Developing Tools for Improved and Predictable Recombinant Protein Production in CHO Cells

Pristovšek, Nuša

Publication date:
2018

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

Developing Tools for Improved and Predictable Recombinant Protein Production in CHO Cells

PhD thesis

Nuša Pristovšek

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

August 2018

Supervisor: **Mikael Rørdam Andersen**, Professor MSO, Department of Biotechnology and Biomedicine, Technical University of Denmark

Co-supervisors: **Helene Faustrup Kildegaard**, Senior Scientist, Novo Nordisk and **Henning Gram Hansen**, Staff Scientist, BRIC, University of Copenhagen

Cover Page Design: Kristina Bastelj

Preface

This thesis is written as a partial fulfillment of the requirements to obtain a PhD degree at The Technical University of Denmark. The thesis was carried out at The Novo Nordisk Foundation Center for Biosustainability at The Technical University of Denmark, in the research group of (at the time) Senior Researcher and co-PI Helene Faustrup Kildegaard. The thesis was additionally supervised by Professor MSO Mikael Rørdam Andersen and (at the time) Researcher Henning Gram Hansen. The project was initiated on 1st of September 2015 and was funded by The Novo Nordisk Foundation and the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 642663. During the period from October 2017 to January 2018 an External Research Stay was conducted at the Austrian Centre of Industrial Biotechnology in Vienna under the supervision of Professor Nicole Borth.

Copenhagen, 31st of August, 2018

Nuša Pristovšek

Acknowledgements

What an interesting journey it has been…

First and foremost a big thank you to my three supervisors for putting up with me for these three years. To Helene - thank you for giving me an opportunity to be a part of the group, for your excellent guidance, for always believing in me and for setting a great example as a successful woman scientist. To Mikael - thank you for your constant encouragements, for your exceptional force of intellect and generosity of spirit (and perhaps most importantly, for liking my puns). To Henning - without your sharp eye, admirable scientific integrity and attention to details I would have never become a researcher I am today - thank you for shaping my 'scientific self'.
Thank you, Gyun Min, for always providing interest and valuable scientific input for my work. Nicole Borth, thank you for your support and for hosting me in your research group. Thank you, Nathan and Hooman, for great overseas collaborations and Peter Rugbjerg for a fruitful across-the-floor collaboration.

To DTU eCHOs (Ankita, Sara and Thomas) - thank you for incredibly fun three years (in and outside the lab), for all the useful tips and discussions, for inspiring me and making me laugh whenever the times were hard. You are experts in both science and foosball. I will miss you!

A warm thanks also to all the other 11 eCHO students from the consortium - what a fun and talented group of young researchers. I hope our paths meet again soon. Anne Tolstrup, thank you for taking the time to discuss my research and always giving me valuable advice.

To all the present and ex CLED group members (Lise, Julie, Daniel, Jae, Daria, Saranya, Nachon, Thomas K., Johan, TK, Manuel, Kai) - thank you for showing me the perks of being a researcher. You are all incredible colleagues with amazing minds - keep up the great work!
To all the CHO Core members - thank you for your kind support and excellent technical guidance throughout my PhD.
Alicia, thank you for always creating a nice atmosphere at CfB and most importantly outside of it and for keeping me sane (PhD ≠ life)!

Toño - muchisimas gracias for always lending me your ear, for your never-ending support, belief and love you have given me. Many times you have helped me gather the strength I didn't know I had within me.

In nenazadnje, hvala moji ljubeči družini, brez katere mi vse to ne bi nikdar uspelo. Vaša podpora in nenehna skrb je moj varen pristan, za kar sem vam neizmerno hvaležna.
Na svetu si, da gledaš sonce.
Na svetu si, da greš za soncem.
Na svetu si, da sam si sonce,
in da s sveta odganjaš ~ sence.
'Tone Pavček'

Abstract

Chinese hamster ovary (CHO) cells are the dominant cell factory for recombinant production of diverse biotherapeutics. The complexity of biotherapeutics and the increasing demand for their rapid, cost-effective and high-quality production have made the biomanufacturing in CHO cells a highly sophisticated process. To meet the demands for better and inexpensive biotherapeutics, further improvements of recombinant protein production in CHO cells seem necessary. In this thesis we present clone screening and gene engineering tools that aim to facilitate improved and more predictable recombinant protein production in the next-generation CHO cell factories.

First, a reporter-based screening method for high production capacity was established, that is able to rank statically-grown clones according to either volumetric productivity (overall yield) or a proxy of specific productivity (per cell yield). By systematically comparing the two clone screening criteria, we demonstrated that they perform similarly in terms of isolating CHO clones with high volumetric productivity in suspension.

Second, a toolbox combining the CRISPR/Cas9 system with a recombinase-based system was developed to generate cell lines with so-called landing pads integrated in defined genomic sites. The toolbox allows targeted gene integration, followed by streamlined recombination of the expression cassette into the landing pad of a genetically identical cell line. This approach is exploited for systematic evaluation of site-specific gene expression modulated by different expression cassette components. We demonstrated that distinct and robust recombinant gene expression can be achieved in defined integration sites in CHO cells.

Third, ectopic expression of target genes for improving the secretory capacity of CHO cells was reviewed and studied. We identified three novel target genes increasing specific productivity of CHO cells, transiently expressing different monoclonal antibodies. We attempted to unveil the mechanism of novel target genes by examining the rate-limiting step in production of these monoclonal antibodies.

Overall, the tools developed in this thesis can in the future enable more rational engineering of CHO cell factories for production of a variety of complex biotherapeutics.

Dansk sammenfatning

Kinesisk hamster ovarie (CHO) celler er den mest dominerende cellefabrik til produktion af forskelligartede rekombinante terapeutiske proteiner. Kompleksiteten af de terapeutiske proteiner, og den stigende efterspørgsel efter CHO cellers hurtige og omkostningseffektive produktion af produkter i høj kvalitet, har gjort fremstillingen til en yderst sofistikeret proces. For at imødekomme kravene til bedre og billigere terapeutiske proteiner er det nødvendigt at forbedre produktionen af rekombinante proteiner i CHO celler yderligere. I denne afhandling præsenterer vi klon screening og genmanipulationsværktøjer, der sigter mod at forbedre og lette forudsigeligheden af produktionen af rekombinante proteiner i næste generations CHO celle fabrikker.

For det første blev der etableret en reporterbaseret screeningsmetode for høj produktionskapacitet, der er i stand til at rangere statisk dyrkede kloner i henhold til enten den volumetriske produktivitet (det samlede udbytte) eller en proxy for specifik produktivitet (udbytte per celle). Ved systematisk sammenligning af de to screeningskriterier viste vi, at de har tilsvarende effektivitet med hensyn til isolering af CHO kloner med høj volumetrisk produktivitet i suspension.

For det andet blev en værktøjskasse, hvori CRISPR/Cas9-systemet blev kombineret med et rekombinase-baseret system, udviklet til at generere cellelinjer med såkaldte landingspuder integreret i definerede genomiske positioner. Værktøjskassen tillader målrettet genintegration, efterfulgt af rekombination af en udtrykkelseskassette i landingspuden i en genetisk identisk cellelinie. Denne fremgangsmåde udnyttes til systematisk evaluering af stedsspecifik genudtrykkelse moduleret af forskellige udtrykkelseskassettekomponenter. Vi demonstrerede, at robust rekombinant genudtrykkelse kan opnås i definerede integrationssteder i CHO celler.

For det tredje undersøgte vi ektopisk udtrykkelse af specifikke gener med det formål at forbedre CHO cellernes sekretoriske kapacitet. Vi identificerede tre nye gener, der øger den specifikke produktivitet i CHO celler, der transient udtrykker forskellige monoklonale antistoffer. Vi forsøgte at afsløre mekanismen for de nye gener ved at undersøge det hastighedsbegrænsende trin i produktionen af disse monoklonale antistoffer.

Samlet set kan værktøjerne der er blevet udviklet i denne afhandling fremover muliggøre mere rationel konstruktion af CHO cellefabrikker til produktion af en række komplekse terapeutiske proteiner.

List of publications

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

I. Using titer and titer normalized to confluence are complementary strategies for obtaining Chinese hamster ovary cell lines with high volumetric productivity of Etanercept

Nuša Pristovšek, Henning Gram Hansen, Daria Sergeeva, Nicole Borth, Gyun Min Lee, Mikael Rørdam Andersen, Helene Faustrup Kildegaard. (2018). *Biotechnology Journal*, 13, 1-10. doi: 10.1002/biot.201700216

II. Systematic evaluation of site-specific recombinant gene expression for programmable mammalian cell engineering

Nuša Pristovšek, Saranya Nallapareddy, Lise Marie Grav, Hooman Hefzi, Nathan E. Lewis, Peter Rugbjerg, Henning Gram Hansen, Gyun Min Lee, Mikael Rørdam Andersen, Helene Faustrup Kildegaard. (2018). (*Manuscript in submission*)

III. Improving the secretory capacity of Chinese hamster ovary cells by ectopic expression of effector genes: Lessons learned and future directions

Henning Gram Hansen, Nuša Pristovšek, Helene Faustrup Kildegaard, Gyun Min Lee. (2017). *Biotechnology Advances,* 35, 64–76. doi: 10.1016/j.biotechadv.2016.11.008.

IV. Transient co-expression of novel effector genes and different IgG1 molecules increases specific productivity of CHO-S cells

Nuša Pristovšek, Henning Gram Hansen, Anne Mathilde Lund, Johan Rojek, Claes Nymand Nilsson, Gyun Min Lee, Mikael Rørdam Andersen, Helene Faustrup Kildegaard. (2018). (*Manuscript in preparation)*

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

V. Multiplex genome editing eliminates the Warburg Effect without affecting oxidative metabolism or growth rate

Hooman Hefzi, Soo Min Noh, Iván M. Monge, Marianne Decker, Johnny Arnsdorf, Stefan Kol, Nuša Pristovšek, Anders Holmgaard Hansen, Sara Petersen Bjørn, Karen Kathrine Brøndum, Elham Maria Javidi, Kristian Lund Jensen, Thomas Kallehauge, Daniel Ley, Patrice Ménard, Helle Munck Petersen, Zulfiya Sukhova, Bjørn G. Voldborg, Lars Nielsen, Gyun Min Lee, Helene Faustrup Kildegaard, Nathan E. Lewis. (2018). *(Manuscript in revision)*

Table of contents

The aim of the thesis

The aim of the thesis was to develop tools addressing different biomanufacturing developments towards improved and predictable recombinant protein production in Chinese hamster ovary (CHO) cells. In order to produce biotherapeutics faster and cheaper, the thesis had the following two overall objectives:

1) To increase the titer of produced biotherapeutics via improvements in specific productivity (q_p) ; and

2) To increase the predictability and tunability of biotherapeutics production, enabling faster product development and speed-to-market.

The first objective has been addressed by developing a method, capable of screening for CHO clones with high q_p and by ectopically expressing target genes (throughout the thesis referred to as effector genes) improving the q_p of CHO cells. The second objective has been addressed by developing a toolbox that allows systematic evaluation of site-specific recombinant gene expression levels in CHO cells in order to achieve predictable and desirable recombinant gene expression.

The structure of the thesis

The thesis is divided into the Introduction, Rationales and Chapters 1-4 and Concluding remarks and future perspectives.

In the **Introduction**, the topics addressed by different thesis chapters are presented in more detail. First, cell line development is discussed briefly, focusing on selection methods for isolating high-producing clones. And second, cell line engineering approaches by gene introduction are discussed due to engineering strategies used in the scope of this thesis. Specific focus is given to: a) CRISPR/Cas9-mediated and recombinase-based targeted gene integration into predefined integration sites, and b) host cell line engineering via effector gene overexpression.

In **Chapter 1** (**Research Article I**), a platform for selecting high-producing clones was established, able to assess two different early reporters under static cultivation of CHO cells: a proxy of q_p and titer. Further, the performance and predictability of the two early reporters for high q_p and titer in suspension cultivation were systematically compared.

Next, **Chapter 2** (**Manuscript in submission II**) presents a CRISPR/Cas9-based toolbox, coupled with recombinase-based system that allows construction of CHO cell lines with a site-specifically integrated landing pad. By recombining different expression cassettes into the landing pad, a systematic evaluation of site-specific recombinant gene expression was performed.

Chapter 3 (**Review Article III**) is a review, discussing the engineering potential of ectopic expression of beneficial effector genes to improve the secretory capacity of CHO cells. Several experimental and cellular factors are reviewed with the objective of guiding the future studies on overexpression of effector genes.

Finally, **Chapter 4** (**Manuscript in preparation IV**) is a study that identifies three novel effector genes that improve the q_p of CHO cells when transiently producing different IgG1 molecules. The work builds on guidelines from Chapter 3 to further examine the conditions enabling the effector genes to increase the secretory capacity of CHO cells.

In the end, **Concluding remarks and future perspectives** provide a summary of the results achieved in this thesis, with the implications made in the field of recombinant protein production in CHO cells.

Introduction

1. Chinese hamster ovary (CHO) cells

Production of recombinant proteins for biotherapeutic use in humans is commonly performed using mammalian cells, such as CHO cells or Human embryonic kidney (HEK293) (1, 2). CHO cells are a well-established and most popular host for biotherapeutics production due to their ability to make human-like post-translational modifications and having a long history of regulatory approval and established bioindustrial pipelines (3). Further, CHO cells are non-permissive for most human viruses and are easily adapted to grow at high density in serum-free suspension conditions and chemically defined media (4). Despite being the mammalian host of choice with sufficient recombinant protein yields, CHO cells still exhibit significant bottlenecks in comparison to bacteria and yeast regarding growth capacity, cultivation time and product yield (5, 6). Even though some biotherapeutics are produced in bacteria due to above-mentioned advantages, the majority of top-selling biotherapeutics requires post-translational modifications (including monoclonal antibodies (IgGs) and erythropoietin), such as glycosylation. The post-translational modification profiles of biotherapeutics significantly influence solubility, bioactivity, stability and immunogenicity (4). Therefore, most biotherapeutics are produced in mammalian cells, accounting for over 70% of biotherapeutics sales revenue (7), with 70% of those products produced in CHO cells.

1.1. Biotherapeutics production

Establishing an upstream (cell line development and bioreactor production) and downstream bioprocesses (product isolation and purification) for a new biotherapeutic requires considerable investment in time, labor and costs. To enable the development of manufacturing pipelines with cost-effective and high-yielding production of biotherapeutics, a steady improvement of CHO cell factories in regard to growth, product yield and quality is a key challenge (2). Further, there is a significant pressure to shorten the development timelines for new biotherapeutics, therefore, predictability and flexibility of production processes are of high importance.

Due to the work conducted in this thesis, the focus of the introduction will be on the upstream process development, as a way to achieve improved and predictable recombinant protein production in CHO cells.

2. CHO cell line development

For industrially relevant production of recombinant proteins, stably producing cell lines have to be established during a process of cell line development (CLD). The CLD process generally consists of: a) transfection of a gene of interest (GOI) and random integration into the genome; b) an enrichment of cells expressing the GOI via antibiotic selection typically combined with selection of cells with high expression levels by gene amplification; c) single-cell cloning; d) screening for desirable, high-producing clones; and finally e) scale-up of selected clones (8). Since these processes are time-consuming and labor-intensive, there is a great demand for optimizing the CLD. Traditionally, improvements were made by empirically optimizing a variety of the extrinsic factors such as: a) expression vectors (8), b) selection and gene amplification processes (9–11), c) production media (12) and d) methods for screening clones with high productivity (13).

Optimization of the CLD and bioprocess has been the focus of the earlier studies, with innovative technologies developed, mostly improving the volumetric productivity (product titer) of recombinant proteins produced in CHO cells. Indeed, the CLD and bioprocess optimizations have greatly improved the recombinant protein titer in CHO cells by more than 100-fold over the past few decades (14), reaching yields well over 10 g/L (15). CHO CLD is currently still predominantly focused on isolating high-producing clones and ensuring monoclonality. Selection of high-producing clones from a heterogenous pool of transfected cells remains one of the most crucial steps in CLD. That is because random integration of GOI generates clones with different integration sites and varying number of gene copies inserted (16), and therefore heterogenous recombinant protein expression levels. However, with the recent emergence and continuously improving Chinese hamster and CHO genomic resources (17–21) and efficient genome editing technologies, targeted genetic manipulation (22), -omics studies (23–26), and modeling efforts (27) have become increasingly important to further characterize and improve the CHO cell factory. To further increase the production characteristics of CHO cells, appropriate solutions beyond the natural biology of CHO cells seem necessary - that is, solutions from rational CHO cell engineering-based approaches.

2.1 Development of selection methods for isolation of high-producing clones

The traditional and straightforward selection methods for isolating high-producing clones include limiting dilution cloning (28), selection via antibiotic resistance (29) and gene amplification (*i.e.* methotrexate/dihydrofolate reductase (MTX/DHFR) system and methionine sulphoximine/glutamine synthetase (MSX/GS) system) (11, 30). However, these

fundamental methods are being replaced or combined with more advanced and high-throughput techniques in the industrial setting. To save time, labor and resources, fluorescence-activated cell sorting (FACS) has become increasingly more utilized for high-throughput isolation of high-producing clones (31). One of the most widely-used methods utilizing FACS is a technique that traps secreted recombinant protein on the surface of the cell at low temperature and detects it with a specific, fluorescently labelled antibody (32). This surface labeling method enables enrichment of cells with high q_p and has been efficiently used in the industry. Increasingly more sophisticated automated selection systems have been developed to facilitate screening of hundreds to thousands of clones, reducing the required labor and significantly increasing the efficiency of selection. Fluorescence-based systems, such as ClonePix and LEAP (33–35), robotic systems, such as Cello and CellSpot (36, 37), and completely automated systems (38) have been developed, achieving remarkably high throughput. However, due to the high cost, the use of these automated systems is mostly limited to the industrial setting. Further, isolated high-producing clones have to be tested in large-volume bioreactors, relevant for the commercial production. It has been shown previously that high-producing clones at small growth scales are not necessarily the best clones for the industrial production (39, 40). Therefore, there is a high demand for clone selection methods that are predictive of clones' commercial performance.

3. CHO cell line engineering

To meet the ever-growing demand for higher quantity, quality and affordability of produced biotherapeutics, further development of more efficient production cell lines is required. Expansion of the bioindustrial pipeline with a wide range of non-IgGs, more difficult-to-express biotherapeutics further challenges both, the manufacturability and timelines to achieve a commercially viable solution. Therefore, the empirical optimization of extrinsic factors during CLD would benefit from the rational engineering of the host cell. By rationally designing the desired phenotype, rather than screening for it, the predictability and further improvement of recombinant protein production in CHO cells can be achieved.

Targeted engineering-based approaches for CHO cells have considerably improved biotherapeutic production and accelerated the manufacturing timelines, by optimizing a variety of cellular factors. Host cell engineering of cell growth, q_p, and protein quality have been performed using a variety of different molecular tools, such as gene overexpression, gene knockout and RNA interference-mediated knock-down (41).

The availability of CHO genomic resources and generated -omics data, such as proteomics, transcriptomics, metabolomics and glycomics (42), have brought a vast potential to further

improve the CHO host by identifying phenotypes of interest and effector genes to engineer. Highly efficient genome editing tools have recently become available to further realize the potential of engineering and facilitate validation and multiplexing of different engineering modifications (43). And finally, recently developed genome-scale models have the potential to be used for better understanding of the complex gene interactions and to predict the impact of the modifications and thereby further guide the engineering efforts (27).

3.1 Engineering of host cells with CRISPR/Cas9-mediated and recombinase-based targeted integration

Targeted GOI integration might not be able to compete with recombinant protein yields achieved by random, multi-copy GOI integration. However, the main advantage of targeted integration seems to be the stability of gene expression and minimal clonal variation (44–47). Minimal differences in gene expression are observed due to diminished variability in gene copy numbers and the genomic integration site (so-called position effect) (48). Site-specific integration of GOI has traditionally been achieved using recombinase-based systems (such as Cre/loxP and Flp/FRT), that enabled integration of so-called landing pads, harboring recombination sites, into transcriptionally active sites of the genome (49, 50). Predictable and sustained gene expression can be achieved when GOIs are integrated into well-characterized integration sites, so-called 'safe harbors'. Further, if these integration sites provide high-level gene expression, they are called transcriptional 'hotspots'. Recombinase-based systems have been used to identify hotspots in the genome by random integration of a GOI vector and subsequent screening for a single-copy integration with high expression levels (51–53). As hotspots seem to be rare in the genome, extensive screening is needed to identify the desirable clone. Furthermore, since these newly discovered hotspots are uncharacterized, testing of expression stability has to be carried out. However, once identified, the master clone with an ideal hotspot can be reused for expression of any GOI using recombinase-mediated cassette exchange (RMCE).

Recently, more advanced genome editing tools have been developed, such as zinc-finger nuclease (ZFNs), transcription activator-like effector nucleases (TALENs) and the latest RNA-guided engineered nucleases, such as CRISPR (clustered regularly interspaced short palindromic repeats)/Cas9 (CRISPR-associated protein 9) system (54). The cost and time-efficient manner together with the ease of use provided by the CRISPR/Cas9 systems offer major advantages over ZFNs and TALENs which require lengthier timeframes due to extensive protein engineering required (54). Originating as a component of prokaryotic adaptive immunity, CRISPR/Cas9 type II system has recently been repurposed into a

versatile, highly efficient and remarkably powerful tool for genome engineering (55–60). The system is based on Cas9, an RNA-guided DNA endonuclease, that upon base pairing between the single guide RNA (sgRNA) and the target DNA site-specifically induces double strand breaks (61). Upon the double strand break, endogenous DNA damage repair pathways are activated, which can be harnessed for different genomic modifications, such as gene knockout, gene correction/knockin or chromosome rearrangements (62). The double strand break is repaired either by non-homologous end joining (NHEJ), that can result in insertions or deletions (indels), or homology-directed repair (HDR). Due to the work conducted in this thesis, only one application of CRISPR/Cas9-mediated genome editing will be discussed here in more detail, *i.e.* site-specific gene integration via HDR mechanism.

When double strand breaks are desired to be repaired via HDR mechanism, an exogenous donor repair template harboring homologous DNA sequences (referred to as homology arms) can be co-delivered together with the CRISPR/Cas9 system. The donor repair template is then used to incorporate a GOI into a defined target site via HDR, specifically homologous recombination (HR) (63). This approach has been successfully used in mammalian cells to achieve site-specific integration of GOIs encoding reporter proteins and/or antibiotic selection markers (16, 55, 64). Nowadays, safe harbor integration sites conferring high and stable gene expression levels can be identified experimentally or predicted *in silico* and retargeted with CRISPR/Cas9-mediated genome editing via HDR. Previous work from our group has indeed demonstrated that stable and homogeneous gene expression with minimal clonal variation can be achieved using CRISPR/Cas9-mediated targeted integration (16, 65). Therefore, targeted integration into predefined, well-characterized integration sites using CRISPR/Cas9-mediated genome editing is an attractive approach for CLD. That is due to potentially no need for comprehensive clone screening and expression stability testing, which would significantly shorten the CLD timelines. CRISPR/Cas9-mediated targeted integration can be coupled with recombinase-based systems facilitating even more flexibility to express GOIs (65–67). Further, the combined approach can be undertaken to screen different GOI expression cassettes in a site-specific manner, as it will be demonstrated in **Chapter 2**.

In conclusion, the development of CRISPR/Cas9-mediated genome editing has transformed the way we approach CHO cell factory engineering (22). CRISPR/Cas9-mediated targeted gene integration into the host genome has enabled increased control over recombinant protein expression, thus accelerating the process of CLD for industrial CHO cell lines. Further, CRISPR/Cas9-mediated targeted engineering opens up opportunities for delivery of

multigene systems and therefore further enhancement of CHO cells in regard to cell productivity, growth, metabolism and product quality.

3.2 Engineering of host cells via effector gene overexpression

Engineering of CHO cells via overexpression of effector genes for improvement of protein production has been frequently exploited in the last decades (41). Expression of beneficial effector genes can be achieved by ectopic overexpression, allowing for the establishment of a gain-of-function phenotype. Many different cellular pathways have been targeted by genetic engineering of CHO cells via effector gene overexpression to improve the recombinant protein production, such as apoptosis (68, 69), metabolism (70–72), cell cycle regulation (73) and glycosylation (74, 75). Another common strategy employed is overexpressing effector genes involved in protein biosynthesis (76–78) and secretory pathway (79–81), in order to increase the cellular productivity and secretion rates of recombinant proteins. Indeed, a recent CHO genome-scale metabolic model (27) predicts that traditionally used approaches to boost productivity in CHO cells, such as an addition of histone deacetylase inhibitors for general increase in mRNA and protein levels, are inefficiently utilizing the available metabolic resources. Instead, rational engineering of the host cells' secretory pathways has a higher potential to improve the secretory capacity (27, 82). With the recent availability of the reference genome resources (17–21), a more complete gene expression and mutation landscape of CHO cell lines can be obtained, identifying novel effectors that could be overexpressed. This could facilitate cell engineering efforts and further improve recombinant protein production.

Chapter 1

A Brute-Force Approach: Screening for Desirable Recombinant Protein Production

Rationale

In early stages of CLD, statically-grown monoclonal CHO cell lines are usually selected as high-producing either according to measurements of product titer or product titer normalized to cell number. Therefore, a relatively high-throughput clone screening method was developed as part of this thesis, assessing titer and confluence in 96-well microtiter plates. A proxy of q_p was established, so-called titer-to-confluence (TTC) ratio, mimicking the titer-to-integral viable cell density (IVCD) ratio (*i.e.* q_p). TTC was introduced as an early reporter in order to increase the selection of clones achieving high final titer through high q_p characteristics. In this study, Etanercept-producing clones were generated and then ranked based on either titer or TTC with the objective to perform a first unbiased comparison of the two early reporters for the clone performance in suspension culture.

The project of developing a q_p-based clone screening method was primarily initiated to generate a panel of different recombinant protein-producing cell lines with high q_p, needed in other studies within this thesis. Therefore, we also applied the presented q_p-based clone screening method for generation of IgG (Rituximab)-producing cell lines without the GS-based gene amplification in order to generate low-copy, high q_p-producing clones. First, stable high-producing clones were needed to discover new safe harbor integration sites for reverse engineering. A stable, relatively high-producing clone was identified with no detectable decrease in q_p or Rituximab gene copies throughout the long-term culturing (throughout the manuscripts referred to as the C6_2 producer). Using this clone, we were able to identify a novel integration site conferring high and stable gene expression. This newly discovered safe harbor integration site was used in **Chapter 2**, where systematic evaluation of site-specific recombinant gene expression was carried out. And second, high-producing clones with possible secretion bottlenecks were needed as screening hosts to identify effector genes that would be beneficial for improving the secretory capacity of CHO cells. For that reason, two stable Rituximab high-producing clones (one of them being the C6_2 producer) were used in **Chapter 4** during the process of characterization of novel, q_p-increasing effector genes.

RESEARCH ARTICLE

Recombinant Proteins

Biotechnology Journal

www.biotechnology-journal.com

Using Titer and Titer Normalized to Confluence Are Complementary Strategies for Obtaining Chinese Hamster Ovary Cell Lines with High Volumetric Productivity of Etanercept

Nuša Pristovšek, Henning Gram Hansen, Daria Sergeeva, Nicole Borth, Gyun Min Lee, Mikael Rørdam Andersen, and Helene Faustrup Kildegaard**

The selection of clonally derived Chinese hamster ovary (CHO) cell lines with the highest production rate of recombinant glycoproteins remains a big challenge during early stages of cell line development. Different strategies using either product titer or product titer normalized to cell number are being used to assess suspension-adapted clones when grown statically in microtiter plates. However, no reported study so far has performed a direct head-to-head comparison of these two early reporters for predicting clone performance. Therefore, a screening platform for high-throughput analysis of titer and confluence of etanercept-producing clones is developed. Then an unbiased comparison of clone ranking based on either titer or titer normalized to confluence (TTC) is performed. Using two different suspension cultivation vessels, the authors demonstrate that titer- or TTC-based ranking gives rise to the selection of clones with similar volumetric productivity in batch cultures. Therefore, using both titer- and TTC-based ranking is proposed, allowing for selection of distinct clones with both high integral of viable cell density (IVCD) and high specific productivity features, respectively. This contributes to selection of a versatile panel of clones that can be further characterized and from which the final producer clone can be selected that best fits the production requirements.

1. Introduction

Chinese hamster ovary (CHO) cells continue to be the cell factory of choice for large-scale biopharmaceutical production.[1] To produce a therapeutic protein with the required quality properties according to FDA and EMA regulations, a stable clonal cell line has to be generated, enabling long-term and reproducible protein expression.[2] Therefore, the classical cell line generation procedure consists of random integration of the transgene into the genome of host cells, followed by selection of transgene containing cells using antibiotics or other transgene selection strategies like methionine sulfoximine (MSX). Subsequently, single-cell cloning and finally, selection of high-producer clones from the bulk of low- and non-producers are performed.[3] Selection of high-producer clones remains one of the major bottlenecks in cell line development (CLD), taking up 4 to 8 months and consuming valuable financial and human resources.[4] Random transgene integration, different transgene copy numbers, and the selection pressure encountered during CLD all contribute to genetically heterogeneous cell populations that

N. Pristovšek, H. G. Hansen, D. Sergeeva, G. M. Lee, H. F. Kildegaard
The Novo Nordisk Foundation Center for Biosustainability
Technical University of Denmark
Kemitorvet, Building 220, 2800 Kgs. Lyngby, Denmark
E-mail: hgra@biosustain.dtu.dk; hef@biosustain.dtu.dk

N. Borth
Department of Biotechnology
University of Natural Resources and Life Sciences
Muthgasse 18, 1190 Vienna, Austria

N. Borth
Austrian Centre of Industrial Biotechnology (ACIB)
Muthgasse 11, 1190 Vienna, Austria

G. M. Lee
Department of Biological Sciences
KAIST, 291 Daehak-ro
Yuseong-gu, Daejeon 305-701, Republic of Korea

M. R. Andersen
Department of Biotechnology and Biomedicine
Technical University of Denmark
Søltofts Plads, Building 221
2800 Kgs. Lyngby, Denmark

DOI: 10.1002/biot.201700216

ADVANCED SCIENCE NEWS
www.advancedsciencenews.com

Biotechnology Journal
www.biotechnology-journal.com

require efficient screening processes. Several strategies have been developed in recent years to improve the ease, throughput and performance of the selection processes, carefully reviewed by Priola et al.[5] These strategies are to a certain degree used in academic labs, but even more so in industrial settings, where highly automated pipelines are employed.[6–8]

One of the main interests during CLD is selecting clones predictive of commercial performance. In order to increase the chances of obtaining clones with desired high-producing performances, the final product yield (maximum titer) is a key parameter to assess in early stages of CLD. High product yield can be achieved via high specific productivity (q_p) and/or high integral of viable cell density (IVCD).[9] Both IVCD and q_p depend on extrinsic factors that differ substantially between small-scale and large-scale cultures.[10] However, both parameters also depend on the intrinsic features of the producer clone that must be selected for during the screening process. When clones progress from small-scale to large-scale cultures, they usually undergo several assessment stages, which include different growth vessels. Early assessment stages are known to not be very predictive of the behavior in the final processes.[11,12] Nonetheless, early predictions represent a critical decision-making step in the selection process of which clones to discard or conversely, which clones with the most promising features to proceed with. Typically, the primary screening and ranking of suspension-adapted clones cultured statically (without shaking) in microtiter plates is based on either the product concentration (titer)[7] or a proxy of q_p.[8,11] To the best of our knowledge, a side-by-side comparison of how these two early reporters perform in terms of selecting CHO clones with high volumetric productivity in suspension cultures has not yet been reported.

In order to investigate the performance of titer and proxy of q_p as early reporters for volumetric productivity, we generated a polyclonal cell line by transfecting CHO-S cells with a plasmid co-expressing glutamine synthetase (GS) and etanercept and then applied MSX-selection. Etanercept is a homodimeric Fc-fusion therapeutic protein that is considered a potentially difficult-to-express protein due to its highly O-glycosylated regions.[13–15] We then developed a screening platform that combines enrichment of high q_p by surface labeling-based fluorescence activated cell sorting (FACS) with high-throughput (HT) and semi-automated confluence analysis of cells cultured statically as well as determination of titer in 96-well plates. A so-called titer-to-confluence (TTC) ratio is introduced as a proxy for q_p for statically cultured cells. Finally, we performed an unbiased evaluation of titer- and TTC-based ranking as early reporters for growth rate as well as for specific and volumetric productivity. We conclude that using titer and TTC as early reporters results in selection of etanercept clones with similar volumetric productivity in batch cultures.

2. Experimental Section

2.1. Plasmid Design and Construction

A plasmid co-expressing GS and etanercept in a pcDNA3.1(+) backbone (PL_GS-etanercept; Figure S1, Supporting Information) was constructed with uracil-specific excision reagent cloning method using flexible assembly sequence tags, uracil-containing primers and Phusion® Hot-Start Flex polymerase (New England Biolabs, Ipswich, MA), as previously described.[16] The GS expression cassette (pSV40-GS-SV40pA) as well as the vector backbone without the zeocin resistance gene expression cassette were directly amplified from a pcDNA3.1/Zeo(+)-GS vector. The etanercept expression cassette was amplified from a pDRIVE5-etanercept expression vector, which had been made by subcloning a DNA fragment with the coding sequence of etanercept (GeneArt, Regensburg, Germany) into the commercial pDRIVE5 vector (InvivoGen, San Diego, CA). The etanercept expression cassette (mCMV-hEF-1a-5'HTLV-etanercept-SV40pA) comprises a promoter, consisting of an mCMV enhancer element, a hEF-1a promoter element, and a 5' HTLV untranslated region (UTR). The etanercept coding sequence is based on the amino acid sequence with accession number 7783 from the IMGT database.[17] The nucleotide sequences of the etanercept and GS-expression cassettes are provided in Table S1 and S2, Supporting Information, respectively. Plasmids used as PCR templates are listed in Table S3, Supporting Information. Assembled PCR fragments were transformed into *Escherichia coli* Mach1 competent cells (Life Technologies, Carlsbad, CA). All constructs were verified with sequencing and purified using NucleoBond Xtra Midi EF (Macherey-Nagel, Düren, Germany) according to manufacturer's instructions.

2.2. Cell Culture, Transfection, and Generation of an Etanercept-Expressing Polyclonal Cell Line

CHO-S suspension cells (Life Technologies) were grown in CD CHO medium (Life Technologies) supplemented with glutamine synthetase expression medium supplement (GSEM; Sigma–Aldrich, St. Louis, MO) and 2 μL/mL anti-clumping agent (Life Technologies). Cells were maintained in 125 mL Erlenmeyer shake flasks (Corning Inc., Acton, MA), incubated at 37 °C, 5% CO_2 at 120 rpm (25 mm shaking amplitude) and passaged every 2–3 days. In order to establish a stable etanercept-expressing polyclonal cell line, 4.5×10^6 cells were transfected with 5.4 μg of plasmid DNA using Amaxa Cell Line Nucleofector Kit V system (Lonza, Basel, Switzerland) and program U-024 on a Nucleofector® 2b device (Lonza), according to the manufacturer's protocol. Cells were then seeded at 4.5×10^5 cells/mL in a 125 mL shake flask and 3 days after transfection, selection for GS expression was initiated by adding 25 μM MSX (Sigma–Aldrich; Cat. M5379). After 14 days of selection, cells were subjected to an increased selection pressure (50 μM MSX).

2.3. Surface Labeling of Etanercept and FACS

To enrich for cells with high q_p, immunostaining of etanercept on the plasma membrane (surface labeling) was performed on the etanercept-expressing cell pool by staining with an anti-human IgG Alexa Fluor 488 antibody (Thermo Fischer Scientific; Cat. A-11013) and 0.4 μg/mL propidium iodide (PI, Thermo Fischer Scientific) at 4 °C, as previously described.[18] Two gates were used for sorting cells on FACS (BD FACSJazz™, BD Biosciences, San Jose, CA): 1) a PI-positive/negative gate

ADVANCED SCIENCE NEWS
www.advancedsciencenews.com

Biotechnology Journal
www.biotechnology-journal.com

dividing dead and viable cells, respectively; 2) a gate enriching for high q_p which was set at top 5% highest Alexa Fluor 488 intensity. Only viable cells were sorted and cells were either gated for top 5% highest Alexa Fluor 488 intensity or not (surface labeled, top 5%-enriched and surface labeled, non-enriched, respectively). Cells were bulk (1×10^5 cells/well) and single-cell sorted into flat-bottom Corning 96-well plate (VWR, Radnor, PA; Cat. 734-0954) and flat-bottom Corning 384-well plates (Sigma–Aldrich; Cat. CLS3542-50EA), respectively. Bulk-sorted cells were expanded in suspension and transferred to a 6-well plate, where cells were seeded at a density of 3×10^5 cells/mL and cultivated under 50 μM MSX selection for 3 days to assess q_p. Single cells were sorted in 30 μL of CD CHO medium, supplemented with GSEM, 1.5% HEPES (Gibco) and 1× Antibiotic-Antimycotic (Gibco). Fifteen days after single-cell sorting, the entire volume of sub-confluent clones incubated statically in the absence of MSX was transferred to 200 μL of CD CHO medium, supplemented with GSEM and 1× Antibiotic-Antimycotic in flat-bottom 96-well plates using an epMotion® 5070 liquid handling workstation (Eppendorf, Hamburg, Germany). Subsequently, the 384-well plates were visually inspected in order to ensure that the majority of the cells had been successfully transferred.

2.4. Titer and Confluence-Based Screening of Clones

Following 2 days of static incubation in 96-well plates, the single-cell sorted cells were re-suspended by pipetting to ensure an even spread of cells on the bottom of the wells. The day after, all clones were sub-confluent and the confluence (the percentage of the well covered by cells) and the number of cells in each well was estimated by image cytometry analysis using the Celígo Imaging Cell Cytometer (Nexcelom Bioscience, Lawrence, MA) with the label-free, bright field confluence or direct cell counting application, respectively. A total of 100 μL of the spent media from the 96-well plates were aspirated and analyzed for etanercept titer as described below. Titer values were divided by confluence in order to compute TTC ratios (TTC = titer[μg/mL]/confluence[%]). Then either TTC or titer alone were used to rank all clones. This initial titer- and TTC-based ranking was used throughout the study in order to evaluate the performance of titer and TTC as early reporters for volumetric productivity.

2.5. Batch Cultivation in 96-Well Plates in Suspension

Selected single-cell derived clones were expanded without MSX selection into pre-sterilized polystyrene 96-square System Duetz half-deepwell (HDW)-plates (CR1496c, Enzyscreen, Haarlem, the Netherlands) capped with autoclaved low-evaporation sandwich Duetz covers (CR1296a, Enzyscreen). Plates were incubated in a humidified S41i shaking incubator (New Brunswick, Eppendorf, Hamburg, Germany) at the following conditions: 37 °C, 5% CO_2, 325 rpm (25 mm shaking amplitude). In order to assess productivity of clones, cells were seeded in a range of 1×10^5–1.5×10^6 cells/mL. Clones below this seeding range were excluded from the analysis. Viable cell density (VCD) and viability were measured daily. In order to make exponential growth possible throughout the experiment, etanercept titers were determined after 2 days. q_p was calculated as described elsewhere.[19] In parallel with the 2-day batch culture experiment, a longer-term batch experiment was performed in 96-HDW plates for the same set of clones. Seeding density was in the range of 1×10^5–7.5×10^5 cells/mL. Clones outside of this seeding range were excluded from the analysis. Final titers of the entire set of clones were measured when 50% of the clones reached a viability below 65%.

2.6. Batch Cultivation in Shake Flasks

The top clones from each ranking (TTC- and titer-based) were further evaluated in duplicates in 125 mL shake flasks without MSX selection. Cells were seeded at 4×10^5 cells/mL in 40 mL CD CHO medium, supplemented with GSEM and 2 μL/mL anti-clumping agent. Cells were incubated at 37 °C, 5% CO_2 at 120 rpm (25 mm shaking amplitude). VCD, viability, and etanercept titers were measured daily. Cultures were discontinued when viability dropped below 50%.

2.7. Viability and VCD

When cells were grown in 96-well HDW plates, viability and VCD were determined by a combined Hoechst and PI stain on the Celígo Imaging cytometer as previously described.[20] Viability and VCD of cells in 6-well plates and 125 mL shake flasks were determined on the NucleoCounter NC-200 Cell Counter (Chemometec, Allerod, Denmark) using Via1-Cassettes and the "Viability and Cell Count Method 2" assay.

2.8. Titer Measurements

Etanercept titers were determined in supernatants by bio-layer interferometry using protein A biosensors and an Octet RED96 system (FortéBio, Pall, Menlo Park, CA) as previously described,[21] with an increased shaking speed of 1000 rpm. The absolute titers of etanercept were calculated using a calibration curve generated from a dilution series of Enbrel (Pfizer, New York City, NY; Lot R51698) and absolute quantification was validated by Coomassie-stained SDS-PAGE gels as previously described (data not shown).[22]

2.9. Statistical Analyses

All statistical analyses were performed using GraphPad Prism 7 software (GraphPad Prism Software Inc., La Jolla, CA). To analyze the correlation between two variables, either linear regression or nonlinear regression was used. Nonlinear regression was used when data points were not equally scattered across the entire y range, and nonlinear regression was performed by applying a straight line or a line through origin model with relative weighting. When nonlinear regression was used, outliers were identified and excluded using the ROUT method,[23] where maximum false discovery rate (FDR) was set

ADVANCED SCIENCE NEWS
www.advancedsciencenews.com

Biotechnology Journal
www.biotechnology-journal.com

to 1% ($Q = 1\%$). For calculating the statistical significance ($p < 0.05$) between the data points of two groups, different unpaired parametric or nonparametric *t*-tests were performed. Nonparametric Mann–Whitney test was used when data points were not normally distributed according to the D'Agostino–Pearson test. When variances from two populations were not equal according to F-test, Welch's correction was used.

3. Results and Discussion

In this study, we set out to compare how product titer as well as product titer normalized to confluence (TTC) of suspension-adapted clones in static cultures perform as early reporters for growth rate, specific and volumetric productivity of CHO cell suspension batch cultures (see **Figure 1** for details on the study design). We hypothesized that the titer-based ranking is biased toward clones with a fast-growing phenotype in static cultures. Since the growth profile changes dramatically when moving clones from static to suspension culturing conditions,[11,12] this type of ranking could give rise to poor predictability of the final clone performance in batch cultures. However, by normalizing titer to confluence (TTC-based ranking), high q_p clones would be selected instead. Since q_p in our experience is generally less prone to variations between growth formats, we hypothesized that TTC-based selection had the potential to not only select clones with high q_p, but also with high volumetric productivity.

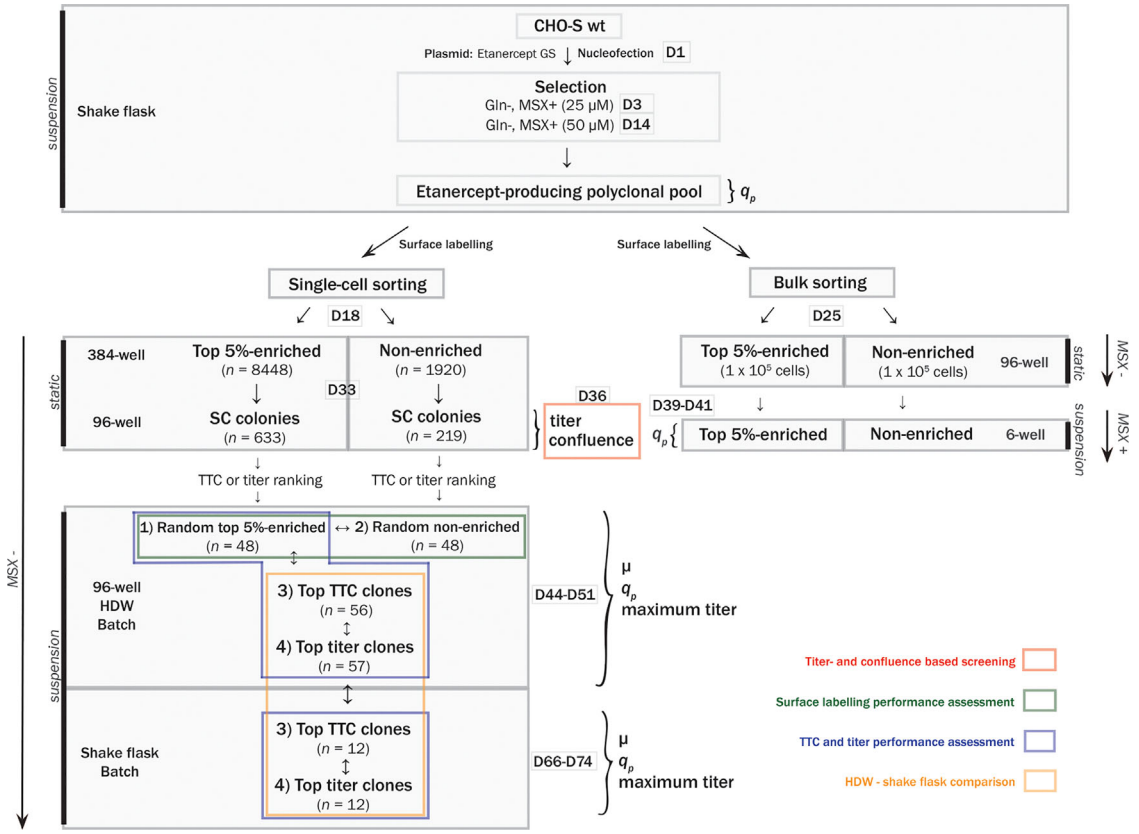


Figure 1. Study design: evaluating surface labeling enrichment and comparing clone performance obtained by TTC- and titer-based selection strategies. An etanercept-producing polyclonal pool was established by nucleofection of CHO-S cells with a plasmid encoding etanercept and glutamine synthetase (GS) on Day 1 and methionine sulfoximine (MSX) selection was subsequently applied. The polyclonal pool was then surface labeled and single-cell sorted and bulk sorted for: a) propidium iodide(PI)-negative cells (non-enriched control) and b) PI negative-cells in the top 5% of anti-IgG fluorescent intensity (top 5%-enriched). Bulk sorted cells were expanded and their specific productivity (q_p) evaluated in a 6-well plate under MSX selection. Following 15 days of static incubation in the absence of MSX, single-cell colonies were transferred from 384-well to 96-well plates. Following 3 days of static incubation without MSX selection, clones were screened for titer and confluence by bio-layer interferometry and image cytometry, respectively. All clones were ranked according to titer and titer-to-confluence (TTC) ratio. Four groups were selected for further evaluation in suspension in 96-half-deepwell (96-HDW) plates: 1) randomly picked top 5%-enriched clones, 2) randomly picked non-enriched clones, 3) top TTC-selected clones, and 4) top titer-selected clones. Surface labeling enrichment was evaluated by comparing q_p between groups 1) and 2). Clone performance (specific cell growth rate (μ), q_p and volumetric productivity) in batch suspension cultures between TTC- and titer-based selection was compared twice without MSX selection. First, between groups 1), 3), and 4) in 96-HDW plates, and second, between a reduced number of highest ranking clones from groups 3) and 4) in shake flasks. The measurements obtained in 96-HDW plates and shake flask batch cultures were compared. SC, single-cell; D1, Day 1.

ADVANCED SCIENCE NEWS
www.advancedsciencenews.com

Biotechnology Journal
www.biotechnology-journal.com

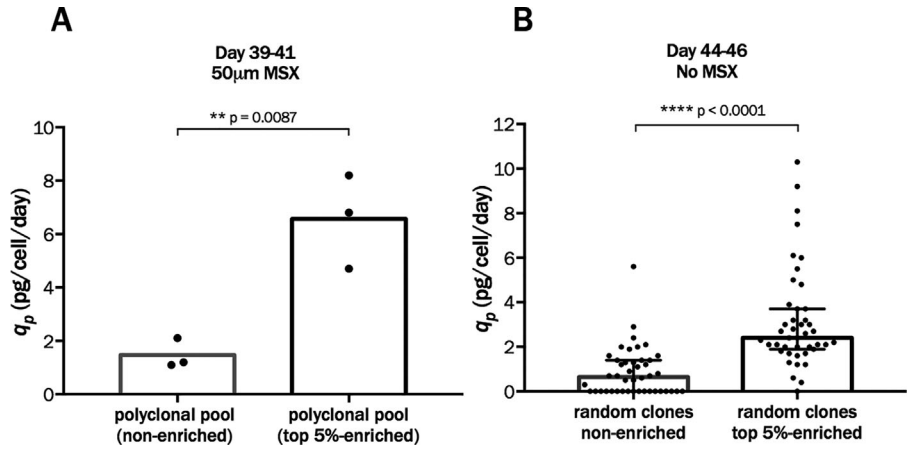


Figure 2. Surface labeling combined with FACS enriches for etanercept-producing clones with high specific productivity (q_p). Etanercept-expressing cell pool was stained with an anti-human IgG Alexa Fluor 488 antibody (surface labeling) and cells were bulk and single-cell sorted for top 5% highest Alexa Fluor 488 fluorescence. A) Evaluation of q_p for bulk sorted non-enriched and top 5%-enriched cells. Subsequent to FACS sorting, cells were recovered and maintained in methionine sulfoximine (MSX)-containing medium and q_p was assessed in 2-days batch cultures (Day 39–41 after nucleofection) in a 6-well plate in technical triplicates. Statistical significance was calculated using unpaired *t*-test (bars represent mean). B) Evaluation of q_p for single-cell sorted, randomly picked non-enriched ($n = 44$) and randomly picked top 5%-enriched ($n = 43$) clones. q_p was assessed in 2-days batch cultures (Day 44–46 after nucleofection) without MSX selection in a 96-half-deepwell (96-HDW) plate. Statistical significance was calculated using Mann–Whitney test (bars represent median with interquartile range (IQR)).

3.1. Enrichment for Etanercept-Producing Clones with High q_p Using Surface Labeling Combined with FACS

After etanercept-producing polyclonal cell line was established during 14 days of MSX selection (Figure S2, Supporting Information), the polyclonal pool was subjected to surface labeling combined with FACS,[18] to enrich for clones with high q_p. A total of 1920 single cells with PI-negative gating (non-enriched cells) and 8448 single cells with double, PI-negative and top 5% highest anti-IgG fluorescence gating (top 5%-enriched cells) were single-cell sorted. Fifteen days after sorting, ≈10% of the single-cell sorted cells were recovered ($n = 852$). Sorting for top 5%-enriched cells gave rise to a threefold increase in q_p both for bulk and single-cell sorted cells (**Figure 2**A and B, respectively). This is consistent with previous reports that have also demonstrated enrichment for clones with high q_p.[6,18,24] Moreover, the inherent heterogeneity in etanercept expression obtained from random genomic integration of plasmids was observed both before and after surface labeling-based enrichment (Figure 2B). When comparing the staining profile at the time of the sorting with obtained q_p values for all surface labeled, single-cell sorted cells (non-enriched and top 5%-enriched), a positive correlation was observed ($R^2 = 0.35$; $R^2 = 0.47$ without outliers; Figure S3, Supporting Information). These results show that surface-labeling combined with FACS is capable of enriching for clones with high q_p despite an imperfect correlation between staining intensity and q_p.

3.2. High-Throughput Titer and Confluence Screening of Etanercept-Producing Clones in Static Culture

Screening efforts encompassing several growth formats are usually required to isolate a stable cell line with the desired properties from inherently heterogeneous cell pools. To assess q_p in the early phase of CLD, we developed a 96-well screening platform where IgG or Fc-fusion protein titer and confluence of suspension-adapted cells are determined in a HT and automation-friendly manner (Figure S4, Supporting Information). These two measurements can be used to calculate the TTC ratio as a proxy for q_p. Similarly, Porter et al.[11] used TTC ratio (referred to as "specific activity") when assessing clones in 96-well plates. Whereas "specific activity" relied on visual inspection of confluence, confluence was determined by image cytometry in the present study (Figure S5, Supporting Information). In addition, a clear linear relationship was observed between confluence and cell numbers (Figure S6, Supporting Information), suggesting that confluence is a valid proxy for cell numbers. A total of 852 surviving clones were cultured statically and screened for titer and confluence within a single day. Approximately 80% of the 852 screened clones had detectable etanercept titers ($n = 672$) and consequently had a positive TTC value assigned. All clones were then ranked twice, either according to their TTC or titer values. This initial titer- and TTC-based ranking was used throughout the study in order to evaluate the performance of titer and TTC as early reporters for volumetric productivity (Table S4, Supporting Information). Calculated TTC values were spanning from 0.1 to 94.5 (arbitrary units) and measured titer values were ranging from 0.05 to 2.8 µg/mL.

3.3. Characterization of Selected Clones in Suspension in 96-Half-Deepwell Plates

To maximize the number of clones being evaluated in suspension, we characterized the selected clones in the previously described System Duetz 96-HDW plates.[20,25] Top

ADVANCED SCIENCE NEWS
www.advancedsciencenews.com

Biotechnology Journal
www.biotechnology-journal.com

56 and 57 clones (approximately top sixth percentile of the clones) in the TTC and titer category, respectively, and 48 randomly picked top 5%-enriched clones were selected for further evaluation (Figure 1). 40 of the 56 TTC-selected clones and 55 of the 57 titer-selected clones survived expansion into 96-HDW plates, whereas 43 of 48 randomly selected clones survived. The difference in survival rate between the TTC- and titer-selected is probably explained by the lower confluence of TTC-selected clones and by the fact that the majority of the non-surviving clones had relatively low confluence (Figure S7, Supporting Information). Sixteen clones were overlapping between the TTC and titer categories giving rise to 24 and 39 unique TTC- and titer-selected clones, respectively.

In 2-day batch cultures in 96-HDW plates, TTC-selected clones on average initially grew slower compared to both titer-selected clones and randomly selected clones (**Figure 3**A). However, clones in the TTC category were more frequently seeded at a lower cell density, which correlated with low growth rates (Figure S8, Supporting Information). By comparing only the clones seeded above 4×10^5 cells/mL, we observed no difference in growth rate between TTC- and titer-selected clones (Figure S9A, Supporting Information). Hence, predominantly TTC-selected clones seeded at a lower cell density ($<4 \times 10^5$ cells/mL) probably experienced a longer lag phase with a possible negative effect on the cell growth rate during the 2 days of culture. Similarly, TTC-selected clones generally had significantly higher q_p compared to titer-selected clones (Figure 3B). However, when comparing the clones with seeding density above 4×10^5 cells/mL, no difference in q_p between the TTC- and titer-selected clones was observed (Figure S9B, Supporting Information). Independent of seeding density, the randomly selected clones had significantly lower q_p compared to both the TTC and titer categories (Figure 3B). While a poor correlation was observed between confluence (static cultures) and specific cell growth rate (suspension cultures) (Figure S10, Supporting Information), a clear correlation was observed between TTC and q_p (Figure 3C). Thus, while growth rate clearly responds to changes in culture conditions consistent with

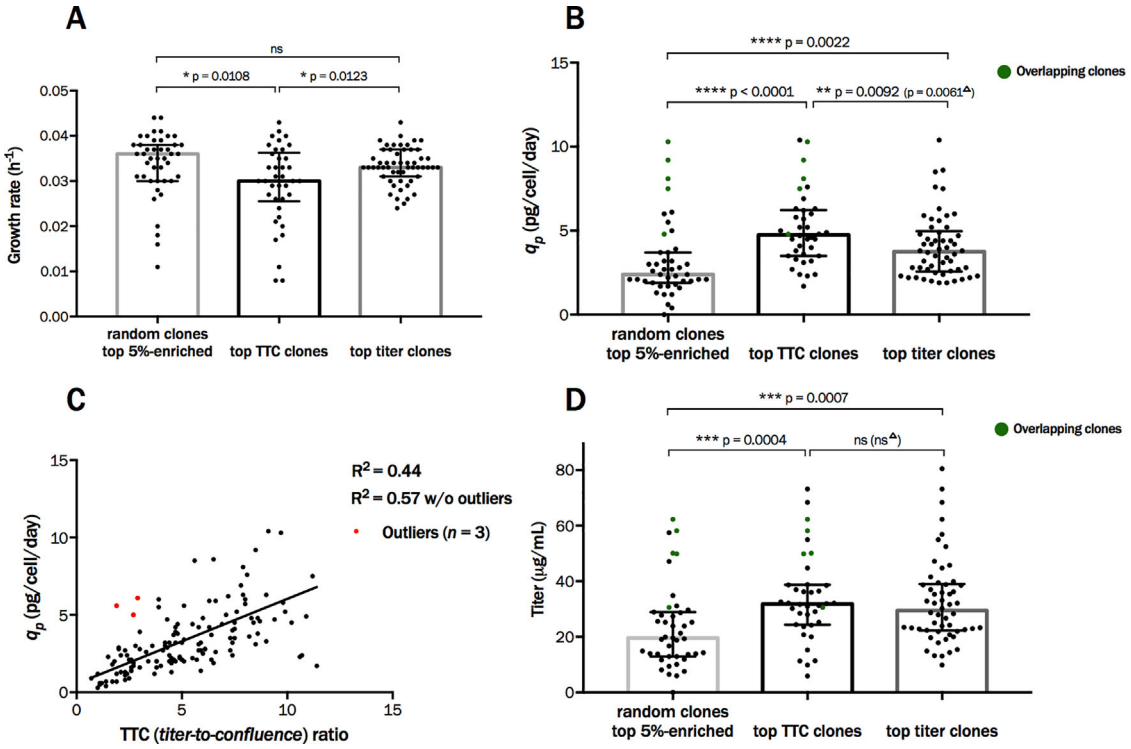


Figure 3. TTC- and titer-selected clones achieve higher specific and volumetric productivity than randomly selected clones in 96-half-deepwell plates. Selected clones were transferred from static culture to suspension in 96-half-deepwell (HDW) plates. Forty-six days after nucleofection, specific cell growth rate (μ) and specific productivity (q_p) were assessed in 2-days batch cultures in 96-HDW plates. In parallel, maximum titer was assessed in 7-days batch cultures in 96-HDW plates. A) Comparison of specific growth rate in a 2-days batch culture experiment between randomly picked top 5%-enriched clones ($n = 43$), top TTC clones ($n = 38$) and top titer clones ($n = 54$). Statistical significance was calculated using Mann–Whitney test (randomly picked top 5%-enriched clones vs. top TTC clones; randomly picked top 5%-enriched clones vs. top titer clones) or Welch's t test (top TTC clones vs. top titer clones) (bars represent mean ± interquartile range (IQR)). B) Comparison of q_p in a 2-days batch culture experiment between randomly picked top 5%-enriched clones, top TTC clones and top titer clones. Statistical significance was calculated using Mann–Whitney test (bars represent median with IQR). C) Correlation analysis of q_p and TTC ratio for all the clones in 96-HDW plates with positive TTC values ($n = 138$) using nonlinear regression with a straight line model. R^2 values without and with relative weighing and elimination of outliers are reported. D) Comparison of maximum titer between randomly picked top 5%-enriched clones, top TTC clones and top titer clones. Statistical significance was calculated using Mann–Whitney test (bars represent median with IQR). Δ, p-value reported for comparison between groups without the clones present in both groups ($n = 16$). Only clones with a seeding density above 1×10^5 cells/mL were analyzed. Overlapping clones marked with green color in panels B and D are clones present in "random clones top 5%-enriched" that are also present in either "top TTC clones" and/or "top titer clones."

ADVANCED SCIENCE NEWS
www.advancedsciencenews.com

Biotechnology Journal
www.biotechnology-journal.com

previous reports [11,12], the q_p of clones seems to be less affected when going from static incubation to suspension culture.

In order to estimate volumetric productivity (maximum titer) in long-term batch suspension cultures in a HT manner, the selected clones were again cultivated in 96-HDW plates. This time, however, all cultures were maintained longer until more than 50% of the clones had viability below 65%. As expected, randomly selected clones had lower maximum titers compared to both the TTC- and titer-selected clones (Figure 3D). However, no significant differences in maximum titers were observed between TTC- and titer-selected clones independent of whether the 16 overlapping clones were included or excluded from the analysis (Figure 3D). In addition, seeding density did not seem to have a clear impact on the maximum titer as opposed to cell growth rate (Figure S8, Supporting Information). Due to the viability-based termination point, viability of clones was different at the end of the experiment. In fact, 47% and 30% of the clones had viability above 65% in the TTC and titer categories, respectively, suggesting that maximum achievable titers were underestimated to a larger degree for the TTC-selected clones compared to the titer-selected clones.

