	0			
8	FUT8_681494	6203	8	0,13%	99,87%		Interesting	2	1	-46		5.2.44
8	BAK1_1544257	28591	40	0,14%	99,86%			1	-12			
8	BAX_1345650	19532	256	1,31%	98,69%	3		1	-48			
9	FUT8_681494	1506	10	0,66%	99,34%		Fail	2	-2	1		5.2.45
9	BAK1_1544257	122	30	24,59%	75,41%			4	0	-2		
9	BAX_1345650	20345	20151	99,05%	0,95%	1		1	0			
10	FUT8_681494	4928	8	0,16%	99,84%		Fail	1	34			5.2.46
10	BAK1_1544257	15237	29	0,19%	99,81%			1	-4			
10	BAX_1345650	1883	1663	88,32%	11,68%	2		1	0			
11	FUT8_681494	5602	20	0,36%	99,64%		Fail	2	1	-10		5.2.47
11	BAK1_1544257	47194	107	0,23%	99,77%			1	-11			
11	BAX_1345650	64554	51628	79,98%	20,02%	2		2	0	-15		
12	FUT8_681494	12826	9	0,07%	99,93%		Fail (3 peak	3	86	-1	-3	5.2.48
12	BAK1_1544257	39337	41	0,10%	99,90%			2	-15	-14		
12	BAX_1345650	40123	557	1,39%	98,61%	3		1	1			
13	FUT8_681494	13636	13587	99,64%	0,36%		Interesting	1	0			5.2.49
13	BAK1_1544257	65211	65011	99,69%	0,31%			1	0			
13	BAX_1345650	58396	57830	99,03%	0,97%	0		1	0			
14	FUT8_681494	12467	31	0,25%	99,75%		Fail	2	-15	1		5.2.50
14	BAK1_1544257	14119	214	1,52%	98,48%			1	-36			
14	BAX_1345650	45232	15460	34,18%	65,82%	2		2	1	0		
15	FUT8_681494	12307	34	0,28%	99,72%		Fail	2	-2	-8		5.2.51
15	BAK1_1544257	48416	25463	52,59%	47,41%		1/2 hit	2	0	-21		
15	BAX_1345650	17770	17522	98,60%	1,40%	1		1	0			
16	FUT8_681494	15267	15235	99,79%	0,21%		Fail	1	0			5.2.52
16	BAK1_1544257	49726	26291	52,87%	47,13%			3	0	-21	-8	
16	BAX_1345650	46494	5561	11,96%	88,04%	0		2	1	0		
17	FUT8_681494	12184	12070	99,06%	0,94%		Interesting	1	0			5.2.53
17	BAK1_1544257	60166	59977	99,69%	0,31%			1	0			
17	BAX_1345650	42141	41780	99,14%	0,86%	0		1	0			
18	FUT8_681494	10270	7003	68,19%	31,81%		Fail	2	0	1		5.2.54
18	BAK1_1544257	49981	49812	99,66%	0,34%			1	0			
18	BAX_1345650	62268	59351	95,32%	4,68%	0		1	0			
19	FUT8_681494	11723	88	0,75%	99,25%		Interesting	1	-8			5.2.55
19	BAK1_1544257	72951	241	0,33%	99,67%			2	-25	-28		
19	BAX_1345650	66204	2056	3,11%	96,89%	3		1	-3			
20	FUT8_681494	9662	24	0,25%	99,75%		Interesting	1	-1			5.2.56
20	BAK1_1544257	25984	134	0,52%	99,48%			1	-61			
20	BAX_1345650	13888	13701	98,65%	1,35%	2		1	0			
21	FUT8_681494	6289	5219	82,99%	17,01%		Fail	2	0	-3		5.2.57
21	BAK1_1544257	188	123	65,43%	34,57%			1	0			
21	BAX_1345650	17636	184	1,04%	98,96%	1		1	-15			
22	FUT8_681494	6556	27	0,41%	99,59%		Fail	1	-18			5.2.58
22	BAK1_1544257	3245	223	6,87%	93,13%			1	-4			
22	BAX_1345650	14151	11769	83,17%	16,83%	2		2	0	1		
23	FUT8_681494	16061	33	0,21%	99,79%		Fail	2	66	-27		5.2.59
23	BAK1_1544257	385	233	60,52%	39,48%			1	0			
23	BAX_1345650	14943	14688	98,29%	1,71%	1		1	0			
24	FUT8_681494	7783	25	0,32%	99,68%		Interesting	1	-7			5.2.60
24	BAK1_1544257	45224	185	0,41%	99,59%			2	-3	-21		
24	BAX_1345650	42639	792	1,86%	98,14%	3		1	1			

19

idx	name										code
25	FUT8_681494	2862	26	0,91%	99,09%	Fail		1	1		
25	BAK1_1544257	442	305	69,00%	31,00%			1	0		5.2.61
25	BAX_1345650	89164	1635	1,83%	98,17%	2		1	1		
26	FUT8_681494	9588	9553	99,63%	0,37%	Fail		1	0		
26	BAK1_1544257	45119	44949	99,62%	0,38%			1	0		5.2.62
26	BAX_1345650	42495	28779	67,72%	32,28%	0	2	0	1		
27	FUT8_681494	19687	62	0,31%	99,69%	Fail		1	-9		
27	BAK1_1544257	54874	300	0,55%	99,45%			1	-11		5.2.63
27	BAX_1345650	41759	5449	13,05%	86,95%	2	2	1	0		
28	FUT8_681494	9225	32	0,35%	99,65%	Fail		1	1		
28	BAK1_1544257	20942	204	0,97%	99,03%		2	-8	75		5.2.64
28	BAX_1345650	74459	44186	59,34%	40,66%	2	2	0	-24		
29	FUT8_681494	11498	27	0,23%	99,77%	Interesting	2	-1	1		
29	BAK1_1544257	68730	216	0,31%	99,69%		2	1	24		5.2.65
29	BAX_1345650	49924	786	1,57%	98,43%	3	1	2			
30	FUT8_681494	13786	25	0,18%	99,82%	Fail		1	-18		
30	BAK1_1544257	448	271	60,49%	39,51%			1	0		5.2.66
30	BAX_1345650	65411	888	1,36%	98,64%	2	1	3			
31	FUT8_681494	12792	12717	99,41%	0,59%	Interesting		1	0		
31	BAK1_1544257	58274	57945	99,44%	0,56%			1	0		5.1.13
31	BAX_1345650	66794	65994	98,80%	1,20%	0	1	0			
32	FUT8_681494	8760	8486	96,87%	3,13%	Fail		1	0		
32	BAK1_1544257	40106	39804	99,25%	0,75%			1	0		5.1.14
32	BAX_1345650	36684	36148	98,54%	1,46%	0	1	0			
33	FUT8_681494	7368	7342	99,65%	0,35%	Interesting		1	0		
33	BAK1_1544257	34697	34596	99,71%	0,29%			1	0		5.1.15
33	BAX_1345650	26603	26351	99,05%	0,95%	0	1	0			
34	FUT8_681494	8093	8069	99,70%	0,30%	Interesting		1	0		
34	BAK1_1544257	45630	45452	99,61%	0,39%			1	0		5.1.16
34	BAX_1345650	44096	43687	99,07%	0,93%	0	1	0			
35	FUT8_681494	17157	17095	99,64%	0,36%	Interesting		1	0		
35	BAK1_1544257	89545	89200	99,61%	0,39%			1	0		5.1.17
35	BAX_1345650	45132	44650	98,93%	1,07%	0	1	0			
36	FUT8_681494	8765	20	0,23%	99,77%	Interesting		1	-54		
36	BAK1_1544257	72744	231	0,32%	99,68%		2	-7	-4		5.2.6
36	BAX_1345650	72633	2253	3,10%	96,90%	3	1	-3			
37	FUT8_681494	9179	9127	99,43%	0,57%	Interesting		1	0		
37	BAK1_1544257	46161	45967	99,58%	0,42%			1	0		6.1.25
37	BAX_1345650	41231	40651	98,59%	1,41%	0	1	0			
38	FUT8_681494	4160	10	0,24%	99,76%	Interesting		1	-16		
38	BAK1_1544257	23810	95	0,40%	99,60%			1	-12		6.1.26
38	BAX_1345650	18404	456	2,48%	97,52%	3	1	1			
39	FUT8_681494	1308	20	1,53%	98,47%	Fail		1	1		
39	BAK1_1544257	220	98	44,55%	55,45%	1/2 hit	2	0	1		6.1.27
39	BAX_1345650	24096	23907	99,22%	0,78%	1	1	0			
40	FUT8_681494	8417	8263	98,17%	1,83%	Fail		1	0		
40	BAK1_1544257	27287	27145	99,48%	0,52%			1	0		6.1.28
40	BAX_1345650	40857	38941	95,31%	4,69%	0	1	0			
41	FUT8_681494	5920	16	0,27%	99,73%	Interesting		1	-15		
41	BAK1_1544257	26295	58	0,22%	99,78%			1	99		6.1.29
41	BAX_1345650	7870	7401	94,04%	5,96%	2	1	0			
42	FUT8_681494	9140	14	0,15%	99,85%	Interesting	2	-23	-15		
42	BAK1_1544257	46960	177	0,38%	99,62%			1	1		6.1.30
42	BAX_1345650	40025	693	1,73%	98,27%	3	1	-3			
43	FUT8_681494	88	31	35,23%	64,77%	Fail	2	0	1		
43	BAK1_1544257	37276	105	0,28%	99,72%			1	-13		6.1.31
43	BAX_1345650	40129	435	1,08%	98,92%	2	1	-21			
44	FUT8_681494	5484	12	0,22%	99,78%	Interesting		1	-33		
44	BAK1_1544257	33504	72	0,21%	99,79%			1	-15		6.1.32
44	BAX_1345650	11987	11884	99,14%	0,86%	2	1	0			
45	FUT8_681494	4063	4037	99,36%	0,64%	Fail		1	0		
45	BAK1_1544257	12987	61	0,47%	99,53%			1	97		6.1.33
45	BAX_1345650	4381	3309	75,53%	24,47%	1	2	0	-2		
46	FUT8_681494	7450	636	8,54%	91,46%	Fail	3	-1	21	1	
46	BAK1_1544257	27805	112	0,40%	99,60%		2	1	-3		6.1.34
46	BAX_1345650	33559	33350	99,38%	0,62%	2	1	0			
47	FUT8_681494	9516	9050	95,10%	4,90%	Fail		1	0		
47	BAK1_1544257	45614	45391	99,51%	0,49%			1	0		6.1.35
47	BAX_1345650	35123	32737	93,21%	6,79%	0	1	0			
48	FUT8_681494	4431	1925	43,44%	56,56%	Fail	3	0	6	1	
48	BAK1_1544257	39068	90	0,23%	99,77%		2	51	66		6.1.36
48	BAX_1345650	34029	570	1,68%	98,32%	2	1	71			