3.4. Characterization of Highest Ranking Clones in Shake Flasks and Comparison to 96-Half-Deepwell Data

To assess the performance of the TTC- and titer-selected clones more thoroughly, the top 12 of the surviving clones in both categories were subjected to a standard batch culture experiment in shake flasks. Four of these clones were overlapping between the two categories, giving rise to eight unique clones in both categories. Clone IDs were assigned to each clone in shake flask batch culture and the clone IDs with the corresponding ranks can be seen in Table S4, Supporting Information.

Specific growth rates of TTC- and titer-selected clones were not significantly different from Day 0 to 2 in shake flasks, whether or not overlapping clones were included in the comparison (**Figure 4**A, left panel; growth curves in Figure S11, Supporting Information). Similarly, no differences in growth rates were observed when comparing the same set of top 12 TTC and titer clones in 96-HDW plates (Figure S12A, Supporting Information). Overall, specific growth rates of the top 12 TTC and titer clones were consistent between the 96-HDW plates and shake flasks except for clones 10 and 11 (Figure 4A, right panel), which were seeded at a lower density in 96-HDW plates (Figure S8, Supporting Information). Interestingly, clones 10 and 11 in shake flasks had comparable cell growth rates to the remaining top 12 TTC and titer-selected clones. This is consistent with the previously mentioned notion that these clones potentially had a longer lag phase in the 96-HDW plates due to lower seeding density. Importantly, we have previously observed similar growth profiles of CHO-S cells between shake flasks and 96-HDW plates.[20] Thus, the 96-HDW plate seems to be a suitable microscale vessel for comparison of clones in batch cultures when seeding densities are aligned.

TTC- and titer-selected clones on average achieved similar q_p in different stages of batch culture (Day 0–2 and Day 3–5), whether or not overlapping clones were included in the comparison (Figure 4B, left panel; growth curves in Figure S11, Supporting Information). Similarly, we observed no differences in q_p when comparing the top 12 TTC and top 12 titer clones in 96-HDW plates (Figure S12B, Supporting Information). Comparison of q_p between HDW plates and shake flasks from Day 0 to Day 2 demonstrated a positive correlation (Figure 4B, right panel) ($R^2 = 0.58$; $R^2 = 0.71$ without one outlier), suggesting q_p was largely unaffected by different seeding densities or formats. However, q_p of the clones was decreasing through the time course of the experiments in the present study (Figure 2 and 4B, right panel). In our attempt to do an unbiased evaluation of TTC- and titer-based selection, MSX selection pressure was not reintroduced after single-cell sorting in order to keep the medium unchanged in all cell culture vessels (Figure 1). The decrease in q_p is likely a consequence of no MSX selection pressure after FACS similar to what has been observed for CHO clonal cell lines obtained by dihydrofolate reductase/methotrexate-based selection[26] and/or the presence of post-transcriptional bottlenecks.[27] Thus, it is important to note that the evaluation of the TTC- and titer-based selection is most likely affected by this loss of productivity. Future studies are warranted to investigate whether TTC- or titer-based selection enriches for clones with stable productivity.

Like q_p, similar average maximum titers were achieved by the TTC- and titer-selected clones (Figure 4C, left panel). When comparing maximum titers obtained in HDW plates and shake flasks, the values correlated moderately (Figure 4C, right panel) ($R^2 = 0.52$). This moderate correlation is likely a combined consequence of difference in seeding density (Figure S8, Supporting Information) and the viability-based termination point in the 96-HDW plates.

Despite not being able to increase the frequency of isolating clones with higher q_p, the two clones with the highest q_p and volumetric productivity in the shake flask experiment were uniquely found by TTC-based selection. These two clones (clone ID: 2 and 11) were ranked as 2nd and 11th highest TTC clone (151st and 557th highest titer clone, respectively) during the initial screening in static culture. As expected, clone 11, having the highest q_p and volumetric productivity also had the lowest growth rate in the early phase (Day 0–2) of the batch culture ($\mu = 0.027\,h^{-1}$) and the lowest accumulated IVCD (1.8×10^7 cells*day/mL). This clone exhibited quite an extraordinary phenotype in comparison to the other highest ranking TTC and titer clones in shake flasks, having a remarkably higher q_p, especially from Day 3 to 5 (Figure 4B, left panel). Interestingly, the two clones (clone ID: 13 and 17) with highest volumetric productivity in the titer category were likewise uniquely found by titer-based selection and ranked as first and fifth highest titer clones (113th and 126th highest TTC clone, respectively) during the initial screening of statically cultured clones. The titer-selected clone with highest volumetric productivity had one of the highest growth rates from Day 0 to 2 ($\mu = 0.039\,h^{-1}$) and the highest accumulated IVCD (3.9×10^7 cells*day/mL) in the entire clone panel. In addition, higher variation in the TTC category was observed for both q_p (Figure 4B, left panel) and volumetric productivity (Figure 4C, left panel). This suggests that clones in the TTC category are phenotypically more diverse compared to titer-selected clones. Overall, these observations support the notion that high volumetric productivity can be achieved through high IVCD of clones selected by titer-based ranking or through high q_p of clones selected by TTC-based ranking.

It has previously been reported that performance in batch cultures does not necessarily predict the performance in final

ADVANCED SCIENCE NEWS
www.advancedsciencenews.com

Biotechnology Journal
www.biotechnology-journal.com

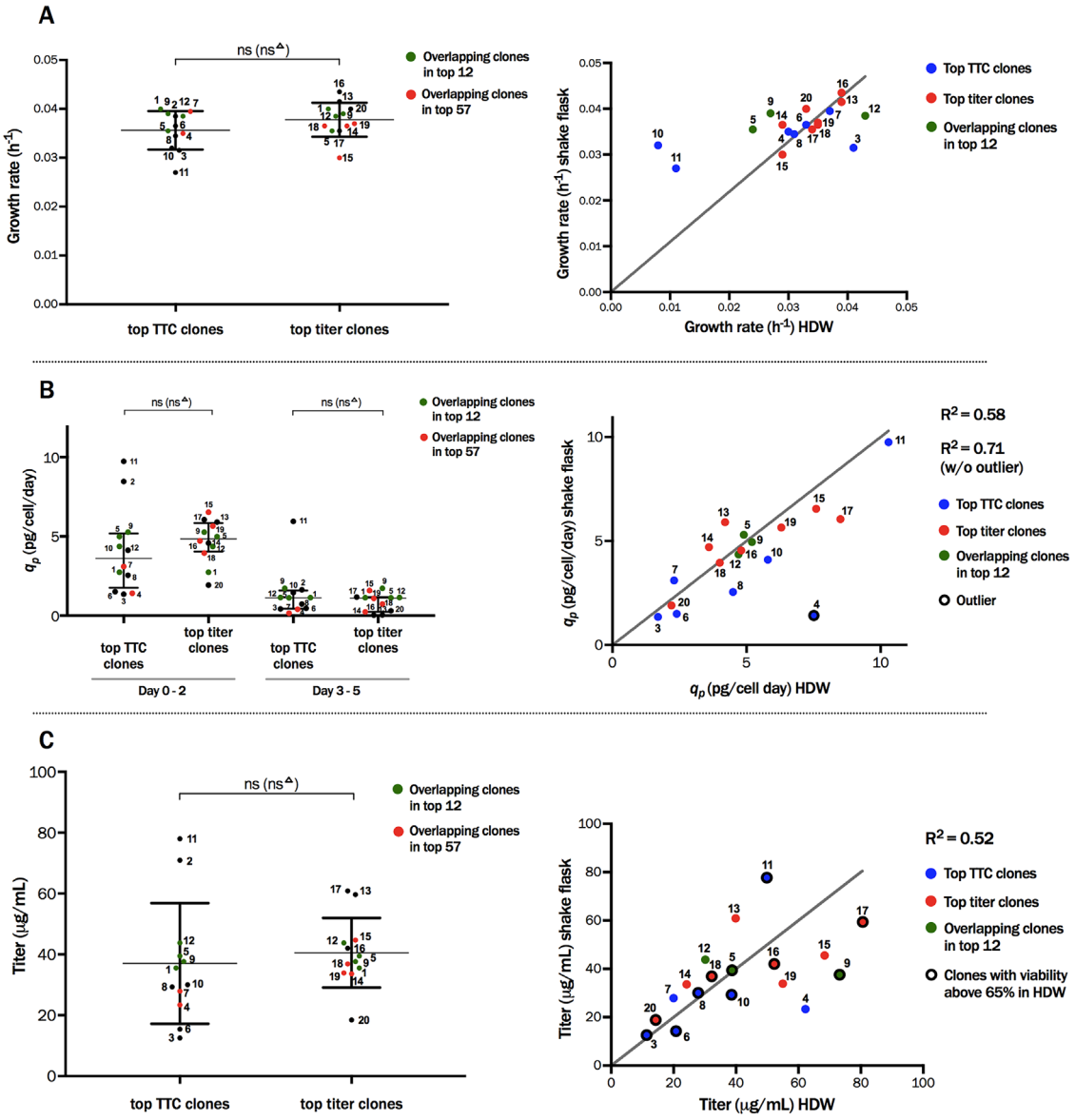


Figure 4. Top TTC- and titer-based clones achieve similar final volumetric productivity in shake flask batch cultures. Top 12 TTC- and titer clones were transferred from 96-half-deepwell (HDW) plates to shake flasks. Sixty-six to seventy-four days after nucleofection, specific cell growth rate (μ), specific productivity (q_p) and maximum titer were assessed in batch cultures. Clone performance was compared between 96-HDW plates and shake flask batch cultures. A, left panel) Comparison of specific growth rate from Day 0 to 2 in shake flask batch cultures between top titer-to-confluence (TTC) clones ($n = 12$) and top titer clones ($n = 12$). Statistical significance was calculated using unpaired t-test (bars represent mean $\pm$ SD). A, right panel) Correlation analysis of specific growth rates of clones from Day 0 to 2 between 96-HDW plates and shake flask batch cultures ($n = 18$), using regression through the origin model. B, left panel) Comparison of q_p from Day 0 to 2 and Day 3 to 5 of shake flask batch cultures between top TTC clones and top titer clones. Statistical significance was assessed using Welch's t-test (Day 0–2) and Mann-Whitney test (Day 3–5) (bars represent median $\pm$ interquartile range (IQR)). B, right panel) Correlation analysis of specific productivity (q_p) of clones from Day 0 to 2 between 96-HDW plates and shake flask batch cultures, using regression through the origin model. R^2 values without and with relative weighing and elimination of an outlier are reported. C, left panel) Comparison of maximum titer in shake flask batch cultures between top TTC clones and top titer clones. Statistical significance was assessed using unpaired t test (bars represent mean $\pm$ SD). C, right panel) Correlation analysis of maximum titer of clones between 96-HDW plates and shake flask batch cultures, using regression through the origin model. R^2 with relative weighing is reported. Δ, p-value for comparison between groups without the clones present in both groups ($n = 4$). Correlation analysis was only performed for clones with seeding density $> 1 \times 10^5$ cells/mL in 96-HDW plates. IDs ranging from 1 to 20 were assigned to clones in shake flask batch cultures and the corresponding titer- and TTC-based ranks can be seen in Table S4, Supporting Information.

fed-batch cultures, where feed and different media formulation are contributing to major changes in culture conditions, influencing cell growth, specific and volumetric productivity of clones.[11,28] Due to these unpredictable differences in clonal behavior between batch culture and fed-batch culture, it might be an advantage to have a panel of phenotypically diverse high-producer clones to test in fed-batch cultures. Such a panel seems achievable by employing a dual titer- and TTC-selection strategy. Accordingly, it seems likely that clones fit-for-purpose for already established commercial protein production processes are present in such diverse clone panels, speeding up the process of cell line commercialization.

4. Conclusions

A thorough analysis of the heterogeneity in a polyclonal pool was accomplished by HT screening of 852 clones, using selection of clones based on either titer or TTC. Despite poor correlation observed between confluence of statically cultured clones and cell growth rate of clones in suspension, we observed that q_p remains largely unaffected. We were able to show that TTC as a proxy for q_p is not a precise reporter for high volumetric productivity in suspension consistent with previous reports.[11,12] However, TTC-based ranking was used to identify clones with high volumetric productivity, which would not have been found if using titer-based ranking. We therefore suggest TTC-based selection as an addition to titer-based selection, because it increases the possibility of selecting clones in the early phase of CLD with high q_p as well as volumetric productivity. Furthermore, the presented screening platform has been developed to isolate etanercept-producing clones and is therefore applicable for other Fc-fusion proteins as well as IgG variants, which currently constitute the majority of biopharmaceuticals on the market.[29]

Abbreviations

CHO, Chinese hamster ovary; CLD, cell line development; FACS, fluorescence-activated cell sorting; GS, glutamine synthetase; HDW, half-deepwell; HT, high-throughput; IVCD, integral of viable cell density; MSX, methionine sulfoximine; PI, propidium iodide; q_p, specific productivity; TTC, titer-to-confluence; VCD, viable cell density.

Supporting Information

Supporting Information is available from the Wiley Online Library or from the author.

Acknowledgments

The authors thank Nachon Charanyanonda Petersen and Karen Kathrine Brøndum for their assistance with the FACS, Sara Petersen Bjørn for providing one of the plasmids and Stefan Kol for his assistance with the titer measurements. The authors also thank Anne Tolstrup for valuable discussions. This work was supported by the Novo Nordisk Foundation (NNF10CC1016517 and NNF16CC0020908). NP, NB, HFK, and MRA are receiving funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 642663.

Conflict of Interest

The authors declare no commercial or financial conflict of interest.

Keywords

cell line development, CHO cells, high-throughput, industrial biotechnology, recombinant proteins, screening

Received: August 4, 2017
Revised: December 15, 2017
Published online: February 9, 2018

[1] J. Zhu, *Biotechnol. Adv.* **2012**, *30*, 1158.
[2] F. M. Wurm, *Nat. Biotechnol.* **2004**, *22*, 1393.
[3] S. M. Noh, M. Sathyamurthy, G. M. Lee, *Curr. Opin. Chem. Eng.* **2013**, *2*, 391.
[4] S. M. Browne, M. Al-Rubeai, *Trends Biotechnol.* **2007**, *25*, 425.
[5] J. J. Priola, N. Calzadilla, M. Baumann, N. Borth, C. G. Tate, M. J. Betenbaugh, *Biotechnol. J.* **2016**, *11*, 853.
[6] S. Shi, R. G. G. Condon, L. Deng, J. Saunders, F. Hung, Y.-S. Tsao, Z. Liu, *J. Vis. Exp.* **2011**, e3010.
[7] K. Lindgren, A. Salmén, M. Lundgren, L. Bylund, Å. Ebler, E. Fäldt, L. Sörvik, C. Fenge, U. Skoging-Nyberg, *Cytotechnology* **2009**, *59*, 1.
[8] J. J. C. Hou, B. S. Hughes, M. Smede, K. M. Leung, K. Levine, S. Rigby, P. P. Gray, T. P. Munro, *N. Biotechnol.* **2014**, *31*, 214.
[9] J. Y. Kim, Y. G. Kim, G. M. Lee, *Appl. Microbiol. Biotechnol.* **2012**, *93*, 917.
[10] F. Li, N. Vijayasankaran, A. Shen, R. Kiss, A. Amanullah, *MAbs* **2010**, *2*, 466.
[11] A. J. Porter, A. J. Racher, R. Preziosi, A. J. Dickson, *Biotechnol. Prog.* **2010**, *26*, 1455.
[12] A. J. Porter, A. J. Dickson, A. J. Racher, *Biotechnol. Prog.* **2010**, *26*, 1446.
[13] S. Houel, M. Hilliard, Y. Q. Yu, N. MeLoughlin, S. Millan Martin, P. M. Rudd, J. P. William, W. Chen, *Anal. Chem.* **2014**, *86*, 567.
[14] P. M. O'Callaghan, M. E. Berthelot, R. J. Young, J. W. A. Graham, A. J. Racher, D. Aldana, *Biotechnol. Prog.* **2015**, *31*, 1187.
[15] B. Hassett, E. Singh, E. Mahgoub, J. O'Brien, S. M. Vicik, B. Fitzpatrick, *MAbs* **2017**, *862*, 0.
[16] A. M. Lund, H. F. Kildegaard, M. B. K. Petersen, J. Rank, B. G. Hansen, M. R. Andersen, U. H. Mortensen, *PLoS ONE* **2014**, *9*, e96693.
[17] M.-P. Lefranc, V. Giudicelli, Q. Kaas, E. Duprat, J. Jabado-Michaloud, D. Scaviner, C. Ginestoux, O. Clément, D. Chaume, *Nucleic Acids Res.* **2005**, *33*, D593.
[18] S. C. Brezinsky, G. Chiang, A. Szilvasi, S. Mohan, R. Shapiro, A. MacLean, W. Sisk, G. Thill, *J. Immunol. Methods* **2003**, *277*, 141.
[19] L. P. Pybus, G. Dean, N. R. West, A. Smith, O. Daramola, R. Field, S. J. Wilkinson, D. C. James, *Biotechnol. Bioeng.* **2014**, *111*, 372.
[20] H. G. Hansen, C. N. Nilsson, A. M. Lund, S. Kol, L. M. Grav, M. Lundqvist, J. Rockberg, G. M. Lee, M. R. Andersen, H. F. Kildegaard, *Sci. Rep.* **2016**, *5*, 18016.
[21] L. M. Grav, J. S. Lee, S. Gerling, T. B. Kallehauge, A. H. Hansen, S. Kol, G. M. Lee, L. E. Pedersen, H. F. Kildegaard, *Biotechnol. J.* **2015**, *10*, 1446.
[22] H. G. Hansen, H. F. Kildegaard, G. M. Lee, S. Kol, *Biotechnol. J.* **2016**, *11*, 1648.

ADVANCED SCIENCE NEWS
www.advancedsciencenews.com

Biotechnology Journal
www.biotechnology-journal.com

[23] H. J. Motulsky, R. E. Brown, *BMC Bioinform*. **2006**, *7*.
[24] J. Pichler, S. Galosy, J. Mott, N. Borth, *Biotechnol. Bioeng.* **2011**, *108*, 386.
[25] W. A. Duetz, B. Witholt, *Biochem. Eng. J*. **2001**, *7*, 113.
[26] S. J. Kim, N. S. Kim, C. J. Ryu, H. J. Hong, G. M. Lee, *Biotechnol. Bioeng*. **1998**, *58*, 73.
[27] H. G. Hansen, N. Pristovšek, H. F. Kildegaard, G. M. Lee, *Biotechnol. Adv.* **2017**, *35*, 64.
[28] D. Reinhart, L. Damjanovic, C. Kaisermayer, R. Kunert, *Appl. Microbiol. Biotechnol*. **2015**, *99*, 4645.
[29] G. Walsh, *Nat. Biotechnol*. **2014**, *32*, 992.

Chapter 2

A Smart Approach: Towards Predictable Recombinant Gene Expression

Rationale

Stable and predictable expression of recombinant genes in mammalian cells is crucial for studies of gene functions in basic research, for gene therapy, as well as for biotherapeutics production in industry. Traditionally, recombinant gene-expressing cell lines are generated by random integration of a recombinant gene into the genome, yielding clones with a wide range of expression, growth and stability characteristics. When applying targeted integration of recombinant gene into defined, safe harbor integration sites, homogenous expression levels with minimal clonal variation have been observed (16, 65). However, recombinant gene expression patterns in a characterized integration site are highly dependent on the expression cassette being used. To the best of our knowledge, no method has been developed that would allow systematic evaluation of site-specific recombinant gene expression levels modulated by different expression cassette components in mammalian cells. Such method would provide an excellent opportunity for the development of CHO cell lines with predictable and desired recombinant gene expression in defined integration sites.

In the present study, we developed a CRISPR/Cas9-based toolbox that allows the construction of CHO cell lines with a site-specifically integrated landing pad, containing a recombinant gene under defined 5' proximal regulatory elements. For this purpose we used three different integration sites - two well-characterized integration sites, used routinely in-house, and one novel integration site, identified in the stable Rituximab high-producing clone, generated with a clone screening method, presented in **Chapter 1**. We then recombined several different expression cassettes (*i.e.* 5' proximal regulatory element-recombinant gene) into the site-specifically integrated landing pad of a genetically identical (isogenic) cell line. Using the developed toolbox, we were able to systematically analyze the interplays between the different layers contributing to the final recombinant gene expression level.

Systematic Evaluation of Site-Specific Recombinant Gene Expression for Programmable Mammalian Cell Engineering

Nuša Pristovšek[1], Saranya Nallapareddy[1], Lise Marie Grav[1], Hooman Hefzi[2,3], Nathan E. Lewis[2,3], Peter Rugbjerg[1], Henning Gram Hansen[1], Gyun Min Lee[1,4], Mikael Rørdam Andersen[5], Helene Faustrup Kildegaard[1]

[1]The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Kemitorvet, Building 220, 2800 Kgs. Lyngby, Denmark
[2]Departments of Pediatrics and Bioengineering, University of California, San Diego, La Jolla, CA 92093, USA
[3]The Novo Nordisk Foundation Center for Biosustainability at the University of California, San Diego School of Medicine, CA 92093, USA
[4]Department of Biological Sciences, KAIST, 291 Daehak-ro, Yuseong-gu, Daejeon 305-701, Republic of Korea
[5]Department of Biotechnology and Biomedicine, Technical University of Denmark, Søltofts Plads, Building 223, 2800 Kgs. Lyngby, Denmark

Abstract

Many branches of biology depend on stable and predictable recombinant gene expression, that has been achieved in the recent years through targeted integration of the recombinant gene into defined integration sites. However, transcriptional levels of recombinant genes in characterized integration sites are controlled by multiple components of the integrated expression cassette. Lack of readily available tools has inhibited meaningful experimental investigation of the interplay between the integration site and the expression cassette components. Here we show in a systematic manner how multiple components contribute to final net expression of recombinant genes in a characterized integration site. We develop a CRISPR/Cas9-based toolbox for construction of mammalian cell lines with site-specifically integrated landing pad, containing a recombinant gene under defined 5' proximal regulatory elements. Generated site-specific recombinant cell lines can be used in a streamlined recombinase-mediated cassette exchange for fast screening of different expression cassettes. Using the developed toolbox, we show that different 5' proximal regulatory elements generate distinct and robust recombinant gene expression patterns in defined integration sites of CHO cells with a wide range of transcriptional outputs. This approach facilitates the generation of user-defined and product-specific gene expression patterns for programmable mammalian cell engineering.

Introduction

The ability to precisely control and maintain the level of recombinant gene expression is crucial to mammalian cell engineering. To ensure that recombinant genes are stably expressed in the host cells, genomic integration is needed, which is typically still conducted in a random or semi-random manner (*i.e.* insertion of cassettes into actively transcribed regions of the genome). When random or semi-random integration is applied, unpredictable and heterogenous transgene expression is observed due to variability in gene copy numbers and the nature of the chromosomal integration site (so-called position effect) (1, 2). Accordingly, laborious screening for the desired expression levels is required (3, 4). In contrast, methods applying targeted integration into predefined loci pave the way toward rational design of cells with predictable transgene expression (5). The use of the type II CRISPR (clustered regularly interspaced short palindromic repeats)/Cas9 (CRISPR-associated protein 9) system has enabled more efficient and precise genome editing (6). Harnessing the homology-directed repair (HDR) mechanism, error-free targeted integration of an exogenous DNA donor template into the genome can be achieved following a double-strand break induced by Cas9 (7). HDR-mediated gene targeting has been used

extensively in mammalian cell engineering with applications ranging from biopharmaceutical production, gene therapy to gene function studies (8). What these fields have in common is that they are all struggling to overcome the obstacles of variable transgene expression, clonal variation, and oncogenic transformation (chromosomal rearrangements).

Therapeutic glycoproteins are traditionally produced using Chinese hamster ovary (CHO) cells, where the final yield and stability of the protein production are important aspects of cost-effective production pipelines (9). To achieve predictable and desirable protein production, protocols for CRISPR/Cas9-based site-specific integration (5, 10–12) and recombination-based site-specific integration (13–18) have been developed, sometimes used in combination (19–21) that ensure controlled genome integration into predefined loci. Recombination-based systems require a prior establishment of master cell lines containing a marker gene with recombinase recognition sites, referred to as landing pads, allowing for recombinase-mediated cassette exchange (RMCE). CRISPR/Cas9-based targeted integration and RMCE result in predictable and desirable recombinant gene expression patterns when employed into previously characterized, safe genomic loci (*i.e.* safe harbors). Safe harbors are genomic sites which are: a) positioned in open chromatin areas that are generally transcriptionally active, b) immune to mono- or biallelic disruption of the target locus upon transgene integration, and c) not receptive to transcriptional perturbations of endogenous genes upon transgene integration (22). There are a few widely used safe harbors provided for engineering in the human genome, such as AAVS1, CCR5 and ROSA26 (23). However, there is a continually growing repository of integration sites supplied for mammalian cell engineering (12, 20, 24), albeit, their proper validation as safe harbors is needed.

Apart from the chromosomal context, the nature of the expression construct used to generate recombinant cell lines has an equally important impact on protein yield and stability (25). A range of vector elements can modulate transgene expression, ranging from chromatin-modifying elements (such as matrix attachment regions (MARs), insulators, ubiquitously acting chromatin opening elements (UCOEs) and stabilizing anti-repressors (STARs)) to natural and engineered promoters (26). Site-specific expression levels with diverse vector elements have been studied in different fields of biology (14, 27, 28). When applying targeted integration into predefined loci, these studies have confirmed that the expression level of the transgene is highly dependent on the interplay between the promoter and the target genomic locus. Since methods that would predict the interaction of the integrated cassette with the genome are not available, both, safe harbor sites and

performance of different expression cassettes in such chromosomal contexts can currently only be verified experimentally.

To the best of our knowledge, no available toolbox has been published so far that would allow streamlined evaluation of site-specific stable gene expression levels under different expression cassettes. To this end, we present a modular toolbox for systematic screening of vector elements in predefined integration sites, using the CHO-S cell line as a model. By studying different expression cassette designs in newly discovered safe harbors using targeted integration, we demonstrate that desired (*e.g.* high or stable) expression levels in defined chromosomal loci are restricted to a specific cassette design. We explore four different layers contributing to the final net expression level, namely clonal background, chromosomal context, vector design, and type of recombinant protein by studying five parental clones, three integration sites, six 5' proximal regulatory elements and three recombinant proteins, respectively. This renders a total of 270 possible combinations between the four features, and we here examine 66 (25%) of those combinations experimentally. A more detailed understanding of the performance of vector elements when embedded in a defined chromosomal context is not only of general interest to the mammalian cell engineering community, but is also critical knowledge when designing suitable expression cassettes for engineering more complex expression patterns. Furthermore, since there is a need to further diversify elements for gene expression, this tool allows for identification of a wide range of robust gene expression patterns providing means toward programming mammalian cells with dynamic functionalities.

Material and methods

Identification of integration sites

Site A was identified in a monoclonal cell line with randomly integrated Rituximab (C6_2), generated with an in-house developed clone screening platform (29). For more information regarding the C6_2 cell line generation, refer to the Supplementary data. The C6_2 cell line was outsourced to Cergentis B.V. (Utrecht, The Netherlands), where integration site A was first identified using the proprietary 'Targeted Locus Amplification' (TLA) technology (30). Site A was re-analyzed with an in-house targeted MiSeq deep-sequencing protocol. The integrated construct and flanking genomic regions were amplified from genomic DNA (gDNA) obtained from the C6_2 cell line and was prepared for Illumina Miseq DNA-sequencing using the Nextera XT v2 set A kit (Illumina) according to the manufacturer's instructions. Sequencing was performed in a pooled run (2 x 150 bp paired-end). The resulting reads were mapped to a draft reference sequence encompassing

the inserted heterologous plasmid sequence using CLC Genomics Workbench (version 8.5). Analysis of broken mapped reads was used to manually reconstruct the sequence junctions and thereby the anticipated sequence of the integrated construct and genomic context. Sites T2 and T9 were predicted based on transcriptomic data and model predictions. Briefly, regions between genes with consistently high expression profiles in a previously published RNA-Seq dataset from CHO-S during exponential and stationary phase (31) were identified. A region between two highly expressed genes (site T2) and between two highly expressed genes, one of which (*Rrm1*) is essential (site T9), were chosen as integration sites, with the assumption that the surrounding chromatin would remain open, and so it would be less likely that transgenes integrated at those sites would be silenced. The assumption of essentiality was based on literature (32) and *in silico* predicted lethality when simulating a *Rrm1* knockout in the genome-scale model of CHO cell metabolism (31). We utilized flux balance analysis (33) to predict essential genes. For this, model simulations used the default uptake and secretion constraints provided previously (31). To account for the possibility that the cells could change which metabolites are being catabolized or secreted, we did not force secretion or uptake of any metabolites in order to exclude false positive predictions of essentiality. The percentile ranks from expression levels of T2 and T9 neighboring genes across the 40 RNA-Seq samples are provided in the Supplementary Figure S1. Integration regions and the accompanying chromosomal and epigenetic information for all three sites, obtained from the new Chinese hamster reference genome (34) and the CHO-Epigenome database (35), respectively, are available in the Supplementary Table S1.

Plasmid design and construction

The CHO codon optimized Cas9 and sgRNA expression vectors with the same design as previously described (36) were used. sgRNA target sequences (available in Supplementary Table S2) with the shortest possible distance to the identified genome-plasmid breakpoints (site A) or to the identified high-expressing locus (site T2 and T9) were designed either using the CRISPy tool (36) (site A) or manually (site T2 and T9). 5' and 3' homology arms were selected to flank each sgRNA target sequence with the length of 750 bp each, and were amplified from CHO-S genomic DNA. The sequences of both homology arms for each site are provided in the Supplementary Table S2. Donor plasmids with the landing pad and donor recombinase-mediated cassette exchange (RMCE) plasmids were constructed via the uracil-specific excision reagent (USER) cloning method as previously described (21) with a minor adjustment. Donor landing pad plasmids were designed with mCherry coding sequence together with EF1α promoter being flanked by *loxP* at the 5' end and *lox2272* at the

3'end (Figure 1a). Consequently, donor RMCE plasmids contained both the GOI and promoter of interest in-between *loxP* at the 5' end and *lox2272* at the 3' end. Plasmids used to PCR amplify expression cassettes of mCherry, eGFP and CHO codon-optimized erythropoietin (EPO) have previously been described (21, 37, 38). Cre recombinase was expressed from the commercial PSF-CMV-CRE recombinase expression vector (Sigma Aldrich; OGS591). The weak and strong Kozak sequences used here have previously been described (38). All the plasmids used in the present study are depicted in the Supplementary Figure S2 and all the primers used are in the Supplementary Table S3. The sequences of the GOIs, promoters, terminators and Kozaks used in this study are listed in the Supplementary Table S4. Assembled PCR fragments were transformed into *E. coli* Mach1 competent cells (Life Technologies). All constructs were verified by sequencing of the expression cassettes (Eurofins Scientific) and purified using NucleoBond Xtra Midi EF (Macherey-Nagel) according to manufacturer's instructions.

Cell cultivation

CHO-S cells (Thermo Fisher Scientific) were maintained in CD CHO medium (Gibco) supplemented with 8mM L-glutamine (Thermo Fisher Scientific) and 2 μL/mL anti-clumping agent (Life Technologies) and cultivated in 125 mL Erlenmeyer flasks. Cells were incubated at 37°C, 5% CO_2 at 120 rpm (25mm shaking amplitude) and passaged every 2-3 days, unless otherwise stated. Viable cell density (VCD) and viability were monitored using the NucleoCounter NC-200 Cell Counter (ChemoMetec, Denmark), using Via1-Cassettes and the "Viability and Cell Count Method 2" assay.

Platform cell line generation using CRISPR/Cas9

CHO-S cells were transfected with the sgRNA plasmid and landing pad donor plasmid (EF1α-mCherry) for the corresponding integration sites, together with a Cas9 plasmid, according to a previously described protocol (21). Stable cell pools were established with G418 selection (500 μg/mL) as previously described (21). To establish monoclonal platform cell lines (PCLs), G418-selected pools were subjected to single-cell cloning using BD FACSJazz cell sorter (BD Biosciences). Double ZsGreen1-DR negative/mCherry positive single cells (*i.e.* cells with targeted integration only) were seeded in flat-bottom 384-well plates (Corning, Sigma Aldrich) with 30 μL of CD CHO medium supplemented with 8 mM L-Glutamine, 1.5% HEPES (Gibco) and 1x Antibiotic-Antimycotic (Gibco). Two weeks after the sorting, surviving single-cell derived colonies were transferred to flat-bottom 96-well plates (Corning) with 200 μL of CD CHO medium supplemented with 8 mM L-Glutamine and 1x Antibiotic-Antimycotic using an epMotion 5070 liquid handler (Eppendorf).

Subsequently, cells were expanded in suspension and verified by mCherry fluorescent analysis, 5' and 3' junction PCR and copy number analysis by qPCR.

Fluorescent level analysis

The generated mCherry PCLs were analyzed using an Xcyto 10 image cytometer (Chemometec), applying unstained suspension cells on the slide and examining mCherry fluorescence intensity in 20,000 - 40,000 individual cells of the population. Only PCLs with homogenous mCherry expression (≥ 98% of mCherry positive cells) were selected.

Confirmation of landing pad integration

For 5' and 3' junction PCRs, crude gDNA was extracted from the cell pellet using QuickExtract™ DNA extraction solution (Epicentre, Illumina) according to the manufacturer's instructions. 1-2 µl of gDNA mixture was used as template for touchdown PCR using 2x Phusion Master Mix (Thermo Fisher Scientific) and the following conditions: 98°C for 30 s; 98°C for 10 s; 68-58°C [-1°/cycle] for 30 s; 72°C for 2 min; 30 x: 98°C for 10 s; 58°C for 30 s; 72°C for 2 min; 72°C for 10 min. PCR primers for 5'/3' junction PCRs are listed in the Supplementary Table S3 and were designed as shown previously (5). PCR products were visualized on 1% agarose gels.

Copy number analysis using qPCR

qPCR was conducted to determine the relative copy numbers of the integrated mCherry gene and the Neomycin selection marker (NeoR). GeneJET Genomic DNA Purification Kit (Thermo Fisher Scientific) was used for extraction of gDNA according to manufacturer's instruction. qPCR was run on a QuantStudio 5 Real-Time PCR System (Agilent Technologies). Amplification was performed under the following conditions: 50°C for 2 min, 95°C for 10 min; 40x: 95°C for 15 s, 60°C for 1 min. Copy numbers of mCherry and NeoR were determined with *C1GALT1C1* (*COSMC*) as an internal control gene for normalization (39) using TaqMan™ Gene Expression Master Mix and custom-made TaqMan™ probes for all three genes (Thermo Fisher Scientific). Primers and probes are listed in the Supplementary Table S5 and were validated by melting curve analysis and primer efficiency test. A delta-delta threshold cycle (ΔΔCt) method was applied to calculate the copy number of mCherry/NeoR integrated using two previously generated PCLs with a single-copy of mCherry landing pad integrated within *COSMC* gene and T2 site (single-copy calibrators) (5, 38). Each experiment included no template control and negative control (CHO-S wt as a

template) with technical triplicates. Only PCLs with a single copy of both, mCherry and NeoR were expanded and banked.

Batch cultivation with single-copy PCL clones

Cells were seeded at density 1 x 10^5 cells/mL in 40 mL volume of CD CHO medium supplemented with 8 mM L-Glutamine and 2 µl/mL anti-clumping agent in 125 mL flasks. VCD and viability were measured daily, as described above. 2 x 10^6 cells were harvested on day three in the mid to late exponential phase by centrifugation (200 g, 5 min, RT). Cell pellets were stored at -80°C. On day three, mCherry fluorescent analysis was performed using Xcyto 10 image cytometer, as described above. Cultures were discontinued when viability dropped below 60%.

Subpopulation generation by recombinase-mediated cassette exchange (RMCE) and bulk sorting

mCherry PCL clones A-D1, T2-C9 and T9-G4 were randomly selected from the panel of single-copy monoclonal PCLs for site A, T2 and T9, respectively. These three clones were used for recombination experiments, unless otherwise stated. PCLs with a VCD of 1 x 10^6 cells/mL were transfected with an RMCE donor (with desired 5' proximal regulatory element-GOI) and a Cre recombinase vector, as previously described (21). After transfection, cells were passaged for two weeks in order to minimize transient transgene expression, followed by bulk sorting of the recombined cell population. When RMCE donor encoded eGFP, double mCherry negative/eGFP positive cells were considered recombined and were bulk-sorted. When RMCE donor encoded EPO, mCherry negative cells were considered recombined and were bulk-sorted. Cells (0.5-1 x 10^5 cells/well) were sorted in flat-bottom 96-well plates using a BD FACSJazz cell sorter. Non-recombined PCLs were used for gating the mCherry negative cells. Bulk-sorted cells were expanded in suspension (to 125 mL flasks) and further used either in a long-term cultivation experiment (eGFP populations) or short-term productivity assay (EPO populations). Independent set of experiments comprising recombination, bulk sorting and long-term cultivation were performed twice in the same original PCL (*i.e.* for A-CP-eGFP, A-SV40-eGFP and T2-CP-eGFP) accounting for two independent replicates for these three conditions.

Copy number analysis on bulk-sorted populations using dPCR

To ensure that bulk-sorted populations on average harbor a single copy of the recombined GOI (eGFP or EPO), relative copy number analysis with digital PCR (dPCR) was performed

using a QuantStudio™ 3D Digital PCR System (Applied Biosystems). Copy numbers of eGFP and EPO were determined using *COSMC* as an internal control gene for normalization, with custom-made TaqMan™ probes used for all three genes (Thermo Fisher Scientific). Primers and probes are listed in the Supplementary Table S5 and were validated by melting curve analysis and primer efficiency test. gDNA of bulk-sorted populations, isolated with GeneJET Genomic DNA Purification Kit, together with the QuantStudio™ 3D Digital PCR Master Mix v2 was loaded onto QuantStudio™ 3D PCR Chips. Amplification was performed under the following conditions: 96°C for 10 min; 39x: 60°C for 2 min, 98°C for 30 s; 60°C for 2 min. A previously generated monoclonal PCL with a single copy of EPO integrated within T9 site was used as a single-copy EPO control (21). Additionally, a titration experiment was performed to ensure the reliability of the obtained results. gDNA of a bulk-sorted sample (A-CP-strong kozak-eGFP) was spiked into CHO-S wt gDNA comprising 0%, 10%, 20%, 50%, and 80% of the total gDNA. The corresponding target-to-reference ratio was subsequently determined for each titration mixture. Each experiment included a no template control. The final analysis of the target-to-reference ratio was performed with the QuantStudio™ 3D Analysis Suite Cloud Software.

Relative mRNA expression level using RT-qPCR

Total RNA was extracted from 1-2 x 10^6 cells using the RNeasy Plus Mini Kit (Qiagen) according to the manufacturer's instructions. RNA concentration was measured with Qubit fluorometric analysis (Life technologies) and the purity/quality was assessed with Nanodrop (Thermo Fisher Scientific) and 1% agarose gel visualization. cDNA was synthesized from 1.5-3 µg of total RNA using the Maxima First Strand cDNA Synthesis Kit for RT-qPCR with dsDNAse treatment (ThermoFisher Scientific). RT-qPCR analyses were performed on the QuantStudio 5 Real-Time PCR System using TaqMan™ Multiplex Master Mix (Thermo Fisher Scientific) in triplexes (GOI and two normalization genes) using the following amplification conditions: 50°C for 2 min, 95°C for 10 min; 40x: 95°C for 15 s, 60°C for 1 min. Custom-made Taqman assays were used for mCherry, eGFP and EPO, as well as normalization genes *Gnb1* and *Fkbp1a* (40). All the primers and probes are listed in the Supplementary Table S5 and were validated by melting curve analysis and primer efficiency test. Using the ΔΔCt method, the relative expression levels of GOIs were calculated by normalization to the geometric mean of expression levels of the two normalization genes. Each experiment included controls with no template and was performed using technical triplicates.

Long-term cultivation

eGFP bulk-sorted populations were recovered after cell sorting and expanded in suspension (6-9 passages), and transferred to 6-well plates (Greiner) after which long-term cultivations were initiated in complete media containing 1x Antibiotic-Antimycotic. Cells were passaged every 2-3 days and kept in culture for two months. Cells from passage 2 (week 1), passage 11 (week 4) and passage 23 (week 8), as counted from the start of the long-term cultivation, were analyzed for eGFP fluorescence levels using FACS and relative eGFP mRNA expression levels using RT-qPCR. The BD FACSJazz cell sorter was calibrated each time using Rainbow Calibration Particles (8 peaks) 3.0 - 3.4 µm (BD Biosciences), followed by the use of constant photomultiplier settings for eGFP and mCherry detection. A total of 100,000-200,000 events were collected in the analysis. Data analysis was performed with FACSDiva Software (BD Biosciences) and FlowJo 10 (FlowJo LLC). Median fluorescence intensities (MFIs) of bulk-sorted populations at each time point were normalized to MFI of the T2-EF1α sample for the ease of representation. For the analysis of eGFP mRNA expression levels, 1-2 x 10^6 cells/mL cells at late exponential phase were centrifuged (200*g*, 5 min, RT) and cell pellets were stored at -80°C. RNA samples were prepared as described above. eGFP mRNA expression levels of bulk-sorted populations at each time point were normalized to the levels of the T2-EF1α sample at the first time point (week 1; internal calibrator).

Short-term productivity assay and titer measurements

A three day batch culture was performed to determine specific productivity of EPO in mid-exponential phase for different bulk-sorted populations. Cells were seeded at density 5 x 10^5 cells/mL in 25 mL volume of CD CHO medium supplemented with 8 mM L-Glutamine and 2 µl/mL anti-clumping agent in 125 mL flasks. VCD and viability were measured daily, as described above. On day three, EPO titers were determined in supernatants by bio-layer interferometry (ForteBio, Pall, Menlo Park, CA) using Streptavidin biosensors (Fortebio 18-5021, Pall), functionalized with the anti-EPO V_HH biotin conjugate (Life Technologies, cat.no: 7103372100), as previously described (37). Absolute EPO titers were calculated using a 7-point calibration curve generated from a dilution series of commercially available EPO (Genscript, Piscat-away, NJ, USA). Specific productivity of EPO was calculated using titer and integral viable cell density (IVCD) values as described elsewhere (41). On day three 2 x 10^6 cells were harvested for mRNA analysis by centrifugation at 200*g* for 5 min at RT and cell pellets were stored at -80°C, and mCherry fluorescence analysis was performed using Xcyto 10 image cytometer, as described above. EPO mRNA expression levels of bulk-sorted populations were normalized to the levels of the T2-EF1α sample (internal calibrator).

Statistical analyses

All statistical analyses were performed using GraphPad Prism 7 software (GraphPad Prism Software Inc., La Jolla, CA). Statistical significance of log2-transformed mCherry mRNA level differences between the integration sites was calculated using an ordinary one-way analysis of variance (ANOVA) with Turkey's multiple comparisons test (n = 5 clones/site) where *p < 0.05, **p < 0.01, ***p < 0.001, and ****p < 0.0001. To analyse the correlation between the two variables, linear regression was used and R^2 together with p value reported.

Results

Carefully studied integration sites constitute preferable genome safe harbors for recombinant gene expression. Therefore, this study set out to identify putative safe harbors in CHO-S cells to use them for systematic characterization of their capacity for stable recombinant gene expression and thereby provide them to the mammalian cell engineering community. Next, to systematically assess recombinant gene expression patterns in these newly defined integration sites, we developed a modular CRISPR/Cas9-based toolbox. Specifically, this toolbox was designed to inspect the interactions between the four layers influencing the final level of recombinant gene expression: cell host, the integration site, the vector construct and the type of recombinant protein. We inspected 66 randomly-selected combinations between the four layers, and named the samples after all the integrated components from each layer.

Identification of putative safe harbor integration sites in the CHO-S genome

To increase the number of well-characterized genomic loci for stable and predictable recombinant expression, we chose a combined strategy of identifying sites; the sites were identified based on either experimental expression stability or *in silico* approach. Site A derives from a low-copy, relatively high-producing cell line with randomly integrated Rituximab, generated using our in-house developed clone screening platform (29). This cell line exhibited stable and relatively high Rituximab productivity over two months of culturing during which no gain or loss in Rituximab gene copy number was observed (Supplementary Figure S3). Proprietary TLA technology was used to identify the genomic loci of integration in this cell line (30), unveiling site A. Full integration of intact Rituximab expression cassette in site A was independently confirmed with targeted MiSeq deep-sequencing (Supplementary Figure S4). The results here validated that the integration event did not disrupt the host genome in the proximity of integration (*i.e.* chromosomal rearrangement)

and therefore this locus can be used for targeted reverse engineering. On the other hand, sites T2 and T9 were selected based on transcriptomic data of neighboring genes and model predictions and were previously used as integration sites in our studies (21, 38). By analyzing published RNA-seq datasets from CHO-S cell line (31) and identifying regions with consistently high expression profiles across 40 samples, sites T2 and T9 were chosen as integration site candidates. We hypothesized that these two sites would be good integration sites owing to consistent chromatin accessibility for transcription. Site T2 is positioned in-between two highly expressed genes and site T9 is in proximity of a highly expressed gene *Rrm1*, that is metabolically essential for DNA synthesis (32), as additionally confirmed by the genome-scale CHO metabolic model (35). The percentile ranks from expression levels of T2 and T9 neighboring genes across the 40 RNA-Seq samples are provided in the Supplementary Figure S1, demonstrating that T2 and T9 neighboring genes are consistently among the highest expressed genes in this cell line. The integration loci information is available in the Supplementary Table S1, together with the corresponding Chinese hamster-derived chromosome numbers and the surrounding epigenetic marks. The features of these three integration sites can in the future help towards defining the general characteristics of safe harbors in CHO cells.

Modular toolbox for generation of mCherry-encoding platform cell lines with a poly(A) trap-based RMCE landing pad

Having identified three sites of interest, a modular CRISPR/Cas9-based toolbox was developed allowing for systematic quantification of recombinant gene expression levels in these sites modulated by different 5' proximal regulatory elements (**Figure 1A**). In the first part of the toolbox, platform cell lines (PCLs) with landing pads flanked by unique *lox* sites are generated for each integration site of interest via HDR and G418 selection, as previously described (21). The PCLs are designed with EF1α-mCherry inside the landing pad and poly(A) outside of the landing pad (poly(A) trapping). By employing a panel of parental clones for each integration site, a number of individual PCLs are generated. This allows for investigating the effects on gene expression originating from the cell host and the integration site, respectively. In the second part of the toolbox, generated PCLs with the poly(A) trap can be subjected to RMCE to replace EF1α-mCherry with desired 5' proximal regulatory elements followed by the gene of interest (GOI). The RMCE facilitates fast integration of the desired expression cassettes into the same PCL, allowing for investigation of the effects originating from the vector elements and type of recombinant protein integrated.

In conclusion, this platform was designed to interrogate site-specific expression levels of any recombinant protein with any 5' proximal regulatory element in a controlled and streamlined manner. Moreover, due to its design, it can be used for direct comparison and benchmarking of different expression contexts and different integration sites of interest.

Validation of generated mCherry-encoding platform cell lines

Following the workflow of the toolbox, we generated PCLs with mCherry integrated in site A, T2 and T9. To account for clonal variation originating from the heterogeneous polyclonal pool, five independent PCLs were generated for each of the three sites. 5' and 3' junction PCRs were performed for confirmation of specific integration into the three integration loci (Supplementary Figure S5). The frequency of clones with correctly targeted integration events post-G418 selection ranged from 21% for site A and 25% for site T2 to 30% for site T9. Following junction PCRs, the PCLs were confirmed to harbor a single copy of the landing pad expression cassette by copy number analysis using qPCR (Supplementary Figure S6). Furthermore, ≥ 98% of the total cell population was mCherry positive for all 15 PCLs, indicating that the mCherry expression is relatively homogeneous, despite the generated PCLs being in different growth phases (Supplementary Figure S7).

Using the established CRISPR/Cas9-based platform, we successfully generated a panel of PCLs with the mCherry expression cassette in the three selected integration sites. These PCLs were selected as cell line models for studying both clonal variation and interactions between integration site and vector regulatory elements for different recombinant proteins.

Minimal clonal variation at defined integration sites

To characterize the effect of the cell host background on site-specific gene expression levels, the 15 generated PCLs (5 PCLs per integration site) were analyzed in batch cultures. Since the expression cassette in the landing pad was identical in all clones (EF1α-mCherry), we were able to examine the two factors contributing to the differences in expression levels: clonal background and the site of integration.

The five PCLs for each integration site demonstrated different growth profiles (see Supplementary Figure S8), probably because they were derived from a genetically heterogeneous cell pool of host cells. Despite different growth profiles, mCherry median fluorescence intensities (MFIs) in the mid exponential phase were uniform between the PCLs of one site as evaluated by the small coefficient of variation (CV) (site A: CV = 10.9%; site T2: CV = 10.9%; site T9: CV = 4.1%) (**Figure 1B**). When comparing the CVs in mCherry MFI within each PCL, we observed much larger values ranging between 42% and 70%. Large CVs

observed in mCherry MFI within the clone indicate the presence of clone-specific fluctuations in mCherry expression on account of transcriptional and translational bursts. However, we observed no differences in clone-specific fluctuations between the three sites, as evaluated by the mean CV of each site. Furthermore, mCherry mRNA expression levels correlated with the MFIs, suggesting that mCherry fluorescence can be used as a proxy for transcription and expression levels (R^2 = 0.62, p = 0.0005) (**Figure 1C**). The variation in mCherry mRNA expression levels observed between the five independent PCLs of each site was relatively small, as seen with the MFI values (site A: CV = 24%; site T2: CV = 15%; site T9: CV = 13%). When comparing mCherry mRNA expression levels between the integration sites, we observed significant differences in how the EF1α promoter responds to the genetic context (p = 0.0015 for A vs T2 and p = 0.0059 for T2 vs T9). Sites A and T9 yielded approximately 1.5-fold higher mCherry expression levels than site T2 (Figure 1B and 1C), indicating the local chromatin environment and possible neighboring enhancers of site A and T9 provide an increased transcriptional output from EF1α promoter.

Overall, these results demonstrated that when controlled, targeted integration into putative safe harbor loci is employed, the generated clones display minimal variation in recombinant gene expression levels, despite having different growth profiles. Moreover, it was reconfirmed that the promoter activity is governed by its spatial positioning (*i.e.* position effect) within a defined genomic landscape. Based on these results, one representative PCL for each integration site was randomly selected from the generated panel for the following recombination experiments.

Proof-of-principle recombination into poly(A) trap elicited tunable eGFP expression patterns

To compare expression from different vector elements in different integration sites, we needed to confirm that EF1α-mCherry can be successfully recombined with another 5' proximal regulatory element-GOI and can be functionally transcribed from the poly(A) trap. To this end, the EF1α-mCherry cassettes in the landing pads of the three representative PCLs were exchanged with four different eGFP constructs. The eGFP constructs were designed with either a composite promoter (CP; comprised of mCMV enhancer, hEF1α and 5'UTR HTLV) or the viral SV40 promoter together with a strong or weak Kozak sequence (38).

One week post-recombination, double mCherry negative/eGFP positive populations were observed, indicating successful recombination and eGFP transcriptional output from the poly(A) trap (**Figure 2A**), thereby confirming the functionality of our platform. When comparing MFIs of the recombined cells, an approximately 13-fold difference was observed

between strong and weak Kozak sequences across all three sites and both promoters (Figure 2A). However, the same constructs recombined in different sites yielded approximately 2-fold higher eGFP fluorescence levels in site A and T9 in comparison to T2, for both promoters with either a weak or strong Kozak sequence. This confirmed the trend observed with the 15 PCLs where site A and T9 gave rise to increased mCherry expression levels with the EF1α promoter.

Next, we wanted to investigate whether the observed double mCherry negative/eGFP positive populations were a result of specific recombination of mCherry with a single copy of eGFP. Only then, these double mCherry negative/eGFP positive populations could be used for analyzing the effect of integration sites and vector constructs on expression levels. Hence, we bulk-sorted five double mCherry negative/eGFP positive populations with strong Kozak sequences and conducted copy number analysis using digital PCR (dPCR). All bulk-sorted populations displayed a 1:1 target-to-reference ratio, confirming that on average these populations harbored a single copy of eGFP (**Figure 2B**). A 1:1 ratio was also found in a confirmed single-copy monoclonal RMCE cell line producing EPO, generated in our previous study (21) (Supplementary Figure S9). Furthermore, a control titration experiment was performed where incrementally increasing amounts of gDNA from A-CP-strong Kozak-eGFP bulk-sorted population were spiked into CHO-S wt gDNA. A linear relationship was confirmed between the eGFP spike-ins and the target-to-reference ratio (R^2 = 0.98, p = 0.0002)(Supplementary Figure S10).

Overall, our proof-of-principle experiment confirmed three important objectives, needed for the following systematic evaluation of site-specific expression patterns. First, poly(A) trapping enabled functional output post-recombination. Second, recombined populations were sufficiently homogenous (*i.e.* single copy) to be reliably used for comparative analysis without the need to establish monoclonal cell lines. Third, different 5' proximal regulatory elements (promoters and Kozak sequences) elicited tunable and context-dependent gene expression patterns demonstrating that stable recombinant gene expression is both construct-dependent and site-dependent.

Observation of site-dependent and construct-dependent eGFP expression patterns

Following the successful proof-of-principle experiment, we set out to conduct a systematic evaluation of site-specific gene expression patterns with all 5' proximal regulatory elements in our panel. We therefore generated eGFP constructs with four different promoters - two viral promoters (CMV and SV40), a constitutive promoter (EF1α) and a composite promoter

(CP), all followed by a strong Kozak sequence. By recombining them into the three representative PCLs, we were able to quantify the combined effect of these promoters and the three identified genomic loci on the expression levels from stably integrated eGFP.

Two weeks post-recombination, the recombined population (*i.e.* mCherry negative/eGFP positive population) was observed in all 12 samples, while samples were largely devoid of the transient eGFP expression (*i.e.* mCherry positive/eGFP positive population) (**Figure 3A**). The percentage of successfully recombined cells varied between 2% to 8%. By comparing the MFI of these recombined populations, the promoter giving rise to the highest expression level was CP in all three sites, followed by the EF1α and SV40 promoters. For all three promoters, sites A and T9 generated approximately a 2-fold higher output than site T2, consistent with the results seen in the analysis of the mCherry expression levels in 15 PCLs (Figure 1C). While CP and SV40 promoter demonstrated higher cell-to-cell variation (wider distribution of the fluorescence levels, *i.e.* CV), constitutive EF1α promoter displayed lower intrinsic noise with a relatively tight and unimodal expression pattern (site A: CV = 37.1%; site T2: CV = 38.8%; site T9: CV = 33.6%). Surprisingly, CMV, generally thought of as one of the strongest exogenous promoters (42, 43), generated the lowest levels of fluorescence in all three sites. While CMV in combination with site T2 generated the strongest MFI when compared to the other two sites, this promoter performed particularly poorly in site T9 with a 30-fold lower MFI compared to CP. Overall, CMV generated the highest cell-to-cell variation in all three sites (site A: CV = 69.5%; site T2: CV = 63.7%; site T9: CV = 108%), indicating that the variation primarily originates from the promoter itself and not the integration site.

To confirm the results seen in the initial fluorescence analysis, where the cell population consisted of a small portion of the recombined cells and a large pool of non-recombined (mCherry positive) cells (Figure 3A), bulk-sorting of the recombined populations was performed. After recovery, copy number analysis was conducted, confirming the presence of a single eGFP copy in all recombined populations (Supplementary Figure S11). Next, eGFP expression levels were determined at the mRNA and protein level (**Figure 3B**). A clear agreement was observed between the fluorescence measurements (Figure 3A and 3B), confirming that the initial MFI measurements prior to bulk sorting give a good proxy for the eGFP expression level. Consistent with the analysis of mCherry expression levels (Figure 1C), a significant linear correlation was observed between eGFP fluorescence and mRNA levels (Figure 3B, $R^2 = 0.59$; $p = 0.0008$). Despite normalizing the data, the discrepancies that we observe between generally lower mRNA levels and higher corresponding protein levels might be on account of slow degradation rates of long-lived proteins such as eGFP that can limit the

ability to track rapid mRNA fluctuations. Both the analyses (at mRNA and protein level) confirmed the strength of the promoters in the three sites, observed at the time of bulk sorting (Figure 3A). Moreover, similar gene expression patterns were generated when RMCE and bulk-sorting were repeated (independent replicates named A-CP_Rep2, T2-CP_Rep2, A-SV40_Rep2), indicating that data generated by the poly(A) trap platform is reproducible and robust.

In conclusion, these results facilitated ranking of the promoter strength in a site-specific manner. A wide range of site- and promoter-specific expression levels was observed whereby elements have been identified for predictable and tunable gene expression patterns. Finally, it has been shown that commonly used promoters, such as EF1α and CMV, function in a position-dependent mechanism.

Similar clone-specific eGFP expression patterns across all three integration sites

To explore if the findings mentioned above are only specific to the particular PCL used, RMCE was performed with the same set of eGFP constructs into two additional PCLs per site. FACS analysis was performed two weeks post-recombination and the generated MFIs were closely following the trend obtained with the original three PCLs (**Figure 3C**). These results further supported our observations with the mCherry-encoding PCLs, that gene expression patterns are prevalently site- and promoter-specific when controlled targeted integration is performed. The cell host background, despite additional round of RMCE-based generation of a subpopulation, did not seem to contribute substantially to the differences in gene expression levels in the three selected putative safe harbor sites.

Long-term stability of eGFP expression achieved in all three integration sites

To explore the long-term stability of recombinant gene expression under the selected set of promoters when integrated in site A, T2 and T9, a two-month cultivation experiment of eGFP-encoding bulk-sorted populations was performed. eGFP fluorescence and mRNA levels were measured in the beginning (week 1), in the middle (week 4) and at the end (week 8) of the cultivation. None of the bulk-sorted populations displayed a demonstrable loss of the eGFP expression during 8 weeks of culturing (**Table 1**), retaining the initial values observed in Figure 3B.

These results indicate that the three integration sites characterized in this study provide robust gene expression patterns over long-term culturing. Therefore, these sites seem to

constitute safe harbors in the CHO-S genome for integration of transgenes, with varying strength of expression levels dependent on the construct integrated.

Site-specific gene expression characterization for a secreted biotherapeutic

To demonstrate the usefulness of our toolbox for recombinant protein production in mammalian cell lines, we aimed to characterize the expression levels of a secreted biopharmaceutical, EPO. The RMCE was performed in the three representative PCLs with EPO constructs comprising three different promoters (CMV, EF1α and CP) and the previously described strong Kozak sequence. Two weeks post-recombination, bulk sorting was performed on mCherry negative cells for all 9 conditions (**Figure 4A**). The mCherry negative cells were deemed as correctly recombined cells expressing EPO and these populations constituted 1% to 5% of all cells. Bulk-sorted cells were recovered and using dPCR analysis the presence of a single copy of EPO was confirmed in these populations (Supplementary Figure S9). This was followed by a three-day batch culture in 125 mL shake flasks where specific EPO productivity and mRNA expression levels were measured in the mid-exponential phase of the culture. Specific EPO productivity and EPO mRNA levels displayed a high level of correlation (**Figure 4B;** $R^2 = 0.98$). The ranking of the promoters for the expression of the secreted EPO was similar to the intracellular eGFP. The strongest promoter in all integration sites was CP, reaching 3 pcd/EPO copy in sites T9 and A. The EF1α promoter construct generated a 3-fold lower output and approximately the same amount of protein in all three sites (1 pcd/EPO copy). Interestingly, the CMV promoter gave rise to higher EPO mRNA levels and specific productivity values in sites A and T2 compared to the levels provided in these sites for eGFP, but remained the weakest promoter in site T9 (7-fold lower pcd than T9-CP-EPO).

These results demonstrate that our toolbox can be successfully used to modulate the expression levels of a secreted biotherapeutic. Furthermore, we were able to show that promoters can generate protein-specific gene expression patterns at characterized genomic loci, advocating for testing of the desired vector constructs for both, the site of interest and the protein of interest.

Discussion

Tight control of recombinant gene expression is a major quest in different fields of biology, ranging from gene therapy, functional genetics studies in basic research, biopharmaceutical production to synthetic biology. In order to adapt to ever changing environmental stimuli, eukaryotic cells have complex systems in place for dynamic and flexible regulation of their

transcriptional and translational processes. Thus, heterologous gene networks with predictable and desirable outputs are difficult to install in eukaryotic cells when random integration is applied. Molecular tools for site-specific integration are therefore crucial to move this field forward; however, increased specificity and efficiency of these tools have to be achieved for routine applications (44). We and others have developed CRISPR/Cas9-based targeted integration protocols for *in vitro* gene editing in mammalian cells (5, 10–12, 19–21). In this study we wanted to further expand the CRISPR/Cas9 applications by developing a CRISPR/Cas9-based modular toolbox for systematic evaluation of site-specific recombinant expression patterns modulated by different 5' proximal regulatory elements. Here, we describe a systematic and streamlined analysis of stable gene expression patterns resulting from the interplay between the four contributing layers: the cell host, the integration locus, the inserted vector elements and the type of recombinant protein being expressed. By precisely quantifying the resulting expression profile, the objective of this toolbox is to pave the way toward generation of mammalian cell lines with fine-tuned and user-defined gene expression patterns.

First, we characterized three candidate safe harbor integration sites in the CHO-S genome, one in an intragenic, specifically intronic region (site A), and two in intergenic regions (site T2 and T9)(Supplementary Table S1) with the objective of further expanding the growing repository of available sites for programmable mammalian cell engineering (12, 20, 24). These sites were selected either on the basis of experimental data with sites showing stable expression or *in silico* analysis identifying sites with high and stable expression profiles of neighboring genes. All three sites identified here are positioned in open chromatin areas that are generally transcriptionally active (Supplementary Table S1). High expression levels of transgenes when integrated into regions with high transcriptional activity and chromatin accessibility were recently reported (45). Being in transcriptionally active regions of the genome, these newly proposed regions are possibly less noisy, exhibiting lower cell-to-cell variability in gene expression patterns. Our study indeed unveiled that all three sites generated relatively robust gene expression patterns with minimal clonal variation and approximately 2-fold higher transcriptional output in sites A and T9 in comparison to T2. We have previously shown that minimal clonal variation can be achieved when targeted integration is applied (5, 21). Moreover, since the chromosome regions surrounding the integration sites seem to support relatively stable long-term transgene expression, the regions around site A, T2 and T9 might be less prone to genomic rearrangements from transcriptionally active regions of euchromatin into closed heterochromatin. We therefore believe that these three sites can be considered safe harbors for programmable transgene

expression in CHO-S cells. However, further studies are warranted to confirm whether these sites give rise to gene expression patterns as robust as seen in other established safe harbors, such as ROSA26.

Second, we developed a toolbox for generation of PCLs using a novel approach with poly(A) trap-based RMCE landing pads. Poly(A) traps have been previously used, however, in a different setting for gene trapping purposes (46). The toolbox allows users to generate PCLs with their own integration site of interest and use them for rapid evaluation of site-specific gene expression patterns resulting from any desired 5' proximal regulatory elements. After PCLs with the sites of interest are established, the timeline to optimize site-specific expression profiles of any given recombinant protein is two to four weeks from transfection (as measured by fluorescence or protein levels, respectively). This brings the timeline of expression optimization closer to that of transient expression-based experiments.

Third, employing the developed toolbox, we analyzed site-specific expression profiles of genes under transcriptional control of some of the commonly used promoters in mammalian cell engineering. Using eGFP and EPO expression cassettes, we observed that the gene expression is heavily influenced by the interplay between the integration site and the promoter of interest, generating up to a 30-fold dynamic range in protein expression. All the promoters tested here were dependent on the site of integration featuring highly genetic context-specific transcriptional regulation. Viral promoters, such as CMV, have previously been shown to be prone to the local epigenetic regulation and silencing (47–51). Our study showed that the otherwise potent CMV promoter gave rise to the overall lowest and the most stochastic transgene expression levels of the four promoters studied here. As the transgene copy number is low in our studied populations (*i.e.* one copy), bursts in transcription/translation can lead to easily detectable differences between otherwise genetically identical cells (52). Therefore, the lower the amount of transcript (*e.g.* CMV at site T9), the higher the intrinsic noise, since low copy number together with low transcription rates can limit the precision of gene regulation (53). On the other hand, we observed that the SV40 viral promoter allowed for relatively high and stable transgene expression levels in our integration loci, suggesting the genomic context of these sites is favorable towards high transcription rates under the SV40 promoter. The SV40 promoter has previously been reported to be more resistant to transcriptional silencing than CMV (43, 54) and is usually included in vector systems to drive the selection marker expression (26). For the purpose of finding high-producing clones it is therefore of utmost importance to carefully devise the expression cassettes of both the GOI and the selection marker by considering the strength, but also general responsiveness of the promoter to the local

environment. The strongest promoters in our study were CP (composed of the mCMV enhancer, hEF1α promoter and 5'UTR HTLV) and EF1α promoter, making them desirable for the applications where high expression levels are needed. However, for multi-gene engineering strategies, their usage might be limited due to a few reasons. Firstly, both CP and EF1α are large promoters that would in turn generate very large multi-gene vectors that might not be efficiently transfected or integrated. Moreover, if either of these promoters would be used successively or if used together (high similarity between the two promoters) for multiple genes, there might be a higher risk of promoter interference when competing for the same endogenous transcription factors (55). This constitutes an important consideration for mammalian synthetic biology, where the more attractive solution might be the integration of a completely synthetic system (*i.e.* a pair of non-mammalian transcription factor and regulatory elements with its cognate binding sites) with minimal host cell interactions (56). All in all, these findings stress the importance of having a toolbox to empirically interrogate site-specific gene expression patterns modulated by a range of desired regulatory elements. The toolbox facilitates site- and application-specific design of expression cassettes, preventing the risk of unpredictable loss of promoter functionality, even from increasingly more utilized built-for-purpose synthetic promoters (55, 57, 58).

Numerous studies have been conducted to improve the levels of stable recombinant expression by optimization of expression constructs (42, 43, 59–65). However, the performance of different vector designs cannot be reliably compared between these studies, since the analyses were conducted: a) in different cell hosts; b) in randomly or semi-randomly integrated, stable cell pools or clones with high variability in gene copy numbers; and c) under antibiotic selection and in overall different culture conditions. Only a few studies so far have explored different vector constructs in a site-specific manner (14, 27, 28), confirming that the strength of the expression cassette is defined by its chromosomal context. However, to the best of our knowledge, no systematic evaluation has been performed wherein the performance of different 5' proximal regulatory elements were benchmarked against each other in a controlled manner. In our system, a single copy of 5' proximal regulatory element-GOI is integrated in a defined, retargeted genomic site. Further, comparisons are done without the presence of antibiotic selection, thereby considerably reducing cell-extrinsic factors causing dynamic transcriptional regulation (66). In fact, by aligning all three main sources of transcriptional regulation (cell-extrinsic, cell type-specific and genomic-loci intrinsic (67)), the contribution of any 5' proximal regulatory element to the final transgene expression level can be precisely determined.

Further studies are warranted to explore how the local genomic environment in the three presented sites interacts with the exogenous promoters, to decipher how the regulatory elements affect the transgene expression in these sites. Similarly, it remains to be explored whether any perturbations have been inflicted upon host cell transcriptome after targeted integration in the three putative safe harbors. Generally, transgene integration into a safe harbor should cause little to no transcriptional alterations to the endogenous genes and should not disrupt the host genome in the proximity of integration (*i.e.* genomic rearrangements). Moreover, the preferred promoters for predictable transgene expression should not be inducing changes in the host cell transcriptome by titrating away the endogenous transcription factors from endogenous genes (68) and should not be as prone to position effects and corresponding epigenetic mechanisms (55).

There are many potential implications of this toolbox, overall enabling identification of parameters determining robust gene expression patterns and fidelity over long-term culturing. By using this toolbox for mammalian cell line engineering, the inherent stochasticity of gene expression can be minimized and predictable transgene expression profiles can be achieved. Predictable and tunable transgene expression profiles are a requirement for programming more complex cell behaviors, such as expressing entire metabolic pathways with dynamic expression patterns and high temporal precision (69), building reliable synthetic circuits (70), or engineering the optimal effector gene expression levels (71). We employed this toolbox to increase the expression levels of a secreted biotherapeutic (EPO), demonstrating the applicability of our workflow for CHO cell line engineering. Similar approaches could be undertaken to optimize the expression context of other recombinant proteins in any integration site of interest across different expression platforms, including a variety of human cells and cell lines.

Acknowledgements

The authors thank Nachon Charanyanonda Petersen for her assistance with FACS, Mads Valdemar Anderson for designing the sgRNA for site A, Sara Petersen Bjørn for cloning the sgRNA plasmid for site A, Mojtaba Samoudi for the re-analysis of TLA sequencing, Daria Sergeeva for the introduction to digital PCR and Heena Dhiman for her assistance with acquiring the epigenetic data from the CHO-Epigenome database.

References

1. Wilson,C., Bellen,H.J. and Gehring,W.J. (1990) Position effects on eukaryotic gene expression. *Annu. Rev. Cell Biol.*, **6**, 679–714.
2. Daboussi,F., Zaslavskiy,M., Poirot,L., Loperfido,M., Gouble,A., Guyot,V., Leduc,S., Galetto,R.,

Grizot,S., Oficjalska,D., *et al.* (2012) Chromosomal context and epigenetic mechanisms control the efficacy of genome editing by rare-cutting designer endonucleases. *Nucleic Acids Res.*, **40**, 6367–6379.

3. Wurm,F.M. (2004) Production of recombinant protein therapeutics in cultivated mammalian cells. *Nat. Biotechnol.*, **22**, 1393.
4. Noh,S.M., Sathyamurthy,M. and Lee,G.M. (2013) Development of recombinant Chinese hamster ovary cell lines for therapeutic protein production. *Curr. Opin. Chem. Eng.*, **2**, 391–397.
5. Lee,J.S., Kallehauge,T.B., Pedersen,L.E. and Kildegaard,H.F. (2015) Site-specific integration in CHO cells mediated by CRISPR/Cas9 and homology-directed DNA repair pathway. *Sci. Rep.*, **5**, 8572.
6. Doudna,J.A. and Charpentier,E. (2014) Genome editing. The new frontier of genome engineering with CRISPR-Cas9. *Science*, **346**, 1258096.
7. Ran,F.A., Hsu,P.D., Wright,J., Agarwala,V., Scott,D.A. and Zhang,F. (2013) Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.*, **8**, 2281–2308.
8. Hsu,P.D., Lander,E.S. and Zhang,F. (2014) Development and applications of CRISPR-Cas9 for genome engineering. *Cell*, **157**, 1262–1278.
9. Bandaranayake,A.D. and Almo,S.C. (2014) Recent advances in mammalian protein production. *FEBS Lett.*, **588**, 253–260.
10. Lee,J.S., Grav,L.M., Pedersen,L.E., Lee,G.M. and Kildegaard,H.F. (2016) Accelerated homology-directed targeted integration of transgenes in Chinese hamster ovary cells via CRISPR/Cas9 and fluorescent enrichment. *Biotechnol. Bioeng.*, **113**, 2518–2523.
11. Scarcelli,J.J., Shang,T.Q., Iskra,T., Allen,M.J. and Zhang,L. (2017) Strategic deployment of CHO expression platforms to deliver Pfizer's Monoclonal Antibody Portfolio. *Biotechnol. Prog.*, **33**, 1463–1467.
12. Zhao,M., Wang,J., Luo,M., Luo,H., Zhao,M., Han,L., Zhang,M., Yang,H., Xie,Y., Jiang,H., *et al.* (2018) Rapid development of stable transgene CHO cell lines by CRISPR/Cas9-mediated site-specific integration into C12orf35. *Appl. Microbiol. Biotechnol.*, **102**, 6105–6117.
13. Kim,M.S. and Lee,G.M. (2008) Use of Flp-mediated cassette exchange in the development of a CHO cell line stably producing erythropoietin. *J. Microbiol. Biotechnol.*, **18**, 1342–1351.
14. Nehlsen,K., Schucht,R., da Gama-Norton,L., Krömer,W., Baer,A., Cayli,A., Hauser,H. and Wirth,D. (2009) Recombinant protein expression by targeting pre-selected chromosomal loci. *BMC Biotechnol.*, **9**, 100.
15. Zhou,H., Liu,Z.-G., Sun,Z.-W., Huang,Y. and Yu,W.-Y. (2010) Generation of stable cell lines by site-specific integration of transgenes into engineered Chinese hamster ovary strains using an FLP-FRT system. *J. Biotechnol.*, **147**, 122–129.
16. Mayrhofer,P., Kratzer,B., Sommeregger,W., Steinfellner,W., Reinhart,D., Mader,A., Turan,S., Qiao,J., Bode,J. and Kunert,R. (2014) Accurate comparison of antibody expression levels by reproducible transgene targeting in engineered recombination-competent CHO cells. *Appl. Microbiol. Biotechnol.*, **98**, 9723–9733.
17. Zhang,L., Inniss,M.C., Han,S., Moffat,M., Jones,H., Zhang,B., Cox,W.L., Rance,J.R. and Young,R.J. (2015) Recombinase-mediated cassette exchange (RMCE) for monoclonal antibody expression in the commercially relevant CHOK1SV cell line. *Biotechnol. Prog.*, **31**, 1645–1656.
18. Baumann,M., Gludovacz,E., Sealover,N., Bahr,S., George,H., Lin,N., Kayser,K. and Borth,N. (2017) Preselection of recombinant gene integration sites enabling high transcription rates in CHO cells using alternate start codons and recombinase mediated cassette exchange. *Biotechnol. Bioeng.*, **114**, 2616–2627.
19. Inniss,M.C., Bandara,K., Jusiak,B., Lu,T.K., Weiss,R., Wroblewska,L. and Zhang,L. (2017) A novel Bxb1 integrase RMCE system for high fidelity site-specific integration of mAb expression cassette in CHO Cells. *Biotechnol. Bioeng.*, **114**, 1837–1846.
20. Gaidukov,L., Wroblewska,L., Teague,B., Nelson,T., Zhang,X., Liu,Y., Jagtap,K., Mamo,S., Tseng,W.A., Lowe,A., *et al.* (2018) A multi-landing pad DNA integration platform for mammalian cell engineering. *Nucleic Acids Res.*, **46**, 4072–4086.
21. Grav,L.M., Sergeeva,D., Lee,J.S., Marín de Mas,I., Lewis,N.E., Andersen,M.R., Nielsen,L.K., Lee,G.M. and Kildegaard,H.F. (2018) Minimizing clonal variation during mammalian cell line

engineering for improved systems biology data generation. *ACS Synth. Biol.*, 10.1021/acssynbio.8b00140.

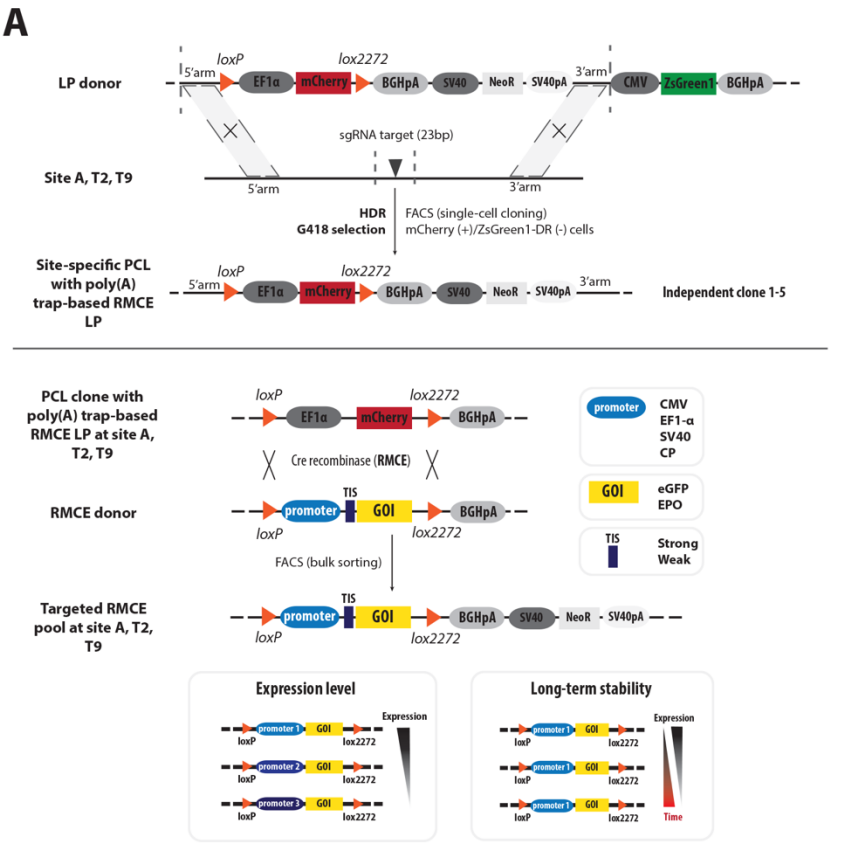
22. Papapetrou,E.P. and Schambach,A. (2016) Gene Insertion Into Genomic Safe Harbors for Human Gene Therapy. *Mol. Ther.*, **24**, 678–684.
23. Sadelain,M., Papapetrou,E.P. and Bushman,F.D. (2011) Safe harbours for the integration of new DNA in the human genome. *Nat. Rev. Cancer*, **12**, 51–58.
24. Cheng,J.K., Lewis,A.M., Kim,D.S., Dyess,T. and Alper,H.S. (2016) Identifying and retargeting transcriptional hot spots in the human genome. *Biotechnol. J.*, **11**, 1100–1109.
25. Kim,J.Y., Kim,Y.-G. and Lee,G.M. (2012) CHO cells in biotechnology for production of recombinant proteins: current state and further potential. *Appl. Microbiol. Biotechnol.*, **93**, 917–930.
26. Romanova,N. and Noll,T. (2018) Engineered and Natural Promoters and Chromatin-Modifying Elements for Recombinant Protein Expression in CHO Cells. *Biotechnol. J.*, **13**, 1700232.
27. Lombardo,A., Cesana,D., Genovese,P., Di Stefano,B., Provasi,E., Colombo,D.F., Neri,M., Magnani,Z., Cantore,A., Lo Riso,P., *et al.* (2011) Site-specific integration and tailoring of cassette design for sustainable gene transfer. *Nat. Methods*, **8**, 861–869.
28. Chen,C.-M., Krohn,J., Bhattacharya,S. and Davies,B. (2011) A Comparison of Exogenous Promoter Activity at the ROSA26 Locus Using a PhiC31 Integrase Mediated Cassette Exchange Approach in Mouse ES Cells. *PLoS One*, **6**, e23376.
29. Pristovšek,N., Hansen,H.G., Sergeeva,D., Borth,N., Lee,G.M., Andersen,M.R. and Kildegaard,H.F. (2018) Using Titer and Titer Normalized to Confluence Are Complementary Strategies for Obtaining Chinese Hamster Ovary Cell Lines with High Volumetric Productivity of Etanercept. *Biotechnol. J.*, **13**, e1700216.
30. de Vree,P.J.P., de Wit,E., Yilmaz,M., van de Heijning,M., Klous,P., Verstegen,M.J.A.M., Wan,Y., Teunissen,H., Krijger,P.H.L., Geeven,G., *et al.* (2014) Targeted sequencing by proximity ligation for comprehensive variant detection and local haplotyping. *Nat. Biotechnol.*, **32**, 1019–1025.
31. Hefzi,H., Ang,K.S., Hanscho,M., Bordbar,A., Ruckerbauer,D., Lakshmanan,M., Orellana,C.A., Baycin-Hizal,D., Huang,Y., Ley,D., *et al.* (2016) A Consensus Genome-scale Reconstruction of Chinese Hamster Ovary Cell Metabolism. *Cell Systems*, **3**, 434–443.e8.
32. Cerqueira,N.M.F.S.A., Fernandes,P.A. and Ramos,M.J. (2007) Ribonucleotide reductase: a critical enzyme for cancer chemotherapy and antiviral agents. *Recent Pat. Anticancer Drug Discov.*, **2**, 11–29.
33. Gutierrez,J.M. and Lewis,N.E. (2015) Optimizing eukaryotic cell hosts for protein production through systems biotechnology and genome-scale modeling. *Biotechnol. J.*, **10**, 939–949.
34. Rupp,O., MacDonald,M.L., Li,S., Dhiman,H., Polson,S., Griep,S., Heffner,K., Hernandez,I., Brinkrolf,K., Jadhav,V., *et al.* (2018) A reference genome of the Chinese hamster based on a hybrid assembly strategy. *Biotechnol. Bioeng.*, **115**, 2087–2100.
35. Feichtinger,J., Hernández,I., Fischer,C., Hanscho,M., Auer,N., Hackl,M., Jadhav,V., Baumann,M., Krempl,P.M., Schmidl,C., *et al.* (2016) Comprehensive genome and epigenome characterization of CHO cells in response to evolutionary pressures and over time. *Biotechnol. Bioeng.*, **113**, 2241–2253.
36. Ronda,C., Pedersen,L.E., Hansen,H.G., Kallehauge,T.B., Betenbaugh,M.J., Nielsen,A.T. and Kildegaard,H.F. (2014) Accelerating genome editing in CHO cells using CRISPR Cas9 and CRISPy, a web-based target finding tool. *Biotechnol. Bioeng.*, **111**, 1604–1616.
37. Kol,S., Kallehauge,T.B., Adema,S. and Hermans,P. (2015) Development of a VHH-Based Erythropoietin Quantification Assay. *Mol. Biotechnol.*, **57**, 692–700.
38. Petersen,S.D., Zhang,J., Lee,J.S., Jakociunas,T., Grav,L.M., Kildegaard,H.F., Keasling,J.D. and Jensen,M.K. (2018) Modular 5'-UTR hexamers for context-independent tuning of protein expression in eukaryotes. *Nucleic Acids Res.*, 10.1093/nar/gky734.
39. Yang,Z., Halim,A., Narimatsu,Y., Jitendra Joshi,H., Steentoft,C., Schjoldager,K.T.-B.G., Alder Schulz,M., Sealover,N.R., Kayser,K.J., Paul Bennett,E., *et al.* (2014) The GalNAc-type O-Glycoproteome of CHO cells characterized by the SimpleCell strategy. *Mol. Cell. Proteomics*, **13**, 3224–3235.
40. Brown,A.J., Gibson,S., Hatton,D. and James,D.C. (2018) Transcriptome-Based Identification of

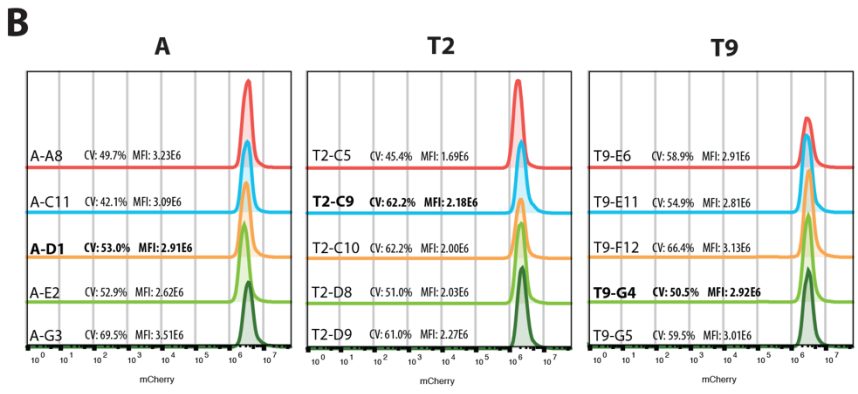
the Optimal Reference CHO Genes for Normalisation of qPCR Data. *Biotechnol. J.*, **13**.

41. Pybus,L.P., Dean,G., West,N.R., Smith,A., Daramola,O., Field,R., Wilkinson,S.J. and James,D.C. (2014) Model-directed engineering of 'difficult-to-express' monoclonal antibody production by Chinese hamster ovary cells. *Biotechnol. Bioeng.*, **111**, 372–385.
42. Ho,S.C.L., Mariati, Yeo,J.H.M., Fang,S.G. and Yang,Y. (2015) Impact of using different promoters and matrix attachment regions on recombinant protein expression level and stability in stably transfected CHO cells. *Mol. Biotechnol.*, **57**, 138–144.
43. Wang,W., Jia,Y.-L., Li,Y.-C., Jing,C.-Q., Guo,X., Shang,X.-F., Zhao,C.-P. and Wang,T.-Y. (2017) Impact of different promoters, promoter mutation, and an enhancer on recombinant protein expression in CHO cells. *Sci. Rep.*, **7**, 10416.
44. Weber,W. and Fussenegger,M. (2010) Synthetic gene networks in mammalian cells. *Curr. Opin. Biotechnol.*, **21**, 690–696.
45. O'Brien,S.A., Lee,K., Fu,H.-Y., Lee,Z., Le,T.S., Stach,C.S., McCann,M.G., Zhang,A.Q., Smanski,M.J., Somia,N.V., *et al.* (2018) Single Copy Transgene Integration in a Transcriptionally Active Site for Recombinant Protein Synthesis. *Biotechnol. J.*
46. Niwa,H., Araki,K., Kimura,S., Taniguchi,S., Wakasugi,S. and Yamamura,K. (1993) An efficient gene-trap method using poly A trap vectors and characterization of gene-trap events. *J. Biochem.*, **113**, 343–349.
47. Jun,S.C., Kim,M.S., Hong,H.J. and Lee,G.M. (2006) Limitations to the development of humanized antibody producing Chinese hamster ovary cells using glutamine synthetase-mediated gene amplification. *Biotechnol. Prog.*, **22**, 770–780.
48. Yang,Y., Mariati, Chusainow,J. and Yap,M.G.S. (2010) DNA methylation contributes to loss in productivity of monoclonal antibody-producing CHO cell lines. *J. Biotechnol.*, **147**, 180–185.
49. Kim,M., O'Callaghan,P.M., Droms,K.A. and James,D.C. (2011) A mechanistic understanding of production instability in CHO cell lines expressing recombinant monoclonal antibodies. *Biotechnol. Bioeng.*, **108**, 2434–2446.
50. Osterlehner,A., Simmeth,S. and Göpfert,U. (2011) Promoter methylation and transgene copy numbers predict unstable protein production in recombinant Chinese hamster ovary cell lines. *Biotechnol. Bioeng.*, **108**, 2670–2681.
51. Ho,S.C.L., Koh,E.Y.C., Soo,B.P.C., Mariati, Chao,S.-H. and Yang,Y. (2016) Evaluating the use of a CpG free promoter for long-term recombinant protein expression stability in Chinese hamster ovary cells. *BMC Biotechnol.*, **16**, 71.
52. Raj,A. and van Oudenaarden,A. (2008) Nature, nurture, or chance: stochastic gene expression and its consequences. *Cell*, **135**, 216–226.
53. Elowitz,M.B., Levine,A.J., Siggia,E.D. and Swain,P.S. (2002) Stochastic gene expression in a single cell. *Science*, **297**, 1183–1186.
54. Ho,S.C.L. and Yang,Y. (2014) Identifying and engineering promoters for high level and sustainable therapeutic recombinant protein production in cultured mammalian cells. *Biotechnol. Lett.*, **36**, 1569–1579.
55. Brown,A.J. and James,D.C. (2016) Precision control of recombinant gene transcription for CHO cell synthetic biology. *Biotechnol. Adv.*, **34**, 492–503.
56. Müller,K., Engesser,R., Schulz,S., Steinberg,T., Tomakidi,P., Weber,C.C., Ulm,R., Timmer,J., Zurbriggen,M.D. and Weber,W. (2013) Multi-chromatic control of mammalian gene expression and signaling. *Nucleic Acids Res.*, **41**, e124.
57. Cheng,J.K. and Alper,H.S. (2016) Transcriptomics-Guided Design of Synthetic Promoters for a Mammalian System. *ACS Synth. Biol.*, **5**, 1455–1465.
58. Brown,A.J., Gibson,S.J., Hatton,D. and James,D.C. (2017) In silico design of context-responsive mammalian promoters with user-defined functionality. *Nucleic Acids Res.*, **45**, 10906–10919.
59. Li,J., Menzel,C., Meier,D., Zhang,C., Dübel,S. and Jostock,T. (2007) A comparative study of different vector designs for the mammalian expression of recombinant IgG antibodies. *J. Immunol. Methods*, **318**, 113–124.
60. Majocchi,S., Aritonovska,E. and Mermod,N. (2014) Epigenetic regulatory elements associate with specific histone modifications to prevent silencing of telomeric genes. *Nucleic Acids Res.*, **42**, 193–204.

61. Saunders,F., Sweeney,B., Antoniou,M.N., Stephens,P. and Cain,K. (2015) Chromatin function modifying elements in an industrial antibody production platform--comparison of UCOE, MAR, STAR and cHS4 elements. *PLoS One*, **10**, e0120096.
62. Zboray,K., Sommeregger,W., Bogner,E., Gili,A., Sterovsky,T., Fauland,K., Grabner,B., Stiedl,P., Moll,H.P., Bauer,A., *et al.* (2015) Heterologous protein production using euchromatin-containing expression vectors in mammalian cells. *Nucleic Acids Res.*, **43**, e102.
63. Kang,S.-Y., Kim,Y.-G., Kang,S., Lee,H.W. and Lee,E.G. (2016) A novel regulatory element (E77) isolated from CHO-K1 genomic DNA enhances stable gene expression in Chinese hamster ovary cells. *Biotechnol. J.*, **11**, 633–641.
64. Ebadat,S., Ahmadi,S., Ahmadi,M., Nematpour,F., Barkhordari,F., Mahdian,R., Davami,F. and Mahboudi,F. (2017) Evaluating the efficiency of CHEF and CMV promoter with IRES and Furin/2A linker sequences for monoclonal antibody expression in CHO cells. *PLoS One*, **12**, e0185967.
65. Chai,Y.-R., Ge,M.-M., Wei,T.-T., Jia,Y.-L., Guo,X. and Wang,T.-Y. (2018) Human rhinovirus internal ribosome entry site element enhances transgene expression in transfected CHO-S cells. *Sci. Rep.*, **8**, 6661.
66. Kaufman,W.L., Kocman,I., Agrawal,V., Rahn,H.-P., Besser,D. and Gossen,M. (2008) Homogeneity and persistence of transgene expression by omitting antibiotic selection in cell line isolation. *Nucleic Acids Res.*, **36**, e111.
67. Voss,T.C. and Hager,G.L. (2013) Dynamic regulation of transcriptional states by chromatin and transcription factors. *Nat. Rev. Genet.*, **15**, 69.
68. Brewster,R.C., Weinert,F.M., Garcia,H.G., Song,D., Rydenfelt,M. and Phillips,R. (2014) The Transcription Factor Titration Effect Dictates Level of Gene Expression. *Cell*, **156**, 1312–1323.
69. Le,H., Vishwanathan,N., Kantardjieff,A., Doo,I., Srienc,M., Zheng,X., Somia,N. and Hu,W.-S. (2013) Dynamic gene expression for metabolic engineering of mammalian cells in culture. *Metab. Eng.*, **20**, 212–220.
70. Xie,M. and Fussenegger,M. (2018) Designing cell function: assembly of synthetic gene circuits for cell biology applications. *Nat. Rev. Mol. Cell Biol.*, **19**, 507–525.
71. Hansen,H.G., Pristovšek,N., Kildegaard,H.F. and Lee,G.M. (2017) Improving the secretory capacity of Chinese hamster ovary cells by ectopic expression of effector genes: Lessons learned and future directions. *Biotechnol. Adv.*, **35**, 64–76.

Table and figures





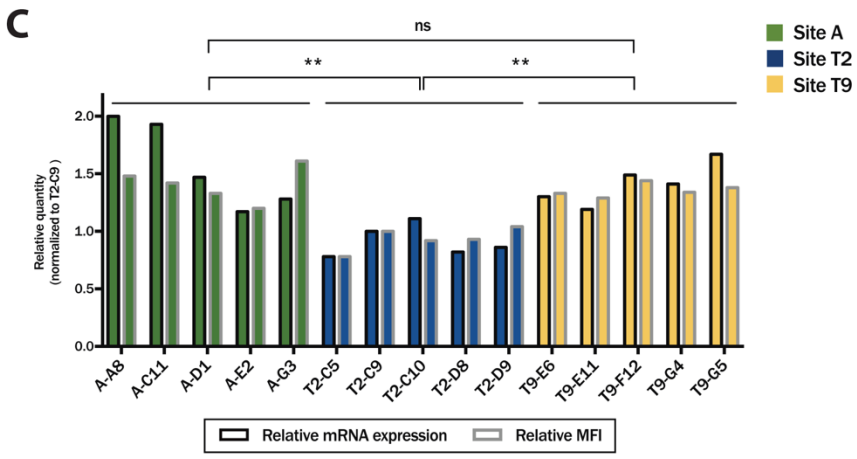


Figure 1: **Generation of mCherry platform cell lines with a poly(A) trap-based recombinase-mediated cassette exchange (RMCE) landing pad in putative safe harbors yields homogeneous mCherry expression patterns across different parental clones**

(A) Five independent mCherry platform cell lines (PCLs) in each of the three putative safe harbors (A, T2 and T9) were generated with CRISPR/Cas9-based targeted integration via homology-directed repair (HDR) using site-specific landing pad (LP) donor and sgRNA plasmid, together with a Cas9 vector. LP donor vector consisted of mCherry and selection marker (NeoR) expression cassettes, flanked by 750 bp long 5' and 3' arms, homologous to the site of interest, and ZsGreen1-DR expression cassette outside of homology arms to exclude random integrants. After HDR-based targeted integration, a 14-day G418-based selection was performed to enrich for cells with a successful integration of the LP donor. This was followed by single-cell cloning of double mCherry positive/ZsGreen1-DR negative cells by FACS to isolate monoclonal cell lines with targeted integration of the LP donor. The LP donors contained *loxP* and *lox2272* recombination sites, flanking EF1α-mCherry, providing a poly(A)-containing LP (*i.e.* poly(A) trapping) for recombinase-mediated cassette exchange (RMCE). EF1α-mCherry was upon co-transfection of Cre recombinase and RMCE donor exchanged with a range of different expression cassettes. RMCE donors consisted of different genes of interest (GOIs) under different 5' proximal regulatory elements (*i.e.* promoters and Kozak/translation initiation sequences (TIS)). Two weeks post-recombination, successfully recombined cells were detected by FACS and bulk sorted (mCherry negative cells for EPO and double mCherry negative/GFP positive cells for eGFP) to isolate an RMCE-generated cell pool comprising targeted integrants. To assess site-specific expression strength resulting from specific regulatory elements, the pools were further analyzed by measuring the GOI's mRNA (by RT-qPCR) and protein levels (by FACS or bio-layer interferometry for eGFP and EPO, respectively). Further, eGFP expression profiles at the mRNA and protein level were measured during two month long-term cultivation to assess the expression stability of both, the integration site and the promoter of interest. (**B**) mCherry fluorescence levels of the generated mCherry-encoding PCLs (n = 5 clones/site) were measured in mid to late exponential phase in batch cultures. Coefficient of variation (CV) and median fluorescence intensity (MFI) are shown for each clone, by analyzing 20,000 - 40,000 cells per clone. (**C**) mCherry mRNA levels were also assessed in mid to late exponential phase in batch cultures in the generated mCherry-encoding PCLs. mCherry mRNA levels were normalized to T2-C9 clone. Statistical significance between relative mRNA expression levels of five clones with targeted site A, T2 and T9 was calculated using unpaired *t* test. A linear regression-based correlation analysis of relative mCherry MFIs from Figure 1b (normalized to mCherry MFI of T2-C9 clone) and relative mCherry mRNA expression levels is shown.

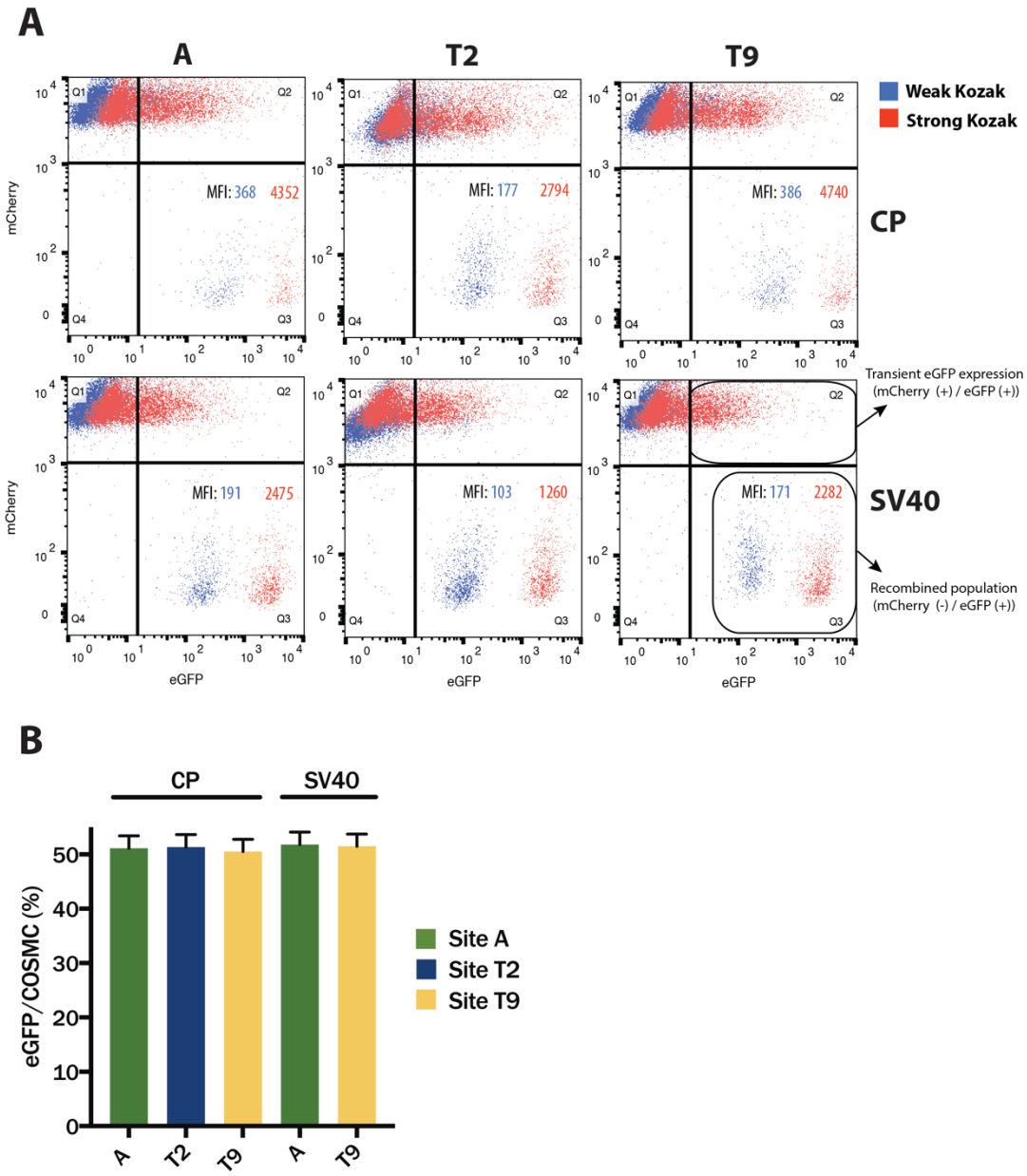
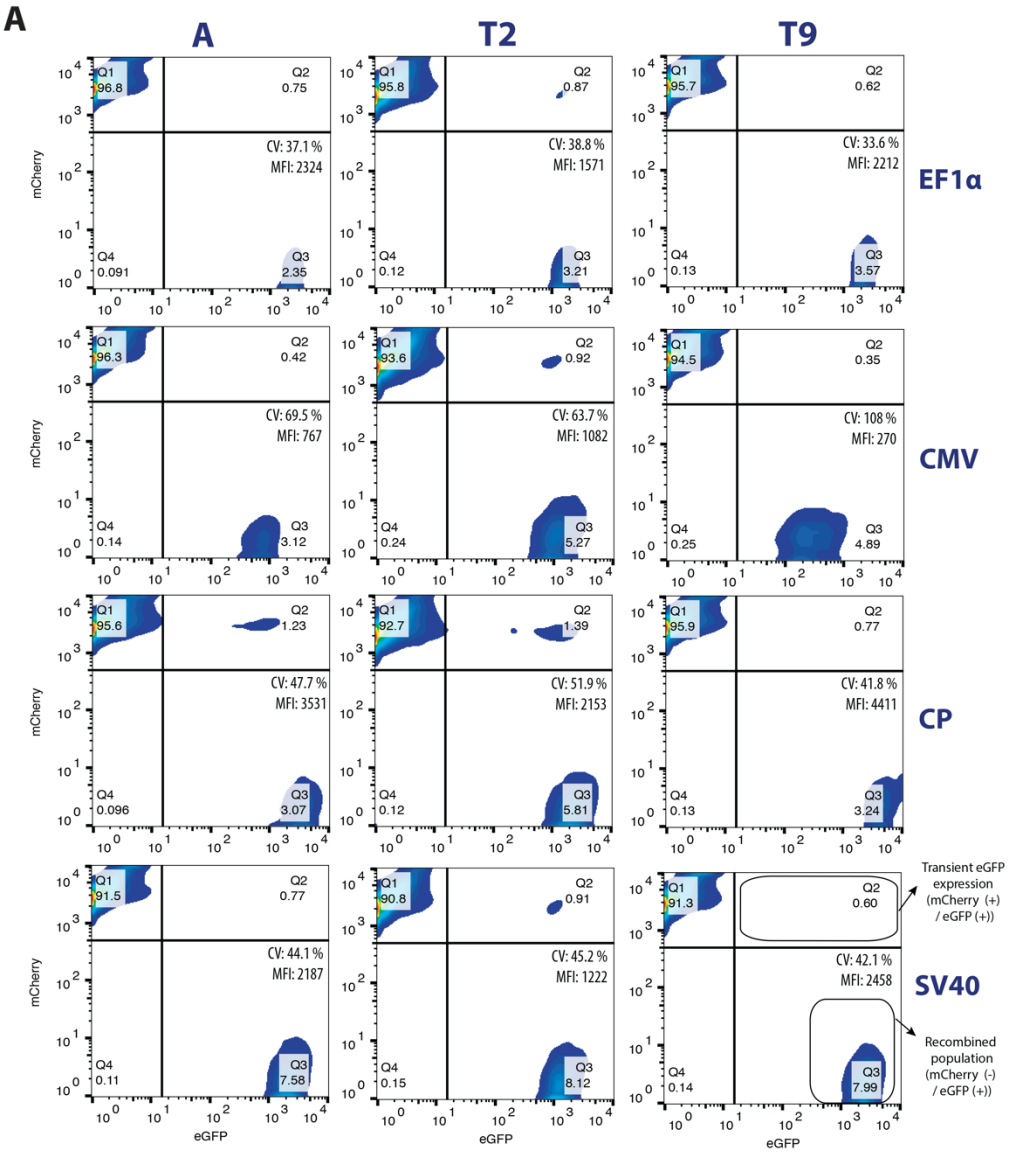
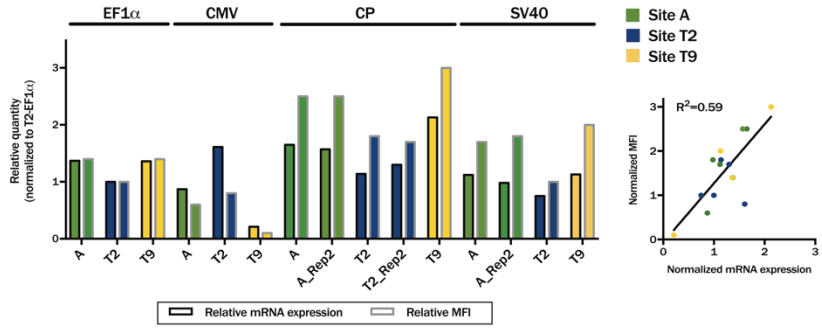


Figure 2: Proof-of-principle RMCE experiment with different eGFP constructs confirms functionality of poly(A) trap and generates single-copy populations with tunable expression patterns

(**A**) Three representative PCLs (A-D1, T2-C9 and T9-G4) were randomly selected for site A, T2 and T9, respectively, and subjected to RMCE with four different eGFP constructs with composite promoter (CP) or SV40 promoter, preceded by either a weak or strong Kozak sequence. One week post-RMCE, mCherry and eGFP fluorescence were measured by FACS (100,000 - 200,000 cells) with MFIs of successfully recombined populations (*i.e.* mCherry negative/eGFP positive cells) reported. (**B**) Recombined populations with eGFP under CP or SV40 promoter and strong Kozak only were bulk sorted and their eGFP copy number measured using digital PCR (dPCR). Using *COSMC* as an internal one-copy control gene, obtained eGFP/COSMC reads (%) are reported for five bulk-sorted populations. The error bars represent 95% confidence intervals, averaged between the technical replicates (n = 2).



B



C

Site	A			T2			T9			eGFP
PCL / Promoter	D1	A8	C11	C9	C10	D8	G4	E6	F12	Normalized output
EF1α	1.4	1.5	1.4	1.0	1.0	1.0	1.4	1.4	1.4	MFI
CMV	0.6	0.6	0.4	0.8	0.6	0.5	0.1	0.2	0.2	MFI
CP	2.5	2.7	2.3	1.8	1.7	1.8	3.0	1.8	2.8	MFI
SV40	1.7	1.7	1.5	1.0	0.7	0.7	2.0	1.5	1.4	MFI

Legend 0 3

Figure 3: Screening of the promoter-eGFP panel generates site-dependent and construct-dependent expression patterns with up to 30-fold dynamic range and minimal clonal variation

(**A**) A panel of eGFP constructs with four differents promoters (EF1α, CMV, CP and SV40 promoter) preceded by a strong Kozak sequence was recombined into A-D1, T2-C9 and T9-G4 PCLs. Two weeks post-RMCE, FACS analysis (100,000 - 200,000 cells) for all 12 conditions was performed assessing MFI and CV of successfully recombined eGFP populations (*i.e.* mCherry negative/eGFP positive cells). (**B**) All 12 recombined populations were bulk sorted and their eGFP fluorescence and mRNA levels measured. eGFP MFIs and mRNA levels were normalized to the values of T2-EF1α population (internal calibrator). A linear regression-based correlation analysis of relative eGFP MFIs and relative eGFP mRNA expression levels is shown. RMCE, bulk sorting as well as eGFP fluorescence and mRNA levels measurements were repeated for three PCL-promoter combinations (Rep2 samples). (**C**) Heatmap with relative MFIs for 36 eGFP recombined populations, measured with FACS (100,000 - 200,000 cells). The panel of eGFP constructs with four promoters and a strong Kozak sequence was recombined into three different parental PCLs for each site. eGFP MFIs of all populations were normalized to the value of T2-C9-EF1α population.

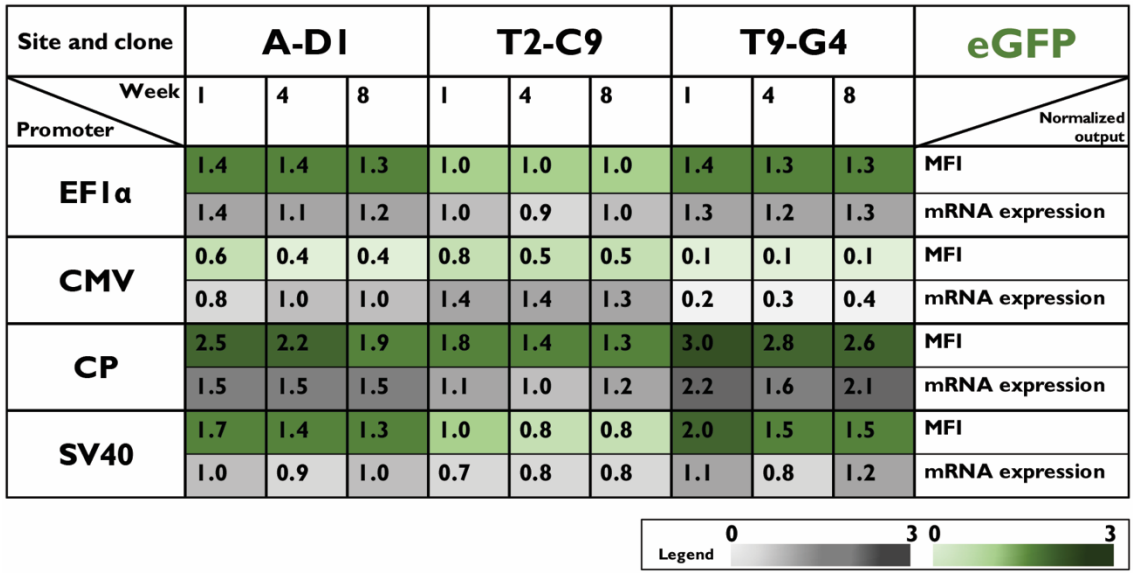


Table 1: Long-term stability of eGFP expression is ensured in all three newly discovered safe harbors regardless of the expression cassette used

Heatmap with relative MFIs and relative mRNA expression levels for 12 eGFP recombined populations (from Figure 3a and b) at week 1, week 4 and week 8 from the start of long-term cultivation. Three bulk-sorted populations (namely A-CP, A-SV40 and T2-CP) were generated twice, independently (for more details refer to Material and Methods, subsection: Subpopulation generation by recombinase-mediated cassette exchange (RMCE) and bulk sorting), and the corresponding MFIs are reported as relative mean MFIs between the two biological replicates. eGFP MFIs and mRNA levels were normalized to the values of T2-C9-EF1α population at the first time point (week 1; internal calibrator).

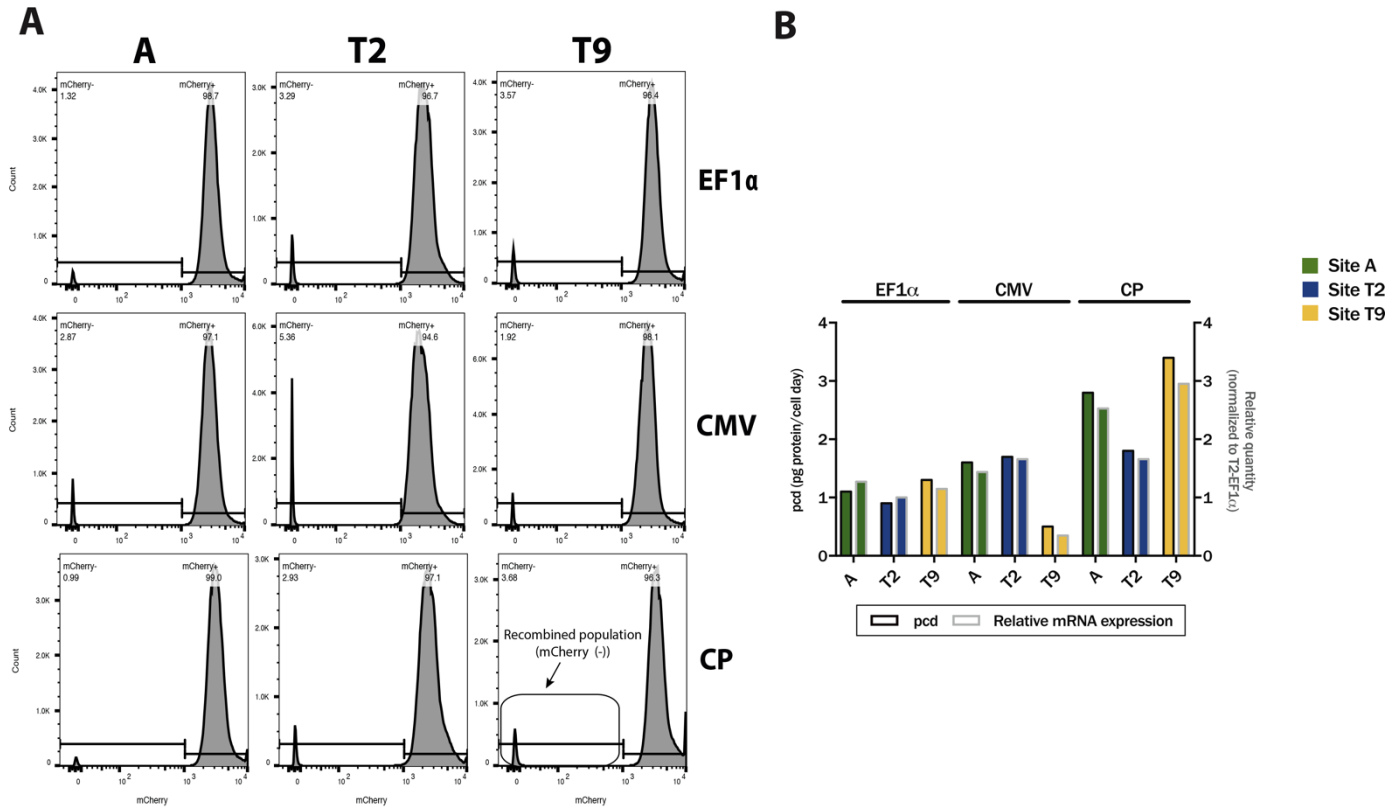


Figure 4: **Expression level characterization for a secreted biotherapeutic (EPO) demonstrates protein-specific patterns under different expression cassettes**

(**A**) A panel of erythropoietin (EPO) constructs with three differents promoters (EF1α, CMV promoter and CP) preceded by a strong Kozak sequence was recombined into A-D1, T2-C9 and T9-G4 PCLs. Two weeks post-RMCE, FACS analysis (100,000 - 200,000 cells) for all 9 conditions was performed assessing percentage of successfully recombined EPO populations (*i.e.* mCherry negative cells). (**B**) All 9 recombined populations were bulk sorted and EPO protein and mRNA levels measured. Specific productivities of EPO (pcd; pg protein/cell day) were correlated with relative EPO mRNA expression levels using linear regression. EPO mRNA levels were normalized to the value of T2-EF1α population.

Chapter 3

Improving Recombinant Protein Production by Ectopic Expression of Effector Genes - A Review

Rationale

In this chapter, a review is presented of the state of the art with respect to engineering effector genes, beneficial for resolving the secretory bottlenecks (q_p) in the production of recombinant proteins in CHO cells. We discuss the different experimental and cellular factors, such as the expression platform, the gene dosage and origin, the clonal variation and the host cell, contributing to and affecting the results observed when overexpressing the effector gene of interest. The review underpins the importance of considering these factors and provides guidance for the future engineering of novel effector genes.

The many factors discussed here are examined and the guidelines applied in a study, presented in **Chapter 4**, where novel effector genes with q_p-increasing function are discovered.

Biotechnology Advances 35 (2017) 64–76

Contents lists available at ScienceDirect

Biotechnology Advances

journal homepage: www.elsevier.com/locate/biotechadv

Research review paper

Improving the secretory capacity of Chinese hamster ovary cells by ectopic expression of effector genes: Lessons learned and future directions

Henning Gram Hansen [a,*], Nuša Pristovšek [a], Helene Faustrup Kildegaard [a], Gyun Min Lee [a,b,**]

[a] The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Kgs. Lyngby, Denmark
[b] Department of Biological Sciences, KAIST, 291 Daehak-ro, Yuseong-gu, Daejeon 305-701, Republic of Korea

ARTICLE INFO

Article history:
Received 13 July 2016
Received in revised form 12 November 2016
Accepted 28 November 2016
Available online 6 December 2016

Keywords:
Cell engineering
Chinese hamster ovary (CHO) cells
Ectopic expression
Endoplasmic reticulum
ER stress
Gene dosage
Product quality
Recombinant protein production
Secretion bottleneck
Specific productivity

ABSTRACT

Chinese hamster ovary (CHO) cells are the preferred cell factory for the production of therapeutic glycoproteins. Although efforts primarily within bioprocess optimization have led to increased product titers of recombinant proteins (r-proteins) expressed in CHO cells, post-transcriptional bottlenecks in the biosynthetic pathway of r-proteins remain to be solved. To this end, the ectopic expression of transgenes (effector genes) offers great engineering potential. However, studies on effector genes have in some cases led to inconsistent results. Whereas this can in part be attributed to product specificity, other experimental and cellular factors are likely important contributors to these conflicting results. Here, these factors are reviewed and discussed with the objective of guiding future studies on effector genes.

Contents

Abbreviations: ATF6c, activating transcription factor 6; CHO, Chinese hamster ovary; DHFR, dihydrofolate reductase; effector gene, a transgene ectopically expressed with the purpose of obtaining a phenotypic change; IE, inducible expression; EPO, erythropoietin; ER, endoplasmic reticulum; HEK, human embryonic kidney; MAb, monoclonal antibody; pcd, pg per cell per day; PDI, protein disulphide isomerase; PTM, post-translational modification; q_p, specific protein productivity; r-protein, recombinant protein; SCS, single-cell sorting; SEAP, secreted alkaline phosphatase; SEE, stable episomal expression; SGE, stable gene expression; TGE, transient gene expression; YY1, transcription factor Yin Yang 1; VEGF, vascular endothelial growth factor; XBP-1S, spliced form of X-box-binding protein 1.

* Correspondence to: H. G. Hansen, DTU Biosustain, Building 220, Kemitorvet, 2800 Kgs. Lyngby, Denmark.
** Correspondence to: G.M. Lee, Department of Biological Sciences, KAIST, 291 Daehak-ro, Yuseong-gu, Daejeon 305-701, Republic of Korea.
E-mail addresses: hgra@biosustain.dtu.dk (H.G. Hansen), gyunminlee@kaist.ac.kr (G.M. Lee).

http://dx.doi.org/10.1016/j.biotechadv.2016.11.008
0734-9750/

1. Introduction

Chinese hamster ovary (CHO) cells are the most frequently used cell host for biopharmaceutical production of glycoproteins (Walsh, 2014). Besides being the host cell used for the first approval of a recombinant biopharmaceutical produced in mammalian cells in 1986 (Wurm, 2004), CHO cells are the preferred choice for a number of reasons. First, CHO cells can easily be adapted for high-density suspension growth in a chemically defined, serum-free medium in large-volume cultures (Kim et al., 2012; Sinacore et al., 2000). Second, gene amplification methods have been established for CHO cells, leading to high specific productivity (q_p) of recombinant protein (r-protein) in stable cell lines (Durocher and Butler, 2009). Third, CHO cells are less prone to virus infection than other mammalian production cell lines and are therefore regarded as a safe host for the production of human therapeutics (Berting et al., 2010). Last, CHO cells and other mammalian cells are the platform of choice for the production of human recombinant glycoproteins because of their ability to correctly make human-like post-translational modifications (PTMs), in particular glycosylation (Butler and Spearman, 2014). Human-like PTMs turn r-protein products into functional drug molecules with reduced immunogenicity, prolonged serum half-life and high pharmacological efficacy in the human body (Walsh and Jefferis, 2006).

The production of r-proteins in CHO cells in optimized bioprocesses can reach q_p of 50–90 pg per cell per day (pcd) (Hacker et al., 2009). As previously pointed out by Khan and Schröder (2008), professional secretory plasma cells are capable of secreting IgM at a rate of 200–400 pcd (Fazekas et al., 1980; Randall et al., 1992). This clearly indicates that nature's physiological limit not yet has been reached and thus, intracellular rate-limiting steps in protein production remain to be resolved. Indeed, post-transcriptional rate-limiting steps in the biosynthetic pathway of r-proteins have been reported multiple times in CHO cells (Johari et al., 2015; Kallehauge et al., 2016; S.J. Kim et al., 1998; Ku et al., 2008; Schröder et al., 1999) as well as in other mammalian cells (Barnes et al., 2004; Fann et al., 1999). The presence of a post-transcriptional bottleneck suggests that there are many opportunities to improve the secretory pathway machinery in CHO cells. Moreover, artificial protein scaffolds such as fusion proteins are becoming more popular in the biopharmaceutical industry with increasing market shares (Aggarwal, 2014). These non-native scaffolds are in general more prone to misfolding (Lee et al., 2007). Thus, the cost-efficient production of these difficult-to-express fusion proteins will most likely require substantial engineering of the folding machinery in the secretory pathway.

Engineering CHO cells by the ectopic expression of transgenes (hereafter referred to as effector genes) is an attractive solution to improve the secretory capacity of CHO cells. In many cases, such engineering efforts have led to positive effects on q_p on a variety of r-proteins (see recent reviews (Fischer et al., 2015; Hussain et al., 2014; Nishimiya, 2013)). This multitude of studies showing positive effects clearly underpins the potential of expressing effector genes. However, as previously pointed out (Hussain et al., 2014; Kim et al., 2012; Mohan et al., 2008), some effector genes are flawed by inconsistent effects. To exemplify this, all published studies on r-protein productivity (volumetric productivity or q_p) in CHO cells with the ectopic expression of the widely studied protein disulphide isomerase (PDI) are listed in Table 1. PDI is an endoplasmic reticulum (ER)-resident enzyme conferring disulphide isomerase activity (Hatahet and Ruddock, 2009). Moreover, PDI forms and reduces disulphide bonds in nascent polypeptides in the lumen of the ER and in parallel inhibits the aggregation of folding intermediates through its function as a chaperone (Appenzeller-Herzog and Ellgaard, 2008). The reported effects of overexpressing PDI on volumetric productivity and q_p vary from a two-fold decrease through no effect to a 1.4-fold increase. This inconsistency, can to some extent, be explained by product specificity, as several different r-proteins have been used as model proteins. In fact, PDI overexpression only increased q_p for one of four monoclonal antibody (MAb) variants in a parallel experimental setup (Pybus et al., 2014). However, many cellular and experimental factors are at play when examining how an effector gene affects volumetric productivity and q_p (Fig. 1). Thus, it is likely that factors other than product specificity are involved in the inconsistency of PDI's effect on volumetric productivity and q_p of r-proteins.