20

#	Name										Ref
49	FUT8_681494	26445	37	0,14%	99,86%	Fail	2	-3	103		6.1.37
49	BAK1_1544257	395	285	72,15%	27,85%		1	0			
49	BAX_1345650	70638	3949	5,59%	94,41%	2	1	-1			
50	FUT8_681494	7797	7279	93,36%	6,64%	Fail	1	0			6.1.38
50	BAK1_1544257	41698	41546	99,64%	0,36%		1	0			
50	BAX_1345650	36594	36330	99,28%	0,72%	0	1	0			
51	FUT8_681494	16646	16099	96,71%	3,29%	Fail	1	0			6.1.39
51	BAK1_1544257	61997	61783	99,65%	0,35%		1	0			
51	BAX_1345650	55053	54625	99,22%	0,78%	0	1	0			
52	FUT8_681494	9088	18	0,20%	99,80%	Fail	1	1			6.1.40
52	BAK1_1544257	314	197	62,74%	37,26%		3	0	-10	-14	
52	BAX_1345650	19179	18823	98,14%	1,86%	1	1	0			
53	FUT8_681494	14273	20	0,14%	99,86%	Interesting	1	-5			6.1.41
53	BAK1_1544257	56069	211	0,38%	99,62%		2	-13	-27		
53	BAX_1345650	71073	1017	1,43%	98,57%	3	1	3			
54	FUT8_681494	16023	15052	93,94%	6,06%	Fail	1	0			6.1.42
54	BAK1_1544257	69838	69620	99,69%	0,31%		1	0			
54	BAX_1345650	71091	70618	99,33%	0,67%	0	1	0			
55	FUT8_681494	12619	3753	29,74%	70,26%	Fail	5	0	-42		6.1.43
55	BAK1_1544257	58285	262	0,45%	99,55%		1	-14			
55	BAX_1345650	74722	904	1,21%	98,79%	2	1	-6			
56	FUT8_681494	8656	11	0,13%	99,87%	Interesting	2	-1	-11		6.1.44
56	BAK1_1544257	44119	189	0,43%	99,57%		1	-10			
56	BAX_1345650	13302	13089	98,40%	1,60%	2	1	0			
57	FUT8_681494	8200	4028	49,12%	50,88%	Fail (1/2 hit)	2	-5	0		6.1.45
57	BAK1_1544257	38095	38009	99,77%	0,23%		1	0			
57	BAX_1345650	23101	500	2,16%	97,84%	1	1	-3			
58	FUT8_681494	10448	10406	99,60%	0,40%	Fail	1	0			6.1.46
58	BAK1_1544257	53576	53391	99,65%	0,35%		1	0			
58	BAX_1345650	49326	43613	88,42%	11,58%	0	2	0	1		
59	FUT8_681494	10664	49	0,46%	99,54%	Interesting	2	-6	-22		6.1.47
59	BAK1_1544257	73435	316	0,43%	99,57%		2	-14	-6		
59	BAX_1345650	65495	808	1,23%	98,77%	3	1	-7			
60	FUT8_681494	11733	4649	39,62%	60,38%	Fail	3	0	-25	-13	6.1.48
60	BAK1_1544257	62139	61935	99,67%	0,33%		1	0			
60	BAX_1345650	52152	20941	40,15%	59,85%	0	2	1	0		
61	FUT8_681494	14188	19	0,13%	99,87%	Interesting	1	-2			6.1.49
61	BAK1_1544257	49583	113	0,23%	99,77%		2	-32	-35		
61	BAX_1345650	88854	1067	1,20%	98,80%	3	1	37			
62	FUT8_681494	5828	4807	82,48%	17,52%	Fail	2	0	1		6.1.50
62	BAK1_1544257	32710	32270	98,65%	1,35%		1	0			
62	BAX_1345650	49791	45844	92,07%	7,93%	0	1	0			
63	FUT8_681494	4717	14	0,30%	99,70%	Fail	2	-11	1		6.1.51
63	BAK1_1544257	39084	121	0,31%	99,69%		1	-2			
63	BAX_1345650	39174	33548	85,64%	14,36%	2	2	0	1		
64	FUT8_681494	10503	7368	70,15%	29,85%	Fail	2	0	-47		6.2.49
64	BAK1_1544257	55445	46	0,08%	99,92%		1	-14			
64	BAX_1345650	70472	484	0,69%	99,31%	2	1	-11			
65	FUT8_681494	11263	10	0,09%	99,91%	Interesting	1	-27			6.2.50
65	BAK1_1544257	75408	40	0,05%	99,95%		1	52			
65	BAX_1345650	71984	517	0,72%	99,28%	3	1	-3			
66	FUT8_681494	8524	13	0,15%	99,85%	Fail	1	-1			6.2.51
66	BAK1_1544257	58122	97	0,17%	99,83%		2	-4	-9		
66	BAX_1345650	18486	16433	88,89%	11,11%	2	1	0			
67	FUT8_681494	16234	18	0,11%	99,89%	Fail	2	-12	-5		6.2.52
67	BAK1_1544257	35979	129	0,36%	99,64%		1	-13			
67	BAX_1345650	119378	36488	30,57%	69,43%	2	2	-12	0		
68	FUT8_681494	5798	7	0,12%	99,88%	Interesting	2	-23	36		6.2.53
68	BAK1_1544257	34124	56	0,16%	99,84%		1	-14			
68	BAX_1345650	5300	4884	92,15%	7,85%	2	1	0			
69	FUT8_681494	9570	9	0,09%	99,91%	Interesting	1	-15			5.2.10
69	BAK1_1544257	49234	104	0,21%	99,79%		2	-14	-8		
69	BAX_1345650	40540	1136	2,80%	97,20%	3	1	-2			
70	FUT8_681494	24300	13	0,05%	99,95%	Fail	1	102			5.2.14
70	BAK1_1544257	67635	95	0,14%	99,86%		1	72			
70	BAX_1345650	7979	7534	94,42%	5,58%	2	1	0			
71	FUT8_681494	25185	16	0,06%	99,94%	Interesting	2	-1	-3		5.2.18
71	BAK1_1544257	91586	146	0,16%	99,84%		1	-11			
71	BAX_1345650	63549	62977	99,10%	0,90%	2	1	0			
72	FUT8_681494	22714	9	0,04%	99,96%	Interesting	2	-2	-15		5.2.27
72	BAK1_1544257	98478	115	0,12%	99,88%		2	1	-11		
72	BAX_1345650	76270	2288	3,00%	97,00%	3	1	-3			

21

#	Probe	Total	Count	Pct1	Pct2	Class	Flag	V1	V2	V3	Ref
73	FUT8_681494	10275	8612	83,82%	16,18%	Fail	2	0	1		
73	BAK1_1544257	70380	70044	99,52%	0,48%		1	0			6.2.25
73	BAX_1345650	64374	63588	98,78%	1,22%	0	1	0			
74	FUT8_681494	5984	12	0,20%	99,80%	Fail	1	-11			
74	BAK1_1544257	45893	157	0,34%	99,66%		2	-4	68		6.2.26
74	BAX_1345650	13514	11563	85,56%	14,44%	2	1	0			
75	FUT8_681494	17581	16	0,09%	99,91%	Interesting	2	31	-6		
75	BAK1_1544257	71989	207	0,29%	99,71%		1	24			6.2.27
75	BAX_1345650	58099	1017	1,75%	98,25%	3	1	-9			
76	FUT8_681494	3335	17	0,51%	99,49%	Interesting	1	1			
76	BAK1_1544257	17773	135	0,76%	99,24%		2	1	-1		6.2.28
76	BAX_1345650	45917	1219	2,65%	97,35%	3	1	1			
77	FUT8_681494	19441	14	0,07%	99,93%	Fail	2	39	-5		
77	BAK1_1544257	389	108	27,76%	72,24%		3	0	-10	1	6.2.29
77	BAX_1345650	5152	4863	94,39%	5,61%	1	1	0			
78	FUT8_681494	11175	11	0,10%	99,90%	Interesting	1	-21			
78	BAK1_1544257	68580	208	0,30%	99,70%		2	-3	-10		6.2.30
78	BAX_1345650	73933	1426	1,93%	98,07%	3	1	1			
79	FUT8_681494	8317	32	0,38%	99,62%	Interesting	1	1			
79	BAK1_1544257	57133	137	0,24%	99,76%		2	-10	2		6.2.31
79	BAX_1345650	62269	3469	5,57%	94,43%	3	1	-8			
80	FUT8_681494	5735	17	0,30%	99,70%	Fail	3	1	-23	-12	
80	BAK1_1544257	349	140	40,11%	59,89%		2	0	-10		6.2.32
80	BAX_1345650	72617	67078	92,37%	7,63%	1	1	0			
81	FUT8_681494	3832	2572	67,12%	32,88%	Fail	2	0	1		
81	BAK1_1544257	14363	14241	99,15%	0,85%		1	0			6.2.33
81	BAX_1345650	12594	12400	98,46%	1,54%	0	1	0			
82	FUT8_681494	7123	10	0,14%	99,86%	Fail	3	-11	-18	1	
82	BAK1_1544257	27803	152	0,55%	99,45%		1	-10			6.2.34
82	BAX_1345650	61826	33080	53,50%	46,50%	2 clones	2	0	1		
83	FUT8_681494	11510	11440	99,39%	0,61%	Interesting	1	0			
83	BAK1_1544257	60886	60503	99,37%	0,63%		1	0			6.2.35
83	BAX_1345650	56542	56054	99,14%	0,86%	0	1	0			
84	FUT8_681494	8550	15	0,18%	99,82%	Interesting	1	-5			
84	BAK1_1544257	49563	157	0,32%	99,68%		2	1	-10		6.2.36
84	BAX_1345650	46514	514	1,11%	98,89%	3	1	-3			
85	FUT8_681494	10351	381	3,68%	96,32%	Interesting	2	1	-19		
85	BAK1_1544257	36367	297	0,82%	99,18%		1	1			6.2.37
85	BAX_1345650	60044	1743	2,90%	97,10%	3	1	-1			
86	FUT8_681494	14014	13951	99,55%	0,45%	Interesting	1	0			
86	BAK1_1544257	64992	64744	99,62%	0,38%		1	0			6.2.38
86	BAX_1345650	64879	64404	99,27%	0,73%	0	1	0			
87	FUT8_681494	13090	31	0,24%	99,76%	Interesting	1	-1			
87	BAK1_1544257	57280	277	0,48%	99,52%		1	3			6.2.39
87	BAX_1345650	32009	31650	98,88%	1,12%	2	1	0			
88	FUT8_681494	17790	25	0,14%	99,86%	Fail	3	-12	-3	1	
88	BAK1_1544257	33376	174	0,52%	99,48%		1	-10			6.2.40
88	BAX_1345650	112350	64969	57,83%	42,17%	2 clones	2	0	-2		
89	FUT8_681494	9810	26	0,27%	99,73%	Fail	2	-11	1		
89	BAK1_1544257	431	203	47,10%	52,90%		2	0	1		6.2.41
89	BAX_1345650	134946	134106	99,38%	0,62%	1	1	0			
90	FUT8_681494	9610	1837	19,12%	80,88%	Fail	2	1	0		
90	BAK1_1544257	65674	65345	99,50%	0,50%		1	0			6.2.42
90	BAX_1345650	114638	113798	99,27%	0,73%	0	1	0			
91	FUT8_681494	11646	33	0,28%	99,72%	Fail	3	-8	-19	1	
91	BAK1_1544257	50784	400	0,79%	99,21%		1	-1			6.2.43
91	BAX_1345650	123193	77252	62,71%	37,29%	2	2	0	1		
92	FUT8_681494	11730	19	0,16%	99,84%	Interesting	2	-8	-23		
92	BAK1_1544257	78820	220	0,28%	99,72%		2	-32	-13		6.2.44
92	BAX_1345650	20577	20351	98,90%	1,10%	2	1	0			
93	FUT8_681494	9723	8342	85,80%	14,20%	Fail	1	0			
93	BAK1_1544257	55002	45923	83,49%	16,51%		1	0			6.2.45
93	BAX_1345650	48889	42436	86,80%	13,20%	0	2	0	2		
94	FUT8_681494	14722	13726	93,23%	6,77%	Fail	1	0			
94	BAK1_1544257	71740	31254	43,57%	56,43%		2	-3	0		6.2.46
94	BAX_1345650	76820	43595	56,75%	43,25%	2 clones	2	0	-15		
95	FUT8_681494	15043	36	0,24%	99,76%	Interesting	2	1	-2		
95	BAK1_1544257	66898	401	0,60%	99,40%		1	1			6.2.47
95	BAX_1345650	71771	2304	3,21%	96,79%	3	2	1	-15		
96	FUT8_681494	20148	424	2,10%	97,90%	Fail	1	1			
96	BAK1_1544257	501	260	51,90%	48,10%		2	0	1		6.2.48
96	BAX_1345650	162801	151035	92,77%	7,23%	1	1	0			

22

Supplementary Table S10: Analysis of indels generated in the multiplexed and FACS enriched clones. The aligned sequences are obtained from Sanger sequencing of the region flanking the sgRNA target sites in *FUT8*, *BAK* and *BAX* respectively. The PAM sequence of the sgRNA target sites is marked in bold.

FUT8 locus

sgRNA target sequence	GTCAGACGCACTGAC------------------------------AAAGT**GGG**
CHO-S	CCCAGAGTCCATGTCAGACGCACTGAC-----------------------AAAGTGGGAACAGAGAAGCAGCCTTC
CHO-S conditioned	CCCAGAGTCCATGTCAGACGCACTGAC-----------------------AAAGTGGGAACAGAGAAGCAGCCTTC
BAK(-13) BAX(-2)	CCCAGAGTCCATGTCAGACGCACTGAC-----------------------AAAGTGGGAACAGAGAAGCAGCCTTC
FUT8(-2,-23) BAK(-19,-14)	CCCAGAGTCCATGTCAGACGCACTGAC-----------------------A-GTGGGAACAGAAGCAGCCTTC
BAX(+88)	CCCAGAGTCCATGTCAGACGCA--------------------------------GCCTTC
FUT8(+59) BAK(+1) BAX(+1)	CCCAGAGTCCATGTCAGACGCTTTACCGGCGTTTACCGGATATGCTCACCGGATTTTTACTCGACTCGGCGGAAAAATATTGGGCCTAAAGAGTGGGAACAGAGAAGCAGCCTTC

BAK locus

sgRNA target sequence	**CCA**ACG-TGGGTTTGACCGGCTTCC
CHO-S	GGCTCTCCTGGGCTTCGGCTACCGCCTGGCCCTGTATGTCTACCAACG-TGGTTTGACCGGCTTCCTGGGCCAGGTGACCTGCTTTTTGG
CHO-S conditioned	GGCTCTCCTGGGCTTCGGCTACCGCCTGGCCCTGTATGTCTACCAACG-TGGTTTGACCGGCTTCCTGGGCCAGGTGACCTGCTTTTTGG
BAK(-13) BAX(-2)	GGCTCTCCTGGGCTTCGGCTACCGCCTGGCCCTGTATGT------------TTGACCGGCTTCCTGGGCCAGGTGACCTGCTTTTTGG
FUT8(-2,-23) BAK(-19,-14)	GGCTCTCCTGGGCTTCGGCTACCGCCTGGCCCTGGCCCTGT------------GACCGGCTTCCTGGGCCAGGTGACCTGCTTTTTGG
BAX(+88)	GGCTCTCCTGGGCTTCGGCTACCGCCTGGCCCTGTATGTCTACC-----------GGCTTCCTGGGCCAGGTGACCTGCTTTTTGG
FUT8(+59) BAK(+1) BAX(+1)	GGCTCTCCTGGGCTTCGGCTACCGCCTGGCCCTGTATGTCTACCAACGGTGGTTTGACCGGCTTCCTGGGCCAGGTGACCTGCTTTTTGG

BAX locus

sgRNA target sequence	GCTGATGGCAACTTCA------------------------------A-CTG**GGG**
CHO-S	ATGTTTGCTGATGGCAACTTCA-----------------------A-CTGGGGCCGGGTTGTTGCC
CHO-S conditioned	ATGTTTGCTGATGGCAACTTCA-----------------------A-CTGGGGCCGGGTTGTTGCC
BAK(-13) BAX(-2)	ATGTTTGCTGATGGCAACTTCA-------------------------TGGGGCCGGGTTGTTGCC
FUT8(-2,-23) BAK(-19,-14)	ATGTTTGCTGATGGCAACTTGATAACGGACTAGCCTTATTTAACTTGCTATTTCTAGCTCTAAAACCAGTTGAAGTTGCCATCAGCGGTGTTTCGTCCTTTCCACAAGATCTGGGGCCGGGTTGTTGCC
BAX(+88)	
FUT8(+59) BAK(+1) BAX(+1)	ATGTTTGCTGATGGCAACTTCA----------------------AACTGGGGCCGGGTTGTTGCC

23

Supplementary Table S11: Analysis of indels generated in the multiplexed and FACS enriched clones. The sequences are obtained from deep sequencing of the region flanking the sgRNA target sites in *FUT8*, *BAK* and *BAX* respectively for the CHO-S conditioned, *BAK(-13) BAX(-2)* double knockout cell line, *FUT8(-2,-23) BAK(-19,-14) BAX(+88)* triple knockout cell line and *FUT8(+59) BAK(+1) BAX(+1)* triple knockout cell line. Deletions are marked with red dashes. Insertions are underscored.

CHO-S conditioned

WT *FUT8* (11824 reads):

TGCCCCATGACTAGGGATACTAATTGAGTACCAGTACATTATCAGTGTGCTCTCCACTTCTCCCCAGAGTCCATTGTCAGACGCACTGACAAAGTGGGAACAGAAGCAGCCTTCCATCCCCATTGAGGAATACATGGTTACAGTTGAAGAACATTTTCAGCTTCTCGAAGGCAGA

WT *BAK* (63151 reads):

CAAGGTGGGCTCTCCGTGATCTATTTGCCCGCAGCCTATTTTAAGAGCGGGCATCAGCTGGGCGCCGTGTGGTGGCTCTCCTGGGCTTCGCGCCGCTGGCCTGCTATGTGCTTACCAAGCTGGTTTGACCGGCTGCACCTGCTTTTGGCGGATATCATACTGCACCATTGCATCG

WT *BAX* (54557 reads):

TGTGGATACTAACTCCCCACGAGAGGTCTTCCGTGTGGCAGCTGCACATGTTTCGTGATGGCAACTTCAACTGGGCCGGGTTGTTGCCCTTTCTGCTGATGCGAGAAACTTGCTGTCTAGGAAACTGGTGCTCAAGGTGGGCAGCTACAGGGCATGGGCTCAGAGATGTACCCTCAAGTCTGAGGACCTGGGGTAGTGAGGTTCAGGGA

BAK(-13) BAX(-2) double knockout cell line

WT *FUT8* (20449 reads):

TGCCCCATGACTAGGGATACTAATTGAGTACCAGTACATTATCAGTGTGCTCTCCACTTCTCCCCAGAGTCCATTGTCAGACGCACTGACAAAGTGGGAACAGAAGCAGCCTTCCATCCCCATTGAGGAATACATGGTTACAGTTGAAGAACATTTTCAGCTTCTCGAAGGCAGA

13 bp deletion *BAK* (80430 reads):

CAAGGTGGGCTCTCCGTGATCTATTTGCCCGCAGCCTATTTTAAGAGCGGGCATCAGCTGGGCGCCGTGTGGTGGCTCTCCTGGGCTTCGGCTACCGGCCGCTGTGT------------TTGACCGGCTGCTCTGGGCCCAGGTTGAGCCTGCTTTTTGGCGGATATCATACTGCACCATTGCATCG

2 bp deletion *BAX* (79004 reads):

TGTGGATACTAACTCCCCACGAGAGGTCTTCCGTGTGGCAGCTGCACATGTTTCGTGATGGCAACTTCA---TGGGGCCGGGTTGTTGCCCTTTCTGCTGATGCGAGAAACTGGTGCTCAAGGTGGGCAGCTACAGGGCATGGGCTCAGAGATGTACCCTCAAGTCTGAGGACCTGGGGTAGTGAGGTTCAGGGA

FUT8(+59) BAK(+1) BAX(+1) triple knockout cell line

59 bp insert *FUT8* (31398 reads):

TGCCCCATGACTAGGGATACTAATTGAGTACCAGTACATTATCAGTGTGCTCTCCACTTCTCCCCAGAGTCCATTGTCAGACGCTTTACCGACTGGCGATATGCTCACCGGGATTTTTACTGCACTGACGATAATGTGGGAAGAGAGCAGCCTTCCATCCCATTGAGGAATACATGGTTACAGTTGAAGAACATTTTCAGCTTCTCGAACGCAGA

24

108

1 bp insert *BAK* (75083 reads):

CAAGGTGGGCTCTCCGTGATCTATTTGCCCGCAGCCTATTTAAGAGCGGCATCAGTTGGGCGCTGTTCGGGGCAGGTGACCTGGTTTTTGCCGGATATCATACTGCACCATTGCATCG

1 bp insert *BAX* (63177 reads):

TGTTGGATACTAACTCCCCACGAGAGGTCTTCTCCGTGTGCGAGCTGACATGTTTGCTGATGGCAACTTCAAGCTGGGCCGGGTTGTTGCCCTTTTCTACTTTGCTAGCAAACTGGTCACCTGGGGCTAGGACCTGGGGGTAGTGAGGTTCAGGGA

FUT8(-2,-23) BAK(-19,-14) BAX(+88) triple knockout cell line

2 bp deletion *FUT8* (5690 reads):

TGCCCCATGCACTAGGGATACTAATTGAGTACCAGTACATTATCAGTGTGCTCGCCACTTCTCCCAGAGTCCATGTCAGACGCACTGACA--GTGGGAACAGAAGCAGCCTTCCATCCCATTGAGGAATACATGGTACACGTTGAAGAACATTTTCAGCTTCTCGAACGCAGA

23 bp deletion *FUT8* (5023 reads):

TGCCCCATGCACTAGGGATACTAATTGAGTACCAGTACATTATCAGTGTGTCGCTCTCCCACTTCTCCCAGAGTCCATGTCAGACGCA-----------------------G-CCTTCCATCCCATTGAGGAATACATGGTACACGTTGAAGAACATTTTCAGCTTCTCGAACGCAGA

19 bp deletion *BAK* (25014 reads):

CAAGGTGGGCTCTCCGTGATCTATTTGCCCGCAGCCTATTTAAGAGCGGCATCAGT-------------------GACCGGCTTCTGGGCCAGGTGAACCTGGTTTTTGGCGGATATCATACTGCACCATTGCATCG

14 bp deletion *BAK* (24219 reads):

CAAGGTGGGCTCTCCGTGATCTATTTGCCCGCAGCCTATTTAAGAGCGGCATCAGT--------------GGCTTCTGGGCCAGGTGACCTGCTTTTTGGCGGATATCATACTGCACCATTGCATCG

88 bp insert *BAX* (91187 reads):

TGTTGGATACTAACTCCCCACGAGAGGTCTTCTCCGTGTGCGAGCTGACATGTTTGCTGATGGCAACTTCAAGCTGGGCCGGGTTGTTGCCCTTTTCTACTTTGCTAGCAAACTGGTCACCTGGGGCTAGGACCTGGGGGTAGTGAGGTTCAGGGA

25

Supplementary Table S12: Mass-to-charge ratio (m/z) of each peak shown in Figure 3B

Glycosylation form	m/z
1	1354.5
2	1744.7
3	1906.7
4	2051.8
5	2197.8
6	2342.9
7	2488.9
8	2854.0
9	3145.1
10	3801.4
11	1760.7
12	2999.1
13	3655.3

26

Supplementary Table S13: Transient transfection of multiplexed and FACS enriched clones. The table presents the raw date from the analysis of growth (viable cell concentration), viability and titer upon transient transfection of the clones with antibody (Rituximab) expression plasmid and cultivation for three days. The Integrated viable cell density and productivity are also presented.