In contrast to PDI, the effect of many effector genes on volumetric productivity or q_p has only been reported once (Hussain et al., 2014; Nishimiya, 2013). Notwithstanding product specificity, it is likely that a considerable number of these effects are conditional – for example, specific to the monoclonal cell line or the expression platform being used. The applicability of such conditional effects is often limited to the research group in question and not to the CHO engineering field in general. Here, cellular and experimental factors that potentially affect the outcome when studying effector genes will be described and discussed. If these factors are appreciated, the risk of unintentionally investigating conditional effects can be minimized and the chance of finding true positive effects can be increased.

2. CHO host cell lines

In 1957, the immortalized, original CHO cell line (the common ancestor for all CHO cell lines) was established from the ovaries of an outbred female Chinese hamster (Puck, 1957; Wurm, 2013). This original cell line has led to a multitude of commercially available and proprietary CHO cell lines (Wurm, 2013). Being an immortalized cell line, the genome of CHO cells is inherently unstable (Frye et al., 2016). Moreover, dihydrofylate reductase deficiency (DHFR) in the widely used DXB11 and DG44 cell lines was achieved by subjecting cells to radiation- and chemical-mediated mutagenesis, respectively (Urlaub et al., 1983; Urlaub and Chasin, 1980). Thus, host CHO cell lines constitute a genomically diverse family in terms of single nucleotide polymorphisms (Lewis et al., 2013), copy-number variations (Kaas et al., 2015) and karyotypes (Wurm and Hacker, 2011). Moreover, it has recently been suggested to regard CHO host cell lines as 'quasispecies', emphasizing the extensive genetic heterogeneity residing in the CHO host cell family (Wurm, 2013).

When CHO host cell lines are compared, they are found to be not only genetically divergent but also phenotypically diverse. For example, it has been shown that the ER size in the CHO-K1 host cell line is larger compared to a DXB11-derived host cell line, and the mitochondrial mass was also found to be higher in CHO-K1 cells (Hu et al., 2013). These phenotypic differences might explain the approximately 10-fold lower q_p observed for the DXB11-derived host cell line compared to CHO-K1 cells, which was obtained for two different MAbs from stable gene-amplified clones (Hu et al., 2013). In a recent CHO bibliome study by Golabgir et al. (2016), a meta-analysis of bioprocess studies showed that the specific growth rate and q_p of DXB11- compared to DG44-derived cell lines were significantly higher and lower, respectively. Although the bibliome data consist of a range of process conditions and experimental setups, both DG44 and DXB11 are DHFR-deficient cell lines. Consequently, the gene-amplification process and clone selection are therefore comparable, warranting the comparison of q_p.

Table 1
Studies on how ectopic expression of protein disulphide isomerase (PDI) affects volumetric productivity or q_p of r-proteins in CHO cells.

R-protein and reference	R-protein expression[a]; clonality[b]	Post-transcriptional bottleneck	Origin of PDI	PDI expression[a]; clonality[c]	Gene dosage	Host cell; cell culture	Effect on titer/q_p
Interleukin-15 (Davis et al., 2000)	SGE; 1 monoclonal cell line	Not investigated	Human	SGE; *quasi*-polyclonal	Single gene dose	CHO-DXB11; 2–3 days batch culture	No effect
Fc-fusion protein (Davis et al., 2000)	Stable expression; 1 monoclonal cell line	Not investigated	Human	SGE; *quasi*-polyclonal and 3 *mono*-monoclonal cell lines	3 clones with different levels of PDI expression	CHO-DXB11; 2–3 days batch culture	Polyclonal: No effect *Mono*-monoclonal: 2-fold decrease in titer for clone with highest PDI expression
MAb (Borth et al., 2005)	SGE; 1 monoclonal cell line	Not investigated	Not re-ported	SGE; *quasi*-polyclonal	Single gene dose	CHO dhfr^{-}; 3 days batch culture	1.37-fold increase in q_p
Thrombopoietin (Mohan et al., 2007)	SGE; 1 monoclonal cell line	Not investigated	CHO	IE; 2 *mono*-monoclonal cell lines	Single gene dose	CHO-DXB11; 3 days batch culture	No effect
MAb (Mohan et al., 2007)	SGE; 1 monoclonal cell line	Not investigated	CHO	IE; 2 *mono*-monoclonal lines	Single gene dose	DG44; 2 days batch culture	1.15–1.27-fold increase in q_p
MAb (Hayes et al., 2010)	SGE; 1 monoclonal cell line	Not investigated	Human	TGE; 1 *mono*-monoclonal line	Single gene dose	CHO-K1SV; 2 days batch culture	No effect
MAb (4 variants) (Pybus et al., 2014)	TGE (episomal-replication); polyclonal	Yes (predicted by modelling)	Human	TGE; polyclonal	Single gene dose	CHO-K1 EB27; 10 days fed-batch culture	No effect for 3 variants and 1.3-fold increase in q_p for 1 variant.
Fc-fusion protein (Johari et al., 2015)	TGE; polyclonal	Yes	Human	TGE; polyclonal	3 gene doses	CHO-S; 3 days batch culture	Gene dosage/effect on q_p: 10%/no effect, 20%/1.2-fold increase (1.4-fold increase in aggregate formation), 40%/1.3-fold increase.
α_1-Antitrypsin (Hansen et al., 2015)	TGE; polyclonal	Not investigated	Mouse	TGE; polyclonal	Single gene dose	CHO-S; 3 days batch culture	1.2-fold decrease in q_p
C1 esterase inhibitor (Hansen et al., 2015)	TGE; polyclonal	Not investigated	Mouse	TGE; polyclonal	Single gene dose	CHO-S; 3 days batch culture	1.6-fold decrease in q_p
MAb (our unpublished data)	TGE; polyclonal	Not investigated	Mouse	TGE; polyclonal	Single gene dose	CHO-S; 3 days batch culture	1.3-fold decrease in q_p

[a] For definition and abbreviation of expression platforms, see Table 2.
[b] Clonality of cells expressing the r-protein before transfection of the effector gene (SGE). r-proteins and effector genes are co-expressed in TGE platforms (polyclonal). For definition of clonality, see Fig. 2.
[c] Clonality of cells expressing PDI. For definition of clonality, see Fig. 2.

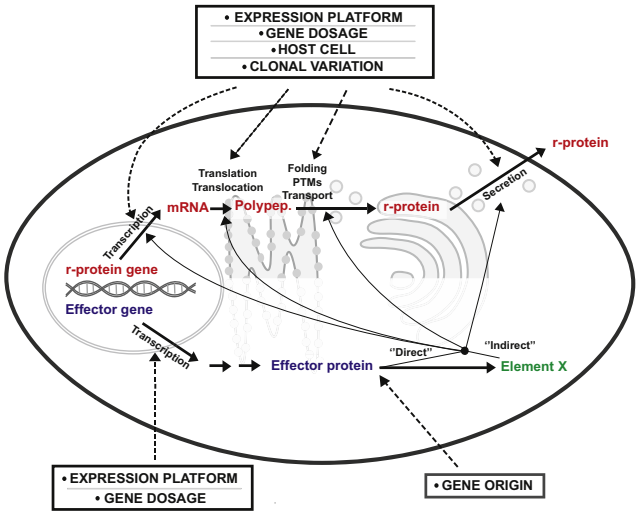


Fig. 1. Cellular and experimental factors affecting the biosynthetic pathway of r-proteins in CHO cells. Graphical representation of a CHO cell co-expressing an r-protein (gene, mRNA, polypeptide and protein in red letters) and an effector gene (gene and protein in blue letters). Effects from external factors are indicated with dashed arrows. Effector protein-mediated effects are shown with thin line arrows, and whether these effects are direct or indirect through interactions with an unknown DNA or protein molecule (element X) are illustrated. Nucleus, ER, Golgi and transport vesicles are depicted in the upper half of the cell. Since localization of effector proteins is gene-dependent, cytoplasmic organelles are not depicted in the lower half of the cell. PTM, post-translational modification.

Moreover, CHO-K1-derived cells were found to grow slower than DG44-derived cells in the bibliome study. The observed differences between CHO-K1, DXB11 and DG44 host cells clearly illustrate the phenotypic diversity that resides within the family of CHO host cells.

Being such a diverse family of cell lines, the following question arises: Are effects on q_p of an effector gene transferable across CHO host cell lines? In the PDI example, at least four CHO host cell lines (DXB11, DG44, CHO-K1 and CHO-S) have been used (Table 1). Because of product specificity and other factors, it is not possible to make any conclusion on the influence of CHO host cell lines on PDI as an effector gene. As for the spliced form of X-box-binding protein 1 (XBP-1S), it has been shown that ectopic XBP-1S expression was able to increase the volumetric productivity of erythropoietin (EPO) in CHO-K1 as well as in the murine NS0 myeloma cell line (Ku et al., 2008). Moreover, Tastanova et al. (2015) found that the CHO-derived transcription factor Yin Yang 1 (YY1) improved the volumetric productivity of a MAb (Rituximab) upon co-transfection in a transient expression setup in CHO-K1 and CHO-S cells. Moreover, YY1 expression also improved the volumetric productivity of MAb in DG44- and DXB11-derived stable cell lines. In a transient expression setup using the human orthologue, the positive effect of YY1 on volumetric productivity was also observed in human embryonic kidney (HEK) cells, human cervical cancer cells (HeLa) and human fibrosarcoma cells (HT-1080), which are immortalized cell lines with genomes vastly different from CHO cells. Future studies will show whether the generic effect of YY1 and XBP-1S between genomically distinct cell lines is an exception that proves the rule.

3. Clonal variation

Phenotypic heterogeneity is observed not only between CHO host cell lines but also within the cell population of CHO host lines. In fact, CHO cells are known for being able to adapt to changes in process conditions, which has been exploited in the industry to generate clonally derived cell lines with enhanced manufacturing capabilities (Frye et al., 2016). Moreover, the majority of cells in host cell lines seem to be intrinsically incapable of high production of MAb, and universally competent cells are likely relatively rare cases (O'Callaghan et al., 2010). Functional heterogeneity residing in the host cell population is typically referred to as clonal variation (Fig. 2). Although not CHO cells, how clonal variation is generated (phenotypic drift) has been elegantly demonstrated in the murine cell line NS0 (Barnes et al., 2006). Three rounds of limiting dilution were performed, and variation in cell growth rate was observed after each round of subcloning. This phenomenon has subsequently been observed in subclones of a CHO-K1-derived host cell line (Davies et al., 2013). The cell growth rate changed for approximately half of the clones during extended culture time, demonstrating that both static (inheritable) phenotypes and phenotypic drift were observed. Moreover, the initial cell growth rate of the subclones varied substantially, which demonstrates the presence of clonal variation within the host cell line. When analysing single cells in a monoclonal cell line culture, large variation in transgene expression has been observed, which could not be attributed to variables such as cell cycle and cell size (Pilbrough et al., 2009). Since this variation was shown to fluctuate within a relatively short period, phenotypic drift may partly originate from non-genetic diversity. However, a comprehensive genome and epigenome characterization of a CHO-K1 host cell line adapted to growth in three different media showed high variation in genome sequence both as a result of media adaptation and under constant culture conditions over time (Feichtinger et al., 2016). Based on these observations and as previously stressed by Frye et al. (2016), absolute genetic homogeneity in a cell culture does not seem achievable because of the genomic plasticity inherent in immortalized mammalian cell lines.

Clonal variation for recombinant CHO cell lines does not originate only from functional heterogeneity in the host cell line, as genetic heterogeneity is also introduced during the generation of recombinant cell lines (Fig. 2). For example, chromosomal aberrations were observed in 10 of 16 stable gene-amplified GFP-expressing cell lines not observed in the DG44 host cell line (Derouazi et al., 2006). In addition, significant differences in specific growth rate and q_p of subclones from a gene-amplified MAb-producing clonally derived DG44 CHO cell line have been observed (N.S. Kim et al., 1998). Thus, recombinant monoclonal cell lines generated by gene amplification are genetically and phenotypically diverse.

Does an observed phenotypic difference between two recombinant cell lines originate from clonal variation or from stable expression of an effector gene? If only one control cell line and one effector gene-expressing cell line (both monoclonal) are being investigated, it is not possible to rule out that an observed phenotypic difference originates from clonal variation (Stockholm et al., 2007). Instead, this would require a number of monoclonal cell lines in both categories to demonstrate that the difference is a consequence of expressing the effector gene. In the PDI example, only single clonally derived producer cell lines have been used as host cell lines when analysing the effect of constitutively expressing PDI (Table 1). These monoclonal producer cell lines have then been used as hosts to generate either *quasi*-monoclonal or *mono*-monoclonal PDI-expressing lines (see Fig. 2 for definitions). Thus, the observed effects of PDI expression are likely biased by clonal variation, as *quasi*-polyclonal and *mono*-monoclonal cell lines in general do not represent the average phenotype in the host cell line. This bias can be addressed if employing polyclonal producer cell lines (see 'Clonality and clonal variation' in Section 7).

In summary, caution must be exercised when investigating the effects of expressing effector genes in clonally derived producer cell lines. At a minimum, *quasi*-polyclonal or *mono*-monoclonal cell lines from two but preferably three different monoclonal producer cell lines should be examined. Alternatively, transient expression or stable episomal expression of the r-protein product circumvents the bias originating from clonal variation (see Section 7).

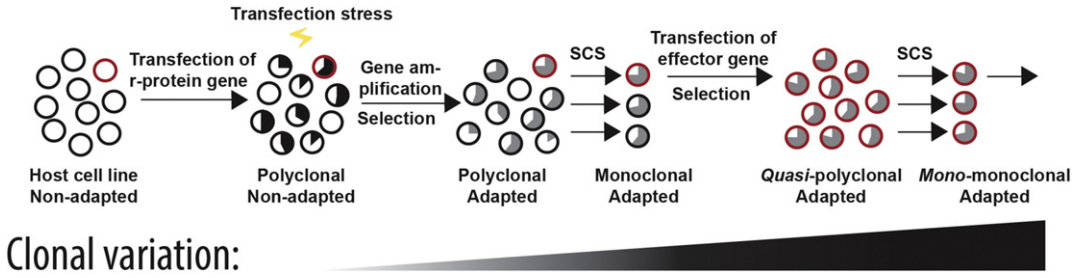


Fig. 2. Clonal variation in stable CHO cell lines. A typical process of generating a clonally derived stable CHO cell line co-expressing an r-protein and an effector gene is illustrated. CHO cells are depicted as circles, and the red circle is a high-producer cell clone selected for transfection of the effector gene. To depict the heterogeneity within the cell pools, the q_p level of each cell is illustrated as black and grey pie charts for transient and constitutive expression of the r-protein, respectively. Clonality of each phase is indicated. *Quasi*-polyclonal and *mono*-monoclonal mean that the polyclonal cell line and monoclonal cell line, respectively, both originate from a monoclonal r-protein producer cell line. Whether cells have adapted to expression of the r-protein is indicated (see 'ER stress' in Table 2). The extent of clonal variation throughout the process is depicted and is defined as the possible phenotypic difference from an average cell in the host cell line. Transfection-mediated stress is depicted (see 'Transfection stress' in Table 2). SCS: single-cell sorting.

4. Effector gene origin

The main application of CHO cells in the biopharmaceutical industry is to produce human-like glycoproteins. Although humanized, some MAb biopharmaceuticals are chimeric molecules of mouse and human amino acid sequences (Ahmadzadeh et al., 2014). Thus, human and, to some extent, murine-derived polypeptide sequences are expressed in a heterologous CHO-based context. In view of this, should the origin of an effector gene be human, CHO or mouse? In the PDI example (Table 1), all three origins have been reported, although human PDI is overrepresented (four of six studies). This observation is consistent with an overall preference for the human origin of effector genes (Fischer et al., 2015).

The argument for using a human effector gene would be that the effector gene protein is expected to directly interact with the r-protein. In other words, it is thought to facilitate a favourable interaction between two autologous (human) molecules that is less likely to take place between two heterologous molecules (direct effect; see Fig. 1). In contrast, the argument for using a CHO effector gene would be that an effector gene is expected to interact with host cell molecules (indirect effect through element X; see Fig. 1). For example, an effector gene-encoded transcription factor and a chaperone are expected to interact with host cell molecules (DNA) and the r-protein, respectively. Thus, the choice of effector gene origin depends on the function of the effector gene.

The influence of effector gene origin on CHO cell line engineering has not been systematically investigated. However, it has been shown that both CHO- and human-derived PDI can improve q_p of human MAb-related r-proteins in CHO cells (Johari et al., 2015; Mohan et al., 2007; Pybus et al., 2014) (Table 1). Moreover, XBP-1S has been shown to increase volumetric productivity and q_p in CHO cells – both the human (Becker et al., 2008; Cain et al., 2013; Pybus et al., 2014) and murine (Hansen et al., 2015; Ku et al., 2008) orthologues. In contrast, only the CHO-derived and not the human orthologue of YY1 was able to increase volumetric productivity in CHO cells and *vice versa* in human cells (Tastanova et al., 2015). As pointed out by the authors, this clearly indicates that interaction with host-specific co-factors was required.

Although there might be a few cases where a human effector gene orthologue is needed, CHO orthologues are more likely to be functional because they will be present in a non-foreign (autologous) cellular context like YY1. Thus, using CHO-derived effector genes likely increases the chances of identifying true positive effects and at the same time minimizing the risk of false negative effects. Furthermore, applying CHO-derived effector genes would enable direct comparisons of the expression level between the recombinant effector gene and the endogenous gene. Accordingly, this would entail comparisons of effector gene doses across experiments and studies (see Section 6).

Cloning CHO and Chinese hamster gene sequences has recently become a more straightforward task owing to the drafts of Chinese hamster and CHO cell line genomes (Brinkrolf et al., 2013; Lewis et al., 2013; Xu et al., 2011). Moreover, efforts to improve the quality of the Chinese hamster genome and to refine annotations are currently ongoing (Kremkow et al., 2015). The availability of correctly annotated Chinese hamster gene sequences is likely to facilitate the autologous, ectopic expression of effector genes in CHO cells in future CHO cell line engineering studies.

5. Secretion bottleneck and ER stress

A non-linear relationship between the transcript level and q_p in CHO cells shows that there is a post-transcriptional bottleneck in the biosynthetic pathway of r-proteins. Such a bottleneck has been reported in CHO cells upon transient expression of an Fc-fusion protein (Johari et al., 2015) and EPO (Ku et al., 2008), although only EPO titer and not q_p was reported for the latter example. In addition, a post-transcriptional bottleneck has been reported in stable MAb-producing CHO cells (S.J. Kim et al., 1998) and NS0 myeloma cells (Barnes et al., 2004). Moreover, a post-translational rate-limiting step (hereafter referred to as a 'secretion bottleneck'; Fig. 3A) has been demonstrated in stable CHO cell lines expressing antithrombin III (Schröder and Friedl, 1997) as well as in baby hamster kidney cells constitutively expressing activated protein C (Fann et al., 1999). In these two cases, a non-linear relationship

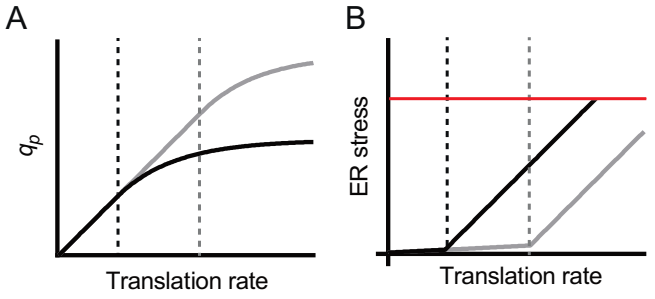


Fig. 3. Secretion bottleneck and ER stress. A) Schematic depiction of a secretion bottleneck upon co-expression of an r-protein and an effector gene. The black solid line shows the relationship between translation rate of an r-protein and q_p. The grey solid line illustrates that expression of certain effector genes gives rise to a positive effect on q_p only when a secretion bottleneck is present. The grey and black dashed lines indicate the onset translation rate for a secretion bottleneck with and without expression of a q_p-increasing effector gene, respectively. B) Schematic graph showing how secretion bottlenecks affect ER stress levels. Same colour code as in panel A is used except for the red line, which indicates the threshold between anti- and pro-apoptotic ER stress levels.

between the intracellular level of r-proteins and q_p was observed, which demonstrates that the bottleneck was downstream of translation and likely also translocation and therefore within the secretory pathway (ER, Golgi and secretory transport vesicles). Most intracellular whole MAb molecules in stable CHO cell lines have been found to be in the early part of the secretory pathway (between ER and *cis*-Golgi) for cell lines with and without a secretion bottleneck (O'Callaghan et al., 2010). The study by O'Callaghan et al. also demonstrated that bottlenecks in the biosynthetic pathway of the same MAb molecule are cell line-specific, irrespective of q_p. For example, in one cell line with q_p of 7 pcd, the folding and assembly rate of MAb was particularly slow, whereas in another cell line with q_p of 8 pcd, secretion was the rate-limiting step.

The unfolded protein response (UPR) is a homeostatic transcriptional program that is induced when the capacity of folding and processing incoming nascent polypeptides in the ER is exceeded (Moore and Hollien, 2012; Walter and Ron, 2011). This protein folding perturbation is called ER stress. The UPR seeks to restore cellular homeostasis, however, if high levels of ER stress conditions persist, the UPR will eventually become pro-apoptotic (Jäger et al., 2012; Moore and Hollien, 2012; Sano and Reed, 2013). ER stress originating from r-protein expression has been observed several times. For example, expression levels of certain r-proteins have been shown to correlate with ER stress levels upon transient expression (Johari et al., 2015; Ku et al., 2010). Once post-transcriptional and/or secretion bottleneck conditions were established, ER stress levels increased abruptly upon a higher expression level of r-protein (Fig. 3B). Moreover, cell lines expressing seemingly difficult-to-express r-proteins have been shown to have increased levels of ER stress compared to cell lines expressing easy-to-express r-proteins (Johari et al., 2015; Le Fourn et al., 2014; Sommeregger et al., 2016). These examples show that ER stress and secretion bottlenecks are intimately linked.

In two studies by Ku et al., a post-transcriptional bottleneck was identified at high gene doses of transiently expressed EPO (Ku et al., 2008, 2010). Ectopic expression of XBP-1S was found to improve the volumetric productivity of EPO only at gene doses causing ER stress and with a post-transcriptional bottleneck phenotype (Ku et al., 2008). Consequently, some ER-related effector genes – such as XBP-1S – can improve volumetric productivity and/or q_p only in conditions where a post-transcriptional bottleneck is present (Fig. 3A). Being an ER-localized protein, the effects of PDI on q_p likely depend on whether such bottlenecks are present. However, the presence of a post-transcriptional bottleneck has been investigated for only two of eleven r-proteins (Table 1). Preferably, non-secretion bottleneck as well as secretion bottleneck conditions should be established when analysing effector genes, which would minimize the risk of obtaining false negative results. Indeed, an inadequate conclusion would have been drawn if only a low gene dosage of EPO had been used by Ku et al. (2008).

Obtaining positive effects on q_p of effector genes in conditions without post-transcriptional bottlenecks is an interesting supposition. To the knowledge of the authors, no studies have methodically investigated this topic. Although not investigated, it seems likely that a post-transcriptional bottleneck was not present when expressing Rituximab in CHO-S cells in the study by Tastanova et al. (2015), as we have obtained a >10-fold higher q_p in a comparable transient expression setup in shake flasks (Hansen et al., 2015). Nevertheless, a three-fold increase in Rituximab titer was obtained upon YY1 overexpression in CHO-S (Tastanova et al., 2015). In a stable CHO-DG44-derived clone expressing MAb, the positive effect of YY1 was also observed, and this effect could not be ascribed to an increase in the cell growth rate or transcript level of MAb heavy and light chains. The increase in q_p combined with an unchanged transcript level suggests that the secretion rate per transcript must be three-fold higher, indicating that the effect is post-transcriptional. Differences in the translation rate have been predicted between stable monoclonal CHO cell lines expressing the same MAb by mathematical modelling (O'Callaghan et al., 2010), suggesting that the effect of YY1 could be an increased translation rate. Alternatively, it could be a decreased degradation of folding intermediates mediated by the ER-associated degradation pathway (Hussain et al., 2014; Merulla et al., 2013) or autophagy- and lysosomal-mediated degradation (Kim et al., 2013). Nevertheless, the YY1 example implies that a secretion bottleneck phenotype does not seem to be a prerequisite for improving q_p.

6. Effector gene dosage

Whether an effector gene is able to improve q_p depends not only on the presence of the effector gene-encoded protein but also on the expression level. In other words, the outcome of expressing an effector gene on q_p depends on the effector gene dosage (Brown and James, 2015; Xiao et al., 2014). In a study by Davis et al. (2000), CHO clones stably expressing an Fc-fusion protein with different expression levels of PDI ('low PDI', 'medium PDI' and 'high PDI') were used (Table 1). Whereas the low PDI clone had no apparent effect on product titer, the medium PDI had increased intracellular levels of the Fc-fusion protein but no effect on titer and the high PDI clone had increased intracellular levels of Fc-fusion protein and a two-fold decrease in titer. If leaving clonal variation out of consideration, this example shows that the folding machinery can be overwhelmed when an ER-localized enzyme exceeds an optimal expression level range. Like ER-resident enzymes, gene dosage titration is also important when investigating transcription factors as effector genes. Johari et al. (2015) observed increased volumetric productivity of two MAb variants when expressing either XBP-1S or cleaved activating transcription factor 6 (ATF6c) at lower gene doses and no effect at the highest gene dosage. Moreover, the gene dosage for both effector genes inversely correlated with cell growth, suggesting that the expression of XBP-1S and ATF6c generated a fitness cost. When Tastanova et al. (2015) investigated how the gene dosage of the transcription factor YY1 affected the volumetric productivity of a variety of MAb molecules and CHO cell lines, an optimal gene dosage range of YY1 was observed. These examples demonstrate the importance of gene dosage titration.

Two complementary strategies were employed to titrate the YY1 gene dosage by transient expression: the filler plasmid (empty vectors) principle and promoters with different strengths (Tastanova et al., 2015). Filler plasmids are used to titrate the effector gene-encoding plasmid (Estes et al., 2015; Rajendra et al., 2012, 2015), whereas promotors with different strengths drive transcription at different rates (Brown and James, 2015; Qin et al., 2010). The filler plasmid principle is a cost-efficient solution in transient expression-based setups, as no additional cloning work is needed besides a single plasmid preparation of the filler plasmid. Moreover, a linear relationship between the transfected plasmid load and mRNA can be expected at relatively low gene doses (Johari et al., 2015; Ku et al., 2008; Rajendra et al., 2015). When controlling the expression level of stably integrated effector genes, using a set of promotors with different strengths is the preferred choice (Brown and James, 2015).

Only a single gene dosage has been used when investigating the effect of PDI expression on nine of eleven r-proteins (Table 1). This means that, in nine of eleven cases, it is unknown whether an optimal or adverse gene dosage was used. This indicates that a substantial number of studies on ectopic expression of effector genes have only employed a single gene dosage, suggesting that the gene dosage space for many studied effector genes remains to be explored.

In summary, a wide range of effector gene doses should preferably be used in the attempt to find an optimal range of expression. Moreover, the expression level of the effector gene and the endogenous gene should be compared to facilitate comparisons across studies, which requires that CHO-derived effector genes be used (see Section 4). These efforts would facilitate higher chances of drawing valid conclusions as well as higher chances of identifying conditions where effector genes improve q_p.

7. Expression platforms

There are several different platforms for the ectopic expression of r-proteins in CHO cells. r-proteins can be expressed either transiently from a non-integrated plasmid or from a gene stably integrated into the genome. Transient gene expression (TGE) is a widely used technology for the rapid production of r-proteins, usually during a 2–10 day batch or fed-batch process (Baldi et al., 2007). TGE is the preferred production method during the early stages of drug development for preclinical studies, as a sufficient quantity of r-protein can be obtained within a short period of time (Kim et al., 2012). However, the lower protein yield achieved with TGE in CHO cells has historically been a major drawback compared to the substantially higher yields obtained with stable gene expression (SGE). Thus far, extensive efforts to improve TGE yields in CHO cells have been made by optimizing the culture environment (Galbraith et al., 2006; Ye et al., 2009), transfection efficiency (Mozley et al., 2014; Rajendra et al., 2012, 2015), vector systems (Cho et al., 2001; Mariati et al., 2010) and host cell line (Cain et al., 2013; Daramola et al., 2014; Macaraeg et al., 2013). Therefore, TGE has become a robust and flexible system, applicable to multiple r-proteins, expression volumes and bioprocesses with substantially increased yields of up to 3 g/L for MAb-producing CHO cells (Liu et al., 2015).

SGE is a prerequisite for the stable, large-scale manufacturing of r-proteins as biologics for clinical applications (Noh et al., 2013; Wurm, 2004). Stable cell lines are typically generated using selection based on a selective marker expressed on the same plasmid as the r-protein (Priola et al., 2016). If desired, the gene encoding the r-protein can be amplified to increase q_p, although gene-amplification in general reduces the stability of volumetric productivity and q_p in extended culture conditions (Chusainow et al., 2009). The two most widely used gene amplification systems are the DHFR system and the glutamine synthetase system (Noh et al., 2013). Nowadays, companies report titers of >10 g/L for MAb production (Gronemeyer et al., 2014). As an alternative to random transgene integration, *piggyBac* or *sleeping beauty*-mediated transposition can be used, where transgene integration into highly transcribed regions of the host genome is favoured. This leads to a generally higher rate of transgene transcription compared to the random integration of plasmids (Ding et al., 2005; Galvan et al., 2009; Wilson et al., 2007). Consequently, higher q_p and stability compared to stable producer clones obtained by random integration have been achieved with MAb titers of up to 7.6 g/L (Matasci et al., 2011; Rajendra et al., 2016).

Inducible expression (IE) platforms support the idea of a regulated, biphasic r-protein expression throughout the production phase, *e.g.*, being turned off during growth and turned on only during the late exponential and stationary phase. The inducible nature of the IE platform provides a powerful expression system for r-proteins that might confer toxicity when being expressed by constitutive promoters. Several systems with either repressor or activator configurations have been developed for mammalian cells to achieve tight control of gene expression. Among many, there are i) antibiotic-based regulation systems, such as the Tet-Off-On system regulated by tetracyclin (Gossen and Bujard, 1992; Mohan et al., 2007), the Pip system regulated by streptogramin (Fussenegger et al., 2000) and the E.REX system regulated by macrolides (Weber et al., 2002); ii) an aptamer-based regulation system (Werstuck and Green, 1998); and iii) the cumate gene-switch (Gaillet et al., 2010; Mullick et al., 2006). All these inducible systems have been successfully used in CHO cell lines, and a 0.24 g/L titer of an Fc-fusion protein has been reported for this system (Gaillet et al., 2010).

Expression from replicating episomes in CHO cells reduces the loss of plasmid and thereby prolongs the nuclear retention time of the plasmid after cell divisions (Van Craenenbroeck et al., 2000). Episomal replication can be achieved using, *e.g.*, the Polyomavirus large T gene (PyLT) and its origin of replication (PyOri) (Heffernan and Dennis, 1991), while plasmid maintenance and segregation can be accomplished using Epstein-Barr virus nuclear antigen-1 (EBNA-1) and its origin of replication (OriP) (Lupton and Levine, 1985; Yates et al., 1984). Using these two sets of complementing viral components, the episomal platform was reported to increase and prolong TGE yields of a growth hormone and MAb in CHO cells in comparison to non-replicating plasmid controls (Codamo et al., 2011; Kunaparaju et al., 2005). Without antibiotic-based selection, expression from replication-proficient episomes can be regarded only as a quasi-stable gene expression platform, as episomes eventually will be lost with a half-life of approximately 8–9 days (Silla et al., 2006). Without selection, titers of up to 2 g/L MAb have been reported (Daramola et al., 2014). However, if antibiotic-based selection is applied, a stable pool of cells can be obtained that stably replicates and segregate episomes for more than two months (Silla et al., 2006). Thus, when combined with selection, the system can be regarded as stable episomal expression (SEE).

Although all the described expression systems have been used successfully to express r-proteins in CHO cells, each expression platform has advantages and disadvantages when used for effector gene expression. In Table 2, four selected expression systems are compared as platforms for analysing how effector genes affect q_p. It is important to note that other expression systems are available – such as stable pools and monoclonal cell lines obtained by lentiviral vector-mediated gene transfer (Oberbek et al., 2011) or stable pools obtained by *piggyBac* transposons (Matasci et al., 2011). In addition, the four expression systems can be combined in different ways – such as using IE (Chung et al., 2004) or TGE (Hayes et al., 2010) in a monoclonal stable producer cell line. Nevertheless, the selected expression platforms serve to highlight important aspects to consider when choosing a platform and interpreting the final data.

7.1. Duration – from transfection to answer

The preparation of plasmids carrying effector and r-protein genes takes approximately the same time for all four expression platforms. However, the time span from transfection to an answer varies substantially. TGE is clearly the fastest track, as experiments are typically done within two to three days (Hansen et al., 2015; Johari et al., 2015; Ku et al., 2008). SEE is also relatively fast, as antibiotic-based selected SEE cells can be obtained within two weeks (Silla et al., 2006). In contrast, both IE and SGE platforms require generation and characterization of monoclonal cell lines that take months to perform (Noh et al., 2013).

7.2. Gene dosage and combining genes

In terms of gene dosage titration (Rajendra et al., 2012) and combining genes (Nishimiya et al., 2013), TGE offers complete flexibility, as a variety of different plasmids (filler plasmid and/or plasmids carrying different effector genes) can be co-expressed. This is not the case for SEE, as the copy numbers of two plasmids with different genes expressed, in some cases, will drift towards selection for the plasmid/gene giving rise to the lowest fitness cost (personal communication with Dr. Mikael Rørdam Andersen, Technical University of Denmark). To avoid this drift, the r-protein gene and effector gene must be co-expressed from the same plasmid. Relatively few cell lines are manageable to maintain in parallel, which decreases the throughput of gene combinations for the IE and SGE platforms. Because the effector gene is constitutively expressed in the described IE and SGE platforms (Table 2), the effector gene dosage is not titratable in these platforms. Notably, recent advances in promoter engineering in CHO cells make it possible to precisely control the gene dosage in TGE and possibly also in SEE, IE and SGE platforms spanning over two orders of magnitude (Brown et al., 2014).

7.3. Transfection stress and variability

A considerable drawback of TGE is the cytotoxic effect from transfection, here termed 'transfection stress' (Fig. 2). Different transfection reagents induce different levels of cytotoxicity in terms of impeded cell

Table 2
Selected expression platforms available for analysing how effector genes affect q_p in CHO cells.

	Transient gene expression (TGE)[a]	Stable episomal expression (SEE)[b]	Inducible expression (IE)[c]	Stable gene expression (SGE)[d]
Time from transfection to answer	**2–3 days**[e]	**~2 weeks**	>2 months	>2 months
Gene dosage easily titrated	**Yes**	No	**Yes**	No
Genes easily combined	**Yes**	No	No	No
Transfection stress and variability	Yes	**No**	**No**	**No**
Cells adapted to ER stress	**No**	**(No)**[f]	**No**	Yes
Secretion bottleneck easily obtained	**Yes**	**(Yes)**[g]	No	**Yes**
Clonality	**Polyclonal**	**Polyclonal**	*Mono*-monoclonal[h]	*Mono*-monoclonal[h]
Clonal variation	**No**	**No**	Yes	Yes
Validation by SGE needed	Yes	Yes	Yes	**No**

[a] Transient, plasmid-based co-expression of r-protein and effector genes. Examples: Hansen et al. (2015), Johari et al. (2015), Ku et al. (2008), and Tastanova et al. (2015).
[b] Stable episomal co-expression of r-protein and effector genes after antibiotic-based selection (without selection resembles TGE). No published examples of co-expression are available, but the expression system is described here: Kunaparaju et al. (2005) and Silla et al. (2006). The r-protein gene and effector gene must be co-expressed from the same plasmid to prevent a drift in copy number between the two genes (see 'Gene dosage and combining genes' in Section 7 for details).
[c] Inducible expression of a stably integrated gene encoding the r-protein in a clonally derived cell line constitutively expressing an effector gene of interest. No reported examples of this specific setup are available to the best of the authors' knowledge, but examples of the expression system are described here: Mohan et al. (2007) and Mohan and Lee (2010).
[d] Constitutive (stable) co-expression of stably integrated genes encoding r-protein and effector genes in clonally derived cell lines. Examples: Dreesen and Fussenegger (2011), Haredy et al. (2013), and Tastanova et al. (2015).
[e] Features regarded as advantages are highlighted in boldface type.
[f] See 'ER stress' in Section 7 for details.
[g] See 'Secretion bottleneck' in Section 7 for details.
[h] *Quasi*-polyclonal cell lines can also be used; however, *mono*-monoclonal cell lines are commonly used. For definition of *quasi*-polyclonal and *mono*-monoclonal cell lines, see Fig. 2.

growth rate and a drop in viability (our own unpublished observations), as well as the inhibition of protein synthesis (Underhill et al., 2003). Notably, transfection stress can be reduced through process or cell engineering optimization (Johari et al., 2015; Macaraeg et al., 2013; Majors et al., 2008). In addition, variability in transfection efficiency is an inherent problem for TGE (Hansen et al., 2015; Liu et al., 2008); however, optimization and careful selection of the transfection method can reduce variability substantially (Davies et al., 2013). For the SEE, IE and SGE platforms, transfection stress is most likely not an issue because of the lengthy (≥2 weeks) antibiotic-based selection processes.

7.4. ER stress

The expression of r-proteins is inducible for the TGE (transfection) and IE (addition of inducing agents) platforms. In contrast, r-proteins are constitutively expressed during the selection processes in the SEE and SGE platforms. Whereas the effector gene and the r-protein are co-expressed during the selection process in the SEE platform, the r-protein is typically first integrated into the genome (monoclonal cell line; Fig. 2) and subsequently the effector gene (*quasi*/*mono*-monoclonal cell line; Fig. 2) in the SGE platform. Thus, in contrast to the TGE, IE and SEE platforms, the SGE platform subjects cells to high expression level of the r-protein before investigating whether the effector gene of interest can alleviate the ER stress originating from r-protein expression. In other words, some cells with a high r-protein expression level and consequently a high level of ER stress will likely not survive during the selection processes (Hu et al., 2013) (Fig. 3B). These highly relevant, stressed cells will therefore not be part of the effector gene test case, in contrast to the TGE and IE platforms. This could introduce a bias towards cells with lower level of ER stress or cells inherently capable of coping with high levels of ER stress. It is important to note that stable producer cells adapted to a permanent increase in ER stress levels can be obtained (Sommeregger et al., 2016); however, the ER stress levels in these clones are not high enough to induce apoptosis. Instead, conditions with pro-apoptotic levels of ER stress are more likely to be established by TGE and IE because of the inducible nature of the two systems. Notably, the SEE platform does not allow for pro-apoptotic levels of ER stress to persist throughout the selection process. However, it does enable an effector gene of interest to alleviate pro-apoptotic levels of ER stress, because the effector gene and r-protein are co-expressed throughout the entire process.

7.5. Secretion bottleneck

Post-transcriptional and secretion bottlenecks have been reported in CHO cells for TGE (Johari et al., 2015; Ku et al., 2008) and SGE platforms (O'Callaghan et al., 2010). To the best of the authors' knowledge, a secretion bottleneck has not been reported in CHO cells for IE platforms, which likely is a result of a presumably low transcription rate of the r-protein gene compared to the TGE and SGE platforms. Naturally, the occurrence of a secretion bottleneck is protein-specific, and it is likely that, for some difficult-to-express proteins, a secretion bottleneck can be readily obtained. A post-transcriptional bottleneck in CHO cells has been reported upon expression from replicating episomes (Pybus et al., 2014); however, this was without antibiotic-based selection and is therefore not regarded as SEE (Table 2). A relatively high average q_p (10 pcd for MAbs) can be obtained using the SEE platform (personal communication with Dr. Meelis Kadaja, Icosagen Cell Factory Ltd., Estonia), implying that post-transcriptional and/or secretion bottleneck conditions can be established using the SEE platform.

7.6. Clonality and clonal variation

When using TGE and SEE platforms, a representative pool of cells from the host cell line is being tested (polyclonality), whereas clonally derived cells are typically used in the IE and SGE platforms (monoclonality). Whilst polyclonal cells represent an average phenotype of all cells in the host cell line, monoclonal cell lines typically represent a favoured phenotype identified within the functionally heterogeneous pool of cells (Fig. 2). Thus, different questions are being put forward when expressing effector genes in polyclonal and monoclonal cells. If q_p is increased upon effector gene expression in polyclonal cells, this finding can likely be transferred to the majority of cells within the host cell line. This is not necessarily the case for clonally derived cell lines expressing effector genes (*mono*-monoclonal; see Section 3). Thus, findings from a clonally derived cell line might be conditional; that is, the effect only applies to the cell line in question (O'Callaghan et al., 2010).

7.7. Validation

Since SGE is the preferred production platform for therapeutic proteins in the industry (Noh et al., 2013), any effector gene should preferably be validated in a SGE context – that is, in high-producer

monoclonal cell lines. For example, effects obtained from TGE are sometimes transferable to a SGE context (Tastanova et al., 2015) and sometimes not (Mohan and Lee, 2010). The need for validation is obviously a drawback of the TGE, SEE and IE platforms. However, if measures described here are taken – such as applying different effector gene doses and establishing conditions with and without post-transcriptional and/or secretion bottlenecks – the success rate of validating effector genes is likely to increase.

In summary, TGE is the preferred platform for screening effector genes because effector genes can readily be combined, gene dosage can easily be titrated and post-transcriptional and/or secretion bottlenecks can be obtained; in addition, TGE is not hampered by clonal variation and adaptation to ER stress (Table 2). A considerable drawback is the presence of transfection stress and transfection variability, which, however, can be addressed to some extent through process optimization. Once combinations of effector genes and gene doses have been identified, they need to be validated, preferably in multiple stable *mono*-monoclonal cell lines derived from different host cell lines.

8. Product quality

Although volumetric productivity and q_p are important measures, the ability to enhance product quality is equally important in the biopharmaceutical industry (Gramer, 2014). Volumetric productivity is the mass of produced r-protein per volume per culture time in cell-free supernatants. However, a significant fraction of the protein mass can be misfolded or incorrectly processed variants of the r-protein (Kunert and Reinhart, 2016). In such cases where product quality is impaired, an increase in product titer does not necessarily correlate with an increase in the yield of bioactive r-protein. Quality attributes defining the overall product quality are molecularly diverse features of the r-protein, such as misfolding/aggregation, incorrectly processed propeptides, enzymatic degradation and amino acid sequence variations (Gramer, 2014). Moreover, glycosylation is probably the most important quality attribute of therapeutic glycoproteins, because the pharmacokinetic effects of undesired glycosylation patterns can be decreased drug efficacy or increased antigenicity (Bertozzi et al., 2009; Butler and Spearman, 2014). Recent advances in glycoengineering and descriptions of how effector genes can be used to modulate glycosylation have been described in recent reviews (Bennun et al., 2016; Dicker and Strasser, 2015; Spahn and Lewis, 2014). Because of scope limitations, only misfolding/aggregation and propeptide processing in relation to effector genes will be described here.

Some secreted proteins are expressed as inactive proprotein precursors containing one or more inhibitory propeptides that need to be proteolytically cleaved off by propeptidases before full activity is achieved (Wiederanders et al., 2003). Since these propeptides in general are essential for protein folding (Chen and Inouye, 2008), propeptide-containing r-proteins need to be expressed as proproteins to prevent misfolding and degradation. However, there are multiple examples of insufficient cleavage of propeptides of r-proteins expressed in CHO cells, leading to the secretion of a mixture of inactive proprotein and mature, correctly processed r-protein (Preininger et al., 1999; Sathyamurthy et al., 2012, 2015; Wasley et al., 1993). In these studies, the ectopic expression of effector genes encoding propeptidases increased the percentage of correctly processed r-proteins. This demonstrates that the propeptidase machinery within CHO cells can be the bottleneck for the production of propeptide-containing r-proteins. Solving this type of bottleneck is unique to propeptide-containing r-proteins and is therefore not generally applicable for enhancing q_p of other types of r-proteins.

Product quality in terms of misfolding and aggregation is intimately linked to the secretory pathway. Schröder et al. (2002) demonstrated that increasing the r-protein expression level of antithrombin III through gene amplification in stable clonally derived CHO cells gave rise to the formation of disulphide-bonded aggregates. Similarly, lowering the gene dosage when transiently expressing an Fc-fusion protein in HEK cells has also been shown to reduce aggregate formation (Estes et al., 2015). In the PDI example (Table 1), the product quality has been investigated in only one of eleven examples. Here, a 1.2-fold increase in q_p of an Fc-fusion protein was observed upon PDI overexpression (Johari et al., 2015). However, PDI overexpression was found to impair product quality through a 1.4-fold increase in the formation of high molecular aggregates. Thus, the overall effect of PDI overexpression was a decrease in the number of correctly folded Fc-fusion protein molecules produced per cell per unit time. This study elegantly illustrates the importance of analysing quality attributes when studying how effector genes affect q_p.

9. Concluding remarks and future directions

A desirable goal of engineering the secretory capacity of CHO cells is to generate a universally competent cell line ('super-CHO') capable of manufacturing all r-proteins in demand. In general, this seems difficult to achieve because of product specificity (McLeod et al., 2011). For example, it has been reported that host cell line subclones being able to produce a MAb at relatively high q_p levels compared to the host cell pool were not able to produce high levels of an Fc-fusion protein (O'Callaghan et al., 2015). Moreover, two similar antibody fragments have been shown to differentially influence the cell's physiology in terms of different proteomic responses in a product-specific manner (Sommeregger et al., 2016). Similarly, effector genes have been shown to increase q_p in a MAb-variant-specific manner (Pybus et al., 2014). However, some effector genes can increase q_p for different types of r-proteins in CHO cells. For example, XBP-1S was able to increase q_p and/or product titer for secreted alkaline phosphatase (SEAP) and vascular endothelial growth factor (VEGF) (Tigges and Fussenegger, 2006), EPO (Ku et al., 2008), MAb (Becker et al., 2008; Cain et al., 2013; Pybus et al., 2014), an Fc-fusion protein (Johari et al., 2015) and α_1-antitrypsin and C1 esterase inhibitor (Hansen et al., 2015). Similarly, YY1 expression was able to increase product titer for SEAP, VEGF and MAb (Tastanova et al., 2015). This type of *general effector genes* seems to be able to traverse product specificity. It is important to note that general effector genes most likely also will fall short when it comes to r-proteins that require specialized modifications, such as the cleavage of propeptides (Sathyamurthy et al., 2015) and γ-carboxylation of clotting factors (Kumar, 2015). Thus, positive effects of general effector genes are likely not generally valid *per se*, but the effect may be retrieved once specialized post-translation modifications are no longer constituting a bottleneck.

The function of proteins encoded by general effector genes is probably conceptually different compared to effector gene-encoded proteins with more product-type specific effects (*i.e.* not general effector genes). In order to traverse product-specificity, the effect of general effector genes is probably multifaceted. Such multifaceted effects can be the result of simultaneously changing the expression level of several genes. Several cellular processes and/or molecules are able to confer such effects; for example phosphorylation (Rajesh et al., 2015), microRNAs (Barron et al., 2011; Hackl et al., 2012), transcription factors (Adachi et al., 2008; Harding et al., 2003) and histone marks (Dahodwala and Sharfstein, 2014). The molecules involved in these processes do not interact directly with the r-protein and the effects are therefore indirect (Fig. 1). An attractive advantage of using general effector genes is the possibility of modulating the expression level of more genes than is currently possible when co-expressing multiple single effector genes. However, some effects from general effector genes may also be adverse due to pleiotropic, undesired regulation of a subset of genes (Stearns, 2010). Thus, investigating and eliminating these potential adverse effects of a general effector gene of interest may improve the desired phenotype.

Although single general effector genes seem to be part of the answer, multiple (general) effector genes are likely a more coherent and robust

solution for increasing the secretory capacity (Harreither et al., 2015; Seth et al., 2007). In fact, multiple reactions within the biosynthetic pathway of MAb seem amenable for improvement (O'Callaghan et al., 2010). This notion is supported by non-limiting examples of studies demonstrating an increase in volumetric productivity or q_p in CHO cells upon the expression of effector genes related to translation and translocation (Le Fourn et al., 2014), ER folding processes (Borth et al., 2005; Chung et al., 2004; Hwang et al., 2003) and protein transport and secretion (Florin et al., 2009; Peng et al., 2011; Peng and Fussenegger, 2009) (Fig. 1). Several reports have demonstrated synergistic effects on the volumetric productivity or q_p of co-expressing multiple genes in CHO cells (Cain et al., 2013; Le Fourn et al., 2014; Mohan and Lee, 2010; Peng and Fussenegger, 2009), demonstrating the potential of combining effector genes. In addition to effector gene overexpression, downregulating or knocking out endogenous genes might be equally important for increasing q_p (Feichtinger et al., 2016; Harreither et al., 2015). Finally, supplementing small-molecule chemical chaperones combined with effector gene overexpression has been shown to exceed the limits of functional heterogeneity present in the CHO-S host cell line (Johari et al., 2015).

When modulating the expression of multiple effector genes, the relative stoichiometry between gene transcripts (gene dosage) becomes an important aspect to consider to obtain an optimal effect on q_p (Xiao et al., 2014). However, q_p is not the only relevant factor: the time integral of viable cell density (accumulated viable cell number) and product quality attributes are important measures to control in production bioprocesses for increasing the product yield (Gramer, 2014; Kim et al., 2012). This is not a simple task, because process parameters (for example metabolite concentration and osmolality) in a bioreactor can change throughout a bioprocess (Justice et al., 2011). These dynamic changes affect the overall physiology and fitness of the cell and are likely to affect transcription rates in a promoter-specific manner (Brown and James, 2015). Such bioprocess-dependent variables can therefore interfere with the gene dosage of single and multiple effector genes. Thus, highly context-specific promoter designs are needed to simultaneously control the expression level of multiple effector genes during production bioprocesses (see recent review (Brown and James, 2015)). In combination with improved designs of vector elements, site-specific integration of expression cassettes into so-called safe harbour sites (Papapetrou et al., 2011) in the CHO genome has the potential of enabling predictable and stable expression levels of effector genes in clonally derived cell lines. Based on the game-changing clustered regularly interspaced short palindromic repeat (CRISPR)/CRISPR-associated (Cas) genome editing technology (Mali et al., 2013), site-specific integration into CHO cells is now possible at a relatively low cost with applicable efficiencies (Lee et al., 2015, 2016).

The global cellular view of systems biology and 'omics-based approaches holds a unique potential to identify pathways and gene networks comprising novel effector genes (Datta et al., 2013; Gutierrez and Lewis, 2015; Kildegaard et al., 2013). These networks would most likely be tedious to identify using a classical reductionist's approach where typically only a few genes are being tested. Indeed, as recently stated by Clarke and Lee (2014), the CHO community is currently generating 'omics data at an ever-increasing rate. This is supported by the following non-exhaustive examples on CHO 'omics studies: genomics (Lewis et al., 2013; Xu et al., 2011), transcriptomics (Becker et al., 2011; Doolan et al., 2008), translatomics (Courtes et al., 2013; Kallehauge et al., 2016, in press), proteomics (Baycin-Hizal et al., 2012; Carlage et al., 2009), metabolomics (Selvarasu et al., 2012; Zang et al., 2011) and integrative 'omics (Clarke et al., 2012). Moreover, a consensus genome-scale reconstruction of CHO cell metabolism has now been established (Hefzi et al., 2016) which potentially opens up new designs of CHO cell lines (Kaas et al., 2014). In fact, using the genome-scale model to assess resource utilization suggests modulating expression of effector genes related to the secretory pathway is a more efficient way of improving q_p compared to typical bioprocess treatments such as sodium butyrate and hypothermia (Hefzi et al., 2016).

As mentioned by Hussain et al. (2014), some CHO 'omics studies have probably been hampered by clonal variation (Fig. 2), because only a few clonally derived cell lines were compared. As previously described, CHO cells – even within a host cell line – are functionally and genomically diverse and only a fraction of the cells are universally competent in terms of r-protein production (O'Callaghan et al., 2010). Moreover, the 'omics profile of gene-amplified clonally derived high producer cells are probably vastly different compared to before transfection, due to, for example, adaptation to ER stress. Thus, an 'omics profile of a high producer might comprise information on the end point of coping with a high secretion load. However, the profile does not necessarily contain any information on the cellular mechanism(s) involved in coping with the initial secretion load that the cell is subjected to right after transfection and/or gene amplification events. Instead, the intrinsic capability for coping with the initial secretion load is more likely to be found in non-transfected cells (non-adapted cells; Fig. 2). In fact, Harreither et al. (2015) were able to show that the capability to obtain high q_p from transient expression is reflected in the native transcriptome of host cell line subclones. Besides being a potential source of novel candidate effector genes, 'omics-based profiling of non-adapted monoclonal cells holds the potential to give fundamental insight into the functional heterogeneity of CHO cells.

Historically, the CHO lineage was not generated for r-protein production but for the investigation of molecular and classical cell genetics (Wurm, 2013). Moreover, CHO cells do not originate from dedicated secretory tissues like plasma cells and β-cells. By default, CHO cells therefore seem to be a suboptimal cell line for r-protein production, with many opportunities for improvement. The significant advances achieved through the ectopic expression of effector genes (Fischer et al., 2015) are likely only the tip of the iceberg, as relatively few of the approximately 20,000 genes in the Chinese hamster genome (Kremkow et al., 2015; Lewis et al., 2013) have been investigated. Furthermore, combining effector genes increases the number of possible test conditions dramatically. However, the number of possible test conditions can be reduced through systems biology- and 'omics-based approaches. Despite valuable guidance from systems biology, several novel effector genes (single genes and combinations) remain to be investigated. In conclusion, the cellular and experimental factors described here will likely aid future investigations of effector genes and thereby accelerate the development of CHO cell lines with optimized secretory capacity.

Acknowledgements

This work was supported by funding from the Novo Nordisk Foundation. N.P. was also funded by the European Union's Horizon 2020 research and innovation program under the Marie Sklodowska-Curie grant agreement (reference 642663).

References

Adachi, Y., Yamamoto, K., Okada, T., Yoshida, H., Harada, A., Mori, K., 2008. ATF6 is a transcription factor specializing in the regulation of quality control proteins in the endoplasmic reticulum. Cell Struct. Funct. 33, 75–89.

Aggarwal, R.S., 2014. What's fueling the biotech engine-2012 to 2013. Nat. Biotechnol. 32, 32–39.

Ahmadzadeh, V., Farajnia, S., Feizi, M.A.H., Nejad, R.A.K., 2014. Antibody humanization methods for development of therapeutic applications. Monoclon. Antib. Immunodiagn. Immunother. 33, 67–73.

Appenzeller-Herzog, C., Ellgaard, L., 2008. The human PDI family: versatility packed into a single fold. Biochim. Biophys. Acta, Mol. Cell Res. 1783, 535–548.

Baldi, L., Hacker, D.L., Adam, M., Wurm, F.M., 2007. Recombinant protein production by large-scale transient gene expression in mammalian cells: state of the art and future perspectives. Biotechnol. Lett. 29, 677–684.

Barnes, L.M., Bentley, C.M., Dickson, A.J., 2004. Molecular definition of predictive indicators of stable protein expression in recombinant NS0 myeloma cells. Biotechnol. Bioeng. 85, 115–121.

Barnes, L.M., Moy, N., Dickson, A.J., 2006. Phenotypic variation during cloning procedures: analysis of the growth behavior of clonal cell lines. Biotechnol. Bioeng. 94, 530–537.