Replicate 1

	Viable cell concentration (mio cells/ml)				Cell viability (%)				Integrated VCD			Titer (µg/ml)				Qrit (pcd)
	t=0	t=1	t=2	t=3	t=0	t=1	t=2	t=3	t=1	t=2	t=3	t=0	t=1	t=2	t=3	
CHO-S	0.53	2.53	3.47	4.82	96.8	94.6	99.0	98.8	1.53	4.53	8.67	0.79	0.89	2.52	5.36	0.63
CHO-S (conditioned)	0.64	3.21	6.34	8.15	96.3	96.3	98.8	99.4	1.92	6.70	13.94	1.10	1.49	4.52	6.22	0.38
BAK(-13) BAX(-2)	0.85	2.41	2.95	5.66	97.2	97.0	99.0	97.3	1.63	4.30	8.61	0.52	1.33	4.79	8.61	1.03
FUT8(-2,-23) BAK(-19,-14) BAX(+88)	0.82	2.61	5.44	6.41	98.1	97.3	99.6	99.3	1.71	5.74	11.66	0.00	1.87	2.88	5.52	0.37
FUT8(+59) BAK(+1) BAX(+1)	0.92	2.59	3.49	6.33	96.1	95.4	98.1	98.6	1.75	4.79	9.70	0.79	2.73	6.59	11.95	1.15

Replicate 2

	Viable cell concentration (mio cells/ml)				Cell viability (%)				Integrated VCD			Titer (µg/ml)				Qrit (pcd)
	t=0	t=1	t=2	t=3	t=0	t=1	t=2	t=3	t=1	t=2	t=3	t=0	t=1	t=2	t=3	
CHO-S	0.65	1.21	2.36	4.50	94.3	95.5	98.5	99.6	0.93	2.72	6.15	0	0.67	2.12	4.47	0.72
CHO-S (conditioned)	0.33	1.81	4.20	6.65	95.9	96.5	97.6	98.9	1.07	4.07	9.49	0	1.44	2.92	4.83	0.40
BAK(-13) BAX(-2)	0.57	1.31	1.97	3.86	96.8	98.2	97.3	99.1	0.94	2.57	5.49	0	1.04	2.79	7.34	1.41
FUT8(-2,-23) BAK(-19,-14) BAX(+88)	1.00	2.38	3.85	6.63	98.0	98.3	99.2	99.0	1.69	4.80	10.05	0	1.76	3.28	5.16	0.40
FUT8(+59) BAK(+1) BAX(+1)	0.66	1.65	2.45	4.35	98.6	96.9	97.3	96.7	1.16	3.21	6.61	0	1.45	5.34	9.33	1.42

27

Supplementary Reference

[1] Hamming, R.W., Error detecting and error correcting codes. *Bell Syst. Tech. J.* 1950, *29*, 147-160.

28

Supplementary Material

Accelerated Homology-Directed Targeted Integration of Transgenes in Chinese Hamster Ovary Cells via CRISPR/Cas9 and Fluorescent Enrichment
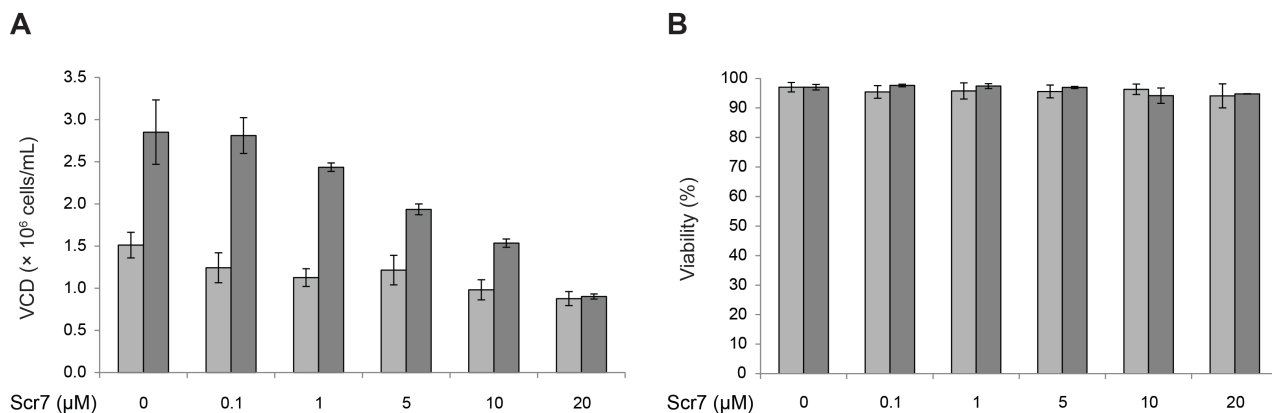
Jae Seong Lee,[1] Lise Marie Grav,[1] Lasse Ebdrup Pedersen,[1] Gyun Min Lee,[1,2] Helene Faustrup Kildegaard[1,*]

[1]The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, 2970 Hørsholm, Denmark

[2]Department of Biological Sciences, KAIST, 291 Daehak-ro, Yuseong-gu, Daejeon 305-701, Republic of Korea

1

FIGURES

A

B

Supplementary Figure S1. Profiles of (A) cell growth (viable cell density, VCD) and (B) viability in the presence of various concentrations of Scr7. CHO cells were inoculated at 6×10^5 cells/mL into 12-well plates containing 1 mL culture medium with or without Scr7. Adjacent light and dark grey bars represent the results on day 1 and day 2 respectively with mean ± standard deviations of at least three independent experiments.

2
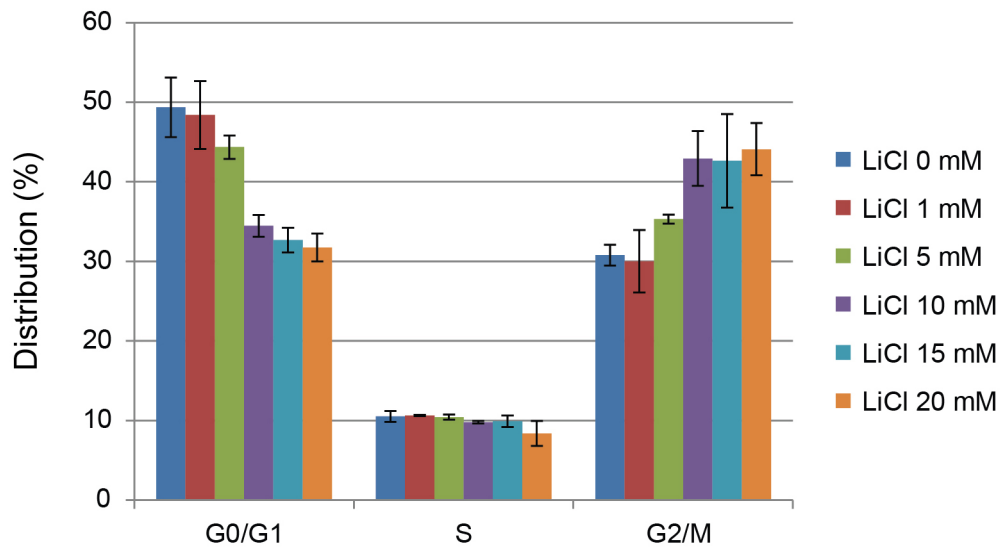
Supplementary Figure S2. Changes in cell cycle distribution in the presence of various concentrations of LiCl. CHO cells were treated with LiCl for 24 h, followed by cell cycle analysis to reveal arrest at G2/M phase. The error bars represent the standard deviations calculated from two independent experiments.

3
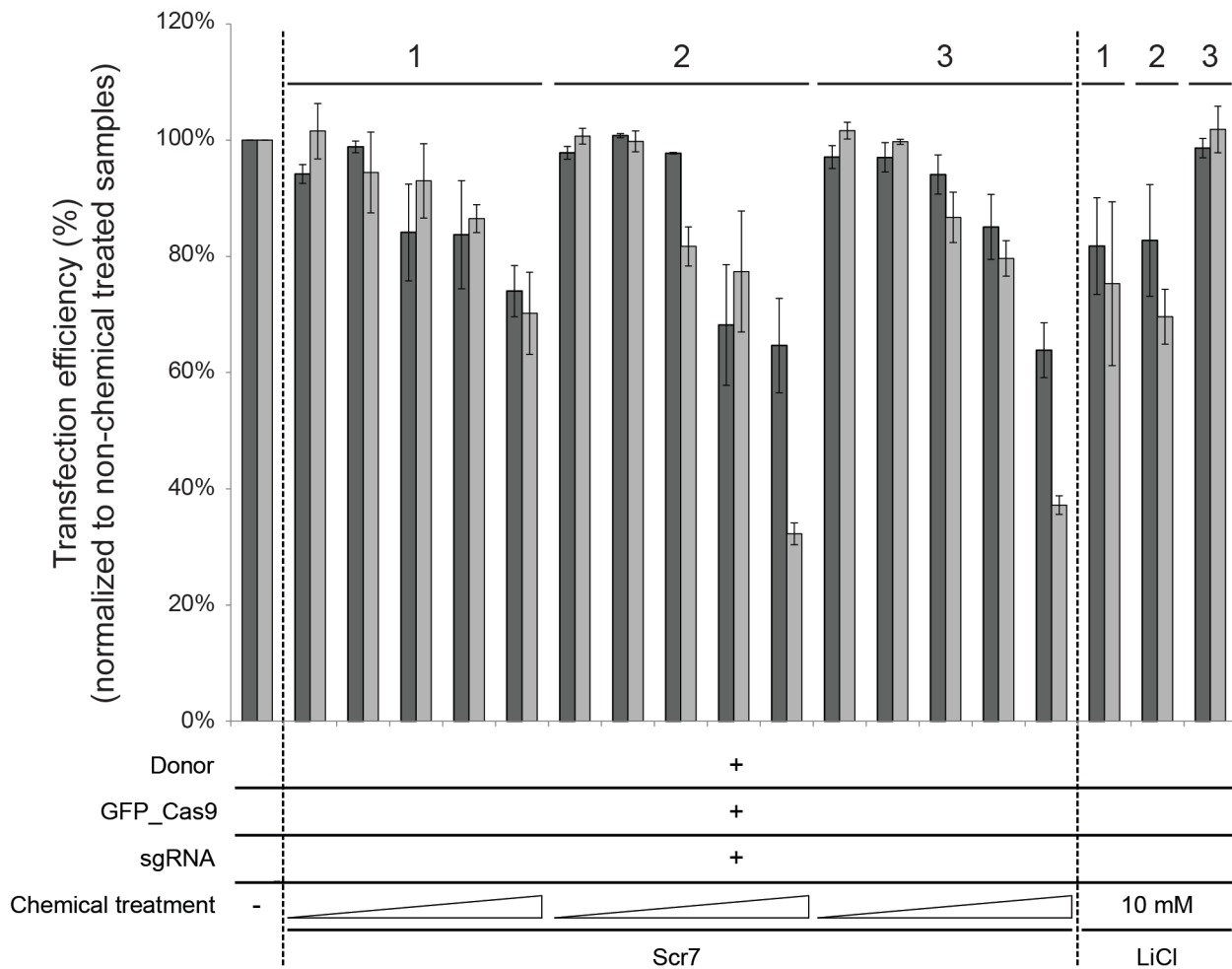
Supplementary Figure S3. The effect of chemical treatment on transfection efficiency. CHO cells were treated with Scr7 or LiCl under three different conditions - 1) immediately after transfection 2) both before and after transfection (3) before transfection only (See Methods) - upon transfection with donor plasmid, expression vectors encoding GFP_Cas9 and sgRNA. Transfection efficiency was analyzed on a basis of mCherry fluorescent populations 48 hr post-transfection. The data are expressed as % transfection efficiency of non-chemical treated samples (set at 100%). Adjacent dark and light grey bars represent the results of two target sites, COSMC and FUT8 respectively with mean ± standard deviations of two independent experiments.

4

TABLES

Supplementary Table S1. Plasmids used in this study

Plasmid name	Description	Reference
GFP_Cas9	GFP 2A peptide-linked Cas9 expression vector	Grav et al., 2015
COSMC sgRNA2	sgRNA expression vector targeting COSMC locus [Target genomic sequence:GAATATGTGAGTGTGGATGGAGG]	Lee et al., 2015b
FUT8 sgRNA3	sgRNA expression vector targeting FUT8 locus [Target genomic sequence: GTCAGACGCACTGACAAAGTGGG]	Grav et al., 2015
COSMC-BamHI-HDR-TI donor	Donor plasmid targeting COSMC [Target GOI: BamHI]	This study
FUT8-BamHI-HDR-TI vector	Donor plasmid targeting FUT8 [Target GOI: BamHI]	This study
COSMC-coEPO-HDR-TI donor	Donor plasmid targeting COSMC [Target GOI: EPO]	This study
COSMC-Rituximab-HDR-TI donor	Donor plasmid targeting COSMC [Target GOI: Rituximab]	This study

Supplementary Table S2. Primer sequences

Primer name	Purpose	Sequence (5'-3')
Donor plasmid		
COSMC 5' arm_fwd (Lee et al., 2015b)	USER PCR primer for COSMC donor plasmid (Homology arm)	AGTCGGTGUGTAATCCATGGAGGAGTTTCT
COSMC 5' arm_rev (Lee et al., 2015b)	USER PCR primer for COSMC-model proteins donor plasmid (Homology arm)	ACGCTGCTUAAGGTCTCCAGATTTTACAGT
COSMC 5' arm_BamHI_rev	USER PCR primer for COSMC-BamHI donor plasmid (Homology arm with BamHI)	ACGCTGCTUGGATCCAAGGTCTCCAGATTTTACAGT
COSMC 3' arm_fwd (Lee et al., 2015b)	USER PCR primer for COSMC-model proteins donor plasmid	AGGTCTGAGUGATTGTCTTAAGCATAGAGTC

5

	(Homology arm)	
COSMC 3' arm_fwd2	USER PCR primer for COSMC-BamHI donor plasmid (Homology arm)	AAGCAGCGUGATTGTCTTAAGCATAGAGTC
COSMC 3' arm_rev (Lee et al., 2015b)	USER PCR primer for COSMC donor plasmid (Homology arm)	AGCGACGUCCTCATTTGCATATATTTGAA
FUT8 5' arm_fwd	USER PCR primer for FUT8-BamHI donor plasmid (Homology arm)	AGTCGGTGUACCCCAGTGTGCTACAGCCCT
FUT8 5' arm_BamHI_rev	USER PCR primer for FUT8-BamHI donor plasmid (Homology arm with BamHI)	ACGCTGCTUGGATCCATGGACTCTGGGGAGAAGTGG
FUT8 3' arm_fwd	USER PCR primer for FUT8-BamHI donor plasmid (Homology arm)	AAGCAGCGUAACAGAAGCAGCCTTCCATCCC
FUT8 3' arm_rev	USER PCR primer for FUT8-BamHI donor plasmid (Homology arm)	AGCGACGUCTCCCATCTCACTCCCCCATC
EF-1α_fwd (Lee et al., 2015b)	USER PCR primer for model proteins donor plasmid (EF-1α)	AAGCAGCGUGTGAGGCTCCGGTGCCC
EF-1α _fwd2	USER PCR primer for Rituximab donor plasmid (Rituximab LC)	AGTGCGAUGTGAGGCTCCGGTGCCC
EF-1α _rev (Lee et al., 2015b)	USER PCR primer for model proteins donor plasmid (EF-1α)	ATGACGTCUTCACGACACCTGAAATGGAA
kozak_co-EPO_fwd	USER PCR primer for EPO donor plasmid (EPO-BGH pA)	AGACGTCAUCGCCACCATGGGAGTGCACG
kozak_Rituximab HC_fwd	USER PCR primer for donor plasmid (Rituximab HC-BGH pA)	AGACGTCAUGACACCATGGGCTGGTCCTG
BGH pA_rev	USER PCR primer for donor plasmid (model protein-BGH pA)	ACTCAGACCUCCATAGAGCCCACCGCATCC
BGH pA_rev3	USER PCR primer for donor plasmid (Rituximab-BGH pA)	ATCGCACUCCATAGAGCCCACCGCATCC
CMV_fwd (Lee et al., 2015b)	USER PCR primer for donor plasmid (mCherry)	ACGTCGCUGTTGACATTGATTATTGACT

6

BGH pA_rev2 (Lee et al., 2015b)	USER PCR primer for donor plasmid (mCherry)	ACGCAAGUCCATAGAGCCCACCGCATCC
pJ204 backbone_fwd (Lee et al., 2015b)	USER PCR primer for donor plasmid (Backbone)	ACTTGCGUAGTGAGTCGAATAAGGGCGACACA AA
pJ204 backbone_rev (Lee et al., 2015b)	USER PCR primer for donor plasmid (Backbone)	ACACCGACUGAGTCGAATAAGGGCGACACCCC A

5'/3' Junction PCR and Out-Out PCR

COSMC genomic fwd [junction] (Lee et al., 2015b)	COSMC amplicon for 5' junction and out-out PCR	TGGTTTCTAGGCTAATGCTTTGA
COSMC genomic rev [junction] (Lee et al., 2015b)	COSMC amplicon for 3' junction and out-out PCR	CCTGCCCCCACAGAAAAGTA
FUT8 genomic fwd [junction]	FUT8 amplicon for 5' junction and out-out PCR	ACACACGGAACAAGTCCAAT
FUT8 genomic rev [junction]	FUT8 amplicon for 3' junction PCR	AAGGAGACAGAAGTGCACAAGT
EF-1α rev [junction] (Lee et al., 2015b)	Amplicon for 5' junction PCR	ATCCTGGCCCGCATTTACAA
BGH fwd [junction]	Amplicon for 3' junction PCR	CGACTGTGCCTTCTAGTT

Deep sequencing

COSMC sgRNA2_F_Nex (Lee et al., 2015b)	COSMC amplicon for MiSeq analysis (sgRNA2 site)	TCGTCGGCAGCGTCAGATGTGTATAAGAGACA GTCCCACCTTGTTCAGGACACT
COSMC sgRNA2_R_Nex (Lee et al., 2015b)	COSMC amplicon for MiSeq analysis (sgRNA2 site)	GTCTCGTGGGCTCGGAGATGTGTATAAGAGAC AGGGATCCATCGCAGCCTTTCTAT
FUT8 sgRNA3_F_Nex (Grav et al., 2015)	FUT8 amplicon for MiSeq analysis (sgRNA3 site)	TCGTCGGCAGCGTCAGATGTGTATAAGAGACA GTGCCCCCATGACTAGGGATA
FUT8 sgRNA3_R_Nex (Grav et al., 2015)	FUT8 amplicon for MiSeq analysis (sgRNA3 site)	GTCTCGTGGGCTCGGAGATGTGTATAAGAGAC AGTCTGCGTTCGAGAAGCTGAAA

7

	qRT-PCR	
EF-1α _1F	EF-1α amplicon for qRT-PCR [Amplicon size: 127 bp]	GGAGAACCGTATATAAGTGCAGTAG
EF-1α _1R	EF-1α amplicon for qRT-PCR [Amplicon size: 127 bp]	AAGGGCCATAACCCGTAAAG
BGHpA_2F	BGHpA amplicon for qRT-PCR [Amplicon size: 103 bp]	TGCCAGCCATCTGTTGTT
BGHpA_2R	COSMC amplicon for qRT-PCR [Amplicon size: 103 bp]	AATGCGATGCAATTTCCTCATT
Vinculin_1F (Lee et al., 2015b)	Vinculin amplicon for qRT-PCR [Amplicon size: 73 bp]	GCTGGTTGCTAAGAGGGAGG
Vinculin_1R (Lee et al., 2015b)	Vinculin amplicon for qRT-PCR [Amplicon size: 73 bp]	ATCAGAGGCAGCTTTCACGG

Supplementary Table S3. Targeted integration efficiency

Target GOI	Size (kb)	Total clones[a]	5'/3' Junction PCR positive	Out/Out PCR positive[b]	Targeting efficiency[c] (%)
EPO	2.1	81	1	1	1.2
Rituximab	5.1	76	1	1	1.3

[a] The number of total single cell sorted clones without mCherry and GFP fluorescence.

[b] Positive indicates the number of clones with expected size of PCR products generated by targeted integration.

[c] Targeting efficiency indicates the percentage of out/out PCR positive clones among total clones investigated.

8

Supplementary Material

Minimized clonal variation for improved systems biology studies

Lise Marie Grav[1], Jae Seong Lee[2], Daria Sergeeva[1], Mikael Rørdam Andersen[3], Lars Keld Nielsen[1,4], Gyun Min Lee[1,5], 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 Biotechnology and Biomedicine, Technical University of Denmark, Kgs. Lyngby, Denmark
[4] Australian Institute for Bioengineering and Nanotechnology, The University of Queensland, Brisbane, Australia
[5] Department of Biological Sciences, KAIST, Daejeon, Republic of Korea