Barron, N., Sanchez, N., Kelly, P., Clynes, M., 2011. MicroRNAs: tiny targets for engineering CHO cell phenotypes? Biotechnol. Lett. 33, 11–21.
Baycin-Hizal, D., Tabb, D.L., Chaerkady, R., Chen, L., Lewis, N.E., Nagarajan, H., Sarkaria, V., Kumar, A., Wolozny, D., Colao, J., Jacobson, E., Tian, Y., O'Meally, R.N., Krag, S.S., Cole, R.N., Palsson, B.O., Zhang, H., Betenbaugh, M., 2012. Proteomic analysis of Chinese hamster ovary cells. J. Proteome Res. 11, 5265–5276.
Becker, E., Florin, L., Pfizenmaier, K., Kaufmann, H., 2008. An XBP-1 dependent bottle-neck in production of IgG subtype antibodies in chemically defined serum-free Chinese hamster ovary (CHO) fed-batch processes. J. Biotechnol. 135, 217–223.
Becker, J., Hackl, M., Rupp, O., Jakobi, T., Schneider, J., Szczepanowski, R., Bekel, T., Borth, N., Goesmann, A., Grillari, J., Kaltschmidt, C., Noll, T., Pühler, A., Tauch, A., Brinkrolf, K., 2011. Unraveling the Chinese hamster ovary cell line transcriptome by next-generation sequencing. J. Biotechnol. 156, 227–235.
Bennun, S.V., Hizal, D.B., Heffner, K., Can, O., Zhang, H., Betenbaugh, M.J., 2016. Systems glycobiology: integrating glycogenomics, glycoproteomics, glycomics, and other 'omics data sets to characterize cellular glycosylation processes. J. Mol. Biol. 428, 3337–3352.
Berting, A., Farcet, M.R., Kreil, T.R., 2010. Virus susceptibility of Chinese hamster ovary (CHO) cells and detection of viral contaminations by adventitious agent testing. Biotechnol. Bioeng. 106, 598–607.
Bertozzi, C.R., Freeze, H.H., Varki, A., Esko, J.D., 2009. Glycans in biotechnology and the pharmaceutical industry. In: Varki, A., Cummings, R.D., Esko, J.D., Freeze, H.H., Stanley, P., Bertozzi, C.R., Hart, G.W., Etzler, M.E. (Eds.), Essentials of Glycobiology, second ed. Cold Spring Harbor Laboratory Press, NY.
Borth, N., Mattanovich, D., Kunert, R., Katinger, H., 2005. Effect of increased expression of protein disulfide isomerase and heavy chain binding protein on antibody secretion in a recombinant CHO cell line. Biotechnol. Prog. 21, 106–111.
Brinkrolf, K., Rupp, O., Laux, H., Kollin, F., Ernst, W., Linke, B., Kofler, R., Romand, S., Hesse, F., Budach, W.E., Galosy, S., Müller, D., Noll, T., Wienberg, J., Jostock, T., Leonard, M., Grillari, J., Tauch, A., Goesmann, A., Helk, B., Mott, J.E., Pühler, A., Borth, N., 2013. Chinese hamster genome sequenced from sorted chromosomes. Nat. Biotechnol. 31, 694–695.
Brown, A.J., James, D.C., 2015. Precision control of recombinant gene transcription for CHO cell synthetic biology. Biotechnol. Adv. 34, 492–503.
Brown, A.J., Sweeney, B., Mainwaring, D.O., James, D.C., 2014. Synthetic promoters for CHO cell engineering. Biotechnol. Bioeng. 111, 1638–1647.
Butler, M., Spearman, M., 2014. The choice of mammalian cell host and possibilities for glycosylation engineering. Curr. Opin. Biotechnol. 30, 107–112.
Cain, K., Peters, S., Hailu, H., Sweeney, B., Stephens, P., Heads, J., Sarkar, K., Ventom, A., Page, C., Dickson, A., 2013. A CHO cell line engineered to express XBP1 and ERO1-Lα has increased levels of transient protein expression. Biotechnol. Prog. 29, 697–706.
Carlage, T., Hincapie, M., Zang, L., Lyubarskaya, Y., Madden, H., Mhatre, R., Hancock, W.S., 2009. Proteomic profiling of a high-producing Chinese hamster ovary cell culture. Anal. Chem. 81, 7357–7362.
Chen, Y.-J., Inouye, M., 2008. The intramolecular chaperone-mediated protein folding. Curr. Opin. Struct. Biol. 18, 765–770.
Cho, M.S., Yee, H., Brown, C., Jeang, K.T., Chan, S., 2001. An oriP expression vector containing the HIV-1 Tat/TAR transactivation axis produces high levels of protein expression in mammalian cells. Cytotechnology 37, 23–30.
Chung, J.Y., Lim, S.W., Hong, Y.J., Hwang, S.O., Lee, G.M., 2004. Effect of doxycycline-regulated calnexin and calreticulin expression on specific thrombopoietin productivity of recombinant Chinese hamster ovary cells. Biotechnol. Bioeng. 85, 539–546.
Chusainow, J., Yang, Y.S., Yeo, J.H.M., Ton, P.C., Asvadi, P., Wong, N.S.C., Yap, M.G.S., 2009. A study of monoclonal antibody-producing CHO cell lines: what makes a stable high producer? Biotechnol. Bioeng. 102, 1182–1196.
Clarke, C., Lee, K.H., 2014. Special focus: an 'omics approach to Chinese hamster ovary based pharmaceutical bioprocessing. Pharm. Bioprocess. 2, 351–353.
Clarke, C., Henry, M., Doolan, P., Kelly, S., Aherne, S., Sanchez, N., Kelly, P., Kinsella, P., Breen, L., Madden, S.F., Zhang, L., Leonard, M., Clynes, M., Meleady, P., Barron, N., 2012. Integrated miRNA, mRNA and protein expression analysis reveals the role of post-transcriptional regulation in controlling CHO cell growth rate. BMC Genomics 13, 656.
Codamo, J., Munro, T.P., Hughes, B.S., Song, M., Gray, P.P., 2011. Enhanced CHO cell-based transient gene expression with the epi-CHO expression system. Mol. Biotechnol. 48, 109–115.
Courtes, F.C., Lin, J., Lim, H.L., Ng, S.W., Wong, N.S.C., Koh, G., Vardy, L., Yap, M.G.S., Loo, B., Lee, D.Y., 2013. Translatome analysis of CHO cells to identify key growth genes. J. Biotechnol. 167, 215–224.
Dahodwala, H., Sharfstein, S.T., 2014. Role of epigenetics in expression of recombinant proteins from mammalian cells. Pharm. Bioprocess. 2, 403–419.
Daramola, O., Stevenson, J., Dean, G., Hatton, D., Pettman, G., Holmes, W., Field, R., 2014. A high-yielding CHO transient system: coexpression of genes encoding EBNA-1 and GS enhances transient protein expression. Biotechnol. Prog. 30, 132–141.
Datta, P., Linhardt, R.J., Sharfstein, S.T., 2013. An 'omics approach towards CHO cell engineering. Biotechnol. Bioeng. 110, 1255–1271.
Davies, S.L., Lovelady, C.S., Grainger, R.K., Racher, A.J., Young, R.J., James, D.C., 2013. Functional heterogeneity and heritability in CHO cell populations. Biotechnol. Bioeng. 110, 260–274.
Davis, R., Schooley, K., Rasmussen, B., Thomas, J., Reddy, P., 2000. Effect of PDI overexpression on recombinant protein secretion in CHO cells. Biotechnol. Prog. 16, 736–743.
Derouazi, M., Martinet, D., Besuchet Schmutz, N., Flaction, R., Wicht, M., Bertschinger, M., Hacker, D.L., Beckmann, J.S., Wurm, F.M., 2006. Genetic characterization of CHO production host DG44 and derivative recombinant cell lines. Biochem. Biophys. Res. Commun. 340, 1069–1077.
Dicker, M., Strasser, R., 2015. Using glyco-engineering to produce therapeutic proteins. Expert. Opin. Biol. Ther. 15, 1501–1516.
Ding, S., Wu, X., Li, G., Han, M., Zhuang, Y., Xu, T., 2005. Efficient transposition of the piggyBac (PB) transposon in mammalian cells and mice. Cell 122, 473–483.
Doolan, P., Melville, M., Gammell, P., Sinacore, M., Meleady, P., McCarthy, K., Francullo, L., Leonard, M., Charlebois, T., Clynes, M., 2008. Transcriptional profiling of gene expression changes in a PACE-transfected CHO DUKX cell line secreting high levels of rhBMP-2. Mol. Biotechnol. 39, 187–199.
Dreesen, I.A.J., Fussenegger, M., 2011. Ectopic expression of human mTOR increases viability, robustness, cell size, proliferation, and antibody production of chinese hamster ovary cells. Biotechnol. Bioeng. 108, 853–866.
Durocher, Y., Butler, M., 2009. Expression systems for therapeutic glycoprotein production. Curr. Opin. Biotechnol. 20, 700–707.
Estes, B., Hsu, Y.R., Tam, L.T., Sheng, J., Stevens, J., Haldankar, R., 2015. Uncovering methods for the prevention of protein aggregation and improvement of product quality in a transient expression system. Biotechnol. Prog. 31, 258–267.
Fann, C.H., Guarna, M.M., Kilburn, D.G., Piret, J.M., 1999. Relationship between recombinant activated protein C secretion rates and mRNA levels in baby hamster kidney cells. Biotechnol. Bioeng. 63, 464–472.
Fazekas, S., St. Groth, D.E., Scheidegger, D., 1980. Production of monoclonal antibodies: strategy and tactics. J. Immunol. Methods 35, 1–21.
Feichtinger, J., Hernández, I., Fischer, C., Hanscho, M., Auer, N., Hackl, M., Jadhav, V., Baumann, M., Krempl, P.M., Schmidl, C., Farlik, M., Schuster, M., Merkel, A., Sommer, A., Heath, S., Rico, D., Bock, C., Thallinger, G.G., Borth, N., 2016. Comprehensive genome and epigenome characterization of CHO cells in response to evolutionary pressures and over time. Biotechnol. Bioeng. 9999, 1–14.
Fischer, S., Handrick, R., Otte, K., 2015. The art of CHO cell engineering: a comprehensive retrospect and future perspectives. Biotechnol. Adv. 33, 1878–1896.
Florin, L., Pegel, A., Becker, E., Hausser, A., Olayioye, M.a., Kaufmann, H., 2009. Heterologous expression of the lipid transfer protein CERT increases therapeutic protein productivity of mammalian cells. J. Biotechnol. 141, 84–90.
Frye, C., Deshpande, R., Estes, S., Francissen, K., Joly, J., Lubiniecki, A., Munro, T., Russell, R., Wang, T., Anderson, K., 2016. Industry view on the relative importance of "clonality" of biopharmaceutical-producing cell lines. Biologicals 6–11.
Fussenegger, M., Morris, R.P., Fux, C., Rimann, M., von Stockar, B., Thompson, C.J., Bailey, J.E., 2000. Streptogramin-based gene regulation systems for mammalian cells. Nat. Biotechnol. 18, 1203–1208.
Gaillet, B., Gilbert, R., Broussau, S., Pilotte, A., Malenfant, F., Mullick, A., Garnier, A., Massie, B., 2010. High-level recombinant protein production in CHO cells using lentiviral vectors and the cumate gene-switch. Biotechnol. Bioeng. 106, 203–215.
Galbraith, D.J., Tait, A.S., Racher, A.J., Birch, J.R., James, D.C., 2006. Control of culture environment for improved polyethylenimine-mediated transient production of recombinant monoclonal antibodies by CHO cells. Biotechnol. Prog. 22, 753–762.
Galvan, D.L., Nakazawa, Y., Kaja, A., Kettlun, C., Cooper, L.J.N., Rooney, C.M., Wilson, M.H., 2009. Genome-wide mapping of PiggyBac transposon integrations in primary human T cells. J. Immunother. 32, 837–844.
Golabgir, A., Gutierrez, J.M., Hefzi, H., Li, S., Palsson, B.O., Herwig, C., Lewis, N.E., 2016. Quantitative feature extraction from the Chinese hamster ovary bioprocess bibliome using a novel meta-analysis workflow. Biotechnol. Adv. 1–13.
Gossen, M., Bujard, H., 1992. Tight control of gene expression in mammalian cells by tetracycline-responsive promoters. Proc. Natl. Acad. Sci. U. S. A. 89, 5547–5551.
Gramer, M.J., 2014. Product quality considerations for mammalian cell culture process development and manufacturing. Adv. Biochem. Eng. Biotechnol. 139, 123–166.
Gronemeyer, P., Ditz, R., Strube, J., 2014. Trends in upstream and downstream process development for antibody manufacturing. Bioengineering 1, 188–212.
Gutierrez, J.M., Lewis, N.E., 2015. Optimizing eukaryotic cell hosts for protein production through systems biotechnology and genome-scale modeling. Biotechnol. J. 10, 939–949.
Hacker, D.L., De Jesus, M., Wurm, F.M., 2009. 25 years of recombinant proteins from reactor-grown cells - where do we go from here? Biotechnol. Adv. 27, 1023–1027.
Hackl, M., Borth, N., Grillari, J., 2012. miRNAs—pathway engineering of CHO cell factories that avoids translational burdening. Trends Biotechnol. 30, 405–406.
Hansen, H.G., Nilsson, C.N., Lund, A.M., Kol, S., Grav, L.M., Lundqvist, M., Rockberg, J., Lee, G.M., Andersen, M.R., Kildegaard, H.F., 2015. Versatile microscale screening platform for improving recombinant protein productivity in Chinese hamster ovary cells. Sci. Rep. 5, 18016.
Harding, H.P., Zhang, Y., Zeng, H., Novoa, I., Lu, P.D., Calfon, M., Sadri, N., Yun, C., Popko, B., Paules, R., Stojdl, D.F., Bell, J.C., Hettmann, T., Leiden, J.M., Ron, D., 2003. An integrated stress response regulates amino acid metabolism and resistance to oxidative stress. Mol. Cell 11, 619–633.
Haredy, A.M., Nishizawa, A., Honda, K., Ohya, T., Ohtake, H., Omasa, T., 2013. Improved antibody production in Chinese hamster ovary cells by ATF4 overexpression. Cytotechnology 65, 993–1002.
Harreither, E., Hackl, M., Pichler, J., Shridhar, S., Auer, N., Łabaj, P.P., Scheideler, M., Karbiener, M., Grillari, J., Kreil, D.P., Borth, N., 2015. Microarray profiling of preselected CHO host cell subclones identifies gene expression patterns associated with increased production capacity. Biotechnol. J. 10, 1625–1638.
Hatahet, F., Ruddock, L.W., 2009. Protein disulfide isomerase: a critical evaluation of its function in disulfide bond formation. Antioxid. Redox Signal. 11, 2807–2850.
Hayes, N.V.L., Smales, C.M., Klappa, P., 2010. Protein disulfide isomerase does not control recombinant IgG4 productivity in mammalian cell lines. Biotechnol. Bioeng. 105, 770–779.
Heffernan, M., Dennis, J.W., 1991. Polyoma and hamster papovavirus large T antigen-mediated replication of expression shuttle vectors in Chinese hamster ovary cells. Nucleic Acids Res. 19, 85–92.
Hefzi, H., Ang, K.S., Hanscho, M., Bordbar, A., Ruckerbauer, D., Lakshmanan, M., Orellana, C.A., Baycin-Hizal, D., Huang, Y., Ley, D., Martínez, V.S., Kyriakopoulos, S., Jiménez, N.E., Zielinski, D.C., Quek, L.-E., Wulff, T., Arnsdorf, J., Li, S., Lee, J.S., Paglia, G., Loira,

N., Spahn, P.N., Pedersen, L.E., Gutierrez, J.M., King, Z.A., Lund, A.M., Nagarajan, H., Thomas, A., Abdel-Haleem, A.M., Zanghellini, J., Kildegaard, H.F., Voldborg, B.G., Gerdtzen, Z.P., Betenbaugh, M.J., Palsson, B.O., Andersen, M.R., Nielsen, L.K., Borth, N., Lee, D.-Y., Lewis, N.E., 2016. A consensus genome-scale reconstruction of Chinese hamster ovary cell metabolism. Cell Syst. 3, 134–443 e8.

Hu, Z., Guo, D., Yip, S.S.M., Zhan, D., Misaghi, S., Joly, J.C., Snedecor, B.R., Shen, A.Y., 2013. Chinese hamster ovary K1 host cell enables stable cell line development for antibody molecules which are difficult to express in DUXB11-derived dihydrofolate reductase deficient host cell. Biotechnol. Prog. 29, 980–985.

Hussain, H., Maldonado-Agurto, R., Dickson, A.J., 2014. The endoplasmic reticulum and unfolded protein response in the control of mammalian recombinant protein production. Biotechnol. Lett. 36, 1581–1593.

Hwang, S.O., Chung, J.Y., Lee, G.M., 2003. Effect of doxycycline-regulated ERp57 expression on specific thrombopoietin productivity of recombinant CHO cells. Biotechnol. Prog. 19, 179–184.

Jäger, R., Bertrand, M.J.M., Gorman, A.M., Vandenabeele, P., Samali, A., 2012. The unfolded protein response at the crossroads of cellular life and death during endoplasmic reticulum stress. Biol. Cell. 104, 259–270.

Johari, Y.B., Estes, S.D., Alves, C.S., Sinacore, M.S., James, D.C., 2015. Integrated cell and process engineering for improved transient production of a "difficult-to-express" fusion protein by CHO cells. Biotechnol. Bioeng. 112, 2527–2542.

Justice, C., Brix, A., Freimark, D., Kraume, M., Pfromm, P., Eichenmueller, B., Czermak, P., 2011. Process control in cell culture technology using dielectric spectroscopy. Biotechnol. Adv. 29, 391–401.

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

Kaas, C.S., Kristensen, C., Betenbaugh, M.J., Andersen, M.R., 2015. Sequencing the CHO DXB11 genome reveals regional variations in genomic stability and haploidy. BMC Genomics 16, 1–9.

Kallehauge, T.B., Li, S., Pedersen, L.E., Ha, T.K., Ley, D., Andersen, M.R., Kildegaard, H.F., Lee, G.M., Lewis, N.E., 2016. Ribosome profiling-guided depletion of an mRNA improves CHO cell growth and recombinant product titers. Sci. Rep. in press.

Kallehauge, T.B., Kol, S., Rørdam Andersen, M., Kroun Damgaard, C., Lee, G.M., Faustrup Kildegaard, H., 2016. Endoplasmic reticulum-directed recombinant mRNA displays subcellular localization equal to endogenous mRNA during transient expression in CHO cells. Biotechnol. J. 11, 1362–1367.

Khan, S.U., Schröder, M., 2008. Engineering of chaperone systems and of the unfolded protein response. Cytotechnology 57, 207–231.

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

Kim, N.S., Kim, S.J., Lee, G.M., 1998. Clonal variability within dihydrofolate reductase-mediated gene amplified Chinese hamster ovary cells: stability in the absence of selective pressure. Biotechnol. Bioeng. 60, 679–688.

Kim, S.J., Kim, N.S., Ryu, C.J., Hong, H.J., Lee, G.M., 1998. Characterization of chimeric antibody producing CHO cells in the course of dihydrofolate reductase-mediated gene amplification and their stability in the absence of selective pressure. Biotechnol. Bioeng. 58, 73–84.

Kim, J.Y., Kim, Y.-G., Lee, G.M., 2012. CHO cells in biotechnology for production of recombinant proteins: current state and further potential. Appl. Microbiol. Biotechnol. 93, 917–930.

Kim, Y.J., Baek, E., Lee, J.S., Lee, G.M., 2013. Autophagy and its implication in Chinese hamster ovary cell culture. Biotechnol. Lett. 35, 1753–1763.

Kremkow, B.G., Baik, J.Y., Macdonald, M.L., Lee, K.H., 2015. CHOgenome.org 2.0: genome resources and website updates. Biotechnol. J. 2011, 931–938.

Ku, S.C.Y., Ng, D.T.W., Yap, M.G.S., Chao, S., 2008. Effects of overexpression of X-box binding protein 1 on recombinant protein production in Chinese hamster ovary and NS0 myeloma cells. Biotechnol. Bioeng. 99, 155–164.

Ku, S.C.Y., Toh, P.C., Lee, Y.Y., Chusainow, J., Yap, M.G.S., Chao, S.H., 2010. Regulation of XBP-1 signaling during transient and stable recombinant protein production in CHO cells. Biotechnol. Prog. 26, 517–526.

Kumar, S.R., 2015. Industrial production of clotting factors: challenges of expression, and choice of host cells. Biotechnol. J. 10, 995–1004.

Kunaparaju, R., Liao, M., Sunstrom, N.-A., 2005. Epi-CHO, an episomal expression system for recombinant protein production in CHO cells. Biotechnol. Bioeng. 91, 670–677.

Kunert, R., Reinhart, D., 2016. Advances in recombinant antibody manufacturing. Appl. Microbiol. Biotechnol. 100, 3451–3461.

Le Fourn, V., Girod, P.-A., Buceta, M., Regamey, A., Mermod, N., 2014. CHO cell engineering to prevent polypeptide aggregation and improve therapeutic protein secretion. Metab. Eng. 21, 91–102.

Lee, G.W., Fecko, J.K., Yen, A., Donaldson, D., Wood, C., Tobler, S., Vunuum, S., Luo, Y., Leonard, M., 2007. Improving the expression of a soluble receptor:Fc fusion protein in CHO cells by coexpression with the receptor ligand. Cell Technology for Cell Products. Springer Netherlands, Dordrecht, pp. 29–39.

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

Lee, J.S., Grav, L.M., Pedersen, L.E., Lee, G.M., Kildegaard, H.F., 2016. Accelerated homology-directed targeted integration of transgenes in Chinese hamster ovary cells via CRISPR/Cas9 and fluorescent enrichment. Biotechnol. Bioeng. 113, 2518–2523.

Lewis, N.E., Liu, X., Li, Y., Nagarajan, H., Yerganian, G., O'Brien, E., Bordbar, A., Roth, A.M., Rosenbloom, J., Bian, C., Xie, M., Chen, W., Li, N., Baycin-Hizal, D., Latif, H., Forster, J., Betenbaugh, M.J., Famili, I., Xu, X., Wang, J., Palsson, B.O., 2013. Genomic landscapes of Chinese hamster ovary cell lines as revealed by the *Cricetulus griseus* draft genome. Nat. Biotechnol. 31, 759–765.

Liu, C., Dalby, B., Chen, W., Kilzer, J.M., Chiou, H.C., 2008. Transient transfection factors for high-level recombinant protein production in suspension cultured mammalian cells. Mol. Biotechnol. 39, 141–153.

Liu, C.Y., Spencer, V., Kumar, S., Liu, J., Chiou, H., Zmuda, J.F., 2015. Attaining high transient titers in CHO cells. Genet. Eng. Biotechnol. News 35, 34–35.

Lupton, S., Levine, A.J., 1985. Mapping genetic elements of Epstein-Barr virus that facilitate extrachromosomal persistence of Epstein-Barr virus-derived plasmids in human cells. Mol. Cell. Biol. 5, 2533–2542.

Macaraeg, N.F., Reilly, D.E., Wong, A.W., 2013. Use of an anti-apoptotic CHO cell line for transient gene expression. Biotechnol. Prog. 29, 1050–1058.

Majors, B.S., Betenbaugh, M.J., Pederson, N.E., Chiang, G.G., 2008. Enhancement of transient gene expression and culture viability using Chinese hamster ovary cells overexpressing Bcl-x(L). Biotechnol. Bioeng. 101, 567–578.

Mali, P., Esvelt, K.M., Church, G.M., 2013. Cas9 as a versatile tool for engineering biology. Nat. Methods 10, 957–963.

Mariati, Ho, S.C.L., Yap, M.G.S., Yang, Y., 2010. Evaluating post-transcriptional regulatory elements for enhancing transient gene expression levels in CHO K1 and HEK293 cells. Protein Expr. Purif. 69, 9–15.

Matasci, M., Bachmann, V., Baldi, L., Hacker, D.L., De Jesus, M., Wurm, F.M., 2011. Rapid recombinant protein production from pools of transposon-generated CHO cells. BMC Proc. 5, P34.

McLeod, J., O'Callaghan, P.M., Pybus, L.P., Wilkinson, S.J., Root, T., Racher, A.J., James, D.C., 2011. An empirical modeling platform to evaluate the relative control discrete CHO cell synthetic processes exert over recombinant monoclonal antibody production process titer. Biotechnol. Bioeng. 108, 2193–2204.

Merulla, J., Fasana, E., Soldà, T., Molinari, M., 2013. Specificity and regulation of the endoplasmic reticulum-associated degradation machinery. Traffic 14, 767–777.

Mohan, C., Lee, G.M., 2010. Effect of inducible co-overexpression of protein disulfide isomerase and endoplasmic reticulum oxidoreductase on the specific antibody productivity of recombinant Chinese hamster ovary cells. Biotechnol. Bioeng. 107, 337–346.

Mohan, C., Park, S.H., Chung, J.Y., Lee, G.M., 2007. Effect of doxycycline-regulated protein disulfide isomerase expression on the specific productivity of recombinant CHO cells: thrombopoietin and antibody. Biotechnol. Bioeng. 98, 611–615.

Mohan, C., Kim, Y.-G., Koo, J., Lee, G.M., 2008. Assessment of cell engineering strategies for improved therapeutic protein production in CHO cells. Biotechnol. J. 3, 624–630.

Moore, K.A., Hollien, J., 2012. The unfolded protein response in secretory cell function. Annu. Rev. Genet. 46, 165–183.

Mozley, O.L., Thompson, B.C., Fernandez-Martell, A., James, D.C., 2014. A mechanistic dissection of polyethylenimine mediated transfection of CHO cells: to enhance the efficiency of recombinant DNA utilization. Biotechnol. Prog. 30, 1161–1170.

Mullick, A., Xu, Y., Warren, R., Koutroumanis, M., Guilbault, C., Broussau, S., Malenfant, F., Bourget, L., Lamoureux, L., Lo, R., Caron, A.W., Pilotte, A., Massie, B., 2006. The cumate gene-switch: a system for regulated expression in mammalian cells. BMC Biotechnol. 6, 43.

Nishimiya, D., 2013. Proteins improving recombinant antibody production in mammalian cells. Appl. Microbiol. Biotechnol. 98, 1–12.

Nishimiya, D., Mano, T., Miyadai, K., Yoshida, H., Takahashi, T., 2013. Overexpression of CHOP alone and in combination with chaperones is effective in improving antibody production in mammalian cells. Appl. Microbiol. Biotechnol. 97, 2531–2539.

Noh, S.M., Sathyamurthy, M., Lee, G.M., 2013. Development of recombinant Chinese hamster ovary cell lines for therapeutic protein production. Curr. Opin. Chem. Eng. 1–7.

Oberbek, A., Matasci, M., Hacker, D.L., Wurm, F.M., 2011. Generation of stable, high-producing CHO cell lines by lentiviral vector-mediated gene transfer in serum-free suspension culture. Biotechnol. Bioeng. 108, 600–610.

O'Callaghan, P.M., McLeod, J., Pybus, L.P., Lovelady, C.S., Wilkinson, S.J., Racher, A.J., Porter, A., James, D.C., 2010. Cell line-specific control of recombinant monoclonal antibody production by CHO cells. Biotechnol. Bioeng. 106, 938–951.

O'Callaghan, P.M., Berthelot, M.E., Young, R.J., Graham, J.W.A., Racher, A.J., Aldana, D., 2015. Diversity in host clone performance within a Chinese hamster ovary cell line. Biotechnol. Prog. 31, 1187–1200.

Papapetrou, E.P., Lee, G., Malani, N., Setty, M., Riviere, I., Tirunagari, L.M.S., Kadota, K., Roth, S.L., Giardina, P., Viale, A., Leslie, C., Bushman, F.D., Studer, L., Sadelain, M., 2011. Genomic safe harbors permit high β-globin transgene expression in thalassemia induced pluripotent stem cells. Nat. Biotechnol. 29, 73–78.

Peng, R.-W., Fussenegger, M., 2009. Molecular engineering of exocytic vesicle traffic enhances the productivity of Chinese hamster ovary cells. Biotechnol. Bioeng. 102, 1170–1181.

Peng, R.W., Abellan, E., Fussenegger, M., 2011. Differential effect of exocytic SNAREs on the production of recombinant proteins in mammalian cells. Biotechnol. Bioeng. 108, 611–620.

Pilbrough, W., Munro, T.P., Gray, P., 2009. Intraclonal protein expression heterogeneity in recombinant CHO cells. PLoS One 4, e8432.

Preininger, A., Schlokat, U., Mohr, G., Himmelspach, M., Stichler, V., Kyd-Rebenburg, A., Plaimauer, B., Turecek, P.L., Schwarz, H.P., Wernhart, W., Fischer, B.E., Dorner, F., 1999. Strategies for recombinant Furin employment in a biotechnological process: complete target protein precursor cleavage. Cytotechnology 30, 1–16.

Priola, J.J., Calzadilla, N., Baumann, M., Borth, N., Tate, C.G., Betenbaugh, M.J., 2016. High-throughput screening and selection of mammalian cells for enhanced protein production. Biotechnol. J. 1–13.

Puck, T.T., 1957. The genetics of somatic mammalian cells. Adv. Biol. Med. Phys. 5, 75–101.

Pybus, L.P., Dean, G., West, N.R., Smith, A., Daramola, O., Field, R., Wilkinson, S.J., James, D.C., 2014. Model-directed engineering of "difficult-to-express" monoclonal antibody production by Chinese hamster ovary cells. Biotechnol. Bioeng. 111, 372–385.

Qin, J.Y., Zhang, L., Clift, K.L., Hulur, I., Xiang, A.P., Ren, B.-Z., Lahn, B.T., 2010. Systematic comparison of constitutive promoters and the doxycycline-inducible promoter. PLoS One 5, e10611.

Rajendra, Y., Kiseljak, D., Manoli, S., Baldi, L., Hacker, D.L., Wurm, F.M., 2012. Role of non-specific DNA in reducing coding DNA requirement for transient gene expression with CHO and HEK-293E cells. Biotechnol. Bioeng. 109, 2271–2278.

Rajendra, Y., Kiseljak, D., Baldi, L., Hacker, D.L., Wurm, F.M., 2015. Transcriptional and post-transcriptional limitations of high-yielding, PEI-mediated transient transfection with CHO and HEK-293E cells. Biotechnol. Prog. 31, 541–549.

Rajendra, Y., Peery, R.B., Barnard, G.C., 2016. Generation of stable Chinese hamster ovary pools yielding antibody titers of up to 7.6 g/L using the piggyBac transposon system. Biotechnol. Prog. 30, 429–442.

Rajesh, K., Krishnamoorthy, J., Kazimierczak, U., Tenkerian, C., Papadakis, A.I., Wang, S., Huang, S., Koromilas, A.E., 2015. Phosphorylation of the translation initiation factor eIF2α at serine 51 determines the cell fate decisions of Akt in response to oxidative stress. Cell Death Dis. 6, e1591.

Randall, T.D., Parkhouse, R.M., Corley, R.B., 1992. J chain synthesis and secretion of hexameric IgM is differentially regulated by lipopolysaccharide and interleukin 5. Proc. Natl. Acad. Sci. U. S. A. 89, 962–966.

Sano, R., Reed, J.C., 2013. ER stress-induced cell death mechanisms. Biochim. Biophys. Acta 1833, 3460–3470.

Sathyamurthy, M., Lee, J.S., Park, J.H., Kim, Y.J., Jeong, J.Y., Jang, J.W., Lee, G.M., 2012. Overexpression of PACEsol improves BMP-7 processing in recombinant CHO cells. J. Biotechnol. 164, 336–339.

Sathyamurthy, M., Kim, C.L., Bang, Y.L., Kim, Y.S., Jang, J.W., Lee, G.M., 2015. Characterization and expression of proprotein convertases in CHO cells: efficient proteolytic maturation of human bone morphogenetic protein-7. Biotechnol. Bioeng. 112, 560–568.

Schröder, M., Friedl, P., 1997. Overexpression of recombinant human antithrombin III in Chinese hamster ovary cells results in malformation and decreased secretion of recombinant protein. Biotechnol. Bioeng. 53, 547–559.

Schröder, M., Körner, C., Friedl, P., 1999. Quantitative analysis of transcription and translation in gene amplified Chinese hamster ovary cells on the basis of a kinetic model. Cytotechnology 29, 93–102.

Schröder, M., Schäfer, R., Friedl, P., 2002. Induction of protein aggregation in an early secretory compartment by elevation of expression level. Biotechnol. Bioeng. 78, 131–140.

Selvarasu, S., Ho, Y.S., Chong, W.P.K., Wong, N.S.C., Yusufi, F.N.K., Lee, Y.Y., Yap, M.G.S., Lee, D.Y., 2012. Combined in silico modeling and metabolomics analysis to characterize fed-batch CHO cell culture. Biotechnol. Bioeng. 109, 1415–1429.

Seth, G., Charaniya, S., Wlaschin, K.F., Hu, W.-S., 2007. In pursuit of a super producer-alternative paths to high producing recombinant mammalian cells. Curr. Opin. Biotechnol. 18, 557–564.

Silla, T., Tagen, I., Geimanen, J., Janikson, K., Abroi, A., Ustav, E., Ustav, M., Mandel, T., 2006. Vectors, Cell Lines and Their Use in Obtaining Extended Episomal Maintenance Replication of Hybrid Plasmids and Expression of Gene Products. WO2006084754 A1.

Sinacore, M.S., Drapeau, D., Adamson, S.R., 2000. Adaptation of mammalian cells to growth in serum-free media. Mol. Biotechnol. 15, 249–257.

Sommeregger, W., Mayrhofer, P., Steinfellner, W., Reinhart, D., Henry, M., Clynes, M., Meleady, P., Kunert, R., 2016. Proteomic differences in recombinant CHO cells producing two similar antibody fragments. Biotechnol. Bioeng. 113, 1902–1912.

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

Stearns, F.W., 2010. One hundred years of pleiotropy: a retrospective. Genetics 186, 767–773.

Stockholm, D., Benchaouir, R., Picot, J., Rameau, P., Neildez, T.M.A., Landini, G., Laplace-Builhé, C., Paldi, A., 2007. The origin of phenotypic heterogeneity in a clonal cell population in vitro. PLoS One 2, e394.

Tastanova, A., Schulz, A., Folcher, M., Tolstrup, A., Puklowski, A., Kaufmann, H., Fussenegger, M., 2015. Overexpression of YY1 increases the protein production in mammalian cells. J. Biotechnol. 219, 72–85.

Tigges, M., Fussenegger, M., 2006. Xbp1-based engineering of secretory capacity enhances the productivity of Chinese hamster ovary cells. Metab. Eng. 8, 264–272.

Underhill, M.F., Coley, C., Birch, J.R., Findlay, A., Kallmeier, R., Proud, C.G., James, D.C., 2003. Engineering mRNA translation initiation to enhance transient gene expression in chinese hamster ovary cells. Biotechnol. Prog. 19, 121–129.

Urlaub, G., Chasin, L.A., 1980. Isolation of Chinese hamster cell mutants deficient in dihydrofolate reductase activity. Proc. Natl. Acad. Sci. U. S. A. 77, 4216–4220.

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

Van Craenenbroeck, K., Vanhoenacker, P., Haegeman, G., 2000. Episomal vectors for gene expression in mammalian cells. Eur. J. Biochem. 267, 5665–5678.

Walsh, G., 2014. Biopharmaceutical benchmarks 2014. Nat. Biotechnol. 32, 992–1000.

Walsh, G., Jefferis, R., 2006. Post-translational modifications in the context of therapeutic proteins. Nat. Biotechnol. 24, 1241–1252.

Walter, P., Ron, D., 2011. The unfolded protein response: from stress pathway to homeostatic regulation. Science 334, 1081–1086.

Wasley, L.C., Rehemtulla, A., Bristol, J.A., Kaufman, R.J., 1993. PACE/furin can process the vitamin K-dependent pro-factor IX precursor within the secretory pathway. J. Biol. Chem. 268, 8458–8465.

Weber, W., Fux, C., Daoud-el Baba, M., Keller, B., Weber, C.C., Kramer, B.P., Heinzen, C., Aubel, D., Bailey, J.E., Fussenegger, M., 2002. Macrolide-based transgene control in mammalian cells and mice. Nat. Biotechnol. 20, 901–907.

Werstuck, G., Green, M.R., 1998. Controlling gene expression in living cells through small molecule-RNA interactions. Science 282, 296–298.

Wiederanders, B., Kaulmann, G., Schilling, K., 2003. Functions of propeptide parts in cysteine proteases. Curr. Protein Pept. Sci. 4, 309–326.

Wilson, M.H., Coates, C.J., George, A.L., 2007. PiggyBac transposon-mediated gene transfer in human cells. Mol. Ther. 15, 139–145.

Wurm, F.M., 2004. Production of recombinant protein therapeutics in cultivated mammalian cells. Nat. Biotechnol. 22, 1393–1398.

Wurm, F., 2013. CHO quasispecies—implications for manufacturing processes. Processes 1, 296–311.

Wurm, F.M., Hacker, D., 2011. First CHO genome. Nat. Biotechnol. 29, 718–720.

Xiao, S., Shiloach, J., Betenbaugh, M.J., 2014. Engineering cells to improve protein expression. Curr. Opin. Struct. Biol. 26, 32–38.

Xu, X., Nagarajan, H., Lewis, N.E., Pan, S., Cai, Z., Liu, X., Chen, W., Xie, M., Wang, W., Hammond, S., Andersen, M.R., Neff, N., Passarelli, B., Koh, W., Fan, H.C., Wang, J., Gui, Y., Lee, K.H., Betenbaugh, M.J., Quake, S.R., Famili, I., Palsson, B.O., Wang, J., 2011. The genomic sequence of the Chinese hamster ovary (CHO)-K1 cell line. Nat. Biotechnol. 29, 735–741.

Yates, J., Warren, N., Reisman, D., Sugden, B., 1984. A cis-acting element from the Epstein-Barr viral genome that permits stable replication of recombinant plasmids in latently infected cells. Proc. Natl. Acad. Sci. U. S. A. 81, 3806–3810.

Ye, J., Kober, V., Tellers, M., Naji, Z., Salmon, P., Markusen, J.F., 2009. High-level protein expression in scalable CHO transient transfection. Biotechnol. Bioeng. 103, 542–551.

Zang, L., Frenkel, R., Simeone, J., Lanan, M., Byers, M., Lyubarskaya, Y., 2011. Metabolomics profiling of cell culture media leading to the identification of riboflavin photosensitized degradation of tryptophan causing slow growth in cell culture. Anal. Chem. 83, 5422–5430.

Chapter 4

Improving Recombinant Protein Production by Ectopic Expression of Effector Genes - A Study

Rationale

Thorough analysis of cellular pathways important for bioprocessing has the potential to unravel novel beneficial effector genes that can be overexpressed to generate superior CHO cell factories. In this chapter, we have screened a panel of candidate effector genes, identified previously through network-based secretory pathway reconstruction and transcriptomic analysis of IgG-producing CHO cell lines. The objective of this study was to find novel effector genes with positive effects on q_p of IgG-producing cells and to understand the cellular mechanisms giving rise to the increase of q_p.

The novel effector genes were examined using the guidelines we proposed and carefully reviewed in **Chapter 3**. We explored a variety of factors, contributing to the observed phenotype, by also investigating the two stable IgG (Rituximab) high-producing clones generated with the clone screening method described in **Chapter 1**.

Transient Co-Expression of Novel Effector Genes and Different IgG1 Molecules Increases Specific Productivity of CHO-S Cells

Nuša Pristovšek[1], Henning Gram Hansen[1], Anne Mathilde Lund[2], Johan Rojek[1], Claes Nymand Nilsson[1], Gyun Min Lee[1,3], Mikael Rørdam Andersen[2], Helene Faustrup Kildegaard[1]

[1]The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Kemitorvet 220, 2800 Kgs. Lyngby, Denmark

[2]Department of Biotechnology and Biomedicine, Technical University of Denmark, Søltofts Plads 223, 2800 Kgs. Lyngby, Denmark

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

Keywords: CHO, secretory capacity, q_p, effector gene engineering, ectopic expression, IgG1

1. Introduction

CHO cells remain the main cell factory for production of glycosylated biotherapeutics, with IgGs (including bispecifics and biosimilars) dominating the market now and most likely in the foreseeable future (1, 2). Protein synthesis and secretion processes are important aspects for the production of biotherapeutics, however, they are often suboptimal in CHO cells for recombinant proteins (r-proteins) due to their ovarian cell type origin, inherently not possessing as high secretory activity as *e.g.* secretory plasma cells (3, 4). Therefore, establishing high-producing cell lines through antibiotic selection and gene amplification (5–7), efficient high-producer clone screening methods (8), cell media optimization (9) and expression vector engineering (10, 11) are needed strategies to achieve increased r-protein synthesis in CHO cells. However, increased r-protein synthesis can quickly result in reaching the limit of r-protein secretion, *i.e.* a secretion bottleneck. Secretion bottleneck is a non-linear relationship between the intracellular levels of r-protein molecules and specific productivity (q_p). The occurrence of the secretion bottleneck is r-protein-specific and is more readily obtained for difficult-to-express (DTE) r-proteins. DTE r-protein is a protein that is difficult to express due to intracellular challenges (*i.e.* transcription, protein folding, post-translational processing, secretion) or to its biophysical properties (12). Cell line engineering efforts have been employed to improve folding, transport and secretion of r-proteins to enhance q_p. Target genes (from here on referred to as effector genes) that are related to the protein secretory pathway have previously been overexpressed, such as protein disulfide isomerase (PDI) (13, 14), Sly1 (15), Munc18b (16) and SNARE proteins (17). Further, r-protein expression can due to high expression levels or difficult expression capability (DTE r-proteins) generate ER stress by exceeding the protein folding and processing capacity of the ER (18–20). Resulting r-protein secretion bottlenecks and increased ER stress can in turn induce the unfolded protein response (UPR) (21). To enhance r-protein production, effector genes with transcription factor functions that are involved in the UPR have previously been overexpressed, such as X-box binding proteins (XBP1) (22–24) and activating transcription factor 4 (ATF4) (25). Collectively, these studies demonstrate that r-protein secretion bottlenecks can be effectively targeted with rational effector gene engineering approaches. Moreover, extensive engineering of cell host is sometimes indeed needed to alleviate expression bottlenecks, especially for production of DTE r-proteins (12). However, it is increasingly more evident that not only r-protein-specific, but also cell line-specific engineering are required for the positive effects of effector genes to

be observed (26, 27), given the intrinsic genetic and functional heterogeneity of CHO cells (28).

In this study, we identified a panel of candidate effector genes through a previously performed network reconstruction of the CHO secretory pathway, combined with CHO transcriptomic data (29). We set out to identify effector genes from the candidate panel that would increase q_p and the volumetric productivity of IgG in CHO cells with the objective of using these genes as engineering targets to improve the secretory capacity of CHO cells. Therefore, we transiently co-expressed a panel of candidate effector genes and an IgG1 molecule (Rituximab), using an established microscale screening platform (30), and identified three novel effector genes that increased q_p of CHO-S cells under the used conditions. Following the identification of the novel effector genes, a thorough investigation of intracellular and experimental conditions supporting the positive q_p phenotype was conducted.

All in all, we demonstrated that positive effects of the novel effector genes identified in this study occur in a relatively narrow range of conditions. Increased q_p of CHO-S cells upon effector gene overexpression was observed for several IgG1 molecules, however, exclusively when IgG1 molecule was transiently expressed and when it was delivered in the expression cassette with a heavy chain (HC) under the CMV promoter and a light chain (LC) under the EF1α promoter. Therefore, this case study serves as a reminder of the many complexities of the ectopic expression of effector genes for increasing the secretory capacity of CHO cells, as exemplified in our review (31).

2. Material and methods

2.1 Plasmid design and construction

All the plasmids were constructed via uracil-specific excision reagent (USER) cloning method, using flexible assembly sequence tags, uracil-containing primers and Phusion U Hot-Start Flex polymerase (New England Biolabs), as previously described (32). Expression cassettes of the three IgG1 plasmids (Rituximab B, Adalimumab and Trastuzumab) consisted of CMV-HC-BGHpA-EF1α-LC-BGHpA expression cassette and the vector backbone with neomycin selection marker. Whereas Rituximab A consisted of the same expression cassette, however, with a vector backbone, lacking a functional mammalian selection marker. Plasmid Rituximab B - swap derives from Rituximab B with heavy chain (HC) and light chain (LC) under the swapped promoters (EF1α-HC-SV40pA-CMV-LC-BGHpA). Plasmids encoding Rituximab were previously described (33). Etanercept plasmids consisted of an expression cassette (CMV or EF1α-Etanercept-BGHpA), with a previously published vector backbone

(34) and Etanercept cds sequence (35). Adalimumab and Trastuzumab coding sequences are based on the amino acid sequence with accession number DB00051 and DB00072, respectively, from the DrugBank database (36) and were CHO codon optimized (GeneArt, Life Technologies). Effector genes were either amplified from cDNA template or synthesized, based on Mus Musculus RefSeq IDs. All effector genes encode amino acid sequences with a known Uniprot ID (more information about the cloning of effector genes in the Supplementary Table S1). Effector genes were cloned into the EF1α-PacI/Nt.BbvCI-SV40pA vector backbone (the empty vector), as previously described (34). The empty vector without any coding sequence was used for all mock transfections. All the plasmids are depicted in the Supplementary Figure S1. Assembled PCR fragments were transformed into *E.coli* Mach1 competent cells (Life Technologies). All constructs were verified by sequencing of the expression cassettes (Eurofins Scientific) and purified using NucleoBond Xtra Midi EF (Macherey-Nagel) according to manufacturer's instructions.

2.2 Cell cultivation

CHO-S cells (Life Technologies) and CHO-S-derived E9_2 and C6_2 stable Rituximab producing cells were grown in 30 mL volume of CD CHO medium (Gibco) supplemented with 8 mM L-Glutamine and 2 μL/mL anti-clumping agent (Gibco) in 125 mL Erlenmeyer flasks (Corning). E9_2 and C6_2 stable producers were established with Rituximab B plasmid, generated using a previously published clone screening platform (35), as described in our manuscript in Chapter 2. CHO-K1-derived stable Rituximab-producing cells (kindly provided by Prof. Gyun Min Lee) were grown in 30 mL volume of PowerCHO medium (Lonza), supplemented with glutamine synthetase expression medium supplement (GSEM; Sigma–Aldrich), 2 μL/mL anti-clumping agent and 25 μM MSX (Sigma–Aldrich; Cat. M5379) in 125 mL Erlenmeyer flasks. Cells were incubated at 37°C, 5% CO_2 at 120 rpm (25mm shaking amplitude) and passaged every 2-3 days.

2.3 Transfection and 3-day batch cultivation in 96-half-deepwell plates

CHO-S cells were co-transfected with effector gene- and r-protein-encoding plasmids in pre-sterilized polystyrene 96-square System Duetz half-deepwell-microplates (CR1496c, Enzyscreen, Haarlem, Netherlands) capped with autoclaved low-evaporation sandwich Duetz covers (CR1296a, Enzyscreen) as described previously (30) with the following exceptions: A 20% : 80% effector gene : r-protein-encoding plasmid mix was employed instead of 50% : 50%. Anti-clumping agent was added one day after transfection, instead of three hours after.

The 96 half-deepwell plates were incubated in an S41i incubator shaker at the following conditions: 325 rpm, 25 mm shaking amplitude, 5% CO_2, 37°C.

2.4 Transfection and 3-day batch cultivation in 6-well plates

One day prior transfection, the medium was changed to remove the anti-clumping agent. For transfection, cells were seeded at density 1 x 10^6 cells/mL in 3 mL volume of CD CHO medium supplemented with 8 mM L-Glutamine (CHO-S cells) or PowerCHO medium, supplemented with GSEM and 25 µM MSX (CHO-K1 cells) in a 6-well flat-bottom plate (Falcon). Every experiment included a pmaxGFP transfection for assessment of transfection efficiency. A total plasmid load was constant, that is 3.75 µg of DNA per well being diluted in OptiPRO™ SFM (Gibco). A plasmid load consisted of 80% model protein and 20% effector gene for transfections in CHO-S wt, and 80% empty vector and 20% effector gene for transfections in stable producing cell lines, unless otherwise stated. Mock samples were transfected with 80% model protein and 20% empty vector in CHO-S wt cells, and with 100% empty vector in stable cell lines. Rituximab was titrated in duplicates (independent transfections) using 0%, 8%, 20%, 33%, 66% and 100% plasmid dose of Rituximab A in a 3.75 µg of total plasmid DNA load, with the remainder of DNA being the empty vector. FreeStyle™ MAX Reagent (Gibco) was diluted in OptiPRO™ SFM according to manufacturer's instructions and added to diluted plasmid mixture for 8 min of incubation. Transfection mixture was added to cells and four hours post-transfection 2 µL/mL of anti-clumping agent was added to cell suspension. Viable cell density (VCD) and viability were measured daily, as described below. Every day 150-300 µL of cell suspension was centrifuged at 1000 g, 5 min and supernatant stored at -80°C for titer measurements. Integral of viable cell density (IVCD) and specific productivity (q_p)(pg protein/cell/day; pcd) for each condition were calculated as described elsewhere (27). On day two GFP fluorescent analysis for transfection efficiency assessment was performed, as described below. 1 x 10^6 cells were harvested on day three by centrifugation at 200 g, 5 min. Cell pellets were stored at -80°C for RNA extraction. Each experiment included two independent mock transfections.

2.5 Viability, VCD and transfection efficiency determination

When cells were grown in 96-half-deepwell plates and 6-well plates, viability and VCD were determined by a combined Hoechst and PI stain on the Celígo image cytometer (Nexcelom Bioscience, Lawrence, MA) as previously described (30). pMaxGFP transfection efficiency was assessed through Hoechst and GFP fluorescence measurements using Celígo image cytometer. Viability and VCD of cells in 125 mL shake flasks were determined on the

NucleoCounter NC-200 Cell Counter (Chemometec, Allerod, Denmark) using Via1-Cassettes and the “Viability and Cell Count Method 2” assay.

2.6 Relative mRNA expression levels using RT-qPCR

Total RNA was extracted from 1 x 10^6 cells using RNeasy Plus Mini Kit (Qiagen) according to the manufacturer’s instructions. RNA concentration was measured with Qubit fluorometric analysis (Life technologies) and the purity/quality was assessed with Nanodrop (Thermo Fisher Scientific) and 1% agarose gel visualization. cDNA was synthesized from 1 µg of total RNA according to manufacturer’s instructions using either TURBO DNA-free™ kit (Invitrogen) followed by qScript™ Flex cDNA kit (Quantabio) or Maxima First Strand cDNA Synthesis Kit for RT-qPCR with dsDNAse treatment (ThermoFisher Scientific). The RT-qPCR was run on the QuantStudio 5 Real-Time PCR System using TaqMan™ Multiplex Master Mix (Thermo Fisher Scientific) in a triplex and the following amplification conditions: 50°C for 2 min, 95°C for 10 min; 40x: 95°C for 15 s, 60°C for 1 min. Custom-made Taqman assays were used for genes encoding Rituximab heavy chain (HC), Rituximab light chain (LC), BiP and HERP, as well as normalization genes *Actb* and *Gapdh*. All the primers and probes are listed in the Supplementary Table S2 and were validated by melting curve analysis and primer efficiency test. BiP and HERP TaqMan probes were previously validated (unpublished data) using tunicamycin (ER stress inducer) treatment. Using the ΔΔCt method, the relative expression levels of genes of interest were calculated by normalization to the expression levels of the two normalization genes. Each experiment included controls with no template and was performed with technical triplicates. mRNA expression levels were normalized to the mean level of two mock samples or to 100% dose of Rituximab in experiments where Rituximab-encoding plasmids were titrated (internal calibrators).

2.7 Titer measurements

Rituximab, Adalimumab, Trastuzumab and Etanercept titers were determined in supernatants by bio-layer interferometry (ForteBio, Pall, Menlo Park, CA) using Protein A biosensors (Fortebio 18-5012, Pall), as previously described (37). Absolute titers of Rituximab, Adalimumab and Trastuzumab were calculated using a calibration curve from a dilution series of purified human IgG (Genscript, Cat: #A01006) and absolute Etanercept titers were calculated using a calibration curve generated from a dilution series of commercially available Enbrel (Pfizer, New York City, NY; Lot R51698).

3. Results

First, we wanted to use a set of effector genes with published positive effects on IgG q_p in CHO cells, either transiently or constitutively expressed, that would serve as positive controls for q_p enhancement (data on 15 'positive control' effector genes available in Supplementary Table S3). The only protein included in this panel with no reported positive effects on q_p for IgG is CHOP, however its positive effect on IgG volumetric productivity was previously reported (38). Second, in our previous study we have manually reconstructed the mouse secretory pathway and by applying a comparative genomic approach, we were able to identify important secretory components and regulators in CHO-K1 genome (29). Using this reconstruction together with the transcriptome data from CHO non-producing and IgG-producing cells, we selected several new candidate engineering targets for improving the q_p of CHO cells for IgG production (data on 20 candidate effector genes available in Supplementary Table S4). The candidate effector genes were selected on the combined basis of their predicted function in the reconstructed secretory pathway and their mRNA fold changes between non-producing and IgG-producing cell lines. And third, we were interested in identifying novel effector genes from the panel of candidate targets that increase overall IgG volumetric yield by increasing the q_p. Further, we set out to understand the cellular mechanism giving rise to the observed increase, so that the novel q_p-increasing targets could be used to engineer CHO cells' secretory capacity.

3.1 Initial screen identifies three novel effector genes that increase q_p of transiently expressed Rituximab in CHO-S cells

A panel of candidate effector genes alongside 'positive control' effector genes and CHOP, all of murine origin, were initially screened in 96-half-deepwell plates when transiently expressing an IgG (Rituximab A) in CHO-S cells (**Figure 1**), using an established screening platform (30). By applying a 80% IgG : 20% effector gene plasmid dosage, the screen uncovered only a few beneficial effector genes from the 'positive control' group, namely *CHOP* and *Xbp1s* with approximately 2.5-fold and 3.5-fold increase in q_p of Rituximab, respectively. This result reconfirms the inconsistency of observed effects across different studies and underpins the importance of understanding that both cellular and experimental factors are able to modulate the outcome when testing effector genes, as exemplified in our review (31). On the other hand, from the panel of newly proposed targets, overexpression of three effector genes demonstrated increased q_p of Rituximab in our initial setting: *Epas1*, *Mafb* and *Nfe2* with approximately 2-fold, 3-fold and 1.5-fold increase, respectively.

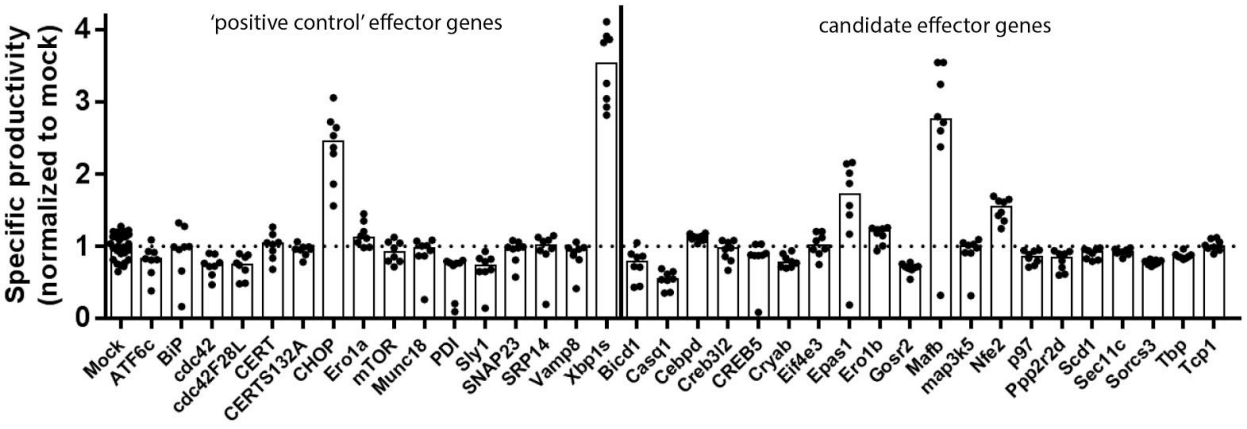


Figure 1: Screen of 'positive control' and candidate effector genes using transient expression-based screening platform in 96-half-deepwell plates

CHO-S cells were transiently co-transfected with Rituximab and effector gene of interest (80% : 20% plasmid dosage, respectively) or Rituximab and empty vector (mock) (80% : 20% plasmid dosage, respectively). Three days post-transfection, q_p was assessed for each condition and normalized to the mock value. Values of each transfection replicate are shown (n = 8).

3.2 Novel effector genes increase q_p exclusively when transiently co-expressed with Rituximab in a CMV-HC and EF1α-LC composition

The three q_p-increasing effector genes, newly identified in the initial screening, were further examined together with the established *Xbp1s* effector gene (positive control) to precisely define the conditions under which their positive effects would be reproducibly observed. We identified and explored several underlying cellular and experimental factors that could be affecting the effector genes: a) different IgG vector compositions, b) different IgG expression platforms (transient and stable), c) different CHO host cell lines, d) different stable producing cell lines, e) different IgG secretion levels and f) different effector gene dosages.

Since the positive effects of these effector genes were first observed when being transiently co-expressed with Rituximab (20% : 80% dosage, respectively) in CHO-S cells, the same experiment was reproduced in a 6-well format using three different Rituximab plasmids. To test whether co-expression of a selection marker gene would have an effect on q_p, we have tested two plasmids, both consisting of the HC- and LC-coding sequences downstream of a CMV and EF1α promoter, respectively (Rituximab A and Rituximab B). Whereas the third plasmid had swapped promoters for the HC- and LC-coding sequences (Rituximab B - swap). A 2- to 5-fold q_p increase was observed for all four overexpressed effector genes for Rituximab in the CMV-HC-EF1α-LC composition (**Figure 2A**), especially prominent for *Epas1* and *Mafb* overexpression. Intriguingly, when promoters were swapped, the three novel effector genes had no positive effect on q_p of Rituximab, while *Xbp1s* retained

approximately 2-fold increase in q_p (Figure 2A). However, it is worth noting that the transfection of a plasmid with HC and LC under swapped promoters contributed to 2-fold higher q_p of Rituximab when comparing the mock samples between the two experiments (data not shown). Next, we set out to investigate if the effector genes would give rise to increase in q_p in stable Rituximab-producing cell lines of different origin. We examined the effect of effector genes in three different monoclonal cell lines from two different cell hosts. However, no increase in q_p was observed when transiently expressing effector genes at different dosages in CHO-S and CHO-K1 clonally-derived cell lines, stably expressing Rituximab at different q_p (6 pcd (clone E9_2), 10 pcd (clone C6_2) and 15 pcd (clone GSR)) (**Figure 2B**), despite achieving similar transfection efficiency than in CHO-S wt (data not shown).

All in all, the increase in q_p for all four effector genes seems to be present only when Rituximab is transiently expressed, and not in clonally-derived Rituximab-producing stable cell lines, even for the already established effector gene, *Xbp1s*. Further, the three novel effector genes increase q_p upon transient co-expression only when the Rituximab expression cassette comprises the CMV-HC and EF1α-LC composition.

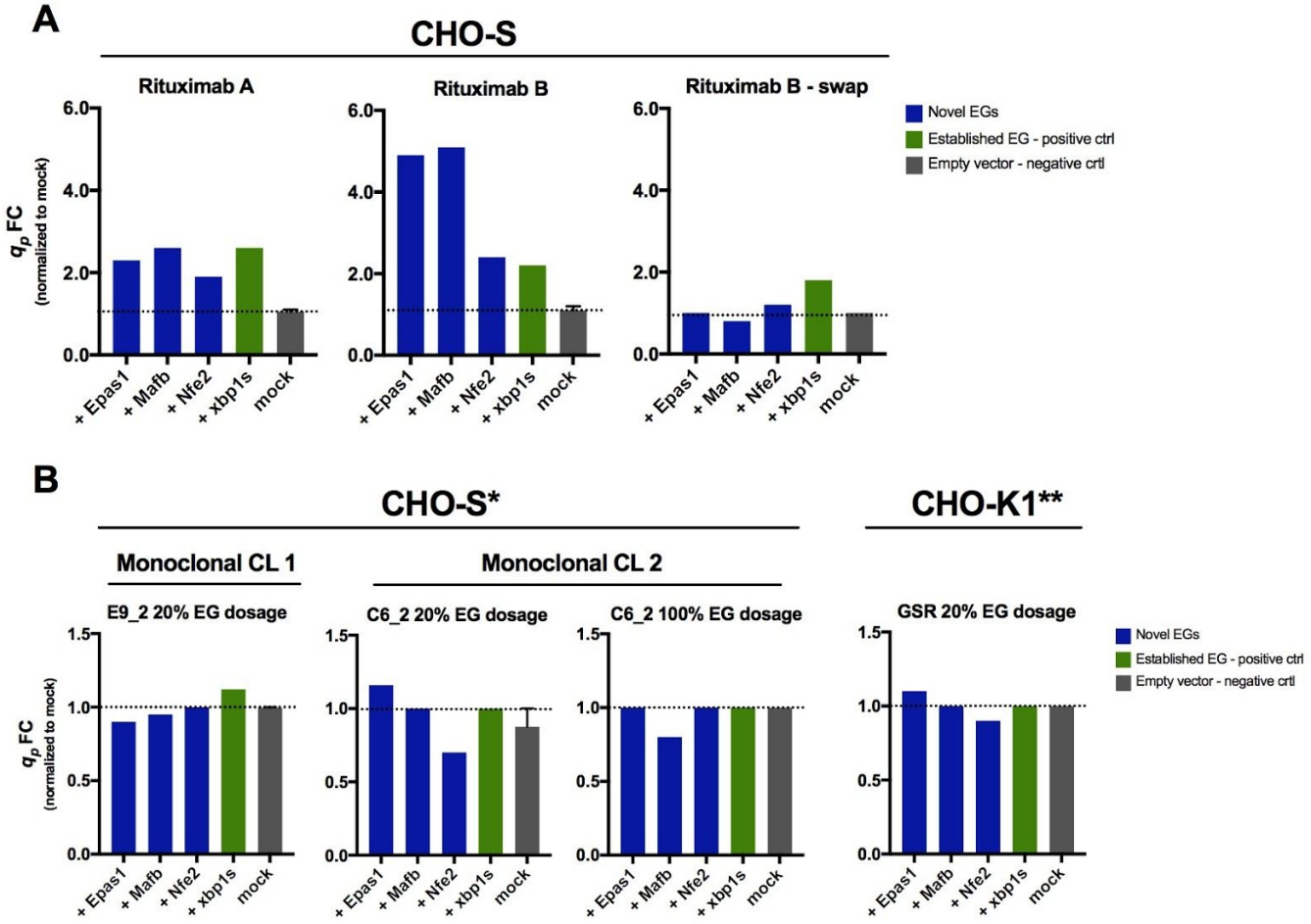


Figure 2: Specific productivity upon co-expression of novel effector genes and Rituximab under different experimental and cellular conditions

(A) Transient co-expression of novel effector genes and Rituximab was performed with different Rituximab expression cassette compositions in CHO-S cells using 6-well plates. Three days post-transfection, q_p was assessed for each condition and normalized to the mock value. Value range from the transfection replicates is shown for the mock sample (n = 2). **(B)** Transient overexpression of effector genes was performed in stable Rituximab-producing cell lines of different origin using 6-well plates. Three days post-transfection, q_p was assessed for each condition and normalized to the mock value. Value range from the transfection replicates is shown for the mock sample (n = 2). CL - cell line; EG - effector gene; FC - fold change. *The monoclonal cell lines were established using Rituximab B plasmid. **The monoclonal cell line was established using a Rituximab plasmid with GS selection marker.

3.3 Increase of q_p via LC mRNA upregulation is observed for *Epas1* and *Mafb* in a promoter-independent way

Since the candidate effector genes were selected on the basis of having potential effects on processes in the secretory pathway, we hypothesised that an increase in q_p was not originating from increased mRNA levels of Rituximab, but rather from processes further downstream. To investigate this, we examined the HC and LC mRNA levels. Surprisingly, when transiently co-expressing Rituximab in a CMV-HC-EF1α-LC composition (Rituximab A and B) and effector genes, 2- to 3.5-fold increase of LC transcript was observed for two novel effector genes, namely *Epas1* and *Mafb*, while the established transcriptional regulator *Xbp1s* increased both, the HC and LC transcripts (**Figure 3A** and **3B**). This first suggested that *Epas1* and *Mafb* increase the q_p of Rituximab via transcriptional regulation of the EF1α promoter. However, when HC and LC were under the swapped promoters, we observed LC upregulation coupled with HC downregulation for *Epas1* and to a lesser extent for *Mafb* (**Figure 3C**), indicating these two effector genes might be upregulating LC in a promoter independent manner. In contrast, in the Rituximab-producing cell line C6_2 overexpression of *Mafb*, *Nfe2* or *Xbps1* did not increase HC or LC transcript levels, however, a marginal LC increase/HC decrease was observed with *Epas1* overexpression (**Figure 3D**).