Figures

Supplementary Fig. 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)
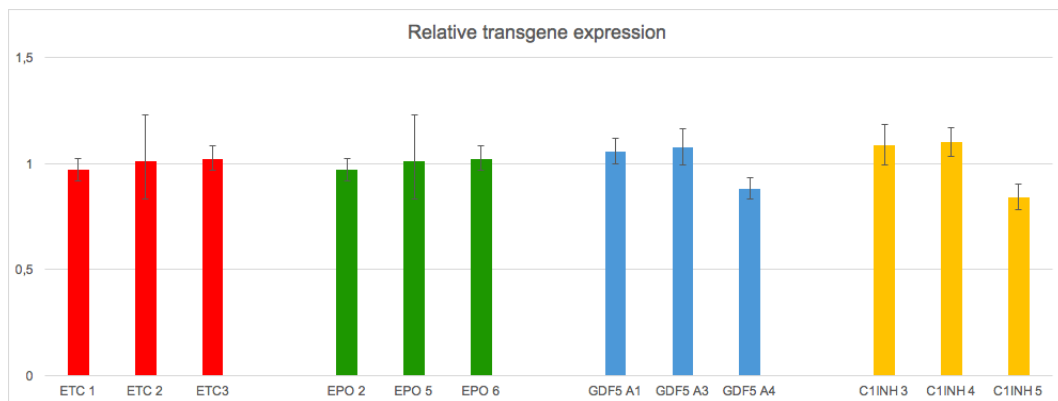
Supplementary Fig. S2 Relative levels of transgene expression in isoCHO-EP-derived subclones, measured by qRT-PCR, normalized to average value of corresponding isoCHO-EP subclones. The error bars represent the standard deviations of technical replicates (n=3).
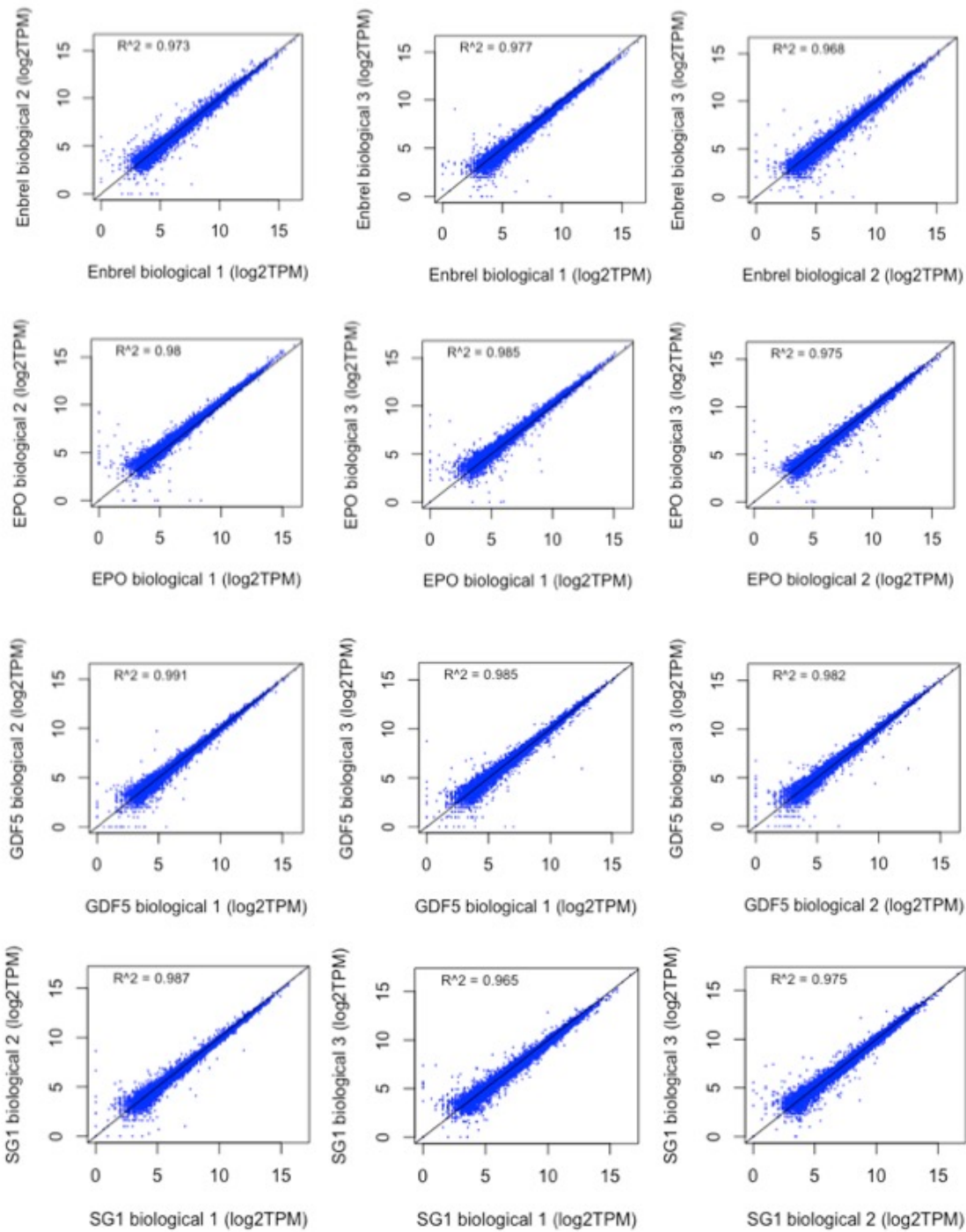
Supplementary Fig. S3 Comparison of expression values for a pair of replicates in dataset 1
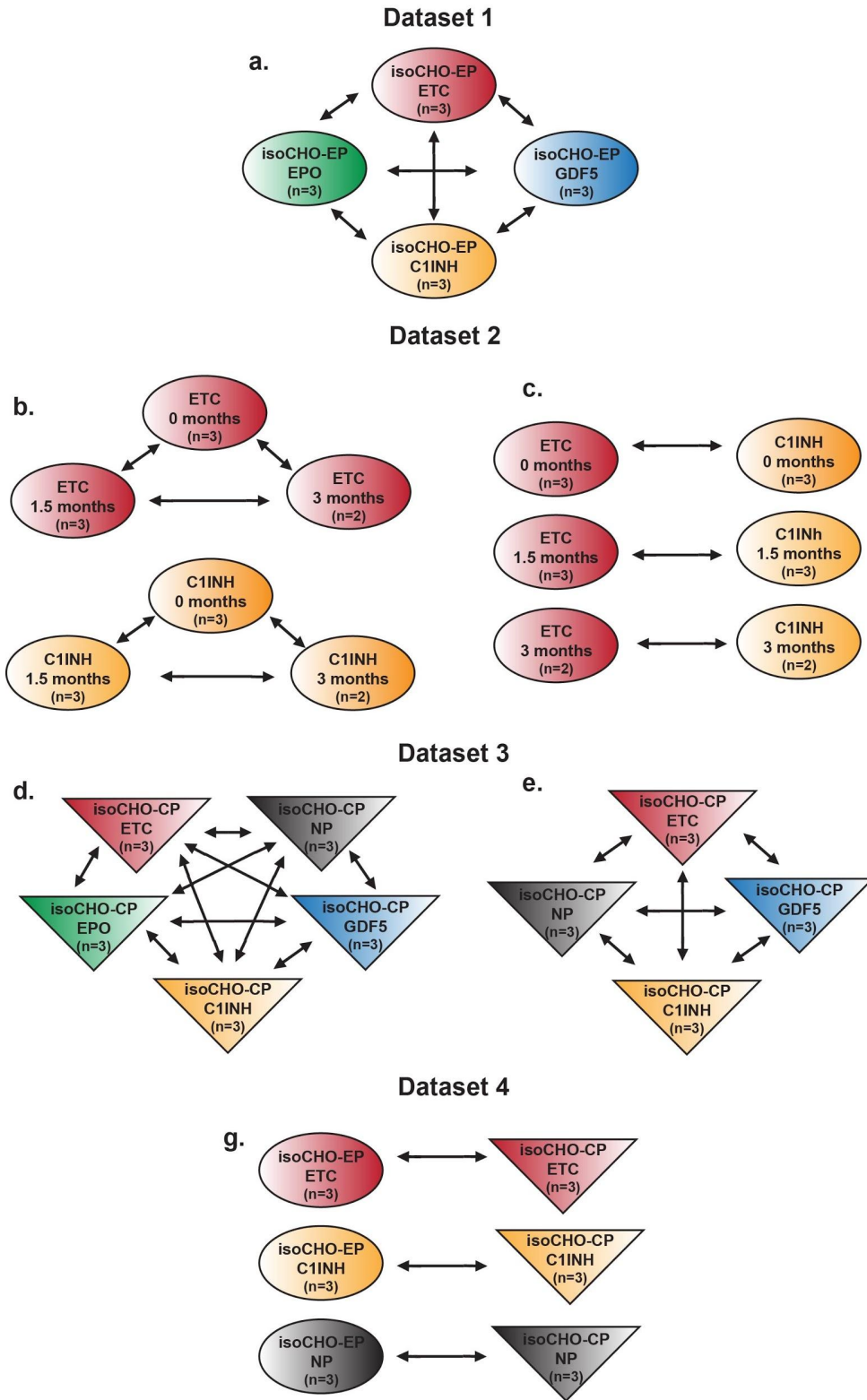
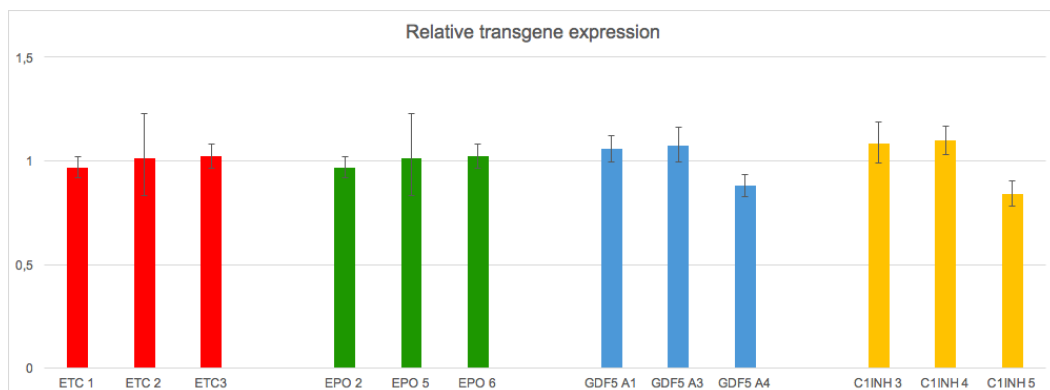**Supplementary Fig. S4 Overview of differential expression analysis datasets**
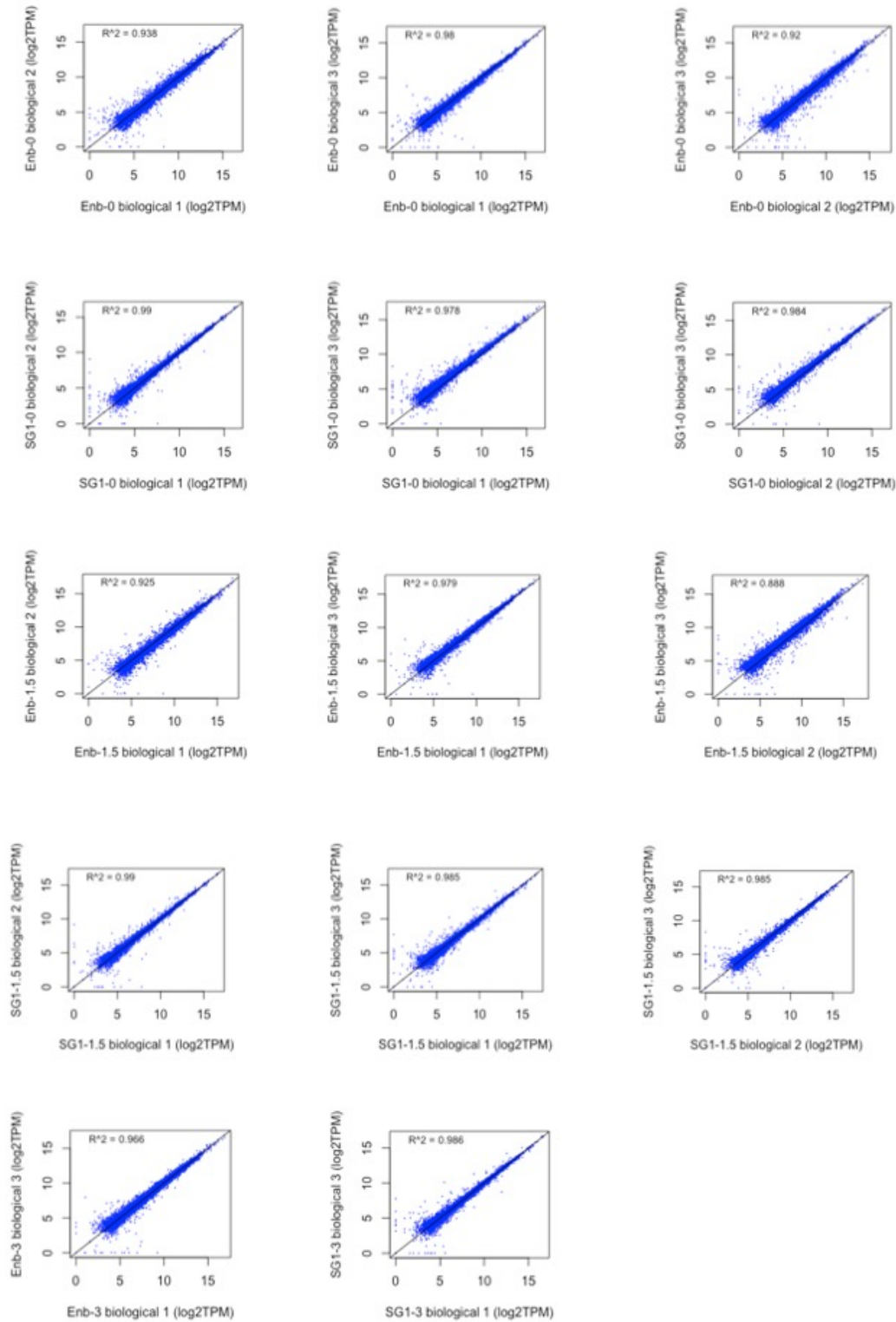
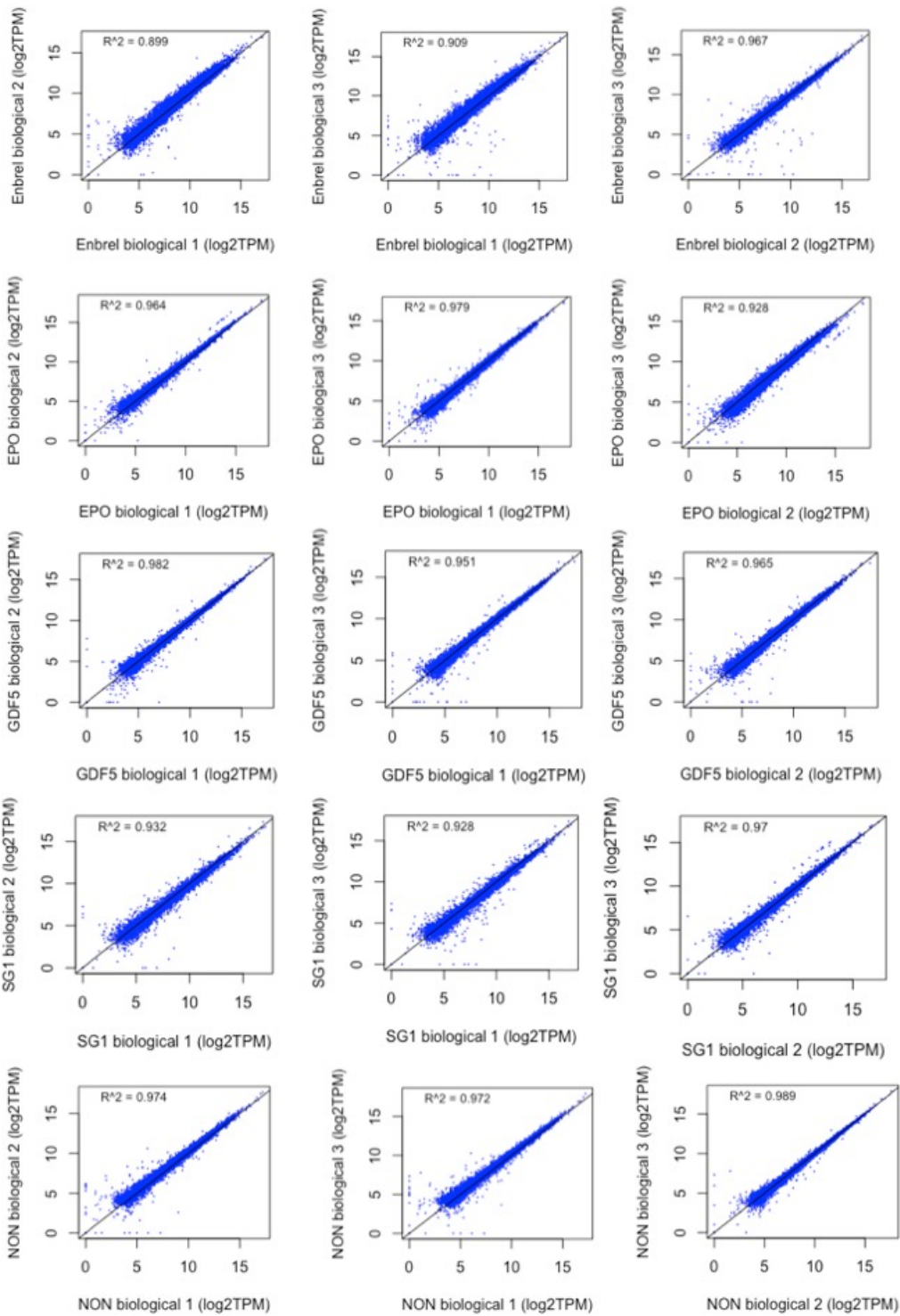**Supplementary Fig. S5** Comparison of expression values for a pair of replicates in dataset 2
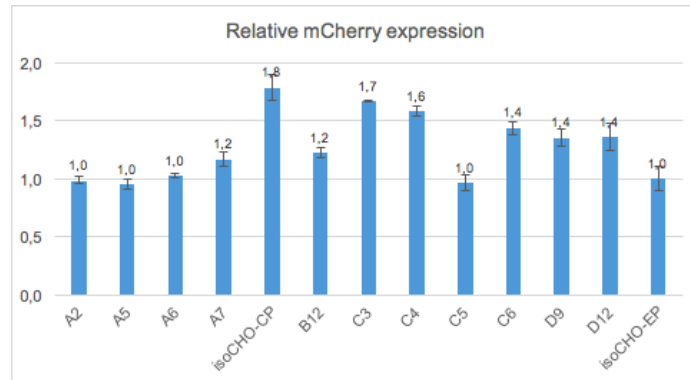
Supplementary Fig. S6. Relative mCherry expression levels of the panel of mCherry-CP clones, normalized to isoCHO-EP**.** Error bars represent the standard deviations of technical replicates (n≥3)
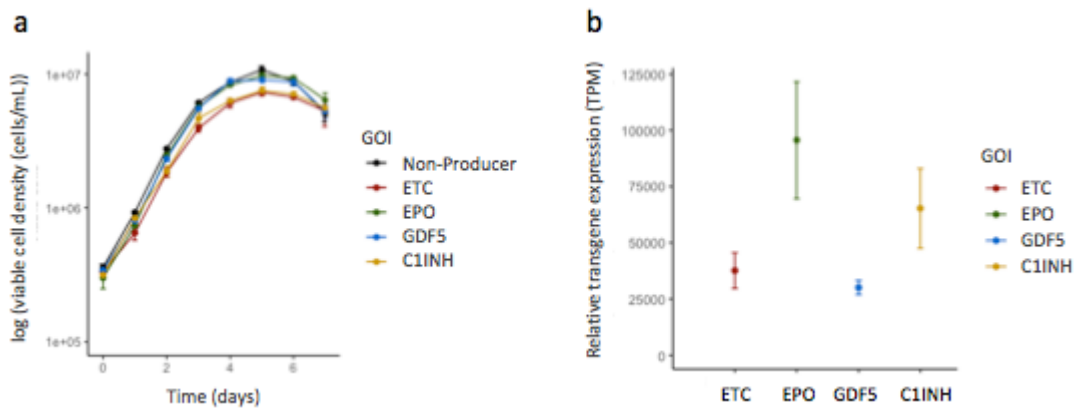

Supplementary Fig. S7. Phenotypes of isoCHO-CP subclones. (a) Viable cell densities of isoCHO-CP subclones expressing ETC, EPO, GDF5 or C1INH. The error bars of each line represent the standard deviations of three isogenic subclones expressing the same GOI (n=3). **(b)** Relative levels of transgene expression, as measured in transcripts per kilobase million (TPM) of ETC, EPO, GDF5 or C1INH. The error bars represent the standard deviations of three isogenic clones expressing the same GOI (n=3).
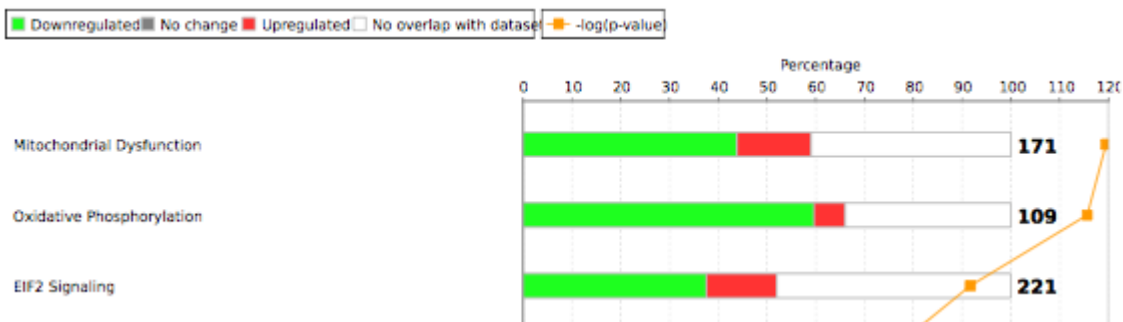