In conclusion, novel effector genes, *Epas1* and *Mafb,* seem to be direct or indirect transcriptional regulators that increase the amount of present LC transcript in a promoter-independent manner. However, their LC-specific upregulation is causing increased q_p only when Rituximab being transiently expressed is in the CMV-HC and EF1α-LC composition (Figure 2A). On the other hand, *Nfe2* seems to be having a different q_p-increasing mechanism that does not seem to rely on transcriptional regulation of LC or HC.

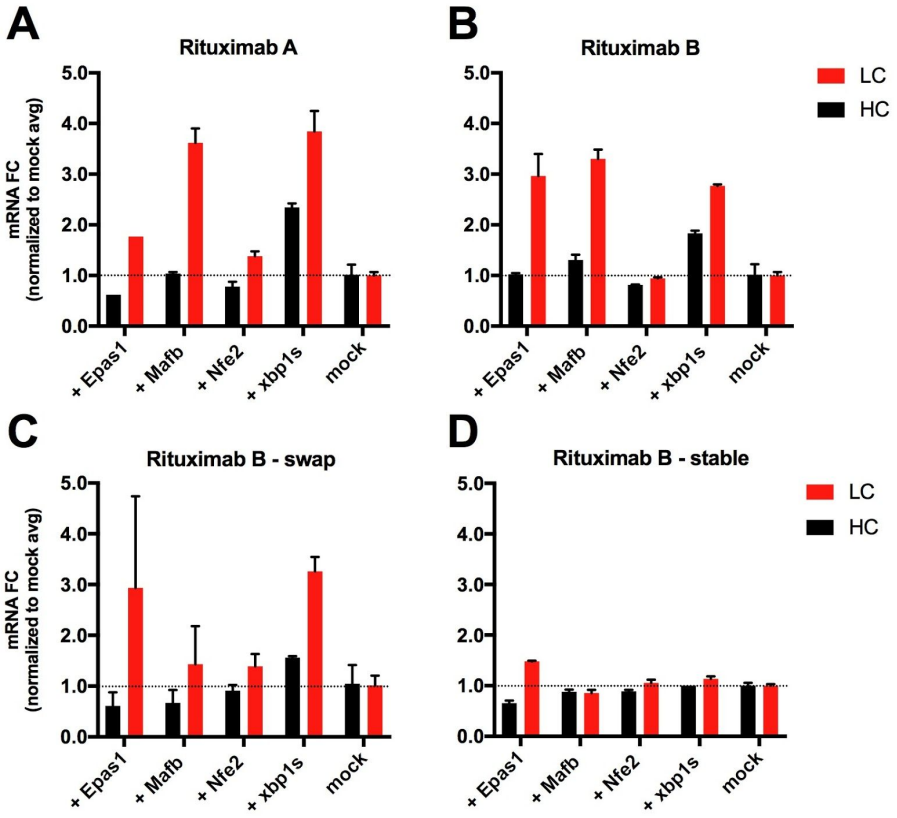


Figure 3: Rituximab heavy chain (HC) and light chain (LC) mRNA transcript levels upon co-expression of novel effector genes and Rituximab

Rituximab HC and LC mRNA transcript levels were measured three days after the transfection of novel effector genes. Transcript levels were normalized to the values of the mock sample. The error bars represent the standard deviations of the technical replicates (n = 3). **(A)** HC and LC mRNA transcript levels upon transient co-expression of novel effector genes and Rituximab A. **(B)** HC and LC mRNA transcript levels upon transient co-expression of novel effector genes and Rituximab B. **(C)** HC and LC mRNA transcript levels upon transient co-expression of novel effector genes and Rituximab B - swap. **(D)** HC and LC mRNA transcript levels upon transient overexpression of novel effector genes in stable Rituximab-producing cell line C6_2. FC - fold change.

3.4 Detection of post-transcriptional bottlenecks upon transient expression of Rituximab

Next, we wanted to investigate whether any transcriptional or post-transcriptional bottlenecks were present when Rituximab was either expressed transiently or constitutively. In our transient expression-based platform, we performed titration using varying Rituximab A dosages ranging from 0% to 100%, while maintaining a fixed total plasmid load. While a linear relationship between the Rituximab plasmid dosage and q_p seemed to be present in the range of 0% to 20%, a non-linear relationship was observed for 33% and 66% plasmid dosage and q_p even decreased at 100% dosage (**Figure 4A**). Increasing HC and LC transcript levels were observed upon increasing Rituximab plasmid dosages, however, in a non-linear

manner (**Figure 4B**). This indicated a possible post-transcriptional bottleneck in production of Rituximab A with an onset after 33% plasmid dosage, where a plateau of q_p is observed with the increasing LC/HC transcript levels (**Figure 4C**). This suggests that a post-transcriptional bottleneck was present when we examined the effect of effector genes on q_p (*i.e.* at 80% Rituximab dosage). Next, to unveil whether or not any bottlenecks were present in the Rituximab-producing C6_2 clonal cell line, we transiently overexpressed Rituximab B in this cell line using a 80% gene dosage (transfection efficiency: 61% GFP-positive cells). No increase in q_p was observed (**Figure 4D**), as well as no increase in LC transcript, while a marginal increase in HC transcript level was demonstrated (**Figure 4E**). This indicated a possible presence of a transcriptional bottleneck upon Rituximab overexpression, suggesting that the rate-limiting step in q_p in the C6_2 clonal cell line is at the transcriptional level.

In conclusion, a transcriptional bottleneck was detected in the Rituximab-producing clonal cell line C6_2 and post-transcriptional bottleneck was detected upon transient expression of Rituximab in CHO-S cells. It seems the conditions surrounding the post-transcriptional bottleneck of transiently expressed Rituximab with CMV-HC and EF1α-LC composition are a prerequisite for observing the effector gene-induced increase of q_p.

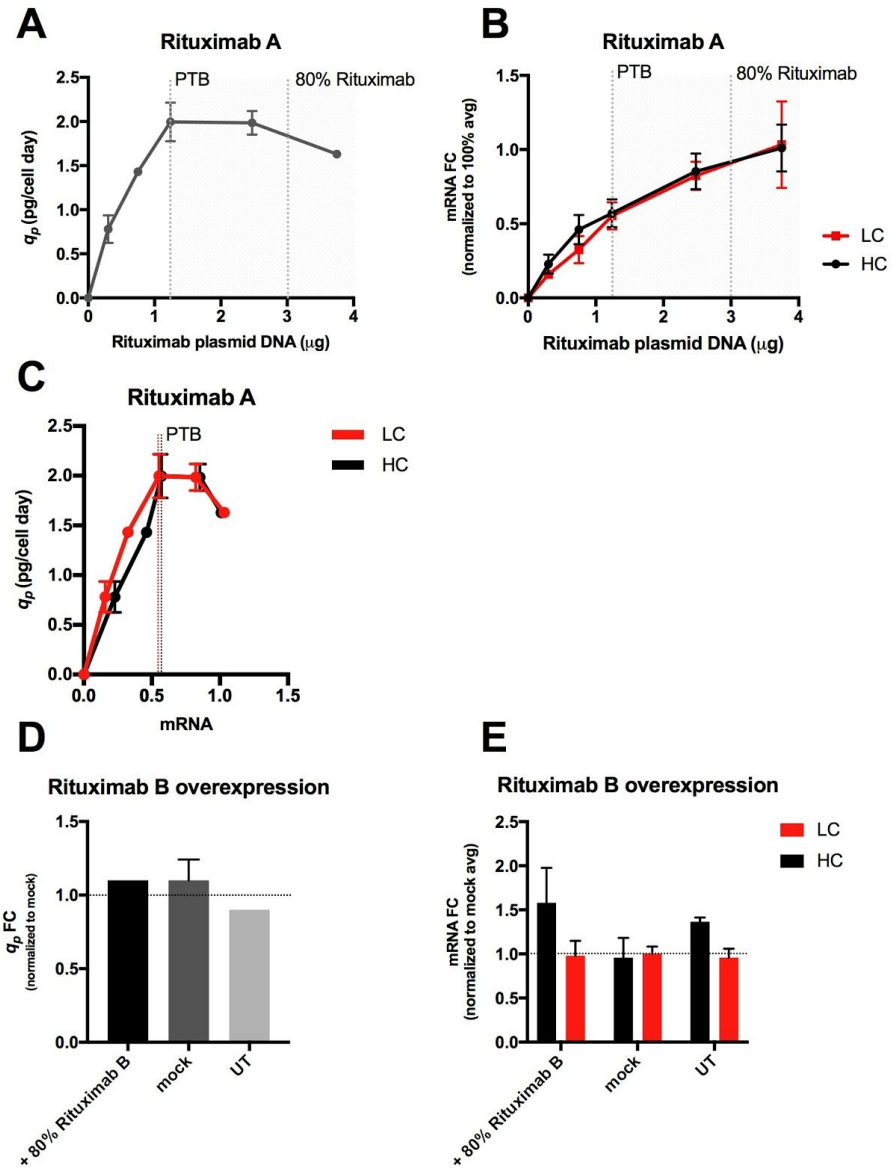


Figure 4: Investigation of intracellular bottlenecks in the production of transiently or constitutively expressed Rituximab

(A) (B) (C) Titration of Rituximab was performed in duplicates (two independent transfections) using 0%, 8%, 20%, 33%, 66% and 100% plasmid dose of Rituximab A in a 3.75 µg of total plasmid DNA load. (A) q_p and (B) Rituximab HC and LC mRNA transcript levels were measured three days post-transfection. Transcript levels were normalized to the values of the 100% sample. Value range from the transfection replicates is shown (n = 2). (C) Determination of a secretion bottleneck of transiently expressed Rituximab A. **(D) (E)** Rituximab B was overexpressed (80% dose) in stable Rituximab-producing cell line C6_2. (D) q_p and (E) Rituximab HC and LC mRNA transcript levels were measured three days post-transfection and normalized to the mock value. q_p value range from the transfection replicates is shown for the mock sample (n = 2). The error bars of the mRNA levels represent the standard deviations of the technical replicates (n = 3). UT - untransfected; PTB - post-transcriptional bottleneck; FC - fold change.

3.5 UPR upregulation observed when Rituximab and the two novel effector genes are transiently co-expressed

After detecting a post-transcriptional bottleneck when Rituximab is expressed transiently and a transcriptional bottleneck when expressed constitutively in stable cell lines, we hypothesized that the expression of Rituximab would give rise to ER stress, and thereby induce the UPR. Therefore, we performed analysis of transcript levels of two UPR-responsive genes, namely BiP and HERP. Transient expression of Rituximab A did not considerably upregulate the UPR levels, with similarly elevated levels of BiP and HERP in the absence and presence of post-transcriptional bottleneck (**Figure 5A**). When co-expressing effector genes and Rituximab B with a post-transcriptional bottleneck-inducing plasmid dosage (80%), we observed an upregulation of HERP when *Epas1*, *Mafb* and *Xbp1s* were overexpressed (**Figure 5B**). Therefore, the UPR seems to be induced upon overexpression of effector genes that in post-transcriptional bottleneck conditions increase the q_p via LC transcript upregulation. When transiently overexpressing Rituximab in the Rituximab-producing C6_2 cell line, the possible presence of transcriptional bottleneck had no UPR-inducing effects (**Figure 5C**). All in all, we demonstrated that UPR levels are upregulated when *Epas1* and *Mafb* increase q_p of Rituximab via LC-inducing transcriptional regulation.

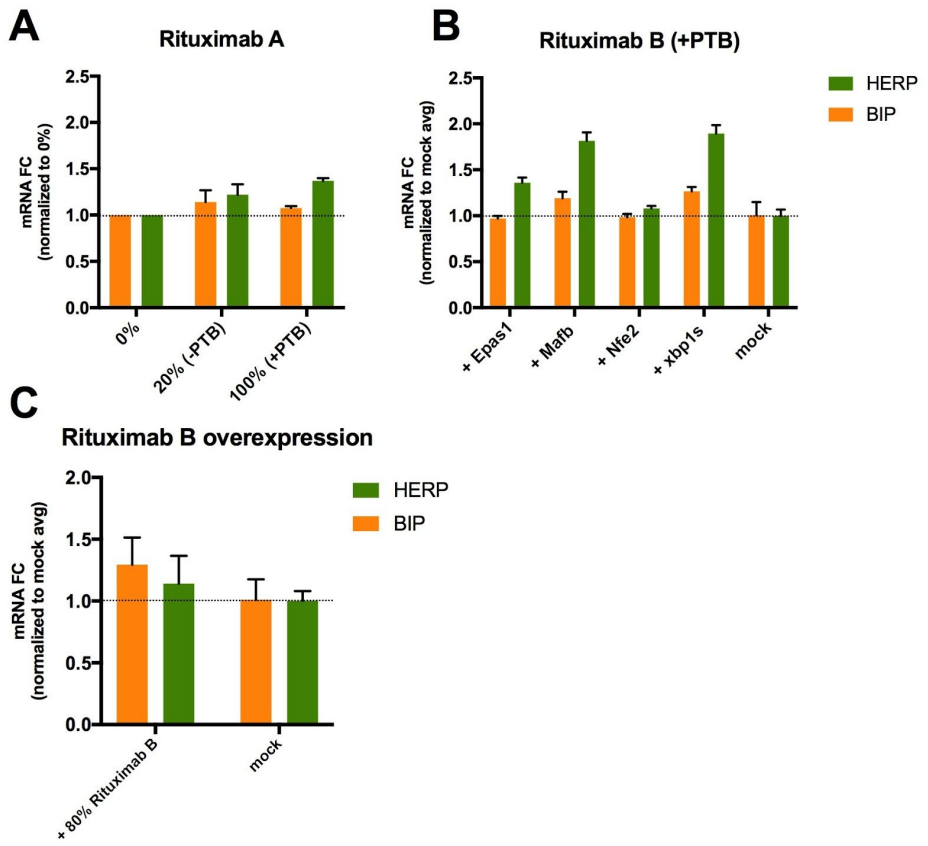


Figure 5: Induction of UPR levels

(A) BiP and HERP mRNA transcript levels were examined upon transient expression of 0%, 20% (*i.e.* absence of post-transcriptional bottleneck) and 100% (*i.e.* presence of post-transcriptional bottleneck) Rituximab plasmid dose. Transcript levels were normalized to the values of the 0% sample. Value range from the transfection replicates is shown (n = 2). **(B)** BiP and HERP mRNA transcript levels were examined upon transient co-expression of novel effector genes and Rituximab B, expressed at the post-transcriptional bottleneck (80% dose). Transcript levels were normalized to the mock sample. The error bars represent the standard deviations of the technical replicates (n = 3). **(C)** BiP and HERP mRNA transcript levels were examined upon transient overexpression of Rituximab B (80% dose) in stable Rituximab-producing cell line C6_2. Transcript levels were normalized to the mock sample. The error bars represent the standard deviations of the technical replicates (n = 3). PTB - post-transcriptional bottleneck; FC - fold change.

3.6 Increase in q_p is observed for different IgG1 molecules, but not for an Fc-fusion protein

We further examined whether the q_p-increasing effect of effector genes was applicable to other therapeutic glycoproteins, such as different IgG1 molecules (Adalimumab and Trastuzumab) and an Fc-fusion protein (Etanercept). We observed an increase in q_p when transiently co-expressing the individual effector genes and either Adalimumab or Trastuzumab (**Figure 6A**), both in the CMV-HC and EF1α-LC expression cassette composition. The effect was most prominently observed for *Mafb* and to a lesser extent for *Epas1* and *Nfe2*. On the other hand, no clear q_p increase was seen for the novel effector genes when co-transfected with either CMV or EF1α promoter-based Etanercept-encoding plasmids, however, a modest effect was observed for the positive control, *Xbp1s* (**Figure 6B**). In conclusion, the q_p-increasing effect of the three novel effector genes seems to be applicable to different IgG molecules of the same subclass (IgG1).

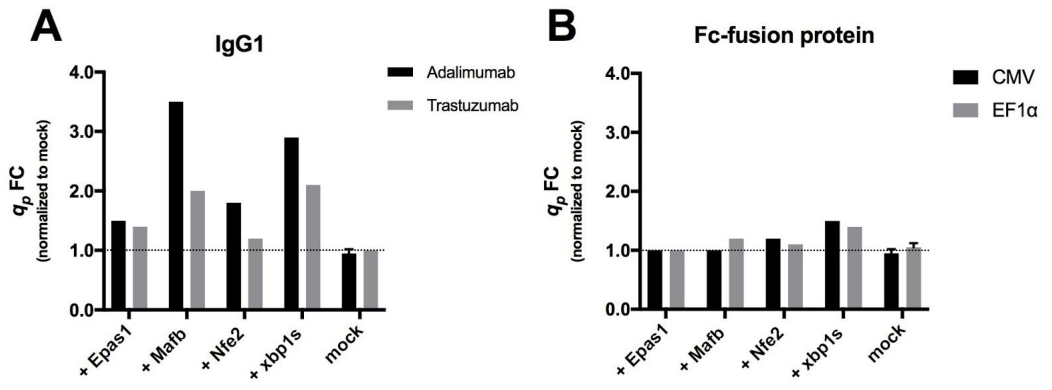


Figure 6: Transient co-expression of novel effector genes and other r-proteins

Transient co-expression of novel effector genes and either **(A)** two different IgG1 molecules (Adalimumab and Trastuzumab) with the CMV-HC and EF1α-LC expression cassette composition or

(B) an Fc-fusion protein (Etanercept) with CMV or EF1α promoter was performed in CHO-S cells using 6-well plates. Three days post-transfection, q_p was assessed for each condition and normalized to the mock value. Value range from the transfection replicates is shown for the mock sample (n = 2). FC - fold change.

4. Discussion

Based on the reconstructed mouse secretory pathway, combined with transcriptome data from CHO non-producing and IgG-producing cells (29), we extracted a panel of candidate effector genes that could be overexpressed and might have the potential to enhance the secretory capacity of CHO cells for IgG production. By ectopically co-expressing this panel of candidate effector genes and Rituximab in our transient expression-based screening platform (30), we identified three novel effector genes (namely *Epas1, Mafb and Nfe2*) with positive effects on our phenotype of interest, q_p. With the objective of elucidating the mechanism by which these novel effector genes give rise to the q_p-increasing effect, we analyzed a host of cellular and experimental factors.

The three novel effector genes reproducibly increased q_p of the cells transiently expressing different IgG1 molecules (Rituximab, Adalimumab, Trastuzumab) delivered in a vector with a CMV-HC-EF1α-LC expression cassette composition. When a Rituximab-encoding plasmid with swapped promoters (EF1α-HC-CMV-LC composition) was transiently transfected into CHO-S cells, no q_p-increasing effects of effector genes were observed. Therefore, it seems that only the CMV-HC-EF1α-LC expression cassette composition created the conditions in which the novel effector genes were able to increase q_p. It has been previously demonstrated that the HC:LC ratio at both the mRNA and polypeptide levels affects the folding and secretion rates of IgG molecules (11, 39). Since a 2-fold increase in q_p was observed for the swapped promoter construct compared to non-swapped promoter construct, the swapped EF1α-HC-CMV-LC composition might generate HC:LC ratios at the mRNA/polypeptide level that are more optimal for IgG folding and secretion. LC might be in excess under the seemingly more potent CMV promoter, resulting in lower HC to LC ratio that was previously reported to improve IgG folding and assembly (40). Next, when analyzing the HC and LC mRNA transcript levels, we observed that *Epas1* and *Mafb* but not *Nfe2*, increased q_p via LC transcript-specific upregulation in a promoter-independent manner. Since the LC-increasing effect was observed regardless of Rituximab promoter composition, this indicates that *Epas1* and *Mafb* function as direct or indirect transcriptional regulators of LC under the given cellular conditions. Meanwhile, *Xbp1s* serving as the 'positive control' effector gene increased both LC and HC transcript levels and increased q_p independent of the Rituximab promoter composition. This is consistent with its function as a general transcription factor

downstream of the UPR transducer IRE1α (41). Moreover, the q_p-increasing effect of *Xbp1s* has been consistent with several previous reports showing an increase in q_p and/or in product titer in different settings and for different r-proteins, such as secreted alkaline phosphatase (SEAP) and vascular endothelial growth factor (VEGF) (22), erythropoietin (EPO) (23), an Fc-fusion protein (18), α1-antitrypsin and C1 esterase inhibitor (30) and IgGs (24, 27, 42).

To test whether the effect of effector genes was not limited only to transient expression in the heterogeneous polyclonal CHO-S pool, we transiently overexpressed effector genes in three stable Rituximab-producing cell lines. These cell lines were of different cell host origin and clonality to address the possibility of both, host-specific and clone-specific effector gene effects. Upon effector gene overexpression, no q_p increase was observed in any of the stable Rituximab-producing cell lines with a range of different q_p (*i.e.* 6–15 pcd). Further, no considerable increase in HC and LC mRNA transcript levels was observed upon effector genes overexpression in the Rituximab-producing C6_2 cell line. There are several possible causes to these observations. First, stable cell lines undergo several stages of selection that are designed to isolate either clones with an innate capability to cope with high levels of r-protein-induced ER stress or clones that have inherently lower ER stress levels (31). Studying the effect of effector genes in stable cell lines is likely biased towards phenotypes inherently adapted to r-protein production. It would therefore be beneficial to integrate the effector genes into the genome of a clonal cell line before the r-protein-encoding gene, as proposed in our review (31). Second, isolated clones might have significantly different endogenous levels of effector genes in comparison to the heterogenous polyclonal host cell line. Therefore, overexpressing effector genes in clonally-derived cell lines might not give rise to an observable effect on q_p with the dosages used in this study; hence transfection of different effector gene-encoding plasmid dosages into these cell lines are warranted. RNA-seq data from our previous studies show that the endogenous levels of novel effector genes are very low or undetectable, both in CHO-S wt cells, as well as in stable cell lines producing different r-proteins (43, 44). And third, previous studies have shown that while the effects of certain effector genes can be observed in different expression platforms (YY1 in (45)), other effector genes might not have the wanted effect in cells transiently and constitutively expressing the r-protein unless rate-limiting conditions in the biosynthetic pathway of r-proteins are present (XBP1 in (23); PDI in (46)). The Rituximab-producing cell lines used in this study possibly have different rate-limiting steps in Rituximab synthesis and/or secretion. Different rate-limiting steps in IgG production might be the reason why only 1 out of 15 'positive control' effector genes (namely *Xbp1s*) gave rise to q_p in our

transient expression-based screening platform, even though they have been previously validated in stable cell lines. Therefore, it would be beneficial if the screening conditions for discovering novel effector genes would as closely as possible capture the environment under which the effector genes are desired to be used, *i.e.* in a stable cell line with an identified rate-limiting step in the secretory pathway.

Since our observations suggest that the three novel effector genes only give rise to an increase in q_p when the r-protein is transiently expressed, we hypothesized that there might be a post-transcriptional bottleneck present when Rituximab was transiently but not constitutively expressed. We were able to identify bottlenecks in the biosynthetic pathway of Rituximab in both expression systems. When transiently expressed, a post-transcriptional bottleneck was detected at 33% or higher Rituximab-encoding plasmid doses, suggesting that a post-transcriptional bottleneck was present when investigating the effect of effector genes. In contrast, a transcriptional bottleneck seemed to be present in the Rituximab-producing cell line C6_2. This possible bottleneck could explain why transient overexpression of the novel effector genes did not give rise to an increase in LC mRNA levels and therefore no detectable effect on q_p. As we previously suggested, the intracellular imbalance of HC and LC polypeptides might be causing suboptimal folding and secretion rates of IgG when Rituximab is transiently expressed. In a recent study, the HC and LC polypeptides were shown to have different rates of intracellular decline, which was dependent on the type of IgG produced (47). The authors further demonstrated that when lower LC decline was observed, higher percentage of total IgG localized within the cis-Golgi, suggesting that the LC excess might result in more efficient transportation of IgG to the Golgi apparatus. It remains to be investigated what is the cause of the intracellular post-transcriptional bottleneck detected in the present study and further, whether or not there is an imbalance between the LC and HC polypeptides. As also previously shown by others (18, 23), our study highlights the importance of determining potential intracellular bottlenecks in the biosynthetic pathway of r-proteins in order to understand under which conditions effector genes have a positive effect on q_p.

We further investigated if the levels of UPR were upregulated when Rituximab was transiently expressed at levels that gave rise to a post-transcriptional bottleneck. We were able to show that the transient expression of Rituximab, with or without the presence of a post-transcriptional bottleneck, did not considerably increase the levels of UPR-responsive genes. However, the overexpression of *Epas1*, *Mafb* and *Xbp1s* when Rituximab was expressed at levels giving rise to a post-transcriptional bottleneck upregulated the mRNA level of the UPR-responsive gene HERP. It remains to be investigated whether the UPR

upregulation is a direct effect of effector gene overexpression or an indirect consequence of their mechanism, *i.e.* q_p-increasing function when Rituximab is at levels giving rise to a post-transcriptional bottleneck. On the other hand, the upregulation of UPR levels was expected for *Xbp1s* as a transcription factor in one of the three UPR pathways (21). Further, transcript levels of UPR-responsive genes were not upregulated upon transient Rituximab overexpression in the Rituximab-producing cell line C6_2. However, we have not assessed the UPR levels in the C6_2 cell line: a) inherently present in comparison to CHO-S wt cells, or b) present upon overexpression of effector genes alone.

As opposed to general effector genes, such as *Xbp1s*, our novel effector genes displayed a limited and conditional q_p-increasing effect, not widely applicable for different expression designs and r-proteins. In fact, only 3 out of 20 candidate effector genes showed positive effects on q_p under the conditions used in this study. This suggests that effector gene engineering prior to integration of r-protein into the genome could be a preferred strategy to enable beneficial effects of novel effector genes (31). Further, our results suggest that a plausible unfavorable HC:LC ratio with the CMV-HC-EF1α-LC expression cassette composition can be improved by vector cassette optimization, giving rise to the increase in q_p.

Despite exploring a variety of cellular and experimental conditions in this study, the underlying q_p- and mRNA(LC)-increasing mechanisms of these novel effector genes have yet to be uncovered. All three novel effector genes are transcription factors, reported to modulate expression levels of genes that are not directly connected to r-protein synthesis or secretion processes. *Epas1* (endothelial PAS domain protein 1) encodes hypoxia-inducible transcription factor alpha (HIF2α) that is one of the key regulators of the cellular response to hypoxia (48). After activation by hypoxia, *Epas1* binds to the hypoxia responsive element (HRE) located in the enhancer and promoter regions of hypoxia-inducible genes (49), involved in metabolism, cell survival, erythropoiesis, and vascular remodeling (50). *Mafb* (v-maf musculoaponeurotic fibrosarcoma oncogene family, protein B) encodes a transcriptional activator or repressor that binds to the Maf recognition element (MARE) by dimerization (51) to regulate lineage-specific gene expression during hematopoiesis (52). *Nfe2* (nuclear factor, erythroid derived 2) is a subunit of the NF-E2 complex that binds the AP-1-like core palindrome present in a number of erythroid and megakaryocytic gene promoters, thereby regulating their maturation and differentiation (53). It would be of interest in future studies to investigate which gene elements or motifs the effector gene-encoded proteins bind to using *e.g.* a chromatin immunoprecipitation (ChIP) assay. Identified genes or gene elements could be further investigated via functional protein

association networks to uncover their possible links to transcriptional regulation of LC/HC transcripts and/or r-protein secretion processes.

Considering the current results, we propose the following multi-layered regulatory mechanism of *Epas1* and *Mafb*: transient expression of Rituximab in a CMV-HC and EF1α-LC expression cassette composition creates conditions under which there is a lack of LC polypeptides. This imbalance between HC and LC polypeptides causes a decreased rate of IgG folding and assembly, which gives rise to a post-transcriptional bottleneck. Upon post-transcriptional bottleneck, *Epas1* and *Mafb* directly or indirectly act as transcriptional regulators to increase the levels of LC (and to a smaller extent decrease the levels of HC) and thereby optimize the HC:LC ratio at the mRNA and polypeptide levels. The optimized ratio results in an increased rate of IgG folding and assembly, which in turn gives rise to an increase in q_p. In turn ER stress and UPR are induced, perhaps due to increased protein folding demand in ER (increased LC mRNA levels), which in turn gives rise to accumulation of incorrectly folded IgG polypeptides. Further investigation is required to demonstrate whether q_p-increasing phenotype is specific to IgG-based post-transcriptional bottlenecks caused by the imbalanced HC and LC levels. In contrast, the mechanism for *Nfe2* is unclear, since its only observed effect in this study was the increase of q_p, which is insufficient to draw any meaningful conclusions.

Future studies are warranted to understand how are the novel effector genes able to alleviate CHO-S wt cells from a post-transcriptional bottleneck originating from transient expression of IgG1 molecules in a CMV-HC and EF1α-LC expression cassette composition. Even though the q_p-increasing effect of effector genes seems to be product-specific, it remains to be explored whether the effect could be observed for other secretory cargoes with similar production bottlenecks present.

5. References

1. Ecker,D.M., Jones,S.D. and Levine,H.L. (2015) The therapeutic monoclonal antibody market. *MAbs*, **7**, 9–14.
2. Grilo,A.L. and Mantalaris,A. (2018) The Increasingly Human and Profitable Monoclonal Antibody Market. *Trends Biotechnol.*, 10.1016/j.tibtech.2018.05.014.
3. de StGroth,S.F. and Scheidegger,D. (1980) Production of monoclonal antibodies: strategy and tactics. *J. Immunol. Methods*, **35**, 1–21.
4. Randall,T.D., Parkhouse,R.M. and Corley,R.B. (1992) J chain synthesis and secretion of hexameric IgM is differentially regulated by lipopolysaccharide and interleukin 5. *Proc. Natl. Acad. Sci. U. S. A.*, **89**, 962–966.
5. Kaufman,R.J. and Sharp,P.A. (1982) Amplification and expression of sequences cotransfected with a modular dihydrofolate reductase complementary dna gene. *J. Mol. Biol.*, **159**, 601–621.
6. Schimke,R.T. (1984) Gene amplification in cultured animal cells. *Cell*, **37**, 705–713.
7. Bebbington,C.R., Renner,G., Thomson,S., King,D., Abrams,D. and Yarranton,G.T. (1992) High-level expression of a recombinant antibody from myeloma cells using a glutamine synthetase gene as

an amplifiable selectable marker. *Biotechnology* , **10**, 169–175.

8. Priola,J.J., Calzadilla,N., Baumann,M., Borth,N., Tate,C.G. and Betenbaugh,M.J. (2016) High-throughput screening and selection of mammalian cells for enhanced protein production. *Biotechnol. J.*, **11**, 853–865.
9. Reinhart,D., Damjanovic,L., Kaisermayer,C. and Kunert,R. (2015) Benchmarking of commercially available CHO cell culture media for antibody production. *Appl. Microbiol. Biotechnol.*, **99**, 4645–4657.
10. Lai,T., Yang,Y. and Ng,S.K. (2013) Advances in Mammalian cell line development technologies for recombinant protein production. *Pharmaceuticals* , **6**, 579–603.
11. Barnes,L.M. and Dickson,A.J. (2006) Mammalian cell factories for efficient and stable protein expression. *Curr. Opin. Biotechnol.*, **17**, 381–386.
12. Alves,C.S. and Dobrowsky,T.M. (2017) Strategies and Considerations for Improving Expression of 'Difficult to Express' Proteins in CHO Cells. *Methods Mol. Biol.*, **1603**, 1–23.
13. Davis,R., Schooley,K., Rasmussen,B., Thomas,J. and Reddy,P. (2000) Effect of PDI overexpression on recombinant protein secretion in CHO cells. *Biotechnol. Prog.*, **16**, 736–743.
14. Borth,N., Mattanovich,D., Kunert,R. and Katinger,H. (2005) Effect of increased expression of protein disulfide isomerase and heavy chain binding protein on antibody secretion in a recombinant CHO cell line. *Biotechnol. Prog.*, **21**, 106–111.
15. Peng,R.-W. and Fussenegger,M. (2009) Molecular engineering of exocytic vesicle traffic enhances the productivity of Chinese hamster ovary cells. *Biotechnol. Bioeng.*, **102**, 1170–1181.
16. Peng,R.-W., Guetg,C., Tigges,M. and Fussenegger,M. (2010) The vesicle-trafficking protein munc18b increases the secretory capacity of mammalian cells. *Metab. Eng.*, **12**, 18–25.
17. Peng,R.-W., Abellan,E. and Fussenegger,M. (2011) Differential effect of exocytic SNAREs on the production of recombinant proteins in mammalian cells. *Biotechnol. Bioeng.*, **108**, 611–620.
18. Johari,Y.B., Estes,S.D., Alves,C.S., Sinacore,M.S. and James,D.C. (2015) Integrated cell and process engineering for improved transient production of a 'difficult-to-express' fusion protein by CHO cells. *Biotechnol. Bioeng.*, **112**, 2527–2542.
19. Le Fourn,V., Girod,P.-A., Buceta,M., Regamey,A. and Mermod,N. (2014) CHO cell engineering to prevent polypeptide aggregation and improve therapeutic protein secretion. *Metab. Eng.*, **21**, 91–102.
20. Sommeregger,W., Mayrhofer,P., Steinfellner,W., Reinhart,D., Henry,M., Clynes,M., Meleady,P. and Kunert,R. (2016) Proteomic differences in recombinant CHO cells producing two similar antibody fragments. *Biotechnol. Bioeng.*, **113**, 1902–1912.
21. Walter,P. and Ron,D. (2011) The unfolded protein response: from stress pathway to homeostatic regulation. *Science*, **334**, 1081–1086.
22. Tigges,M. and Fussenegger,M. (2006) Xbp1-based engineering of secretory capacity enhances the productivity of Chinese hamster ovary cells. *Metab. Eng.*, **8**, 264–272.
23. Ku,S.C.Y., Ng,D.T.W., Yap,M.G.S. and Chao,S.-H. (2008) Effects of overexpression of X-box binding protein 1 on recombinant protein production in Chinese hamster ovary and NS0 myeloma cells. *Biotechnol. Bioeng.*, **99**, 155–164.
24. Becker,E., Florin,L., Pfizenmaier,K. and Kaufmann,H. (2008) An XBP-1 dependent bottle-neck in production of IgG subtype antibodies in chemically defined serum-free Chinese hamster ovary (CHO) fed-batch processes. *J. Biotechnol.*, **135**, 217–223.
25. Ohya,T., Hayashi,T., Kiyama,E., Nishii,H., Miki,H., Kobayashi,K., Honda,K., Omasa,T. and Ohtake,H. (2008) Improved production of recombinant human antithrombin III in Chinese hamster ovary cells by ATF4 overexpression. *Biotechnol. Bioeng.*, **100**, 317–324.
26. O'Callaghan,P.M., McLeod,J., Pybus,L.P., Lovelady,C.S., Wilkinson,S.J., Racher,A.J., Porter,A. and James,D.C. (2010) Cell line-specific control of recombinant monoclonal antibody production by CHO cells. *Biotechnol. Bioeng.*, **106**, 938–951.
27. Pybus,L.P., Dean,G., West,N.R., Smith,A., Daramola,O., Field,R., Wilkinson,S.J. and James,D.C. (2014) Model-directed engineering of 'difficult-to-express' monoclonal antibody production by Chinese hamster ovary cells. *Biotechnol. Bioeng.*, **111**, 372–385.
28. Davies,S.L., Lovelady,C.S., Grainger,R.K., Racher,A.J., Young,R.J. and James,D.C. (2013) Functional heterogeneity and heritability in CHO cell populations. *Biotechnol. Bioeng.*, **110**,

260–274.

29. Lund,A.M., Kaas,C.S., Brandl,J., Pedersen,L.E., Kildegaard,H.F., Kristensen,C. and Andersen,M.R. (2017) Network reconstruction of the mouse secretory pathway applied on CHO cell transcriptome data. *BMC Syst. Biol.*, **11**, 37.
30. Hansen,H.G., Nilsson,C.N., Lund,A.M., Kol,S., Grav,L.M., Lundqvist,M., Rockberg,J., Lee,G.M., Andersen,M.R. and Kildegaard,H.F. (2015) Versatile microscale screening platform for improving recombinant protein productivity in Chinese hamster ovary cells. *Sci. Rep.*, **5**, 18016.
31. Hansen,H.G., Pristovšek,N., Kildegaard,H.F. and Lee,G.M. (2017) Improving the secretory capacity of Chinese hamster ovary cells by ectopic expression of effector genes: Lessons learned and future directions. *Biotechnol. Adv.*, **35**, 64–76.
32. Lund,A.M., Kildegaard,H.F., Petersen,M.B.K., Rank,J., Hansen,B.G., Andersen,M.R. and Mortensen,U.H. (2014) A Versatile System for USER Cloning-Based Assembly of Expression Vectors for Mammalian Cell Engineering. *PLoS One*, **9**, e96693.
33. Kol,S., Kallehauge,T.B., Adema,S. and Hermans,P. (2015) Development of a VHH-Based Erythropoietin Quantification Assay. *Mol. Biotechnol.*, **57**, 692–700.
34. Hansen,B.G., Salomonsen,B., Nielsen,M.T., Nielsen,J.B., Hansen,N.B., Nielsen,K.F., Regueira,T.B., Nielsen,J., Patil,K.R. and Mortensen,U.H. (2011) Versatile enzyme expression and characterization system for Aspergillus nidulans, with the Penicillium brevicompactum polyketide synthase gene from the mycophenolic acid gene cluster as a test case. *Appl. Environ. Microbiol.*, **77**, 3044–3051.
35. Pristovšek,N., Hansen,H.G., Sergeeva,D., Borth,N., Lee,G.M., Andersen,M.R. and Kildegaard,H.F. (2018) Using Titer and Titer Normalized to Confluence Are Complementary Strategies for Obtaining Chinese Hamster Ovary Cell Lines with High Volumetric Productivity of Etanercept. *Biotechnol. J.*, **13**, e1700216.
36. Wishart,D.S., Feunang,Y.D., Guo,A.C., Lo,E.J., Marcu,A., Grant,J.R., Sajed,T., Johnson,D., Li,C., Sayeeda,Z., *et al.* (2018) DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res.*, **46**, D1074–D1082.
37. Grav,L.M., Lee,J.S., Gerling,S., Kallehauge,T.B., Hansen,A.H., Kol,S., Lee,G.M., Pedersen,L.E. and Kildegaard,H.F. (2015) One-step generation of triple knockout CHO cell lines using CRISPR/Cas9 and fluorescent enrichment. *Biotechnol. J.*, **10**, 1446–1456.
38. Nishimiya,D., Mano,T., Miyadai,K., Yoshida,H. and Takahashi,T. (2013) Overexpression of CHOP alone and in combination with chaperones is effective in improving antibody production in mammalian cells. *Appl. Microbiol. Biotechnol.*, **97**, 2531–2539.
39. Dinnis,D.M. and James,D.C. (2005) Engineering mammalian cell factories for improved recombinant monoclonal antibody production: lessons from nature? *Biotechnol. Bioeng.*, **91**, 180–189.
40. Schlatter,S., Stansfield,S.H., Dinnis,D.M., Racher,A.J., Birch,J.R. and James,D.C. (2005) On the optimal ratio of heavy to light chain genes for efficient recombinant antibody production by CHO cells. *Biotechnol. Prog.*, **21**, 122–133.
41. Yoshida,H., Matsui,T., Yamamoto,A., Okada,T. and Mori,K. (2001) XBP1 mRNA is induced by ATF6 and spliced by IRE1 in response to ER stress to produce a highly active transcription factor. *Cell*, **107**, 881–891.
42. Cain,K., Peters,S., Hailu,H., Sweeney,B., Stephens,P., Heads,J., Sarkar,K., Ventom,A., Page,C. and Dickson,A. (2013) A CHO cell line engineered to express XBP1 and ERO1-Lα has increased levels of transient protein expression. *Biotechnol. Prog.*, **29**, 697–706.
43. Singh,A., Kildegaard,H.F. and Andersen,M.R. (2018) An Online Compendium of CHO RNA-Seq Data Allows Identification of CHO Cell Line-Specific Transcriptomic Signatures. *Biotechnol. J.*
44. Grav,L.M., Sergeeva,D., Lee,J.S., Marín de Mas,I., Lewis,N.E., Andersen,M.R., Nielsen,L.K., Lee,G.M. and Kildegaard,H.F. (2018) Minimizing clonal variation during mammalian cell line engineering for improved systems biology data generation. *ACS Synth. Biol.*, 10.1021/acssynbio.8b00140.
45. Tastanova,A., Schulz,A., Folcher,M., Tolstrup,A., Puklowski,A., Kaufmann,H. and Fussenegger,M. (2016) Overexpression of YY1 increases the protein production in mammalian cells. *J. Biotechnol.*, **219**, 72–85.

46. Hayes,N.V.L., Smales,C.M. and Klappa,P. (2010) Protein disulfide isomerase does not control recombinant IgG4 productivity in mammalian cell lines. *Biotechnol. Bioeng.*, **105**, 770–779.
47. Kaneyoshi,K., Uchiyama,K., Onitsuka,M., Yamano,N., Koga,Y. and Omasa,T. (2018) Analysis of intracellular IgG secretion in Chinese hamster ovary cells to improve IgG production. *J. Biosci. Bioeng.*, 10.1016/j.jbiosc.2018.06.018.
48. Covello,K.L. and Simon,M.C. (2004) HIFs, hypoxia, and vascular development. *Curr. Top. Dev. Biol.*, **62**, 37–54.
49. Pugh,C.W., Tan,C.C., Jones,R.W. and Ratcliffe,P.J. (1991) Functional analysis of an oxygen-regulated transcriptional enhancer lying 3' to the mouse erythropoietin gene. *Proc. Natl. Acad. Sci. U. S. A.*, **88**, 10553–10557.
50. Semenza,G.L. (2000) HIF-1 and human disease: one highly involved factor. *Genes Dev.*, **14**, 1983–1991.
51. Kataoka,K., Noda,M. and Nishizawa,M. (1994) Maf nuclear oncoprotein recognizes sequences related to an AP-1 site and forms heterodimers with both Fos and Jun. *Mol. Cell. Biol.*, **14**, 700–712.
52. Sieweke,M.H., Tekotte,H., Frampton,J. and Graf,T. (1996) MafB Is an Interaction Partner and Repressor of Ets-1 That Inhibits Erythroid Differentiation. *Cell*, **85**, 49–60.
53. Casteel,D., Suhasini,M., Gudi,T., Naima,R. and Pilz,R.B. (1998) Regulation of the erythroid transcription factor NF-E2 by cyclic adenosine monophosphate-dependent protein kinase. *Blood*, **91**, 3193–3201.

Concluding remarks and future perspectives

With an ever increasing market share and global demand, the biotherapeutics production in mammalian cell lines is steadily growing. The advancements made in upstream bioprocessing have led to improvement in speed and efficiency of generating robust and high-producing cell lines for large-scale production of biotherapeutics. IgGs currently still account for the majority of the approved biotherapeutics, but in the years to come, the next-generation biotherapeutics, such as alternative scaffolds and difficult-to-produce proteins, will be increasingly more prevalent (41, 83, 84). Further, production of biosimilars, biotherapeutics similar to the already approved drug, will expand as many patents are expiring (85). With the advent of precision medicine, recombinant production of tailored biotherapeutics, fitting the individual patient, might ultimately come to fruition (86).

In the scope of this thesis, clone screening and genome editing tools were developed with the aim to achieve improved and predictable recombinant protein production in CHO cells and accommodate the inevitable expansion of biotherapeutics portfolio.

In **Chapter 1**, we established a clone screening method enabling measurements of confluence and product titer in a relatively high-throughput manner. We further benchmarked a q_p-based clone ranking, using an early stage q_p proxy (TTC ratio), against an established titer-based clone ranking. In a systematic and unbiased comparison, we demonstrated that both clone selection strategies deliver unique clonal cell lines with relatively high product titer in suspension, however, with a low predictive power. Only one other study so far has performed in-depth evaluation of the clone selection criteria (40), despite the fact that clone screening is one of the most crucial stages of early CLD. Hopefully, studies like ours will encourage the field to systematically examine the clone screening methods used today, to improve the understanding of different selection criteria and their prediction power. In addition, to demonstrate the usefulness of the established screening workflow, we have generated a stable, high IgG-producing cell line (C6_2 producer) that was used in several studies throughout the thesis and elsewhere, as well as for the identification of a novel, safe harbor integration site in CHO-S genome.

In **Chapter 2**, we developed a CRISPR/Cas9-based toolbox for generation of site-specific recombinant cell lines that can be utilized for systematic screening of different expression cassette components. The mammalian cell engineering community can now decipher how multiple components contribute to the final expression of recombinant genes in characterized integration sites. The toolbox therefore facilitates optimized mammalian cell engineering and a detailed characterization of proposed safe harbors across different

mammalian expression platforms. Future work could focus on boosting the productivity of biotherapeutics, by creating synthetic circuits that are dynamically responsive to culture conditions; or for effector gene engineering by precisely defining and controlling the optimal expression level of multiple effector genes in stable producing cell lines. All in all, this toolbox could pave the way toward CHO multigene engineering approaches, currently exceeding our ability to coordinately balance and test the expression of multiple single effector genes.

There are three fundamental approaches to increase the productivity of a recombinant protein: host cell engineering, vector design engineering, and the optimization of the cell culture processes (83). **Chapter 3** has covered current state of the art of host cell engineering by ectopically expressing the effector genes improving the secretory capacity of CHO cells. Several cellular and experimental factors were described, affecting the outcomes of such engineering approaches. Whereas, in **Chapter 4**, a study was performed where we screened a panel of candidate effector genes to identify novel, q_p-increasing effectors. Alongside the candidate panel, we have screened a 'positive control' panel with effector genes from published studies and demonstrated that only a few had reproducible effects on q_p in conditions used in our study. We alleviated the secretion of transiently expressed IgG1 molecules by overexpressing three novel effector genes, and then attempted to determine which rate-limiting step in IgG1 production was relieved upon effector gene overexpression. In the scope of the thesis, we have not yet conclusively determined what is the underlying mechanism of the novel effector genes, and whether or not the effects are cell line- and/or protein-specific. However, we have shown that effector gene studies require a fundamental understanding of the recombinant protein of interest, as well as a well-designed set of experiments to discover and effectively alleviate its secretion bottlenecks. Furthermore, given the existence of product-interacting effector genes, cell process-specific or non-specific (pleiotropic) effector genes, an overexpression of a single effector gene may not be sufficient for improving the recombinant protein production. Instead, coordinated expression of multiple effector genes is likely to be a preferred option for future cell engineering strategies. Traditionally, several non-specific effectors have been used to universally enhance different cell functions (*e.g.* transactivators). However, these effectors may simultaneously and unpredictably affect many cellular processes that are irrelevant or even adverse to product-specific recombinant production. Therefore, effector genes that alleviate specific, rate-limiting cellular processes (*e.g.* secretory vesicle formation, UPR induction) are the logical engineering targets in a product- and process-specific context.

In conclusion, coordinated application of the tools developed in this thesis supports rational engineering of next-generation CHO cell factories and has the potential to substantially improve the production of a wide range of biotherapeutics.

References

1. Wurm,F.M. (2004) Production of recombinant protein therapeutics in cultivated mammalian cells. *Nat. Biotechnol.*, **22**, 1393.
2. Barnes,L.M. and Dickson,A.J. (2006) Mammalian cell factories for efficient and stable protein expression. *Curr. Opin. Biotechnol.*, **17**, 381–386.
3. Kim,J.Y., Kim,Y.-G. and Lee,G.M. (2012) CHO cells in biotechnology for production of recombinant proteins: current state and further potential. *Appl. Microbiol. Biotechnol.*, **93**, 917–930.
4. Walsh,G. (2010) Biopharmaceutical benchmarks 2010. *Nat. Biotechnol.*, **28**, 917–924.
5. Schmidt,F.R. (2004) Recombinant expression systems in the pharmaceutical industry. *Appl. Microbiol. Biotechnol.*, **65**, 363–372.
6. Collins,J.H. and Young,E.M. (2018) Genetic engineering of host organisms for pharmaceutical synthesis. *Curr. Opin. Biotechnol.*, **53**, 191–200.
7. Zhu,J. (2012) Mammalian cell protein expression for biopharmaceutical production. *Biotechnol. Adv.*, **30**, 1158–1170.
8. Lai,T., Yang,Y. and Ng,S.K. (2013) Advances in Mammalian cell line development technologies for recombinant protein production. *Pharmaceuticals* , **6**, 579–603.
9. Kaufman,R.J. and Sharp,P.A. (1982) Amplification and expression of sequences cotransfected with a modular dihydrofolate reductase complementary dna gene. *J. Mol. Biol.*, **159**, 601–621.
10. Schimke,R.T. (1984) Gene amplification in cultured animal cells. *Cell*, **37**, 705–713.
11. Bebbington,C.R., Renner,G., Thomson,S., King,D., Abrams,D. and Yarranton,G.T. (1992) High-level expression of a recombinant antibody from myeloma cells using a glutamine synthetase gene as an amplifiable selectable marker. *Biotechnology* , **10**, 169–175.
12. Li,F., Vijayasankaran,N., Shen,A.Y., Kiss,R. and Amanullah,A. (2010) Cell culture processes for monoclonal antibody production. *MAbs*, **2**, 466–479.
13. Priola,J.J., Calzadilla,N., Baumann,M., Borth,N., Tate,C.G. and Betenbaugh,M.J. (2016) High-throughput screening and selection of mammalian cells for enhanced protein production. *Biotechnol. J.*, **11**, 853–865.
14. Kuo,C.-C., Chiang,A.W., Shamie,I., Samoudi,M., Gutierrez,J.M. and Lewis,N.E. (2018) The emerging role of systems biology for engineering protein production in CHO cells. *Curr. Opin. Biotechnol.*, **51**, 64–69.
15. Huang,Y.-M., Hu,W., Rustandi,E., Chang,K., Yusuf-Makagiansar,H. and Ryll,T. (2010) Maximizing productivity of CHO cell-based fed-batch culture using chemically defined media conditions and typical manufacturing equipment. *Biotechnol. Prog.*, **26**, 1400–1410.
16. Lee,J.S., Kallehauge,T.B., Pedersen,L.E. and Kildegaard,H.F. (2015) Site-specific integration in CHO cells mediated by CRISPR/Cas9 and homology-directed DNA repair pathway. *Sci. Rep.*, **5**, 8572.
17. Xu,X., Nagarajan,H., Lewis,N.E., Pan,S., Cai,Z., Liu,X., Chen,W., Xie,M., Wang,W., Hammond,S., *et al.* (2011) The genomic sequence of the Chinese hamster ovary (CHO)-K1 cell line. *Nat. Biotechnol.*, **29**, 735–741.
18. Hammond,S., Swanberg,J.C., Kaplarevic,M. and Lee,K.H. (2011) Genomic sequencing and analysis of a Chinese hamster ovary cell line using Illumina sequencing technology. *BMC Genomics*, **12**, 67.
19. Lewis,N.E., Liu,X., Li,Y., Nagarajan,H., Yerganian,G., O'Brien,E., Bordbar,A., Roth,A.M., Rosenbloom,J., Bian,C., *et al.* (2013) Genomic landscapes of Chinese hamster ovary cell

lines as revealed by the Cricetulus griseus draft genome. *Nat. Biotechnol.*, **31**, 759.
20. Brinkrolf,K., Rupp,O., Laux,H., Kollin,F., Ernst,W., Linke,B., Kofler,R., Romand,S., Hesse,F., Budach,W.E., *et al.* (2013) Chinese hamster genome sequenced from sorted chromosomes. *Nat. Biotechnol.*, **31**, 694.
21. Rupp,O., MacDonald,M.L., Li,S., Dhiman,H., Polson,S., Griep,S., Heffner,K., Hernandez,I., Brinkrolf,K., Jadhav,V., *et al.* (2018) A reference genome of the Chinese hamster based on a hybrid assembly strategy. *Biotechnol. Bioeng.*, **115**, 2087–2100.
22. Lee,J.S., Grav,L.M., Lewis,N.E. and Faustrup Kildegaard,H. (2015) CRISPR/Cas9-mediated genome engineering of CHO cell factories: Application and perspectives. *Biotechnol. J.*, **10**, 979–994.
23. Becker,J., Hackl,M., Rupp,O., Jakobi,T., Schneider,J., Szczepanowski,R., Bekel,T., Borth,N., Goesmann,A., Grillari,J., *et al.* (2011) Unraveling the Chinese hamster ovary cell line transcriptome by next-generation sequencing. *J. Biotechnol.*, **156**, 227–235.
24. Baycin-Hizal,D., Tabb,D.L., Chaerkady,R., Chen,L., Lewis,N.E., Nagarajan,H., Sarkaria,V., Kumar,A., Wolozny,D., Colao,J., *et al.* (2012) Proteomic analysis of Chinese hamster ovary cells. *J. Proteome Res.*, **11**, 5265–5276.
25. Clarke,C., Henry,M., Doolan,P., Kelly,S., Aherne,S., Sanchez,N., Kelly,P., Kinsella,P., Breen,L., Madden,S.F., *et al.* (2012) Integrated miRNA, mRNA and protein expression analysis reveals the role of post-transcriptional regulation in controlling CHO cell growth rate. *BMC Genomics*, **13**, 656.
26. Courtes,F.C., Lin,J., Lim,H.L., Ng,S.W., Wong,N.S.C., Koh,G., Vardy,L., Yap,M.G.S., Loo,B. and Lee,D.-Y. (2013) Translatome analysis of CHO cells to identify key growth genes. *J. Biotechnol.*, **167**, 215–224.
27. Hefzi,H., Ang,K.S., Hanscho,M., Bordbar,A., Ruckerbauer,D., Lakshmanan,M., Orellana,C.A., Baycin-Hizal,D., Huang,Y., Ley,D., *et al.* (2016) A Consensus Genome-scale Reconstruction of Chinese Hamster Ovary Cell Metabolism. *Cell Systems*, **3**, 434–443.e8.
28. Coller,H.A. and Coller,B.S. (1986) Poisson statistical analysis of repetitive subcloning by the limiting dilution technique as a way of assessing hybridoma monoclonality. *Methods Enzymol.*, **121**, 412–417.
29. Lanza,A.M., Kim,D.S. and Alper,H.S. (2013) Evaluating the influence of selection markers on obtaining selected pools and stable cell lines in human cells. *Biotechnol. J.*, **8**, 811–821.
30. Cacciatore,J.J., Leonard,E.F. and Chasin,L.A. (2013) The isolation of CHO cells with a site conferring a high and reproducible transgene amplification rate. *J. Biotechnol.*, **164**, 346–353.
31. Kumar,N. and Borth,N. (2012) Flow-cytometry and cell sorting: an efficient approach to investigate productivity and cell physiology in mammalian cell factories. *Methods*, **56**, 366–374.
32. Brezinsky,S.C.G., Chiang,G.G., Szilvasi,A., Mohan,S., Shapiro,R.I., MacLean,A., Sisk,W. and Thill,G. (2003) A simple method for enriching populations of transfected CHO cells for cells of higher specific productivity. *J. Immunol. Methods*, **277**, 141–155.
33. Hanania,E.G., Fieck,A., Stevens,J., Bodzin,L.J., Palsson,B.Ø. and Koller,M.R. (2005) Automated in situ measurement of cell-specific antibody secretion and laser-mediated purification for rapid cloning of highly-secreting producers. *Biotechnol. Bioeng.*, **91**, 872–876.
34. Hou,J.J.C., Hughes,B.S., Smede,M., Leung,K.M., Levine,K., Rigby,S., Gray,P.P. and Munro,T.P. (2014) High-throughput ClonePix FL analysis of mAb-expressing clones using the UCOE expression system. *N. Biotechnol.*, **31**, 214–220.

35. Love,J.C., Ronan,J.L., Grotenbreg,G.M., van der Veen,A.G. and Ploegh,H.L. (2006) A microengraving method for rapid selection of single cells producing antigen-specific antibodies. *Nat. Biotechnol.*, **24**, 703.
36. Lindgren,K., Salmén,A., Lundgren,M., Bylund,L., Ebler,A., Fäldt,E., Sörvik,L., Fenge,C. and Skoging-Nyberg,U. (2009) Automation of cell line development. *Cytotechnology*, **59**, 1–10.
37. Harriman,W.D., Collarini,E.J., Cromer,R.G., Dutta,A., Strandh,M., Zhang,F. and Kauvar,L.M. (2009) Multiplexed Elispot assay. *J. Immunol. Methods*, **341**, 127–134.
38. Shi,S., Condon,R.G.G., Deng,L., Saunders,J., Hung,F., Tsao,Y.-S. and Liu,Z. (2011) A high-throughput automated platform for the development of manufacturing cell lines for protein therapeutics. *J. Vis. Exp.*, 10.3791/3010.
39. Porter,A.J., Dickson,A.J. and Racher,A.J. (2010) Strategies for selecting recombinant CHO cell lines for cGMP manufacturing: realizing the potential in bioreactors. *Biotechnol. Prog.*, **26**, 1446–1454.
40. Porter,A.J., Racher,A.J., Preziosi,R. and Dickson,A.J. (2010) Strategies for selecting recombinant CHO cell lines for cGMP manufacturing: improving the efficiency of cell line generation. *Biotechnol. Prog.*, **26**, 1455–1464.
41. Fischer,S., Handrick,R. and Otte,K. (2015) The art of CHO cell engineering: A comprehensive retrospect and future perspectives. *Biotechnol. Adv.*, **33**, 1878–1896.
42. Kildegaard,H.F., Baycin-Hizal,D., Lewis,N.E. and Betenbaugh,M.J. (2013) The emerging CHO systems biology era: harnessing the 'omics revolution for biotechnology. *Curr. Opin. Biotechnol.*, **24**, 1102–1107.
43. Grav,L.M., Lee,J.S., Gerling,S., Kallehauge,T.B., Hansen,A.H., Kol,S., Lee,G.M., Pedersen,L.E. and Kildegaard,H.F. (2015) One-step generation of triple knockout CHO cell lines using CRISPR/Cas9 and fluorescent enrichment. *Biotechnol. J.*, **10**, 1446–1456.
44. Nehlsen,K., Schucht,R., da Gama-Norton,L., Krömer,W., Baer,A., Cayli,A., Hauser,H. and Wirth,D. (2009) Recombinant protein expression by targeting pre-selected chromosomal loci. *BMC Biotechnol.*, **9**, 100.
45. Kim,M.S. and Lee,G.M. (2008) Use of Flp-mediated cassette exchange in the development of a CHO cell line stably producing erythropoietin. *J. Microbiol. Biotechnol.*, **18**, 1342–1351.
46. Zhou,H., Liu,Z.-G., Sun,Z.-W., Huang,Y. and Yu,W.-Y. (2010) Generation of stable cell lines by site-specific integration of transgenes into engineered Chinese hamster ovary strains using an FLP-FRT system. *J. Biotechnol.*, **147**, 122–129.
47. Campbell,M., Corisdeo,S., McGee,C. and Kraichely,D. (2010) Utilization of site-specific recombination for generating therapeutic protein producing cell lines. *Mol. Biotechnol.*, **45**, 199–202.
48. Wiberg,F.C., Rasmussen,S.K., Frandsen,T.P., Rasmussen,L.K., Tengbjerg,K., Coljee,V.W., Sharon,J., Yang,C.-Y., Bregenholt,S., Nielsen,L.S., *et al.* (2006) Production of target-specific recombinant human polyclonal antibodies in mammalian cells. *Biotechnol. Bioeng.*, **94**, 396–405.
49. Huang,Y., Li,Y., Wang,Y.G., Gu,X., Wang,Y. and Shen,B.F. (2007) An efficient and targeted gene integration system for high-level antibody expression. *J. Immunol. Methods*, **322**, 28–39.
50. Kito,M., Itami,S., Fukano,Y., Yamana,K. and Shibui,T. (2002) Construction of engineered CHO strains for high-level production of recombinant proteins. *Appl. Microbiol. Biotechnol.*, **60**, 442–448.
51. Derouazi,M., Martinet,D., Besuchet Schmutz,N., Flaction,R., Wicht,M., Bertschinger,M.,

Hacker,D.L., Beckmann,J.S. and Wurm,F.M. (2006) Genetic characterization of CHO production host DG44 and derivative recombinant cell lines. *Biochem. Biophys. Res. Commun.*, **340**, 1069–1077.

52. Turan,S. and Bode,J. (2011) Site-specific recombinases: from tag-and-target- to tag-and-exchange-based genomic modifications. *FASEB J.*, **25**, 4088–4107.
53. Turan,S., Galla,M., Ernst,E., Qiao,J., Voelkel,C., Schiedlmeier,B., Zehe,C. and Bode,J. (2011) Recombinase-mediated cassette exchange (RMCE): traditional concepts and current challenges. *J. Mol. Biol.*, **407**, 193–221.
54. Kim,H. and Kim,J.-S. (2014) A guide to genome engineering with programmable nucleases. *Nat. Rev. Genet.*, **15**, 321.
55. Mali,P., Yang,L., Esvelt,K.M., Aach,J., Guell,M., DiCarlo,J.E., Norville,J.E. and Church,G.M. (2013) RNA-guided human genome engineering via Cas9. *Science*, **339**, 823–826.
56. Cong,L., Ran,F.A., Cox,D., Lin,S., Barretto,R., Habib,N., Hsu,P.D., Wu,X., Jiang,W., Marraffini,L.A., *et al.* (2013) Multiplex genome engineering using CRISPR/Cas systems. *Science*, **339**, 819–823.
57. Jinek,M., East,A., Cheng,A., Lin,S., Ma,E. and Doudna,J. (2013) RNA-programmed genome editing in human cells. *Elife*, **2**, e00471.
58. Jiang,W., Bikard,D., Cox,D., Zhang,F. and Marraffini,L.A. (2013) RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nat. Biotechnol.*, **31**, 233–239.
59. Hwang,W.Y., Fu,Y., Reyon,D., Maeder,M.L., Tsai,S.Q., Sander,J.D., Peterson,R.T., Yeh,J.-R.J. and Joung,J.K. (2013) Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nat. Biotechnol.*, **31**, 227–229.
60. Cho,S.W., Kim,S., Kim,J.M. and Kim,J.-S. (2013) Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. *Nat. Biotechnol.*, **31**, 230.
61. Jinek,M., Chylinski,K., Fonfara,I., Hauer,M., Doudna,J.A. and Charpentier,E. (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*, **337**, 816–821.
62. Gaj,T., Gersbach,C.A. and Barbas,C.F.,3rd (2013) ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends Biotechnol.*, **31**, 397–405.
63. Ran,F.A., Hsu,P.D., Wright,J., Agarwala,V., Scott,D.A. and Zhang,F. (2013) Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.*, **8**, 2281–2308.
64. Lee,J.S., Grav,L.M., Pedersen,L.E., Lee,G.M. and Kildegaard,H.F. (2016) Accelerated homology-directed targeted integration of transgenes in Chinese hamster ovary cells via CRISPR/Cas9 and fluorescent enrichment. *Biotechnol. Bioeng.*, **113**, 2518–2523.
65. Grav,L.M., Sergeeva,D., Lee,J.S., Marín de Mas,I., Lewis,N.E., Andersen,M.R., Nielsen,L.K., Lee,G.M. and Kildegaard,H.F. (2018) Minimizing clonal variation during mammalian cell line engineering for improved systems biology data generation. *ACS Synth. Biol.*, 10.1021/acssynbio.8b00140.
66. Inniss,M.C., Bandara,K., Jusiak,B., Lu,T.K., Weiss,R., Wroblewska,L. and Zhang,L. (2017) A novel Bxb1 integrase RMCE system for high fidelity site-specific integration of mAb expression cassette in CHO Cells. *Biotechnol. Bioeng.*, **114**, 1837–1846.
67. Gaidukov,L., Wroblewska,L., Teague,B., Nelson,T., Zhang,X., Liu,Y., Jagtap,K., Mamo,S., Tseng,W.A., Lowe,A., *et al.* (2018) A multi-landing pad DNA integration platform for mammalian cell engineering. *Nucleic Acids Res.*, **46**, 4072–4086.
68. Tey,B.T., Singh,R.P., Piredda,L., Piacentini,M. and Al-Rubeai,M. (2000) Influence of bcl-2 on cell death during the cultivation of a Chinese hamster ovary cell line expressing a chimeric antibody. *Biotechnol. Bioeng.*, **68**, 31–43.
69. Ifandi,V. and Al-Rubeai,M. (2008) Regulation of Cell Proliferation and Apoptosis in

CHO-K1 Cells by the Coexpression of c-Myc and Bcl-2. *Biotechnol. Prog.*, **21**, 671–677.
70. Park,H., Kim,I.H., Kim,I.Y., Kim,K.H. and Kim,H.J. (2000) Expression of carbamoyl phosphate synthetase I and ornithine transcarbamoylase genes in Chinese hamster ovary dhfr-cells decreases accumulation of ammonium ion in culture media. *J. Biotechnol.*, **81**, 129–140.
71. Kim,S.H. and Lee,G.M. (2007) Functional expression of human pyruvate carboxylase for reduced lactic acid formation of Chinese hamster ovary cells (DG44). *Appl. Microbiol. Biotechnol.*, **76**, 659–665.
72. Wlaschin,K.F. and Hu,W.-S. (2007) Engineering cell metabolism for high-density cell culture via manipulation of sugar transport. *J. Biotechnol.*, **131**, 168–176.
73. Mazur,X., Fussenegger,M., Renner,W.A. and Bailey,J.E. (1998) Higher productivity of growth-arrested Chinese hamster ovary cells expressing the cyclin-dependent kinase inhibitor p27. *Biotechnol. Prog.*, **14**, 705–713.
74. Zhang,X., Lok,S.H. and Kon,O.L. (1998) Stable expression of human alpha-2,6-sialyltransferase in Chinese hamster ovary cells: functional consequences for human erythropoietin expression and bioactivity. *Biochim. Biophys. Acta*, **1425**, 441–452.
75. Weikert,S., Papac,D., Briggs,J., Cowfer,D., Tom,S., Gawlitzek,M., Lofgren,J., Mehta,S., Chisholm,V., Modi,N., *et al.* (1999) Engineering Chinese hamster ovary cells to maximize sialic acid content of recombinant glycoproteins. *Nat. Biotechnol.*, **17**, 1116–1121.
76. Borth,N., Mattanovich,D., Kunert,R. and Katinger,H. (2005) Effect of increased expression of protein disulfide isomerase and heavy chain binding protein on antibody secretion in a recombinant CHO cell line. *Biotechnol. Prog.*, **21**, 106–111.
77. Ohya,T., Hayashi,T., Kiyama,E., Nishii,H., Miki,H., Kobayashi,K., Honda,K., Omasa,T. and Ohtake,H. (2008) Improved production of recombinant human antithrombin III in Chinese hamster ovary cells by ATF4 overexpression. *Biotechnol. Bioeng.*, **100**, 317–324.
78. Pybus,L.P., Dean,G., West,N.R., Smith,A., Daramola,O., Field,R., Wilkinson,S.J. and James,D.C. (2014) Model-directed engineering of 'difficult-to-express' monoclonal antibody production by Chinese hamster ovary cells. *Biotechnol. Bioeng.*, **111**, 372–385.
79. Tigges,M. and Fussenegger,M. (2006) Xbp1-based engineering of secretory capacity enhances the productivity of Chinese hamster ovary cells. *Metab. Eng.*, **8**, 264–272.
80. Peng,R.-W. and Fussenegger,M. (2009) Molecular engineering of exocytic vesicle traffic enhances the productivity of Chinese hamster ovary cells. *Biotechnol. Bioeng.*, **102**, 1170–1181.
81. Le Fourn,V., Girod,P.-A., Buceta,M., Regamey,A. and Mermod,N. (2014) CHO cell engineering to prevent polypeptide aggregation and improve therapeutic protein secretion. *Metab. Eng.*, **21**, 91–102.
82. Hansen,H.G., Pristovšek,N., Kildegaard,H.F. and Lee,G.M. (2017) Improving the secretory capacity of Chinese hamster ovary cells by ectopic expression of effector genes: Lessons learned and future directions. *Biotechnol. Adv.*, **35**, 64–76.
83. Alves,C.S. and Dobrowsky,T.M. (2017) Strategies and Considerations for Improving Expression of 'Difficult to Express' Proteins in CHO Cells. *Methods Mol. Biol.*, **1603**, 1–23.
84. Kintzing,J.R., Filsinger Interrante,M.V. and Cochran,J.R. (2016) Emerging Strategies for Developing Next-Generation Protein Therapeutics for Cancer Treatment. *Trends Pharmacol. Sci.*, **37**, 993–1008.
85. Huzair,F. and Kale,D. (2015) Biosimilars and the long game. *Trends Biotechnol.*, **33**,

250–252.
86. Collins,F.S. and Varmus,H. (2015) A New Initiative on Precision Medicine. *N. Engl. J. Med.*, **372**, 793–795.

Using Titer and Titer Normalized to Confluence are Complementary Strategies for Obtaining Chinese Hamster Ovary Cell Lines with High Volumetric Productivity of Etanercept

Supplementary data

Nuša Pristovšek[1], Henning Gram Hansen[1], Daria Sergeeva[1], Nicole Borth[2, 3], Gyun Min Lee[1,4], Mikael Rørdam Andersen[5] and Helene Faustrup Kildegaard[1]

[1]The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Kemitorvet 220, 2800 Kgs. Lyngby, Denmark
[2]Department of Biotechnology, University of Natural Resources and Life Sciences, Muthgasse 19, 1190 Vienna, Austria
[3]Austrian Centre of Industrial Biotechnology (ACIB), Muthgasse 11, 1190 Vienna, Austria
[4]Department of Biological Sciences, KAIST, 291 Daehak-ro, Yuseong-gu, Daejeon 305-701, Republic of Korea
[5]Department of Biotechnology and Biomedicine, Technical University of Denmark, Søltofts Plads, Building 221, 2800 Kgs. Lyngby, Denmark

Table S1: Nucleotide sequence of the etanercept expression cassette.

	5'-3': mCMV-hEF-1a-5'HTLV-etanercept(cds)-BGHpA
Etanercept	GAGTCAATGGGAAAAACCCATTGGAGCCAAGTACACTGACTCAATAGGGACTTTCCATTGGGTTTTGCCCAGTACATAAGGTCAATAGGGGGTGAGTCAACAGGAAAGTCCCATTGGAGCCAAGTACATTGAGTCAATAGGGACTTTCCAATGGGTTT TGCCCAGTACATAAGGTCAATGGGAGGTAAGCCAATGGGTTTTTCCCATTACTGACATGTATACTGAGTCATTAGGGACTTTCCAATGGGTTTTGCCCAGTACATAAGGTCAATAGGGGTGAATCAACAGGAAAGTCCCATTGGAGCCAAGTACACTG AGTCAATAGGGACTTTCCATTGGGTTTTGCCCAGTACAAAAGGTCAATAGGGGGTGAGTCAATGGGTTTTTCCCATTATTGGCACATACATAAGGTCAATAGGGGTGACTAGTCAGTGGGCAGAGCGCACATCGCCCACAGTCCCCGAGAAGTTGGGG GGAGGGGTCGGCAATTGAACCGGTGCCTAGAGAAGGTGGCGCGGGGTAAACTGGGAAAGTGATGTCGTGTACTGGCTCCGCCTTTTTCCCGAGGGTGGGGGAGAACCGTATATAAGTGCAGTAGTTGCCGTGAACGTTCTTTTTCGCAACGGGTTTGC CGCCAGAACACAGCTGAAGCTTCGAGGGGCTCGCATCTCTCCTTCACGCGCCCGCCGCCCTACCTGAGGCCGCCATCCACGCCGGTTGAGTCGCGTTCTGCCGCCTCCCGCCTGTGGTGCCTCCTGAACTGCGTCCGCCGTCTAGGTAAGTTTAAAGCTCA GGTCGAGACCGGGCCTTTGTCCGGCGCTCCCTTGGAGCCTACCTAGACTCAGCCGGCTCTCCACGCTTTGCCTGACCCTGCTTGCTCAACTCTACGTCTTTGTTTCGTTTTCTGTTCTGCGCCGTTACAGATCCAAGCTGTGACCGGCGCCTACCTGAGA TCACCGGTCACCATGGCGCCCGTCGCCGTCTGGGCCGCGCTGGCCGTCGGACTGGAGCTCTGGGCTGCGGCGCACGCCTTGCCCGCCCAGGTGGCATTTACACCCTACGCCCCGGAGCCCGGGAGCACATGCCGGCTCAGAGAATACTATGACCAGACAG CTCAGATGTGCTGCAGCAAATGCTCGCCGGGCCAACATGCAAAAGTCTTCTGTACCAAGACCTCGGACACCGTGTGTGACTCCTGTGAGGACAGCACATACACCCAGCTCTGGAACTGGGTTCCCGAGTGCTTGAGCTGTGGCTCCCGCTGTAGCTCTG ACCAGGTGGAAACTCAAGCCTGCACTCGGGAACAGAACCGCATCTGCACCTGCAGGCCCGGCTGGTACTGCGCGCTGAGCAAGCAGGAGGGGTGCCGGCTGTGCGCGCCGCTGCGCAAGTGCCGCCCGGGCTTCGGCGTGGCCAGACCAGGAACTGAAA CATCAGACGTGGTGTGCAAGCCCTGTGCCCCGGGGACGTTCTCCAACACGACTTCATCCACGGATATTTGCAGGCCCCACCAGATCTGTAACGTGGTGGCCATCCCTGGGAATGCAAGCATGGATGCAGTCTGCACGTCCACGTCCCCCACCCGGAGTAT GGCCCCAGGGGCAGTACACTTACCCCAGCCAGTGTCCACACGATCCCAACACACGCAGCCAACTCCAGAACCCAGCACTGCTCCAAGCACCTCCTTCCTGCTCCCAATGGGCCCCAGCCCCCCAGCTGAAGGGAGCACTGGCGACGAGCCCAAATCTTGT GACAAAACTCACACATGCCCACCGTGCCCAGCACCTGAACTCCTGGGGGGACCGTCAGTCTTCCTCTTCCCCCCAAAACCCAAGGACACCCTCATGATCTCCCGGACCCCTGAGGTCACATGCGTGGTGGTGGACGTGAGCCACGAAGACCCTGAGGTCA AGTTCAACTGGTACGTGGACGGCGTGGAGGTGCATAATGCCAAGACAAAGCCGCGGGAGGAGCAGTACAACAGCACGTACCGTGTGGTCAGCGTCCTCACCGTCCTGCACCAGGACTGGCTGAATGGCAAGGAGTACAAGTGCAAGGTCTCCAACAAA GCCCTCCCAGCCCCCATCGAGAAAACCATCTCCAAAGCCAAAGGGCAGCCCCGAGAACCACAGGTGTACACCCTGCCCCCATCCCGGGAGGAGATGACCAAGAACCAGGTCAGCCTGACCTGCCTGGTCAAAGGCTTCTATCCCAGCGACATCGCCGTGG AGTGGGAGAGCAATGGGCAGCCGGAGAACAACTACAAGACCACGCCTCCCGTGCTGGACTCCGACGGCTCCTTCTTCCTCTACAGCAAGCTCACCGTGGACAAGAGCAGGTGGCAGCAGGGGAACGTCTTCTCATGCTCCGTGATGCATGAGGCTCTGC ACAACCACTACACACAGAAGAGCCTCTCCCTGTCTCCGGGTAAATGATAAGAATTCTGCAGATATCCAGCACAGTGGCGGCCGCTCGAGTCTAGCTGGCCAGACATGATAAGATACATTGATGAGTTTGGACAAACCACAACTAGAATGCAGTGAAAA AAATGCTTTATTTGTGAAATTTGTGATGCTATTGCTTTATTTGTAACCATTATAAGCTGCAATAAACAAGTTAACAACAACAATTGCATTCATTTTATGTTTCAGGTTCAGGGGGAGGTGTGGGAGGTTTTTTAAAGCAAGTAAAACCTCTACAAAT GTGGTATGG

Table S2: Nucleotide sequence of the glutamine synthetase (GS) expression cassette.

	5'-3': pSV40-GS(cds)-SV40pA
Glutamine synthetase (GS)	CTGTGGAATGTGTGTCAGTTAGGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCAGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTC AGCAACCATAGTCCCGCCCCTAACTCCGCCCATCCCGCCCCTAACTCCGCCCAGTTCCGCCCATTCTCCGCCCCATGGCTGACTAATTTTTTTTATTTATGCAGAGGCCGAGGCCGCCTCTGCCTCTGAGCTATTCCAGAAGTAGTGAGGAGGCTTTTTT GGAGGCCTAGGCTTTTGCAAAAAGCTCCCGGGAGCTTGTATATCCATTTTCGATCGATATGGCCACCTCAGCAAGTTCCCACTTGAACAAAAACATCAAGCAAATGTACTTGTGCCTGCCCCAGGGTGAGAAAGTCCAAGCCATGTATATCTGGGTTG ATGGTACTGGAGAAGGACTGCGCTGCAAAACCCGCACCCTGGACTGTGAGCCCAAGTGTGTAGAAGAGTTACCTGAGTGGAATTTTGATGGCTCTAGTACCTTTCAGTCTGAGGGCTCCAACAGTGACATGTATCTCAGCCCTGTTGCCATGTTTCGGG ACCCCTTCCGCAGAGATCCCAACAAGCTGGTGTTCTGTGAAGTTTTCAAGTACAACCGGAAGCCTGCAGAGACCAATTTAAGGCACTCGTGTAAACGGATAATGGACATGGTGAGCAACCAGCACCCCTGGTTTGGAATGGAACAGGAGTATACTCTG ATGGGAACAGATGGGCACCCTTTTGGTTGGCCTTCCAATGGCTTTCCTGGGCCCCAAGGTCCGTATTACTGTGGTGTGGGCGCAGACAAAGCCTATGGCAGGGATATCGTGGAGGCTCACTACCGCGCCTGCTTGTATGCTGGGGTCAAGATTACAGGA ACAAATGCTGAGGTCATGCCTGCCCAGTGGGAATTCCAAATAGGACCCTGTGAAGGAATCCGCATGGGAGATCATCTCTGGGTGGCCCGTTTCATCTTGCATCGAGTATGTGAAGACTTTGGGGTAATAGCAACCTTTGACCCCAAGCCCATTCCTGGG AACTGGAATGGTGCAGGCTGCCATACCAACTTTAGCACCAAGGCCATGCGGGAGGAGAATGGTCTGAAGCACATCGAGGAGGCCATCGAGAAACTAAGCAAGCGGCACCGGTACCACATTCGAGCCTACGATCCCAAGGGGGGCCTGGACAATGCCCG TCGTCTGACTGGGTTCCACGAAACGTCCAACATCAACGACTTTTCTGCTGGTGTCGCCAATCGCAGTGCCAGCATCCGCATTCCCCGGACTGTCGGCCAGGAGAAGAAAGGTTACTTTGAAGACCGCCGCCCCTCTGCCAATTGTGACCCCTTTGCAGTG ACAGAAGCCATCGTCCGCACATGCCTTCTCAATGAGACTGGCGACCAGCCCTTCCAATACAAAAACTAACCGCGGAACTTGTTTATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTCACAAATAAAGCATTTTTTTCACTGCA TTCTAGTTGTGGTTTGTCCAAACTCATCAATGTATCTTATCATGTCTG

Table S3: Plasmids used as PCR-templates for USER-cloning.

Plasmid name	Description of plasmid	Template or source
pcDNA3.1/Zeo(+)-GS	pcDNA™3.1/Zeo(+)-backbone with Amp resistance and expression cassette encoding glutamine synthetase (GS) (pSV40-GS–SV40pA)	Kindly provided by Gyun Min Lee (KAIST)
pDRIVE5-etanercept	pDRIVE5-backbone with an expression cassette encoding etanercept (mCMV-hEF-1a-HTLV-etanercept-SV40pA)	Kindly provided by Sara Petersen Bjørn (DTU Biosustain)

Due to the size of **Table S4,** it will be displayed at the end of this document (clone ranking and performance in static cultures are available online).

Figure S1: Schematic representation of the plasmid co-expressing GS and etanercept: PL_GS-etanercept

The cloning design of PL_GS-etanercept is described in the Material and Methods section.

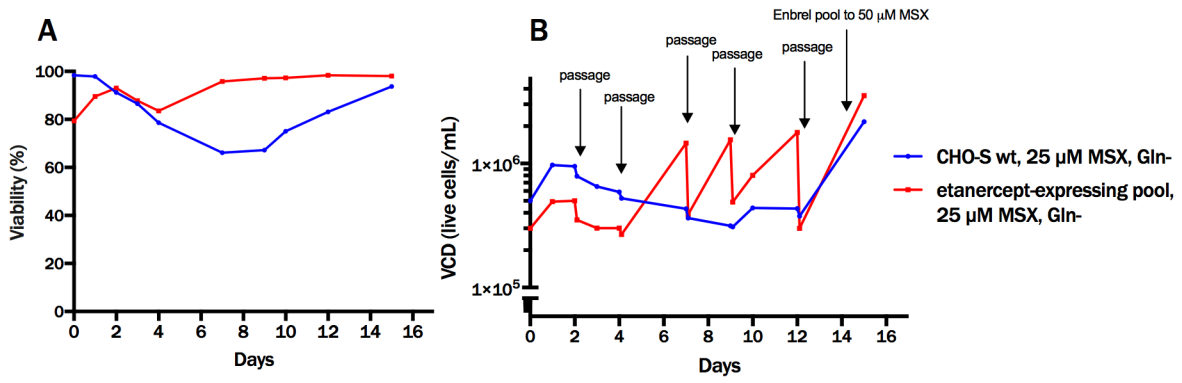


Figure S2. Establishing an etanercept-producing polyclonal pool using methionine sulfoximine (MSX) selection

Three days post-nucleofection, cells were subjected to 25 µM MSX selection for 14 days in order to establish a polyclonal pool of etanercept-producing CHO-S cells. (**A-B**) Viability and viable cell density of the etanercept (Enbrel) polyclonal pool and CHO-S wild-type (wt) cells during 14 days of selection by MSX and no glutamine (Gln-). Every 2-3 days, the selection media was replaced (*i.e.* passaging) until cells recovered.

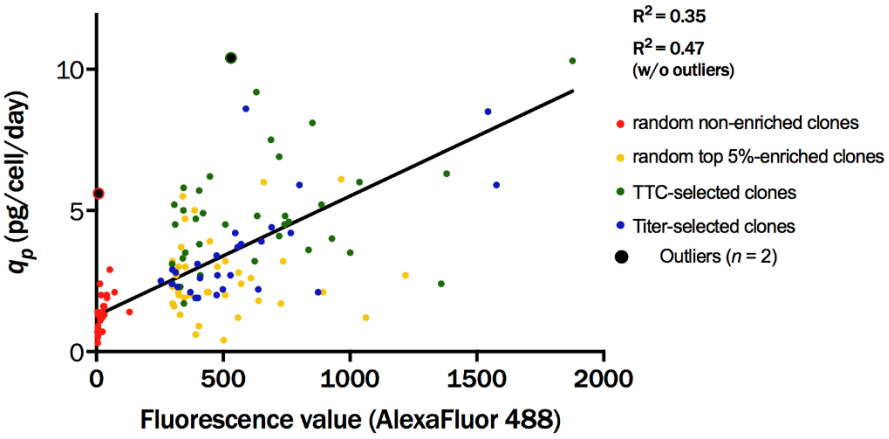


Figure S3. Correlation analysis of FACS-based staining intensity of single cells and q_p in 96-half-deepwell microplates

Correlation analysis of AlexaFluor 488 fluorescence intensity values and average q_p from day 0 to 2 for all the clones with seeding density > 1 x 10^5 cells/mL in 96-half-deepwell (HDW) microplates (n = 125). Nonlinear regression with a straight-line model and relative weighing was applied (R^2 without and without elimination of the outliers is shown).

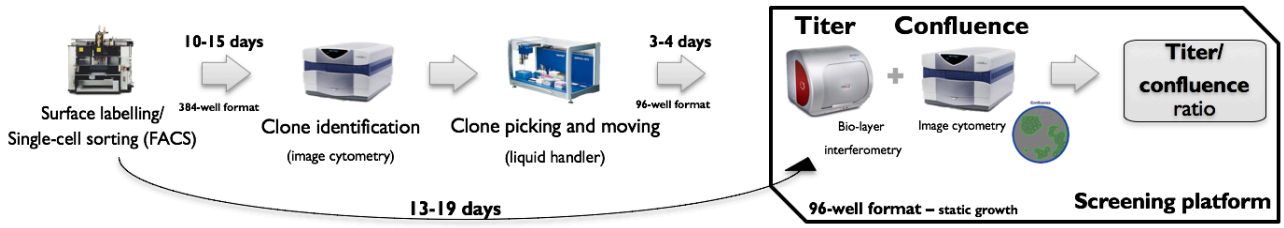


Figure S4. Platform for high-throughput screening of titer and confluence for static cultures of suspension-adapted cells

Following surface labelling and single-cell sorting to 384-well plates using FACS, clones are incubated statically for 10-15 days. Surviving single-cell colonies are identified using image cytometry and moved to 96-well flat-bottom plates using a liquid handler. After 3-4 days of static incubation, sub-confluent single-cell derived clones are screened for titer and confluence, using bio-layer interferometry and image cytometry, respectively. Clones are then ranked according to either *titer-to-confluence* (TTC) ratio or titer.

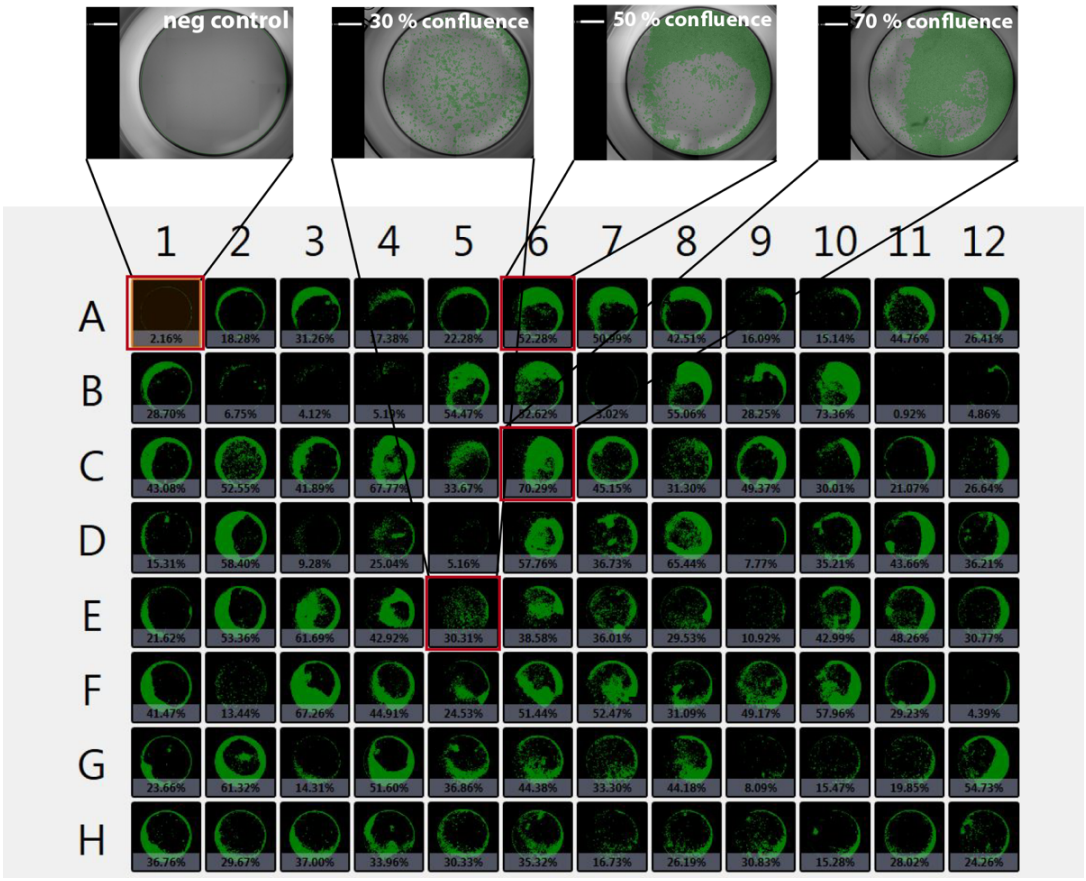


Figure S5. Confluence measurement of suspension-adapted clones grown statically in 96-well microplates.

Confluence measurements of sub-confluent clones grown statically in 96-well microplates were performed on the Celígo imaging cell cytometer using the label-free, bright field confluence application. Detected cells are shown in green color and the area of cells covering the well is defined as confluence. Highlighted wells show a range of confluence values: A1 – negative control (no cells seeded into the well), E5 – 30% confluence, A6 –

50% confluence and C6 – 70% confluence. The shown 96-well microplate is representative of all analysed plates and clones.

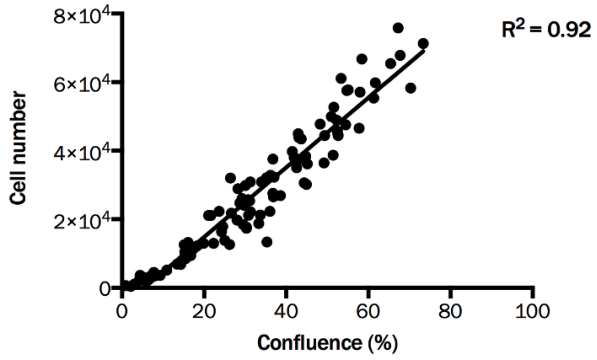


Figure S6. Correlation analysis between confluence values and cell counts of clones in 96-well microplates

Correlation analysis of confluence values and cell counts of clones from 96-well microplate in Figure S4 (n = 96) was performed using linear regression. Cell numbers were estimated by the Celígo imaging cell cytometer using the direct cell counting bright field application.

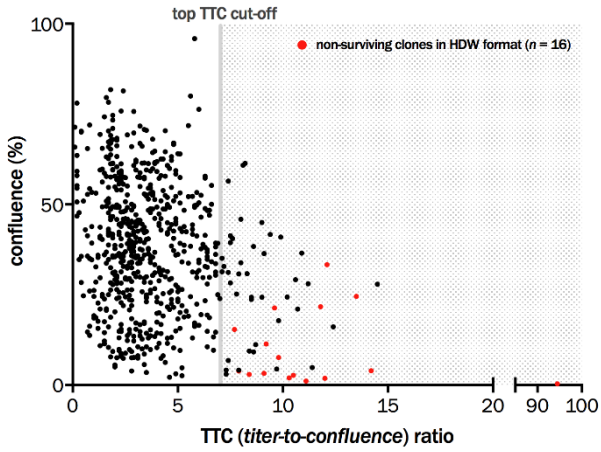


Figure S7. Confluence and TTC ratio of all analysed clones

Confluence values of clones grown statically and the corresponding *titer-to-confluence* (TTC) ratios are shown. Clones above the TTC cut-off (*i.e.* 7) were selected for further examination in 96-half-deepwell (96-HDW) microplates (n = 56). Clones, marked in red, did not survive the transfer into 96-HDW microplates (n = 16). Only clones with positive TTC values are shown (n = 672).

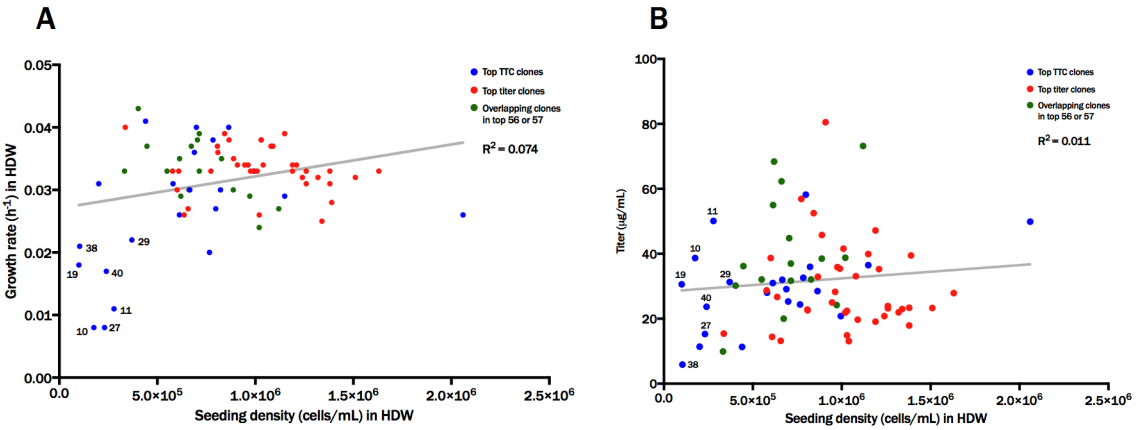


Figure S8. Correlation analysis between seeding density and specific growth rate or maximum titer of clones in suspension in 96-half-deepwell microplates

Clones were selected and evaluated as described in Figure 3. (**A**) Correlation analysis between initial seeding density and specific growth rate (µ) in the exponential phase for top TTC- and titer-selected clones (n = 76) in 96-half-deepwell (96-HDW) microplates, using linear regression. (**B**) Correlation analysis between initial seeding density and maximum titer for top TTC- and titer-selected clones in 96-HDW microplates using linear regression.

(1) IDs of selected clones indicate the TTC-based ranks and can be seen in Table S4.

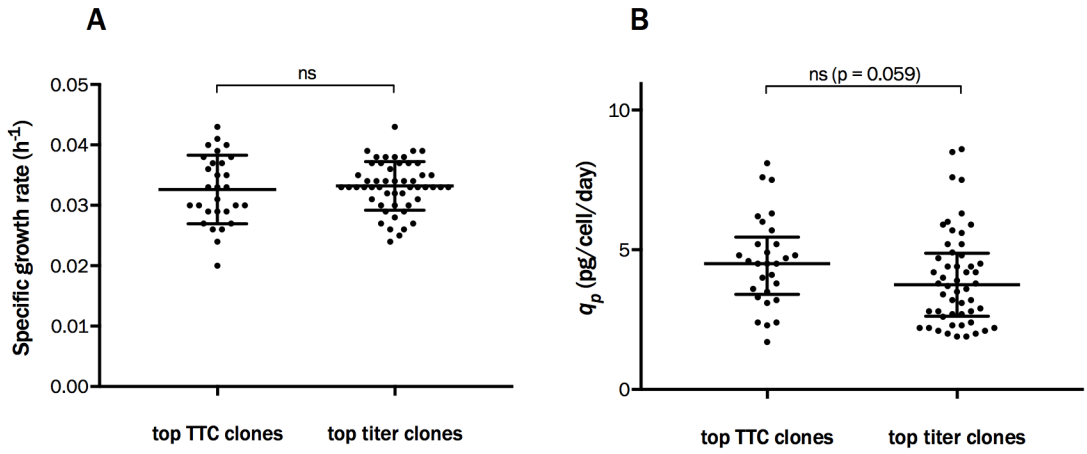


Figure S9. Performance of top TTC- and titer-selected clones with seeding density above 4 x 10⁵ cells/mL in suspension in 96-half-deepwell microplates

Top ranking *titer-to-confluence*- (TTC) and titer-selected clones based on static cultures were evaluated in suspension in 96-half-deepwell (HDW) microplates, as described in Fig. 3. Only clones with a seeding density above 4 x 10⁵ cells/mL are shown here. (**A**) Comparison of specific growth rate of 96-HDW batch cultures between top TTC clones (n = 29) and top titer clones (n = 52) is shown. Statistical significance was calculated using Welch's t test (bars represent mean ± SD). (**B**) Comparison of specific productivity (q_p) of 96-HDW batch cultures between top TTC clones and top titer clones. Statistical significance was calculated using Mann-Whitney test (bars represent median ± interquartile range (IQR)).

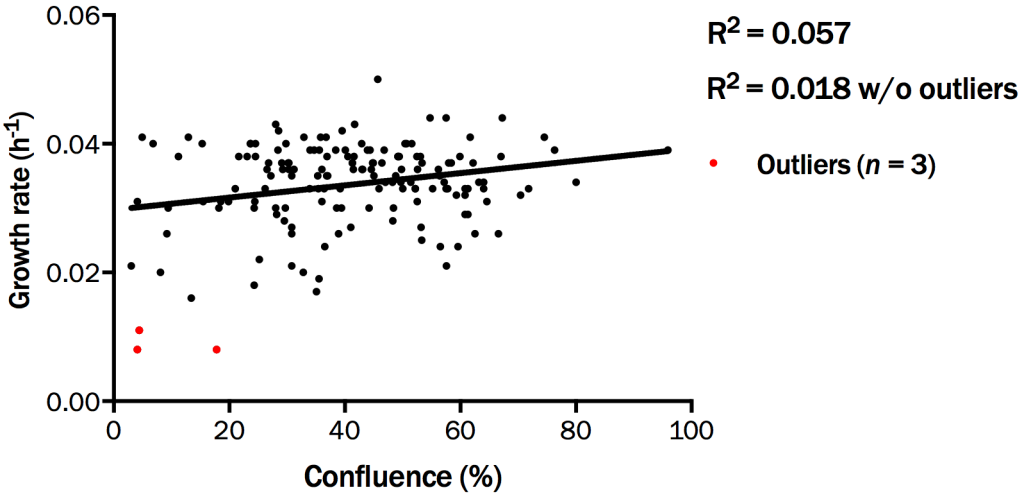


Figure S10. Correlation analysis of confluence and specific cell growth rate: Static versus suspension culture

Selected clones were transferred from static culture to suspension in 96-half-deepwell (HDW) microplates, where specific cell growth rate (μ) was assessed in two-day batch cultures. Correlation of confluence values under static incubation and μ in suspension is shown for all the clones with seeding density > 1 x 10^5 cells/mL in 96-HDW microplates (n = 158) using nonlinear regression with a straight line model. R^2 without and with relative weighing and elimination of outliers is reported.

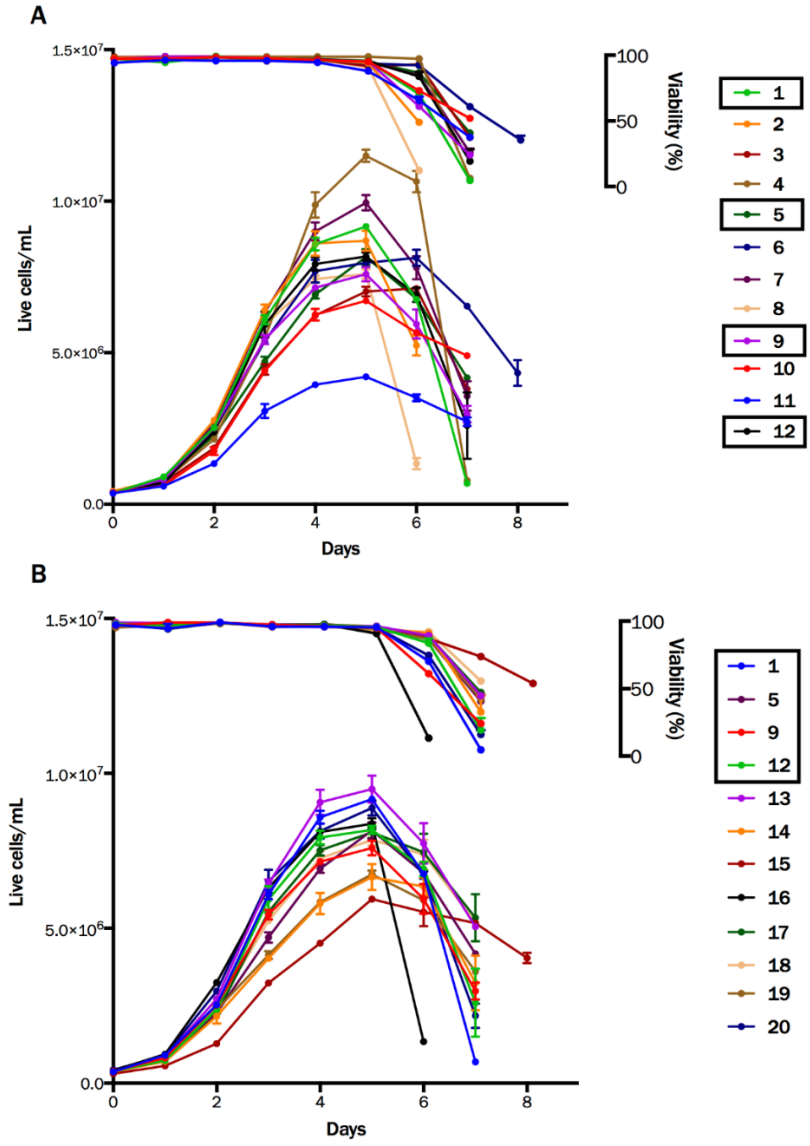


Figure S11. Viable cell density and viability of top 12 TTC- and titer-selected clones in shake flask batch cultures

Top 12 TTC- and titer-selected clones were evaluated in shake flask batch cultures as described in Figure 4. (**A**) Viability and viable cell density of top 12 *titer-to-confluence* (TTC) clones throughout the batch culture with corresponding clone IDs are shown. (**B**) Viability and viable cell density of top 12 titer clones throughout the batch culture with corresponding clone IDs. Encircled clones (clone 1, 5, 9 and 12) are overlapping clones present in both groups (n = 4).

(1) IDs ranging from 1-20 were assigned to clones in shake flask batch cultures and the corresponding titer- and TTC-based ranks can be seen in Table S4.

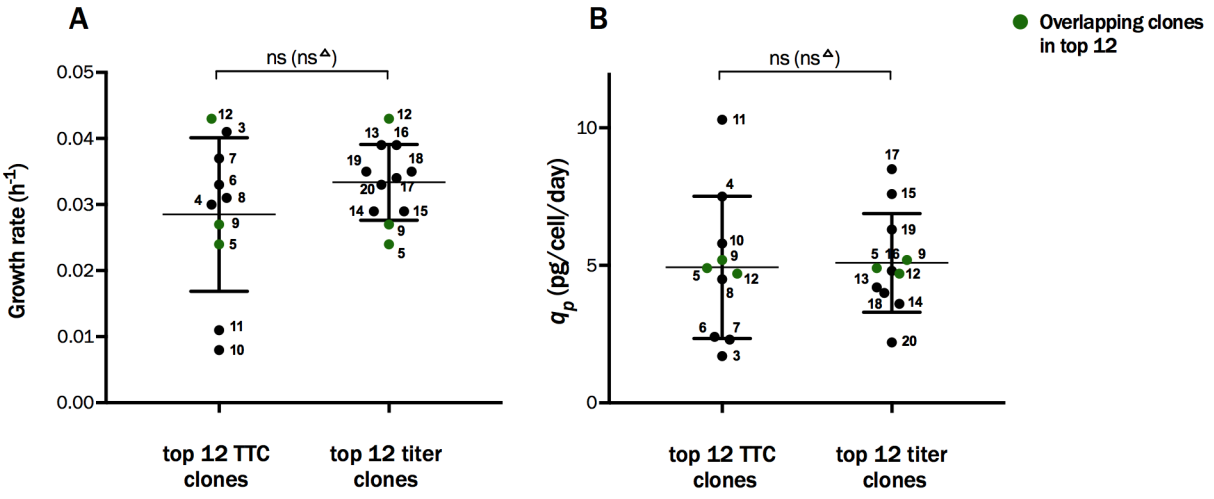


Figure S12. Performance of top 12 TTC- and titer-selected clones in suspension in 96-half-deepwell microplates

Top 12 ranking *titer-to-confluence*- (TTC) and titer-selected clones based on static cultures were evaluated in suspension in 96-half-deepwell (HDW) microplates, where specific cell growth rate (μ) and specific productivity (q_p) were assessed in two-day batch cultures. (**A**) Comparison of specific growth rate of 96-HDW batch cultures between top TTC clones (n = 12) and top titer clones (n = 12) with corresponding clone IDs are shown. Statistical significance was calculated using Welch's t test (bars represent mean ± SD). (**B**) Comparison of q_p of 96-HDW batch cultures between top TTC clones and top titer clones. Statistical significance was calculated using unpaired t test (bars represent mean ± SD).

(1) Only clones with a seeding density above 1 x 10⁵ cells/mL were analyzed. (2) IDs ranging from 1-20 were assigned to clones in shake flask batch cultures and the corresponding titer- and TTC-based ranks can be seen in Table S4.

Table S4: Clone ranking and performance in 96-half-deepwell microplates and shake flask batch cultures

Random non-enriched clones		Static 96-well format (3-day endpoint)			Suspension, 96-HDW format (Day0-2)				Suspension, 96-HDW format (Day 7)
Clone name	Fluorescence value intensity (AF488) (sorting)	Confluence (%)	Titer (µg/mL)	TTC value	Seeding density (Day0)	Titer (µg/mL)	µ (h^{-1})	q_p *(pcd)*	Max titer (µg/mL)
E1_2	53	35.5%	0.35	2.0	5.11E+05	4.1	0.019	2.9	25.2
E2_2	4	30.8%	0.00	0.0	6.97E+05	0.0	0.035	0	1.5
E3_2	8	59.6%	0.56	1.9	3.53E+05	6.8	0.024	5.6	27.8
E4_2	40	57.5%	0.97	3.4	5.86E+05	8.2	0.044	2	20.6
E5_2	5	56.5%	0.00	0.0	4.11E+05	0.0	0.024	0	0.0
E6_2	18	45.7%	0.42	1.8	2.82E+05	4.6	0.05	2	9.6
E7_2		1.0%	0.00	0.0					
E8_2	9	30.8%	0.00	0.0	2.15E+05	0.0	0.021	0	0.0
E9_2	8	38.9%	0.00	0.0	5.79E+05	1.1	0.026	0.5	3.2
E10_2	15	35.6%	0.42	2.4	8.77E+05	6.3	0.039	1.2	12.2
E11_2	12	35.4%	0.34	1.9	7.69E+05	2.4	0.033	0.7	8.8
E12_2	6	27.2%	0.00	0.0	2.33E+05	0.0	0.035	0	0.0
F1_2	24	44.6%	0.57	2.6	6.15E+05	4.2	0.036	1.4	11.0
F2_2	5	28.5%	0.16	1.1	7.86E+05	2.5	0.042	0.5	7.6
F3_2	5	28.4%	0.06	0.4	7.81E+05	0.0	0.039	0	0.0
F4_2	6	41.3%	0.15	0.7	9.67E+05	4.6	0.037	0.9	7.8
F5_2	15	43.9%	0.51	2.3	5.21E+05	3.1	0.039	1.1	6.7
F6_2	8	50.7%	0.05	0.2	6.92E+05	0.0	0.04	0	0.0
F7_2	6	35.8%	0.38	2.1	6.83E+05	5.4	0.041	1.3	8.8
F8_2	8	18.5%	0.00	0.0	7.22E+05	0.0	0.031	0	0.0
F9_2	24	34.7%	0.29	1.7	8.76E+05	3.8	0.039	0.7	7.0
F10_2	5	52.6%	0.69	2.6	1.18E+06	8.7	0.036	1.4	13.2
F11_2	131	29.8%	0.88	5.9	6.13E+05	5.0	0.04	1.4	14.4
F12_2	12	18.2%	0.00	0.0	4.68E+05	0.0	0.03	0	0.0
G1_2	4	56.2%	0.39	1.4	8.14E+05	3.2	0.036	0.7	9.4
G2_2	23	49.4%	0.47	1.9	9.25E+05	6.1	0.038	1.2	13.0
G3_2	42	41.6%	0.57	2.7	8.21E+05	8.5	0.038	1.9	27.1
G4_2	5	57.6%	0.00	0.0	3.71E+05	0.0	0.021	0	0.0
G5_2	8	40.1%	0.00	0.0	2.48E+05	0.0	0.039	0	0.0
G6_2	7	53.1%	0.28	1.1	1.20E+06	4.2	0.038	0.6	8.6
G7_2	72	23.1%	0.58	5.0	4.91E+05	5.4	0.038	2.1	10.8
G8_2	3	43.1%	0.00	0.0	5.55E+05	0.0	0.036	0	0.0
G9_2	5	46.8%	0.05	0.2	9.04E+05	0.0	0.039	0	0.0
G10_2		2.7%	0.00	0.0					
G11_2	5	36.9%	0.00	0.0	9.41E+05	0.0	0.038	0	1.5
G12_2	31	12.9%	0.29	4.5	7.26E+05	6.2	0.041	1.3	20.5
H1_2	27	39.5%	0.59	3.0	5.91E+05	5.9	0.042	1.6	12.4
H2_2	6	26.8%	0.00	0.0	6.71E+05	0.0	0.037	0	0.0
H3_2	7	26.5%	0.00	0.0	1.02E+06	0.0	0.036	0	1.6
H4_2		2.8%	0.00	0.0					
H5_2	31	32.9%	0.42	2.6	8.96E+05	9.3	0.041	1.6	19.6
H6_2	14	30.2%	0.32	2.1	4.93E+05	6.6	0.037	2.4	13.0
H7_2	13	74.5%	0.00	0.0	9.40E+05	0.0	0.041	0	1.5
H8_2	4	28.0%	0.00	0.0	7.73E+05	0.0	0.043	0	0.0
H9_2	7	36.0%	0.00	0.0	1.01E+06	0.0	0.036	0	0.0
H10_2	6	44.8%	0.51	2.3	1.21E+06	5.3	0.037	0.8	9.0
H11_2		4.7%	0.00	0.0					
H12_2	5	24.5%	0.13	1.0	1.03E+06	1.6	0.04	0.3	17.6

did not survive in HDW format

Table S4: Clone ranking and performance in 96-half-deepwell microplates and shake flask batch cultures - continuation

Random top 5%-enriched clones		Static 96-well format (3-day endpoint)			Suspension, 96-HDW format (Day0-2)				Suspension, 96-HDW format (Day 7)
Clone name	Fluorescence value intensity (AF488) (sorting)	Confluence (%)	Titer (µg/mL)	TTC value	Seeding density (Day0)	Titer (µg/mL)	µ (h^{-1})	q_p (pcd)	Max titer (µg/mL)
E1_6	346.0	21.62%	0.48	4.5	7.05E+05	7.7	0.038	1.9	28.7
E2_6	325.0	53.36%	1.01	3.8	1.04E+06	11.5	0.037	2	23.9
E3_6	560.0	61.69%	0.92	3.0	6.28E+05	10.6	0.041	2.8	19.1
E4_6	299.0	42.92%	0.90	4.2	8.33E+05	16.3	0.04	3.2	29.7
E5_6	368.0	30.31%	0.57	3.8	7.66E+05	7.6	0.036	2	12.9
E6_6	315.0	38.58%	0.44	2.3	5.58E+05	6.9	0.03	2.7	14.0
E7_6	301.0	36.01%	0.26	1.5	3.96E+05	3.7	0.031	2.3	10.1
E8_6	387.0	29.53%	0.40	2.7	6.49E+05	13.5	0.028	5	47.1
E9_6	323.00	10.92%	0.00	0.0	4.33E+04				
E10_6	609.0	42.99%	0.72	3.3	9.20E+05	12.1	0.036	2.6	18.8
E11_6	404.0	48.26%	0.59	2.5	8.27E+05	3.5	0.034	0.9	9.4
E12_6	852.0	30.77%	1.21	7.9	7.98E+05	8.1	0.027	8.1	58.2
F1_6	323.0	41.47%	0.54	2.6	8.27E+05	8.7	0.036	2.1	16.7
F2_6	341.0	13.44%	0.26	3.9	6.00E+05	9.0	0.016	5.5	19.9
F3_6	639.0	67.26%	0.59	1.7	5.83E+05	6.6	0.044	1.8	13.8
F4_6	570.0	44.91%	0.55	2.4	9.15E+05	11.6	0.037	2.4	19.6
F5_6	502.0	24.53%	0.17	1.4	1.05E+06	2.4	0.038	0.4	6.5
F6_6	324.0	51.44%	0.93	3.6	1.08E+06	16.2	0.034	3	27.8
F7_6	736.0	52.47%	1.14	4.4	9.18E+05	16.4	0.038	3.2	25.6
F8_6	414.0	31.09%	0.91	5.9	8.69E+04				
F9_6	334.0	49.17%	1.14	4.6	7.33E+05	14.4	0.038	3.7	28.9
F10_6	477.0	57.96%	1.23	4.2	8.52E+05	13.6	0.037	3	23.9
F11_6	894.0	29.23%	0.67	4.6	9.92E+05	10.9	0.036	2.1	27.4
F12_6	1877.0	4.39%	0.21	9.7	2.79E+05	10.3	0.011	10.3	50.1
G1_6	301.0	23.66%	0.32	2.7	6.05E+05	5.8	0.04	1.7	13.7
G2_6	659.0	61.32%	1.20	3.9	5.05E+05	14.4	0.033	6	25.3
G3_6	365.0	14.31%	0.19	2.7	6.16E+04				
G4_6	922.00	51.60%	0.00	0.0	7.94E+05	0.0	0.04	0	0.0
G5_6	351.0	36.86%	0.52	2.8	6.02E+05	8.3	0.035	3	19.3
G6_6	508.0	44.38%	1.03	4.7	9.79E+05	18.0	0.039	3.2	31.1
G7_6		33.30%	2.01	12.1					
G8_6	1218.0	44.18%	0.90	4.1	7.07E+05	7.8	0.03	2.7	14.9
G9_6	349.0	8.09%	0.17	4.3	1.44E+05	2.3	0.02	4.7	8.1
G10_6	447.0	15.47%	0.23	3.0	4.00E+05	6.6	0.031	3.9	21.3
G11_6	727.0	19.85%	0.40	4.0	6.51E+05	4.5	0.031	1.7	13.9
G12_6	508.0	54.73%	1.30	4.7	6.61E+05	9.2	0.044	2	25.3
H1_6	442.0	36.76%	1.18	6.4	7.90E+05	10.1	0.041	2.1	34.9
H2_6	436.0	29.67%	0.38	2.5	5.62E+05	4.7	0.03	2.1	14.3
H3_6	965.0	37.00%	0.54	2.9	5.74E+05	16.0	0.035	6.1	57.5
H4_6	329.0	33.96%	0.38	2.3	8.14E+05	5.7	0.039	1.3	11.7
H5_6	306.0	30.33%	0.58	3.8	7.85E+05	6.4	0.037	1.6	12.9
H6_6	392.0	35.32%	0.21	1.2	1.55E+06	5.0	0.035	0.6	7.6
H7_6	348.0	16.73%	0.19	2.3	8.09E+04				
H8_6	1063.0	26.19%	0.13	1.0	4.43E+05	2.2	0.033	1.2	6.0
H9_6	634.0	30.83%	1.29	8.3	2.06E+06	33.1	0.026	4.8	49.9
H10_6	558.0	15.28%	0.28	3.7	7.96E+05	5.8	0.04	1.2	11.9
H11_6	688.0	28.02%	1.57	11.2	6.62E+05	17.8	0.03	7.5	62.3
H12_6	631.0	24.26%	1.03	8.5	1.00E+05	2.4	0.018	9.2	30.6

overlaps with top TTC clone
n = 6

did not survive in HDW format

Table S4: Clone ranking and performance in 96-half-deepwell microplates and shake flask batch cultures – continuation

Top TTC clones			Static 96-well format (3-day endpoint)			Suspension, 96-HDW format (Day0-2)				Suspension, 96-HDW format (Day 7)		Suspension, shake flask (Day0-2)				Suspension, shake flask (Day 7)	
TTC Ranking	Clone name	Fluorescence value in	Confluence (%)	Titer (µg/mL)	TTC value	Seeding density (Day0)	Titer (µg/mL)	µ (h^{-1})	q_p *(pcd)*	Max titer (µg/mL)	Clone ID in batch	µ (h^{-1}) rep1	µ (h^{-1}) rep2	q_p (pcd) rep1	q_p (pcd) rep2	Max titer (µg/mL) rep1	Max titer (µg/mL) rep2
	A10_2		0.3%	0.16	94.5												
1	A7_5	361	27.9%	2.03	14.5	5.39E+04					1	0.039	0.041	2.4	2.6	35.1	36.0
	A11_2		4.0%	0.28	14.2												
	E12_9		24.5%	1.65	13.5												
2	A9_6	372	16.1%	1.00	12.4	2.73E+04					2	0.038	0.039	6.7	6.7	73.8	68.1
	G7_6		33.3%	2.01	12.1												
	B4_2		1.9%	0.11	12.0												
	D11_9		21.7%	1.28	11.8												
3	B12_6	345	4.9%	0.28	11.4	4.40E+05	3.9	0.041	1.7	11.3	3	0.031	0.032	1.2	1.2	12.7	12.3
4	H11_6	688	28.0%	1.57	11.2	6.62E+05	17.8	0.03	7.5	62.3	4	0.035	0.035	1.1	0.9	22.6	24.2
	B4_4		1.1%	0.06	11.1												
5	A5_9	420	36.5%	1.99	10.9	1.02E+06	16.4	0.024	4.9	38.8	5	0.035	0.036	4.8	4.9	41.1	37.8
6	H11_7	1359	21.0%	1.13	10.7	9.95E+05	10.5	0.033	2.4	20.8	6	0.039	0.034	1.1	1.1	15.1	15.7
7	H12_8	330	29.1%	1.55	10.6	6.74E+05	7.4	0.037	2.3	20	7	0.038	0.041	2.6	2.5	29.3	26.5
	A6_5		2.7%	0.14	10.5												
	D3_9		2.0%	0.10	10.3												
8	H11_9	742	24.4%	1.25	10.2	5.80E+05	9.8	0.031	4.5	28	8	0.034	0.035	2.3	2.3	29.8	30.3
9	A3_5	307	41.0%	2.03	9.9	1.12E+06	17.4	0.027	5.2	73.2	9	0.037	0.041	4.2	4.2	36.3	39.0
	A9_9		7.6%	0.37	9.8												
10	G5_5	344	17.8%	0.87	9.8	1.76E+05	2.2	0.008	5.8	38.7	10	0.034	0.03	3.5	3.7	29.2	29.5
11	F12_6	1877	4.4%	0.21	9.7	2.79E+05	7.3	0.011	10.3	50.1	11	0.026	0.028	9.2	8.3	78.9	77.1
	B5_4		21.3%	1.02	9.6												
12	H2_5	392	41.7%	1.96	9.4	4.03E+05	10.3	0.043	4.7	30.2	12	0.039	0.038	3.3	3.4	48.9	38.7
	H9_4		11.4%	0.53	9.2												
	B8_2		3.3%	0.15	9.1												
13	B8_5	530	36.4%	1.65	9.1	3.33E+05	15.1	0.033	10.4	9.9							
14	H7_7	340	24.3%	1.09	9.0	8.22E+05	10.3	0.03	3.3	36							
15	B11_4		45.0%	2.02	9.0	6.14E+05	15.3	0.035	6.3	55							
16	D10_8	758	11.2%	0.48	8.7	7.84E+05	16.8	0.038	4.6	32.6							
17	H1_5	406	38.4%	1.65	8.6	7.14E+05	13.2	0.039	3.8	31.7							
18	C10_10	744	9.2%	0.39	8.6	6.13E+05	9.8	0.026	4.8	31							
19	H12_6	631	24.3%	1.03	8.5	1.00E+05	2.4	0.018	9.2	30.6							
20	F2_7	298	23.7%	1.00	8.5	8.64E+05	13.6	0.04	3.1	28.5							
21	H7_5	508	9.4%	0.40	8.4	6.66E+05	10.9	0.03	4.5	32							
	A12_2		2.9%	0.12	8.4												
22	H9_6	634	30.8%	1.29	8.3	2.06E+06	33.1	0.026	4.8	49.9							
23	A9_10	836	61.3%	2.53	8.2	9.72E+05	14.4	0.029	3.6	24.2							
24	B10_4		60.8%	2.47	8.1	6.20E+05	15.5	0.029	7.6	68.4							
25	F2_8	405	33.9%	1.36	8.0	5.50E+05	12.6	0.033	5.7	32.1							
26	E10_8	1037	45.9%	1.83	8.0	7.14E+05	18.5	0.033	6	37							
27	E6_5	1381	4.1%	0.16	7.9	2.31E+05	4.2	0.008	6.3	15.3							
	C7_4		3.9%	0.15	7.9												
28	E12_6	852	30.8%	1.21	7.9	7.98E+05	22.0	0.027	8.1	58.2							
29	E12_7	721	25.2%	0.98	7.8	3.70E+05	7.1	0.022	6.9	31.3							
	G1_7		15.4%	0.60	7.7												
30	H1_9	887	40.5%	1.55	7.6	7.05E+05	17.4	0.038	5.2	44.8							
31	H1_4	350	41.3%	1.55	7.5	4.47E+05	7.0	0.037	3.5	36.2							
32	C7_8	310	39.4%	1.47	7.5	8.88E+05	15.7	0.03	4.5	38.5							
33	B9_6	720	28.2%	1.06	7.5	1.15E+06	17.2	0.029	4.1	36.5							
34	B5_5	928	56.4%	2.09	7.4	8.28E+05	13.6	0.035	4	32.1							
35	C10_7	299	31.2%	1.15	7.4	6.89E+05	7.7	0.036	2.4	29.1							
36	E1_9	624	6.8%	0.25	7.4	6.99E+05	11.3	0.04	3.2	25.3							
37	B3_6	1000	4.1%	0.15	7.3	2.01E+05	2.8	0.031	3.5	11.4							
38	B7_6	409	3.0%	0.11	7.3	1.04E+05	0.9	0.021	2.7	5.9							
39	G1_9	447	32.8%	1.18	7.2	7.66E+05	13.8	0.02	6.2	24.4							
40	A12_5	343	35.1%	1.25	7.1	2.40E+05	3.1	0.017	5	23.7							
	Average		22.9%	1.03	10.8												

overlaps with titer clone

did not survive in HDW format/no fluorescence data

Top 12 selected clones for shake flask batch

n = 19

Table S4: Clone ranking and performance in 96-half-deepwell microplates and shake flask batch cultures – continuation