Supplementary Fig. S8 Relative levels of transgene expression in isoCHO-CP-derived subclones, measured by qRT-PCR, normalized to average value of corresponding isoCHO-CP subclones. The error bars represent the standard deviations of technical replicates (n=3).
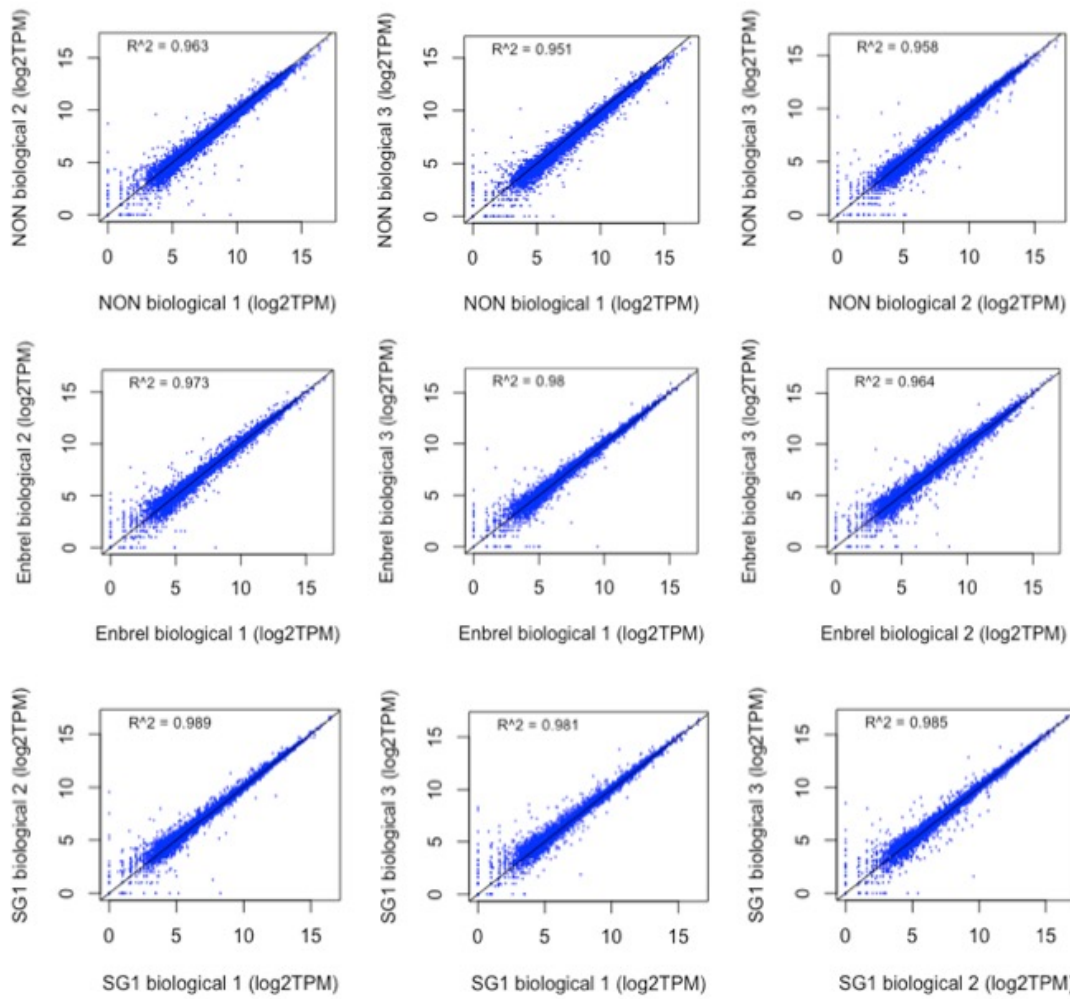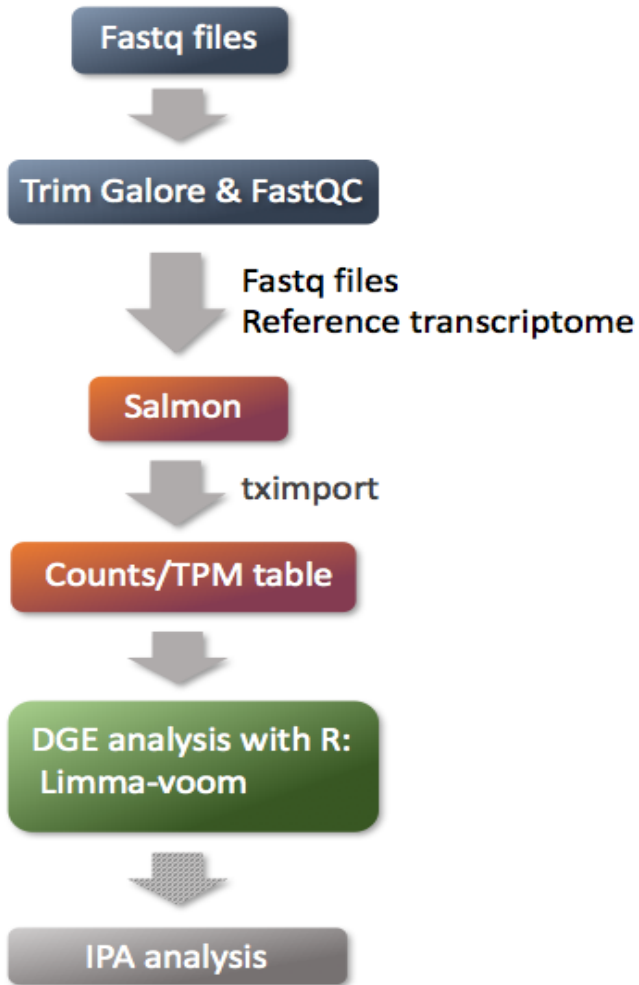
Supplementary Fig. S9 Comparison of expression values for a pair of replicates in dataset 3 and 4

Supplementary Fig. S10 Canonical pathways enrichment analysis

Supplementary Fig.S11 Overview of transcriptomics pipeline

Tables

Supplementary 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

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

Clone	CHO-S WT	A2	A5	A6	A7	isoCHO-CP	B12	C3	C4	C6	D9	D12	C5
mCherry positive population	0,2%	97,6%	92,0%	97,2%	87,4%	93,2%	77,1%	84,3%	80,9%	83,6%	85,6%	94,4%	93,4%
mCherry intensity	6	139	165	160	148	205	127	180	171	176	169	136	188

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

```
ATGAGACTGCCCAAGCTGCTGACCTTCCTGCTGTGGTATCTGGCCTGGCTGGACCTGGAATTCATCTGCACCGTGCTGGGCGCTCCCGATCTGGGACAGAGGCCT
CAGGGAACCAGACCCGGACTGGCTAAGGCCGAGGCCAAAGAGAGGCCTCCCCTGGCCAGAAACGTGTTCAGACCTGGCGGCCACTCTTACGGCGGAGGCGCGCCA
CCAATGCCAACGCCAGAGCTAAGGGCGGCACCGGACAGACAGGTGGCCTGACCCAGCCTAAGAAGGACGAGCCCAAGAAGCTGCCTCCTAGACCAGGCGGCCCC
TGAGCCTAAGCCTGGACATCCTCCACAGACCAGACAGGCCACCGCCAGAACCGTGACCCCTAAGGGACAGCTGCCTGGCGGAAAGGCCCCTCCTAAGGCTGGCT
CTGTGCCCTCCAGCTTTCTGCTGAAGAAGGCCAGAGAGCCTGGCCCCCCTAGAGAGCCCAAAGAGCCCTTCAGACCCCCCCCTATCACCCCCCCACGAGTACATGCT
GTCCCTGTACCGGACCCTGTCTGACGCCGATCGGAAGGGCGGAAACTCCTCCGTGAAGCTGGAAGCCGGCCTGGCCAACACCATCACCAGCTTCATCGACAAGG
GCCAGGACGACAGGGGTCCCGTCGTGCGGAAGCAGAGATACGTGTTCGACATCTCCGCCCTGGAAAAGGACACGGCCTGCTGGGAGCCGAGCTGCGGATCCTGAG
AAAGAAGCCTTCCGACACCGCCAAGCCTGCTGCTCCTGGCGGAGGTAGAGCTGCCCAGCTGAAGCTGTCCAGCTGCCCTTCTGGCAGACAGCCTGCCGCTCTGCT
GGATGTGCGATCTGTGCCAGGACTGGACAGGCTCCGGATGGGAGGTGTTCGATATCTGGAAGCTGTTCCGCAACTTCAAGAACTCCGCCCAGCTGTGCCTGGAAC
TGGAAGCTTGGGAGAGGGGCAGAGCCGTGGATCTGAGAGGCCTGGGCTTCGACAGAGCCGCTAGACAGGTGCACGAGAAGGCCCTGTTTCTGGTGTTCGGCC
GGACCAAGAAGCGGGACCTGTTCTTCAACGAGATCAAGGCCAGATCCGGCCAGGATGACAAGACCGTGTACGAGTACCTGTTCTCCCAGCGGCGGAAGCGGAG
AGCCCCTCTGGCTACAAGACAGGGCAAGCGGCCCTCCAAGAACCTGAAGGCCCGGTGCTCTAGAAAGGCCCTGCACGTGAACTTCAAGGACATGGGCTGGGAC
GACTGGATCATTGCCCCCCCTGGAATACGAGGCCTTCCACTGCGAGGGCCTGTGCGAGTTCCCTCTGAGATCCCACCTGGAACCCACCAACCACGCCGTGATCCAG
ACCCTGATGAACTCCATGGACCCCGAGTCCACCCCCCCTACCTGTTGTGTGCCTACCCGGCTGTCCCCCATCTCCATCCTGTTCATCGACTCCGCCAACAACGTGGT
GTACAAGCAGTACGAGGACATGGTGGTGGAATCCTGCGGCTGCCGGTGA
```

Supplementary Table S4. Primers for plasmids construction and insert PCR

lox-mCherryOri vector construction (USER primers)

Primer name	Sequence
LoxP-kozak_mcherry_LA_fwd	agtcggtgUATAACTTCGTATAGCATACATTATACGAAGTTATCGCCACCATGGTGAGCA
Lox2272-mcherry_O4_rev	AGACTGTGUataacttcgtataaagtatcctatacgaagttatCTACTTGTACAGCTCGT
BGH pA_O4_fwd	ACACAGTCUCTGTGCCTTCTAGTTGCC
BGH pA_O5_rev	ACGCAAGUCCATAGAGCCCACCGCAT

Landing pad vectors construction (USER primers)

Primer name	Sequence
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

RMCE donor vectors construction (USER primers)

Primer name	Sequence
LoxP-kozak_EPO_LA_fwd	agtcggtgUATAACTTCGTATAGCATACATTATACGAAGTTATCGCCACCATGGGAGTGC
Lox2272-EPO_O5_rev	ACGCAAGUataacttcgtataaagtatcctatacgaagttatTCATCTATCGCCGGTCC
LoxP-kozak_ETC_LA_fwd	agtcggtgUATAACTTCGTATAGCATACATTATACGAAGTTATAGCACCATGGCGCCCGT
Lox2272-ETC_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

Insert PCR

Primer name	Sequence
EF1a-mcherry junction fwd	CCTCAGACAGTGGTTCAAAGT
BGH pA rev	AGATGGCTGGCAACTAGAAG

Supplementary Table S5. qRT-PCR primers and probes

qRT-PCR (SYBR Green assay)

Gene	Fwd primer	Rev primer
mCherry	AGGACGGCGAGTTCATCTA	CCCATGGTCTTCTTCTGCATTA
GAPDH	TTGTCATCAACGGGAAGG	GTGAAGACGCCAGTAGATT

TaqMan assays

Gene	Fwd primer	Rev primer	Probe	Dye
ETC	CAGCCGGAGAACAACTACAA	CATCACGGAGCATGAGAAGA	TACAGCAAGCTCACCGTGGACAAG	FAM-MGB
EPO	CTGGAAAGATACCTGCTGGAAG	AGGCGTAGAAGTTCACTTTGG	CCAAAGAGGCCGAGAACATCACCA	FAM-MGB
GDF5	GTGATCCAGACCCTGATGAAC	GTCGATGAACAGGATGGAGATG	TACCTGTTGTGTGCCTACCCGG	FAM-MGB
C1INH	GGATGGAGCCCTTTCACTTTA	GGATGACCAGGCTCAGATTATG	TCATCGACCAGACCCTGAAGGCTA	FAM-MGB

Supplementary Table S6. RNA-seq datasets

Dataset	Samples	Sequencing kit
1	**isoCHO-EP**: ETC (n=3), EPO (n=3), GDF5 (n=3), C1INH (n=3)	mid-output
2	**isoCHO-EP**: ETC-0 months (n=3), ETC-1.5 months (n=3), ETC-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**: ETC (n=3), EPO (n=3), GDF5 (n=3), C1INH (n=3), non-producer (n=3)	high-output
4	**isoCHO-EP**: ETC (n=3), C1INH (n=3), non-producer (n=3) **isoCHO-CP**: ETC (n=3), C1INH (n=3), non-producer (n=3)	high-output