Top titer clones			Static 96-well format (3-day endpoint)			Suspension, 96-HDW format (Day0-2)				Suspension, 96-HDW format (Day 7)		Suspension, shake flask (Day0-2)				Suspension, shake flask (Day 7)	
Titer ranking	Clone name	Fluorescence value into	Confluence (%)	Titer (µg/mL)	TTC value	Seeding density (Day0)	Titer (µg/mL)	µ (h^{-1})	q_p (pcd)	Max titer (µg/mL)	Clone ID in batch	µ (h^{-1}) rep1	µ (h^{-1}) rep2	q_p (pcd) rep1	q_p (pcd) rep2	Max titer (µg/mL) rep1	Max titer (µg/mL) rep2
1	D6_4		95.9%	2.79	5.8	1.15E+06	24.2	0.039	4.2	39.9	13	0.043	0.04	5.3	5.4	60.6	61.2
2	A9_10	836	61.3%	2.53	8.2	9.72E+05	14.4	0.029	3.6	24.2	14	0.035	0.038	4.2	3.7	35.7	31.5
3	B10_4		60.8%	2.47	8.1	6.20E+05	15.5	0.029	7.6	68.4	15	0.029	0.031	5.6	5.7	43.5	45.9
4	B7_4		76.3%	2.30	6.0	8.43E+05	21.6	0.039	4.8	52.5	16	0.044	0.043	4.2	4.2	42.6	41.4
5	G7_8	1544	80.0%	2.23	5.6	9.09E+05	32.7	0.034	8.5	80.5	17	0.037	0.034	5	5.5	60.3	59.1
6	B5_5	928	56.4%	2.09	7.4	8.28E+05	13.6	0.035	4	32.1	18	0.036	0.037	3.4	3.2	36.6	37.2
7	A7_5	361	27.9%	2.03	14.5	5.39E+04					1	0.039	0.041	2.4	2.6	35.1	36.0
8	A3_5	307	41.0%	2.03	9.9	1.12E+06	17.4	0.027	5.2	73.2	9	0.037	0.041	4.2	4.2	36.3	39.0
9	B11_4		45.0%	2.02	9.0	6.14E+05	15.3	0.035	6.3	55	19	0.038	0.036	4.7	4.7	35.4	32.4
	G7_6		33.3%	2.01	12.1												
10	A5_9	420	36.5%	1.99	10.9	1.02E+06	16.4	0.024	4.9	38.8	5	0.035	0.036	4.8	4.9	41.1	37.8
11	C5_4		71.8%	1.96	5.5	6.09E+05	5.6	0.033	2.2	14.4	20	0.04	0.04	1.6	1.7	16.0	20.9
12	H2_5	392	41.7%	1.96	9.4	4.03E+05	10.3	0.043	4.7	30.2	12	0.038	0.039	3.3	3.4	48.9	38.7
13	E10_8	1037	45.9%	1.83	8.0	7.14E+05	18.5	0.033	6	37							
14	H5_9	1577	55.2%	1.83	6.6	9.76E+05	24.5	0.033	5.9	35.9							
15	D2_5	313	57.8%	1.82	2.0	1.01E+06	12.0	0.033	2.8	41.6							
16	E3_8	766	57.2%	1.81	6.3	1.19E+06	12.0	0.034	4.2	47.2							
17	H1_5	406	38.4%	1.65	8.6	7.14E+05	13.2	0.039	3.8	31.7							
18	B8_5	530	36.4%	1.65	9.1	3.33E+05	15.1	0.033	10.4	9.9							
	E12_9		24.5%	1.65	13.5												
19	F11_7	408	50.0%	1.65	6.6	1.63E+06	12.4	0.033	2.6	27.9							
22	C6_4		64.0%	1.63	5.1	7.73E+05	17.8	0.033	5.6	56.9							
23	C4_7		67.0%	1.61	4.8	8.66E+05	17.8	0.038	4.4	32.9							
24	H2_7		52.5%	1.61	6.1	1.26E+06	13.5	0.031	2.8	23.2							
25	E7_4		64.0%	1.60	5.0	9.46E+05	13.5	0.034	3.2	25							
26	D11_7	570	66.6%	1.60	4.8	6.37E+05	24.2	0.026	3.8	26.7							
27	E7_7	589	48.4%	1.58	6.5	6.01E+05	10.2	0.03	8.6	38.7							
28	E6_8	528	53.3%	1.58	5.9	1.34E+06	10.2	0.025	2.7	23							
29	H11_6	688	28.0%	1.57	11.2	6.62E+05	17.8	0.03	7.5	62.3							
30	H1_4	350	41.3%	1.55	7.5	4.47E+05	3.5	0.037	3.5	36.2							
31	G2_8	557	48.3%	1.55	6.4	1.39E+06	12.3	0.028	3.7	39.5							
32	H12_8	330	29.1%	1.55	10.6	6.74E+05	7.4	0.037	2.3	20							
33	H1_9	887	40.5%	1.55	7.6	7.05E+05	17.4	0.038	5.2	44.8							
34	H4_4	474	70.4%	1.55	4.4	1.24E+06	10.8	0.032	2	20.8							
35	D4_8	801	48.8%	1.53	6.3	8.89E+05	13.3	0.035	5.9	45.8							
36	A8_10	391	47.0%	1.52	6.5	1.04E+06	7.8	0.034	1.9	13.1							
37	B9_2	254	50.4%	1.50	6.0	3.37E+05	4.0	0.04	2.5	15.4							
38	A10_10	874	52.2%	1.50	5.7	1.38E+06	12.9	0.033	2.1	23.4							
39	F4_8	547	63.2%	1.48	4.7	9.63E+05	11.6	0.034	4.2	28.3							
40	G2_9	651	62.2%	1.48	4.7	1.08E+06	21.0	0.037	3.9	33.1							
41	C7_8	310	39.4%	1.47	7.5	8.88E+05	15.7	0.03	4.5	38.5							
42	E11_4		49.8%	1.47	5.9	8.09E+05	12.4	0.036	3.2	22.6							
43	G4_4		57.5%	1.45	5.1	1.19E+06	10.2	0.033	2	19.1							
44	H6_8	398	49.7%	1.45	5.8	1.21E+06	16.1	0.034	3.1	35.3							
45	D9_4		52.2%	1.45	5.5	5.78E+05	10.2	0.033	4.4	28.7							
46	A3_7	638	59.3%	1.44	4.9	1.32E+06	12.3	0.032	2.2	22							
47	A10_4		46.4%	1.42	6.1	8.07E+05	10.8	0.037	2.8	22.8							
48	H8_4	477	49.2%	1.42	5.8	1.03E+06	13.3	0.038	2.7	22.4							
49	D2_6	321	58.4%	1.42	4.9	1.09E+06	11.6	0.037	2.3	19.7							
50	A4_7	297	64.6%	1.41	4.4	1.38E+06	10.8	0.031	2.4	17.9							
51	B3_5	371	59.9%	1.38	4.6	1.03E+06	10.2	0.038	2.1	14.9							
52	C12_9	690	39.2%	1.36	6.9	9.90E+05	17.8	0.033	4.4	35.4							
53	E2_8	401	60.8%	1.36	4.5	1.26E+06	11.3	0.033	1.9	23.9							
54	F2_8	405	33.9%	1.36	8.0	5.50E+05	12.6	0.033	5.7	32.1							
55	A4_10	474	53.2%	1.36	5.1	6.57E+05	7.7	0.027	3.4	13.2							
56	B10_9	498	60.8%	1.35	4.5	1.51E+06	14.1	0.032	2.2	23.3							
57	C8_7	299	62.5%	1.35	4.3	1.02E+06	10.2	0.026	2.9	21.9							
		Average	52.4%	1.68	6.8												

overlaps with the TTC clone
n = 19

did not survive in HDW format/no fluorescence data

Top 12 selected clones for shake flask batch

Systematic Evaluation of Site-Specific Recombinant Gene Expression for Programmable Mammalian Cell Engineering

Supplementary data

Nuša Pristovšek[1], Saranya Nallapareddy[1], Lise Marie Grav[1], Hooman Hefzi[2,3], Nathan E. Lewis[2,3], Peter Rugbjerg[1], Henning Gram Hansen[1], Gyun Min Lee[1,4], Mikael Rørdam Andersen[5], Helene Faustrup Kildegaard[1]

[1]The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Kemitorvet, Building 220, 2800 Kgs. Lyngby, Denmark

[2]Department of Pediatrics, University of California, San Diego, La Jolla, CA 92093, USA

[3]The Novo Nordisk Foundation Center for Biosustainability at the University of California, San Diego School of Medicine, CA 92093, USA

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

[5]Department of Biotechnology and Biomedicine, Technical University of Denmark, Søltofts Plads, Building 223, 2800 Kgs. Lyngby, Denmark

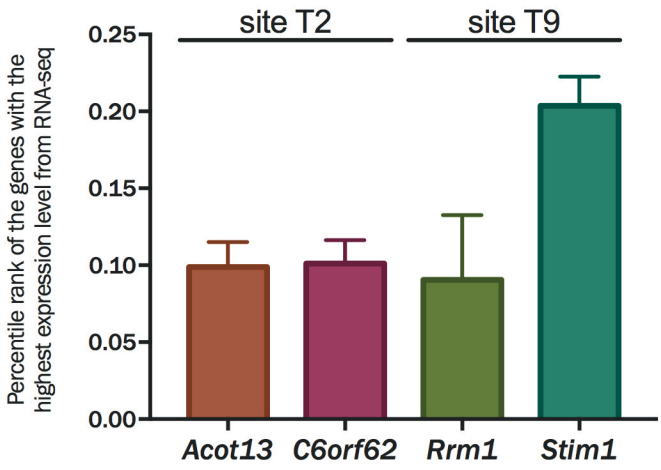
Material and methods

Generation of C6_2 cell line

Plasmid used for generation of C6_2 cell line consisted of the Rituximab expression cassette (CMV-HC-BGHpA-EF1α-LC-BGHpA) and the vector backbone with the selection marker (SV40-NeoR-SV40pA) and is depicted in Figure S3. CHO-S cells (Thermo Fisher Scientific) were transfected with Rituximab plasmid using FuGENE® HD transfection reagent (Promega), followed by G418 selection (500 µg/mL; Sigma Aldrich). After stable polyclonal pools were established, single-cell cloning, clone screening for high-producers, titer and specific productivity measurement were performed as described in our previous study (1). Copy number of integrated Rituximab were measured as previously described (2) by using EF1α promoter and BGHpA terminator as target genes and Vinculin as a control gene for normalization. Next, a 2-month long-term cultivation in 125 mL Erlenmeyer flasks (Corning Inc.) was performed. The C6_2 cell line was passaged every 2-3 days and the supernatants 0, 4 and 8 weeks after the start point of cultivation were stored at -80°C. Rituximab titer was subsequently determined in the supernatant samples as previously described (3). Specific Rituximab productivity was calculated based on Rituximab titer and integral viable cell density measurements as previously described (4) with the objective of assessing stability of production.

Supplementary Figure S1: Percentile ranks from expression levels of neighboring genes of T2 and T9 site

Published RNA-seq datasets from CHO-S cell line (5) were analyzed and regions identified with consistently high expression profiles across 40 samples, thereby selecting new candidate integration sites (site T2 and T9). The percentile ranks from expression levels of T2 and T9 neighboring genes across all samples of CHO-S cell line are provided. The error bars represent the standard deviation across all the samples (n = 40).



Supplementary Table S1: Integration site coordinates according to old (CHO-K1) and new PICR assembly, their chromosomal locations and chromatin states

Old assembly (orientation of integration)	Internal site name	New PICR assembly (6)	Chromosome location in Chinese hamster genome (6)	Genomic location	Expression level of corresponding gene in CHO-S cell line (7)	Upstream gene	Downstream gene	Chromatin state within integration site according to CHO-Epigenome database (8)
NW_003614466.1: 715154 (+)	**Site A**	Picr_5:33588893	2	Intron 2 of Polk	Intermediate expression	Col4A3bp	Pdhb	Weak enhancer (state number 6) - enriched mark: H3K4me1
NW_003613881.1: 1456617 (+)	**Site T2**	Picr_36:8978007	3	Intergenic	/	Acot13	Uncharacterized protein C6orf62 homolog	Strong transcription (state number 4) – enriched mark: H3K36me3
NW_003614758.1: 453083 (+)	**Site T9**	Picr_29:19091106	3	Intergenic	/	Stim1	Rrm1	Weak enhancer (state number 6), only at one time point (lag phase of the culture) – enriched mark: H3K4me1

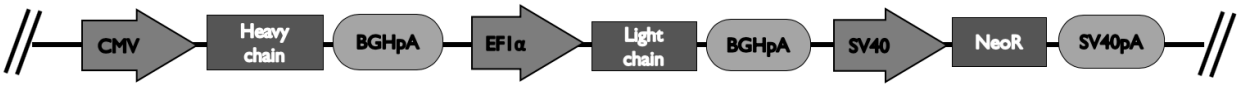
Supplementary Table S2: sgRNA target sequences and 750 bp long 5' and 3' homology arm sequences for all three sites

Site	sgRNA target sequence	5' homology arm	3' homology arm
A	AACATAGTAACCCCAAAGAGAGG	GAAGCTTACCCCCAGGATTTCCTGATACTACTTTAGAAAAAAACTAGCTTTAATATATTGGCATCAATATATTATTTTTCCTCTCTGCCTCCTTTTCTTTCTTTTTTTAATTTTTTATTAATTTATTTTTTACAGTCCAATCCCAGTTCCCCCTCCCTCCTGTCCTCCTGTTCCTCCTACTCCCCCTGCATCTGCTCCTTGGAGAGGGTAAGGCCTCCCCTTTTCCTTTCTTGAGGGGAAAATTTTTAAGGCAGCTAAGATCCTTCTCCTGAAGGCACTCACTGTTAGTGATCTAATACAATAAGGGCTGTGGTGAAAACAAGTACAAGAAGATTAATTCTCTGGCACAGAAGAGAAAGAGCCAGATGACTTTCAAGGGATGTTTGAGTTGACTCTTGAAAGCTTGTGACAGAGCTCAGAGAGGTGTTCAAAACAGAGCCAGCCAGGGTCAGAGTCCCTCAGGTCATTCAGAGTTGGAGACTCGGCAAGAGTAGGCCTGCAGACATTCAGCCTCTGCTACAGACAGATGTGACTTAGGGCCTTTTCTAAGAGAGGAATTGCAGTGTTATTACCCAGCTAGTGGACAGCTAATGTGTTCCTCCTGAAGAAGACTATGAAACAAGTATTCTGTAGAGTGAGGGCAAGATGGAAGACCAGCATGGGGGTGGGAGGGGGTACATTATACCTTTGTGGTTCTCCACATAAGGTAGGTTGCTTTCATTTAAGAGATCACTTGCTTTGGGCTGTTTT	ATGCCTCCATTGATTCTTTCAGTGGCACCAGACATTGTTCCATGTTCCGTGTGTTGTGCTTGCTTTATTGCTGTTTTCTCTTATGATTTACCTTAATGGTATAAGCAACTTCCTTACAGGTTGCTACAATTAAAACAATCACATAGCTACCCTTTGCATGGTAAGCTGGGAAAGCACACATACAACATTAGTAGAGGCAGTCCATAAAAAGCAAGCTAAGGATACTTTTGCTATATAATCTCCCTTTTTTTGTTTGTTTGTTTGTTTTTCAAGTTAGGGTTTCTCTGTGGCTTTGGAGGCTGTCTTGGAACTAGCTCTTGTAGACCAGGCTGGTCTTGAACTCACAGAGATCTCCCTGCCTGTGCCTCCCGAGTGCTGGGACTAAAGGCATGCGCCACCACTGCCCAGTTTATATAATCCCTTTTTTTTTTTTCACTCATCTAGTTTATTTCTATGTTTATTTCTCTTTTATCTTTTAATCCAATAATTTGTCTTCAAAGAGGCCAATATTTTTAAAGATTATAATTCTGCATTGTTTATATTTGACATATATATTAAAATTTTATGAAAAATAAGCTTTCATCTTCAGTTTTCAATTGGTGTTTGTATAGAATAACCATCTCAGACGTTAGCTGTGGTCCTGGCTAGACTTATGTCAACCTGCCAGAAGCCAGAGTCAGGGAAGAGGAACTCCCAATTGAGAAAATGACTCAATAAGATCAAGCTGTAGTCAAGTCTATAAGGTATTTTCT
T2	GAGATCTGTGGGCCGAGTAA	GTTTCTACACCATCCTAAAGAAGTTGAGTGTTTGGAATAGAAAAAAATATTAAAAGTGGGGATCTTACATTGTAAAATTAACCAGAGTGTCAAGAGTTCTAAAAGCAGAACTCAATTGTGCAATGAACATAAGGAAAGACTACTGTATAGCTTTTTTATTTTTACTTTTTTGTTTTTTGTTTTTTGTTTTTTTTTTTTTTTCTTTTAAATGAAGAAAGGCTTTGCTTAAGGGTTGCATACTTTTATTGGAGTAAATCGGAATGATCCTGCTCCATTGGAGTAAAACTAGTGCTTACTGGTTTCCATTTGTATTTAGCTTCTGGTTGGAATTTGGAAAAAAAAAATGAAACCTAAATAAAATAGGTGAAAGCTTCCTGGCTATTCAGATGAATATATAAAATCTGCA	AGTCCACAGATTCTGAATGAAAGAAAAATATTTATTTGGCAGAACATCCACAATTTAAAGATATGAGGAACAGGATACTGCTATCAGATCACTACCACCCAAGAGAGCTAGAGGTTGCTCTCAGTCCACACATAGCACCTGAGGCCCTGCCCTGAAGGGACTGGTACCTATTGGCTGAATTGAATTTCCATAACAAATCCACCAGCTTGCTTCCTGAGCACCACCACTGGTTTTGAAGGTTCCCTAAGGCCAGGCATTGTTCTACAGCTGTCTTTAACAGCTTGGAGTGGAAGTAGACATCCAGGTTGAGAAATTTGCCCAAGCTCATATAAGTAGAGGATTTAAATCACCAGATGCCAGTGGCTTTTTAAAATCCGTTTTACTGTTACACTGTTACTGTTGAA

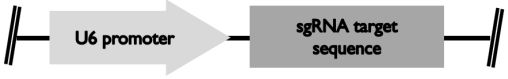
		ATGTTAACTTTTTTATCCCATGTCTTTCATCACTTAAGGTTT ATTCAGAAAAAAAGAACTGGCAGTTGTAACTTGAACAAG AACCTAAGACAGAGTTGGTCAGCTTCAATAGCTCCCTGCC CTTTGCTTTGTTTCATAGAAAAGGAAAAAATAACCATTCTT AATGTAAAAGGGCATTTTTACATGTTCTCTGATTTTGTTCT CTGACTAGTTCGAAGGTAGAATCAATTTGATACAACATTTT CAAAGTTTCCTTTTTTTTGTTGTTTTTCCAGACAGGGTTTCTC TGTAGCTTTGGAGACTGTCCTGGTACTAGCTCTGTAGATC AGACTGGCCTCTAATTCA	CATATTAGGTAATTCAGTAGCCATTTTAACTCGGGATGATC AGAAAAGTTGGCACGAACAATATGGGATGTTCTGGTGTA ACACATTAAGATTTTAACATTTTGCACATACTTTTCTCTTTG CACATATTTGCCCAACCATTGCATTTGTTGCACACATTTACT TTTTGCCCTTTGGTACTTATGGCCATGTTACAAGTAAGTAT AGGATTTGGTGATGGAGGGGTGTTCCTGGTTACAATATCT TACAGTATTATTACCAAGCTTCCAACATGACAGTATACCAT TCATGTTCTAGTTAACTGTCTCAGAAAATTCAGTTTGTCTA TGGTTAGCTTGACTTTGTA
T9	GCCCACTACA GGTTCGGGC G	GATAAACTATTCTTCATTTGCCAGGGGTGAACGAAACTCC AGTGCACATTTCTCAGCTCTATAGCTGAGAGATACCTCATA AAACATAGGATACTTAGGTAAATTTGAGTTTCAGGTGAAG ATTTTTTAAGTATAATCATGTCTTGCACAATGTTTGGGATT CATTTTTATTGATTAAATCTGACAGTCCTTTTTTTTTCTTTCC TTTAAAAAACTTTTTTTTTTTTTTTAAAGACAGGGTCTTATGT ATCCTAAGGTGACCTTGAACTTGCCTTGTAGATGAGTCTG GCTTTGAGCCCCTAATACCTACCTCCACCTCTCAAATCCTA GGATTACAGGTGTATACCACTCTACCCATTTGACAATTCTA CTTTTGAGCCATCTTAAAGGTAAAAAGTACACTTATTTATC ATATGCCTGTGTGTGGAAGTCAGAACAACTTGTGGTAGTT GGTTCTTTCCTTCATGGGGGCCCTGGGGGTTAGACTCCAA TTGTCAGGCTTGCTGGTAAATGCCTTATTACCACCTTGATG TGGAGGTACAAATAATGATTCACAGGACTAGGTTCAAATA TTAAATATAATTTATTCCGGGGAGAAGTCACTTACAGAAG AAGCAGACAGCTGCTGCAGGTTCCAGCTCTAAGGATCCGC ACAGGCTACTCCCTGTGGCCCAACCAGACGGAAGTGAGA GCCACCTGCTGGGACCCCAGACCTAAACTCACTCCCAGTC AGGTCCTCACCTGTAACCAC	TGGATATCTCACTACACTTTGATGGCCCAATCTTTTTTTAAA AGCATATATTTATTGAATAAAATATGAGTTCATATTGGCAT CCTCATGCATGTCTGGCAAGTACTTGGCTATTATTCACTTC CTCCTCTAAATCCCATCCATTCTTGGGTCCATAAACCTATCA TCTTAGTTTCTCCTCATTTCAAAAGAAATTAGGAAACAGAA GTTTGTAACAGGAAATGGATTTATTTATAGTTTTGCCACTG GGGAGAAAAAATAACTCCATCCTGTTCTCTAACATGCTGG TAGTCTTTTACAGTTTATTGAAAACTATGAGACGGGATGG AGAGGTGGTTCAGAGGTTAAGAGCAATGGCTGTTCTTCCA GAGGTCCTGAGTTCAATTCCCAGCAACCACATAGTAGCTC ACAGCTATCTGTAATGAGATCTGGTGTGCAAGCATACATG CCGGCAGAACACTGTATACAAAATAAATAAATCTTAAAAA AAAAAAGAAAACTATGAGACAAAGACAAGAAGGTATGTA GCTTTGCTGTGAGACAGTTCAACATGGCTTTTCTGGGGTT GTGAGGAGGTCACTAGTTTCATATGACCTGCAGTATCTCA GTTCTCAGCAGTTCTTGTTTAGCATGGAGGGGGTCTCTCTT TACAATTTTGCCATGGTTCCAGCCCTTTCTCAAGCATGGAA TCACTGCTATATTAGTGTCCTCTTTGTTATGGGGTAAAAAC CCAGAGACCATTCAAGCAAC

Supplementary Figure S2: Depictions of plasmids, used in this study

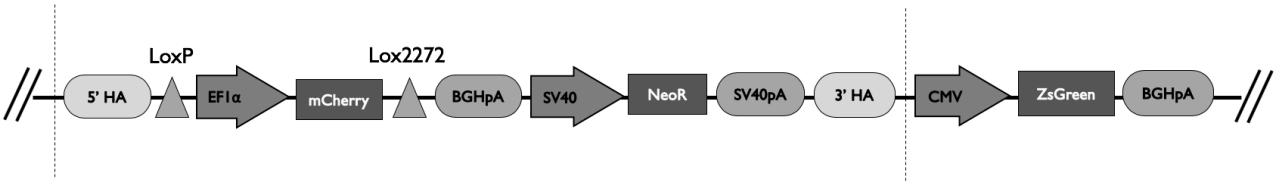
A – Rituximab plasmid for generation of C6_2 cell line



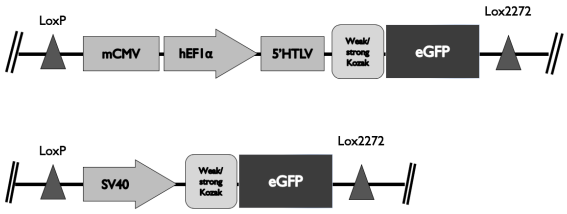
B – sgRNA plasmid



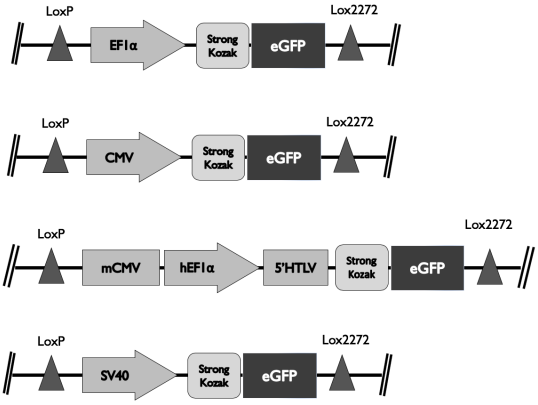
C – Landing pad donor plasmid



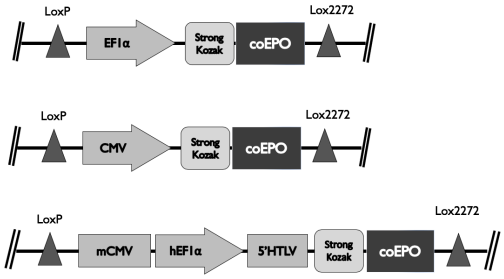
D – RMCE eGFP donor plasmids, used for proof-of-principle experiment



E – A panel of RMCE eGFP donor plasmids



F – A panel of RMCE EPO donor plasmids



Supplementary Table S3: All the primers used in this study

A – Cloning primers used for constructing the landing pad donor plasmids

Primer	Sequence
A 5HA LA_fwd	AGTCGGTGUGAAGCTTACCCCCAGGATTTCC
A 5HA LB_rev	ACGCTGCTUAAAACAGCCCAAAGCAAGTGAT
A 3HA LD_fwd	AGGTCTGAGUATGCCTCCATTGATTCTTTC
A 3HA O1_rev	AGCGACGUAGAAAATACCTTATAGACTTGACTA
T2 5HA LA fwd	AGTCGGTGUGTTTCTACACCATCCTAAAGA
T2 5HA LB rev	ACGCTGCTUTGAATTAGAGGCCAGTCTGAT
T2 3HA LD fwd	AGGTCTGAGUAGTCCACAGATTCTGAATGAA
T2 3HA O1 rev	AGCGACGUTACAAAGTCAAGCTAACCATAG
T9 5HA LA fwd	AGTCGGTGUGATAAACTATTCTTCATTTGC

T9 5HA LB rev	ACGCTGCTUGTGGTTACAGGTGAGGACC
T9 3HA LD fwd	AGGTCTGAGUTGGATATCTCACTACACTTTG
T9 3HA O1 rev	AGCGACGUGTTGCTTGAATGGTCTCTGG
pJ204 backbone O5 fwd	ACTTGCGUAGTGAGTCGAATAAGGGCGACACAAA
pJ204 backbone LA rev	ACACCGACUGAGTCGAATAAGGGCGACACCCCA
Lox2272-mCherry O4 rev	AGACTGTGUATAACTTCGTATAAAGTATCCTATACGAAGTTAT CTACTTGTACAGCTCGT
LoxP-EF1a LB fwd	AAGCAGCGUATAACTTCGTATAGCATACATTATACGAA GTTATGTGAGGCTCCGGTGCC
BGHpA fwd O4	ACACAGTCUCTGTGCCTTCTAGTTGCCA
NeoR LD rev	ACTCAGACCUCAGACATGATAAGATACATTG
CMV O1 fwd	ACGTCGCUGTTGACATTGATTATTGACT
BGHpA O5 rev	ACGCAAGUCCATAGAGCCCACCGCATCC

B – Cloning primers used for constructing RMCE eGFP donor plasmids

Primer	**Sequence**
CP fwd	AGTTATAGUCAATGGGAAAAACCCATTGGAGCC
CP strg Kozak rev	ATGGTGGCGUAGGCGCCGGTCACAGCTT
BB CP strg Kozak fwd	ACGCCACCAUGCATGTGAGCAAGGGCGA
BB CP rev	ACTATAACUTCGTATAATGTATGCTATACGAAG
CP weak Kozak rev	AAAAAGUAGGCGCCGGTCACAGCTT
BB CP weak Kozak fwd	ACTTTTUTATGCATGTGAGCAAGGG
SV40 fwd	AGTTATCUGTGGAATGTGTGTCAGTTAG
SV40 strg Kozak rev	AAAATGGAUATACAAGCTCCCGGGAGCTT
BB SV40 strg kozak fwd	ATCCATTTUCGGCCACCATGCATGTGAGC
BB SV40 rev	AGATAACUTCGTATAATGTATGCTATACGAAG
SV40 weak Kozak rev	AACGAAAAUGGATATACAAGCTCCCGG
BB SV40 weak Kozak fwd	ATTTTCGTUTTTTATGCATGTGAGCAA
EF1α fwd	AGTTATGUGAGGCTCCGGTGCCCGTC
EF1α strg Kozak rev	AGTGCTGUGGACTTTACCTTCTTTTTTTGAAA
BB EF1α strg Kozak fwd	ACAGCACUGCCACCATGCATGTGAGC
BB EF1α rev	ACATAACUTCGTATAATGTATGCTATACGAAG
CMV fwd	AGTTATGTUGACATTGATTATTGACTAGT
CMV strg Kozak rev	ATTTCGAUAAGCCAGTAAGCAGTGGG
BB CMV strg Kozak fwd	ATCGAAAUGCCACCATGCATGTGAGC
BB CMV rev	AACATAACUTCGTATAATGTATGCTATACGAAG
eGFP RMCE donor w strg Kozak fwd	ACGCCACCAUGCATGTGAGCAAGGGCGA
eGFP RMCE donor w strg Kozak rev	AATAACUTCGTATAATGTATGCTATACGAAG

C – Cloning primers used for constructing RMCE EPO donor plasmids

Primer	**Sequence**
LoxP CP LA fwd	AGTCGGTGUATAACTTCGTATAGCATACATTATACGAAGTTAT AGTCAATGGGAAAAACCCATTGG
CP LC rev	ATGACGTCUGTAGGCGCCGGTCACAGC
LoxP CMV LA fwd	AGTCGGTGUATAACTTCGTATAGCATACATTATACGAAGTTAT GTTGACATTGATTATTGACTAGT
CMV LC rev	ATGACGTCUATTTCGATAAGCCAGTAAGC
EF1α LC rev	ATGACGTCUTCACGACACCTGAAATGGA
LoxP EF1α LA fwd	AGTCGGTGUATAACTTCGTATAGCATACATTATACGAA GTTATGTGAGGCTCCGGTGC
pJ204 backbone O5 fwd	ACTTGCGUAGTGAGTCGAATAAGGGCGACACAAA
pJ204 backbone LA rev	ACACCGACUGAGTCGAATAAGGGCGACACCCCA
Kozak coEPO LC fwd	AGACGTCAUCGCCACCATGGGAGTGCACG
Lox2272 coEPO O5 rev	ACGCAAGUATAACTTCGTATAAAGTATCCTATACGAAGTTATTCATCTATCGCCGGTCC

D – Cloning primers used for 5'/3' junction PCRs

Primer	Sequence
A 5' junction fwd	GCCGCATGACCTTGTTCAAA
5' junction general rev	ATCCTGGCCCGCATTTACAA
3' junction general fwd	CTGGACGAAGAGCATCAGGG
A 3' junction rev	TCTGGCACAAGATGTAATGCTG
T2 5' junction fwd	CTCTTTCCCCGAGTTGCTGT
T2 3' junction rev	CTGCCATCAGTCACTAGCAGT
T9 5'junction fwd	GCATGCACAGAGAGGGACAT
T9 3'junction rev	CCCTCTGCAACTGCTAACCA

Supplementary Table S4: Sequences of all genes of interest (GOIs), promoters, terminators and Kozaks used in this study

Promoters	Sequence
EF1α	GTGAGGCTCCGGTGCCCGTCAGTGGGCAGAGCGCACATCGCCCACAGTCCCCGAGAAGTTGGGGGGAGGGGTCGGCAATTGA ACCGGTGCCTAGAGAAGGTGGCGCGGGGTAAACTGGGAAAGTGATGTCGTGTACTGGCTCCGCCTTTTTCCCGAGGGTGGGG GAGAACCGTATATAAGTGCAGTAGTCGCCGTGAACGTTCTTTTTCGCAACGGGTTTGCCGCCAGAACACAGGTAAGTGCCGTGT GTGGTTCCCGCGGGCCTGGCCTCTTTACGGGTTATGGCCCTTGCGTGCCTTGAATTACTTCCACCTGGCTCCAGTACGTGATTCT TGATCCCGAGCTGGAGCCAGGGGCGGGCCTTGCGCTTTAGGAGCCCCTTCGCCTCGTGCTTGAGTTGAGGCCTGGCCTGGGCG CTGGGGCCGCCGCGTGCGAATCTGGTGGCACCTTCGCGCCTGTCTCGCTGCTTTCGATAAGTCTCTAGCCATTTAAAATTTTTGA TGACCTGCTGCGACGCTTTTTTTCTGGCAAGATAGTCTTGTAAATGCGGGCCAGGATCTGCACACTGGTATTTCGGTTTTTGGGG CCGCGGGCGGCGACGGGGCCCGTGCGTCCCAGCGCACATGTTCGGCGAGGCGGGGCCTGCGAGCGCGGCCACCGAGAATCG GACGGGGGTAGTCTCAAGCTGGCCGGCCTGCTCTGGTGCCTGGCCTCGCGCCGCCGTGTATCGCCCCGCCCTGGGCGGCAAGG CTGGCCCGGTCGGCACCAGTTGCGTGAGCGGAAAGATGGCCGCTTCCCGGCCCTGCTCCAGGGGGCTCAAAATGGAGGACGC GGCGCTCGGGAGAGCGGGCGGGTGAGTCACCCACACAAAGGAAAGGGGCCTTTCCGTCCTCAGCCGTCGCTTCATGTGACTCC ACGGAGTACCGGGCGCCGTCCAGGCACCTCGATTAGTTCTGGAGCTTTTGGAGTACGTCGTCTTTAGGTTGGGGGGAGGGGTT TTATGCGATGGAGTTTCCCCACACTGAGTGGGTGGAGACTGAAGTTAGGCCAGCTTGGCACTTGATGTAATTCTCCTTGGAATT TGCCCTTTTTGAGTTTGGATCTTGGTTCATTCTCAAGCCTCAGACAGTGGTTCAAAGTTTTTTTCTTCCATTTCAGGTGTCGTGA
CP (mCMV-hEF1α-5'HTLV)	AGTCAATGGGAAAAACCCATTGGAGCCAAGTACACTGACTCAATAGGGACTTTCCATTGGGTTTTGCCCAGTACATAAGGTCAA TAGGGGGTGAGTCAACAGGAAAGTCCCATTGGAGCCAAGTACATTGAGTCAATAGGGACTTTCCAATGGGTTTTGCCCAGTAC ATAAGGTCAATGGGAGGTAAGCCAATGGGTTTTTCCCATTACTGACATGTATACTGAGTCATTAGGGACTTTCCAATGGGTTTT GCCCAGTACATAAGGTCAATAGGGGTGAATCAACAGGAAAGTCCCATTGGAGCCAAGTACACTGAGTCAATAGGGACTTTCCA TTGGGTTTTGCCCAGTACAAAAGGTCAATAGGGGGTGAGTCAATGGGTTTTTCCCATTATTGGCACATACATAAGGTCAATAGG GGTGACTAGTCAGTGGGCAGAGCGCACATCGCCCACAGTCCCCGAGAAGTTGGGGGGAGGGGTCGGCAATTGAACCGGTGCC TAGAGAAGGTGGCGCGGGGTAAACTGGGAAAGTGATGTCGTGTACTGGCTCCGCCTTTTTCCCGAGGGTGGGGGAGAACCGT ATATAAGTGCAGTAGTTGCCGTGAACGTTCTTTTTCGCAACGGGTTTGCCGCCAGAACACAGCTGAAGCTTCGAGGGGCTCGCA TCTCTCCTTCACGCGCCCGCCGCCCTACCTGAGGCCGCCATCCACGCCGGTTGAGTCGCGTTCTGCCGCCTCCCGCCTGTGGTGC CTCCTGAACTGCGTCCGCCGTCTAGGTAAGTTTAAAGCTCAGGTCGAGACCGGGCCTTTGTCCGGCGCTCCCTTGGAGCCTACC TAGACTCAGCCGGCTCTCCACGCTTTGCCTGACCCTGCTTGCTCAACTCTACGTCTTTGTTTCGTTTTCTGTTCTGCGCCGTTACAG ATCCAAGCTGTGACCGGCGCCTAC
SV40 early promoter	CTGTGGAATGTGTGTCAGTTAGGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTA GTCAGCAACCAGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCA TAGTCCCGCCCCTAACTCCGCCCATCCCGCCCCTAACTCCGCCCAGTTCCGCCCATTCTCCGCCCCATGGCTGACTAATTTTTTTTA TTTATGCAGAGGCCGAGGCCGCCTCTGCCTCTGAGCTATTCCAGAAGTAGTGAGGAGGCTTTTTTGGAGGCCTAGGCTTTTGCA AAAAGCTCCCGGGAGCTTGTATATCCATTTTCG
CMV	GTTGACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCCGCGTTAC ATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGT AACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCA TATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTATGGGACTTT CCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAATGGGCGTGGATAGC GGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCC AAAATGTCGTAACAACTCCGCCCCATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCTCT GGCTAACTAGAGAACCCACTGCTTACTGGCTTATCGAAAT

Terminators	Sequence
BGHpA	CTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTC CTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGG GGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGATGCGGTGGGCTCTATGG
SV40pA	CTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTC CTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGG GGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGATGCGGTGGGCTCTATGG
GOIs and Kozaks	**Sequence**
mCherry	ATGGTGAGCAAGGGCGAGGAGGATAACATGGCCATCATCAAGGAGTTCATGCGCTTCAAGGTGCACATGGAGGGCTCCGTGAAC GGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGCACCCAGACCGCCAAGCTGAAGGTGACCAAGG GTGGCCCCCTGCCCTTCGCCTGGGACATCCTGTCCCCTCAGTTCATGTACGGCTCCAAGGCCTACGTGAAGCACCCCGCCGACATC CCCGACTACTTGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAGGACGGCGGCGTGGTGACCGTG ACCCAGGACTCCTCCCTGCAGGACGGCGAGTTCATCTACAAGGTGAAGCTGCGCGGCACCAACTTCCCCTCCGACGGCCCCGTAA TGCAGAAGAAGACCATGGGCTGGGAGGCCTCCTCCGAGCGGATGTACCCCGAGGACGGCGCCCTGAAGGGCGAGATCAAGCAG AGGCTGAAGCTGAAGGACGGCGGCCACTACGACGCTGAGGTCAAGACCACCTACAAGGCCAAGAAGCCCGTGCAGCTGCCCGG CGCCTACAACGTCAACATCAAGTTGGACATCACCTCCCACAACGAGGACTACACCATCGTGGAACAGTACGAACGCGCCGAGGGC CGCCACTCCACCGGCGGCATGGACGAGCTGTACAAGTAG
eGFP	ATGCATGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAA GTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCC CGTGCCCTGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGACT TCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGA GGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGC ACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGAT CCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCT GCCCGACAACCACTACCTGAGCACCCAGTCCAAGCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTC GTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGTAG
NeoR	ATGGTGAGCAAGGGCGAGGAGGATAACATGGCCATCATCAAGGAGTTCATGCGCTTCAAGGTGCACATGGAGGGCTCCGTGAAC GGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGCACCCAGACCGCCAAGCTGAAGGTGACCAAGG GTGGCCCCCTGCCCTTCGCCTGGGACATCCTGTCCCCTCAGTTCATGTACGGCTCCAAGGCCTACGTGAAGCACCCCGCCGACATC CCCGACTACTTGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAGGACGGCGGCGTGGTGACCGTG ACCCAGGACTCCTCCCTGCAGGACGGCGAGTTCATCTACAAGGTGAAGCTGCGCGGCACCAACTTCCCCTCCGACGGCCCCGTAA TGCAGAAGAAGACCATGGGCTGGGAGGCCTCCTCCGAGCGGATGTACCCCGAGGACGGCGCCCTGAAGGGCGAGATCAAGCAG AGGCTGAAGCTGAAGGACGGCGGCCACTACGACGCTGAGGTCAAGACCACCTACAAGGCCAAGAAGCCCGTGCAGCTGCCCGG CGCCTACAACGTCAACATCAAGTTGGACATCACCTCCCACAACGAGGACTACACCATCGTGGAACAGTACGAACGCGCCGAGGGC CGCCACTCCACCGGCGGCATGGACGAGCTGTACAAGTAG
ZsGreen	ATGGCCCAGTCCAAGCACGGCCTGACCAAGGAGATGACCATGAAGTACCGCATGGAGGGCTGCGTGGACGGCCACAAGTTCGTG ATCACCGGCGAGGGCATCGGCTACCCCTTCAAGGGCAAGCAGGCCATCAACCTGTGCGTGGTGGAGGGCGGCCCCTTGCCCTTC GCCGAGGACATCTTGTCCGCCGCCTTCATGTACGGCAACCGCGTGTTCACCGAGTACCCCCAGGACATCGTCGACTACTTCAAGAA CTCCTGCCCCGCCGGCTACACCTGGGACCGCTCCTTCCTGTTCGAGGACGGCGCCGTGTGCATCTGCAACGCCGACATCACCGTGA GCGTGGAGGAGAACTGCATGTACCACGAGTCCAAGTTCTACGGCGTGAACTTCCCCGCCGACGGCCCCGTGATGAAGAAGATGA CCGACAACTGGGAGCCCTCCTGCGAGAAGATCATCCCCGTGCCCAAGCAGGGCATCTTGAAGGGCGACGTGAGCATGTACCTGCT GCTGAAGGACGGTGGCCGCTTGCGCTGCCAGTTCGACACCGTGTACAAGGCCAAGTCCGTGCCCCGCAAGATGCCCGACTGGCA CTTCATCCAGCACAAGCTGACCCGCGAGGACCGCAGCGACGCCAAGAACCAGAAGTGGCACCTGACCGAGCACGCCATCGCCTC CGGCTCCGCCTTGCCCAAGCTTCCGCGGAGCCATGGCTTCCCGCCGGCGGTGGCGGCGCAGGATGATGGCACGCTGCCCATGTCT TGTGCCCAGGAGAGCGGGATGGACCGTCACCCTGCAGCCTGTGCTTCTGCTAGGATCAATGTGTAG
EPO	ATGGGAGTGCACGAGTGTCCTGCTTGGCTGTGGCTGCTGCTGTCCCTGCTGTCTCTGCCTCTGGGACTGCCTGTGCTGGGCGCTCC TCCTAGACTGATCTGCGACTCCCGGGTGCTGGAAAGATACCTGCTGGAAGCCAAAGAGGCCGAGAACATCACCACCGGCTGCGC CGAGCACTGCTCCCTGAACGAGAATATCACCGTGCCCGACACCAAAGTGAACTTCTACGCCTGGAAGCGGATGGAAGTGGGCCA GCAGGCTGTGGAAGTGTGGCAGGGACTGGCTCTGCTGAGCGAGGCTGTGCTGAGAGGACAGGCCCTGCTCGTGAACTCCTCCCA GCCTTGGGAACCCCTGCAGCTGCACGTGGACAAGGCTGTGTCCGGCCTGAGATCCCTGACCACCCTGCTGAGAGCACTGGGAGCC CAGAAAGAGGCCATCTCTCCACCTGACGCCGCCTCTGCTGCTCCTCTGAGAACCATCACCGCCGACACCTTCAGAAAGCTGTTCCG GGTGTACTCCAACTTCCTGCGGGGCAAGCTGAAGCTGTACACCGGCGAGGCTTGCCGGACCGGCGATAGATGA
Strong Kozak	GCCACC
Weak Kozak	TTTTTT

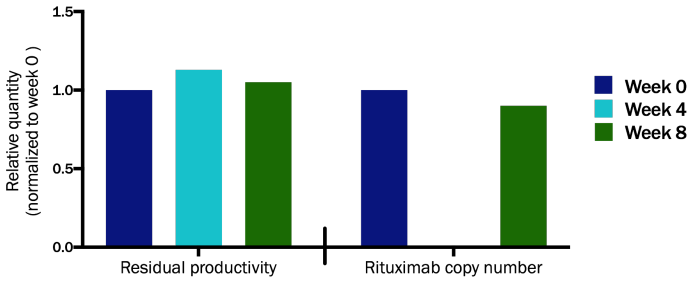
Supplementary Table S5: TaqMan assays (primers and probes) for qPCR

Gene	Forward Primer	Reverse Primer	TaqMan Probe	Dye	Primer Efficiency
mCherry	GACTACTTGAAGCTGTCCTTCC	CGCAGCTTCACCTTGTAGAT	TTCAAGTGGGAGCGCGTGATGAA	FAM-MGB	96.4%
NeoR	GCTATCAGGACATAGCGTTGG	GATACCGTAAAGCACGAGGAAG	ACCCGTGATATTGCTGAAGAGCTTGG	FAM-MGB	101%

eGFP	GCACAAGCTGGAGTACAACTA	TGTTGTGGCGGATCTTGAA	AGCAGAAGAACGGCATCAAGGTGA	FAM-MGB	98.75%
COSMC	ACCCGAACCAGGTAGTAGAA	ACATGTCCAAAGGCCCTAAG	AGTGACAGCCATATTGGAACAGCATCC	VIC-MGB	101%
coEPO	CTGGAAAGATACCTGCTGGAAG	AGGCGTAGAAGTTCACTTTGG	CCAAAGAGGCCGAGAACATCACCA	FAM-MGB	94.19%
Fkbp1a	CTCTCGGGACAGAAACAAGC	GACCTACACTCATCTGGGCTAC	ATGCTAGGCAAGCAGGAGGTGATC	VIC-MGB	100%
Gnb1	CCATATGTTTCTTTCCCAATGGC	AAGTCGTCGTACCCAGCAAG	ACTGGTTCAGACGATGCTACGTGC	ABY-QSY	99%

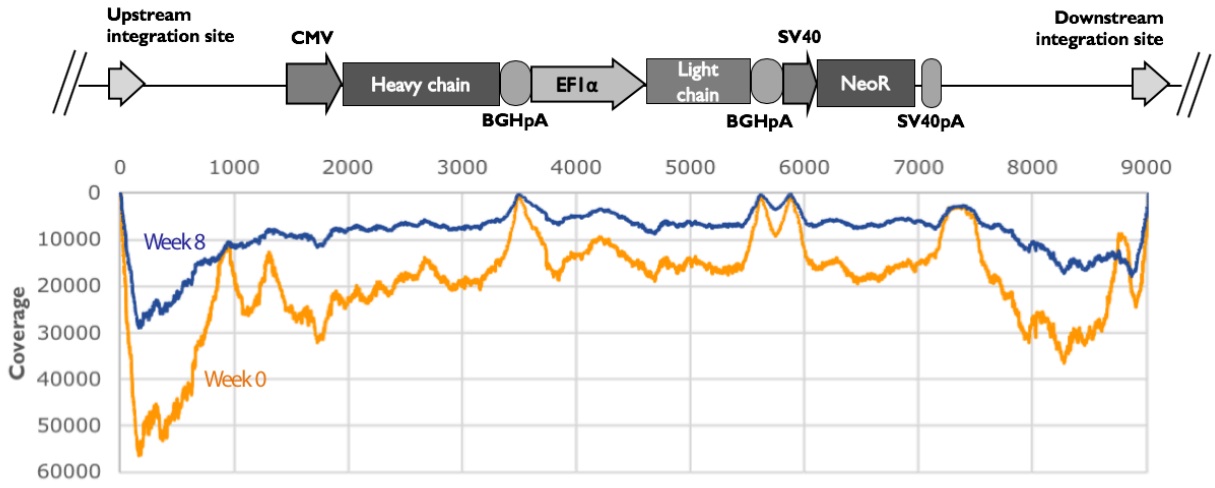
Supplementary Figure S3: Residual productivity and copy numbers of Rituximab in randomly integrated high-producing CHO-S monoclonal cell line throughout the 2-month stability assay

Specific productivity and copy number of Rituximab were measured at week 0, 4 (only specific productivity) and 8 from the start of the 2-month stability assay and normalized to week 0 values (RQ = 1).



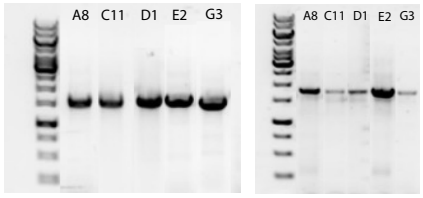
Supplementary Figure S4: Targeted MiSeq deep-sequencing of two PCR product encompassing integrated construct and flanking genomic regions of site A

At week 0 and 8 from the start of the 2-month stability assay, genomic DNA of C6_2 cell line was extracted and the region surrounding site A, predicted from Cergentis's TLA sequencing (9), was PCR amplified. Obtained 9-kb PCR products were MiSeq deep-sequenced and aligned to the integrated Rituximab plasmid.

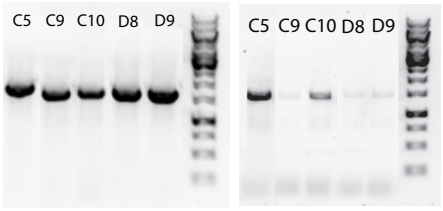


Supplementary Figure S5: 5' and 3' junction PCRs for confirmation of specific integration of mCherry landing pad in all three integration sites for all 15 selected platform cell lines (PCLs)

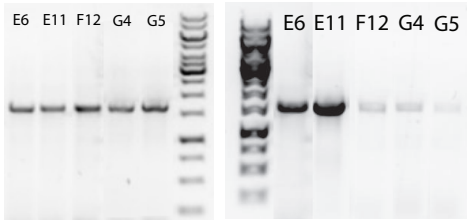
A – 5' junction (left) and 3' junction (right) PCRs for validation of integration in site A for five selected PCLs



B – 5' junction (left) and 3' junction (right) PCRs for validation of integration in site T2 for five selected PCLs

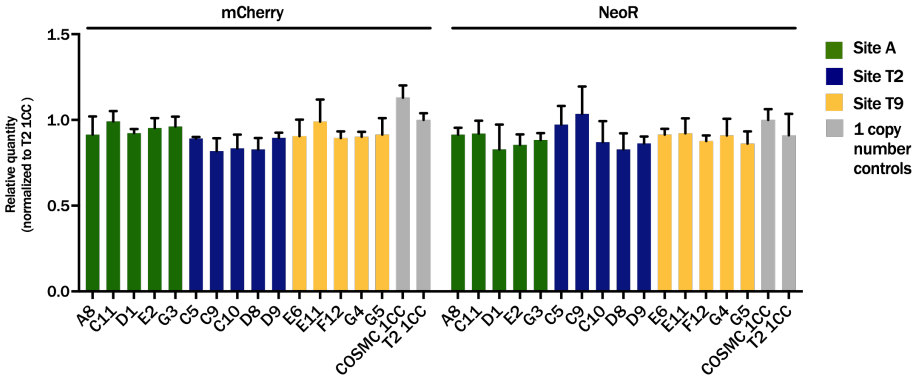


C – 5' junction (left) and 3' junction (right) PCRs for validation of integration in site T9 for five selected PCLs



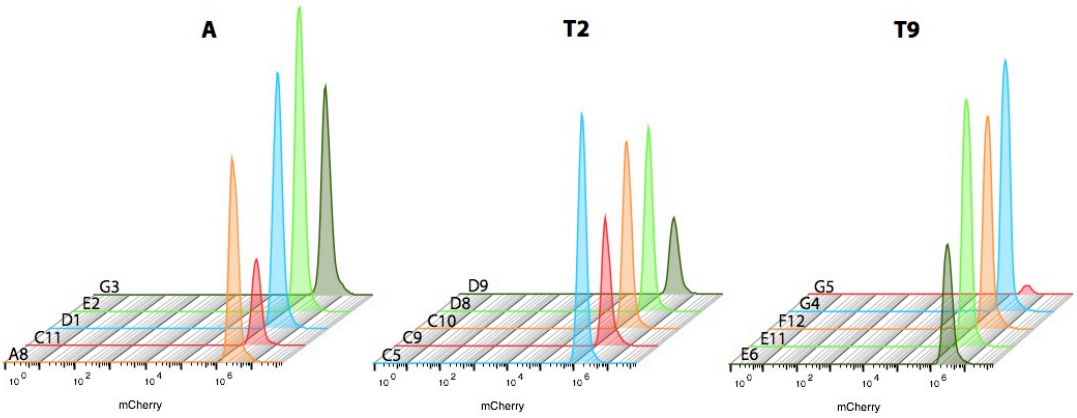
Supplementary Figure S6: mCherry and NeoR copy number in 15 selected PCLs

mCherry and NeoR copy numbers were measured in genomic DNA of 15 selected PCLs, where integration of mCherry landing pad in site A, T2 and T9 was confirmed by 5' and 3' junction PCR (Figure S4). *COSMC* was used as a control gene for normalization and PCLs' copy numbers were compared to two control samples, verified in our previous studies (10, 11) to contain one copy of mCherry landing pad integrated in COSMC and T2 site, respectively. The error bars represent the standard deviation of the technical replicates (n = 3).



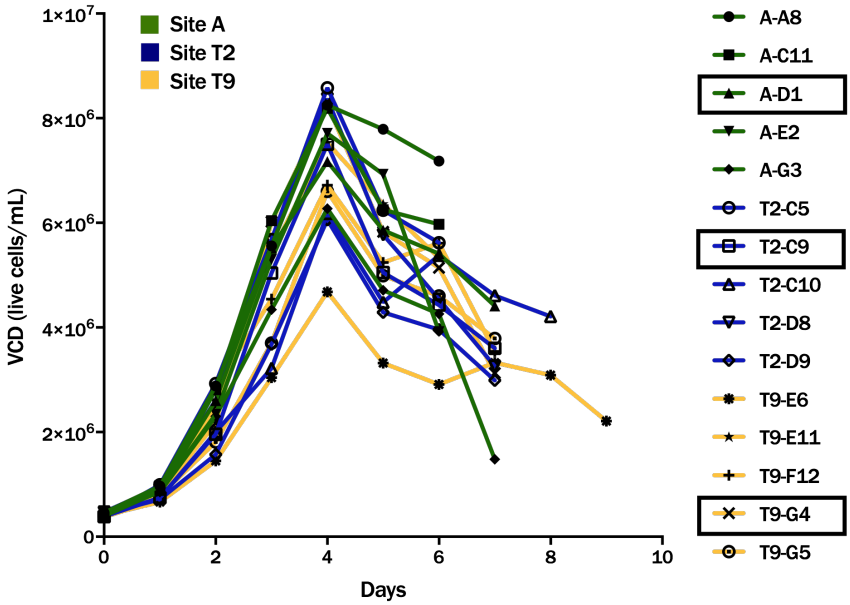
Supplementary Figure S7: mCherry fluorescence homogeneity in 15 selected PCLs at the time of validation

mCherry fluorescence was measured in unstained suspension PCL cells using Xcyto 10 image cytometer and was considered homogenous when ≥ 98% cells were mCherry positive.



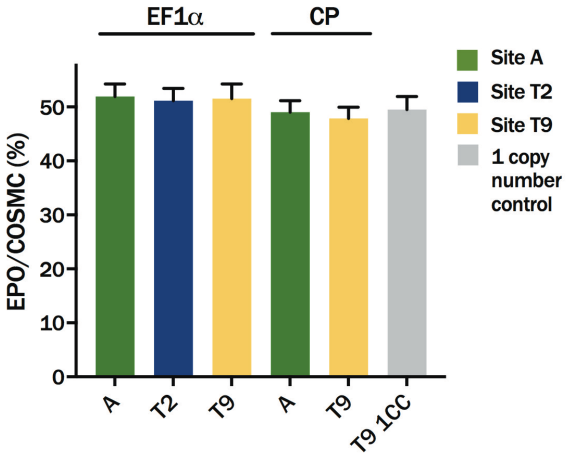
Supplementary Figure S8: Growth curves of 15 selected PCLs during batch cultivation

15 PCLs were cultivated under batch conditions for 6-9 days (*i.e.* until the viability dropped below 60% of cells). Viable cell density (VCD) and viability were measured daily. Encircled clones, A-D1, T2-C9 and T9-G4, are the representative PCL clones for site A, T2 and T9, respectively, used for the majority of recombination experiments.



Supplementary Figure S9: Target-to-reference ratios for bulk-sorted samples recombined with codon optimized EPO (coEPO) using digital PCR (dPCR)

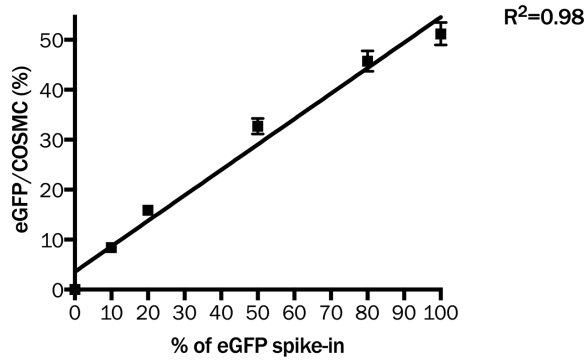
A-D1, T2-C9 and T9-G4 PCLs were recombined with RMCE EPO donor plasmids and bulk sorted by FACS (mCherry negative cells). Copy number of EPO was measured in recovered bulk-sorted populations using dPCR, being reported as coEPO/COSMC reads (%), with *COSMC* being used as one-copy control gene. coEPO/COSMC ratio of bulk-sorted populations was compared to the ratio of a control sample (monoclonal cell line, verified in our previous study to contain one copy of EPO in site T9 (12). The error bars represent 95% confidence intervals, averaged between the technical replicates (n = 2).



Supplementary Figure S10: Target-to-reference ratios for titration experiment with A-CP eGFP bulk-sorted sample using dPCR

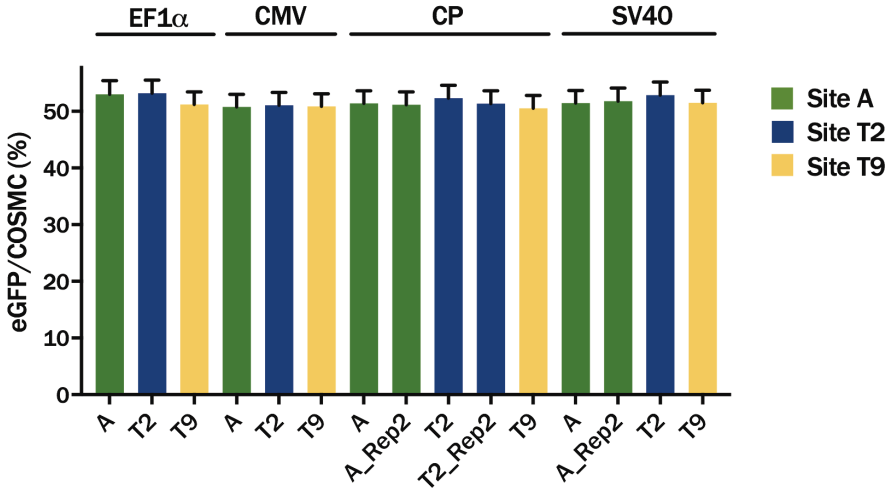
Titration experiment was performed using dPCR, with genomic DNA of A-CP eGFP bulk-sorted sample being incrementally spiked in genomic DNA of CHO-S wt (*i.e.* in 0%, 10%, 20%, 50%, 80% and 100% of the total DNA). eGFP/COSMC reads (%) are reported for each titration mix, with *COSMC* being used as one-copy

control gene. The error bars represent 95% confidence intervals, averaged between the technical replicates (n = 2).



Supplementary Figure S11: Target-to-reference ratios for bulk-sorted samples recombined with eGFP using dPCR

A-D1, T2-C9 and T9-G4 PCLs were recombined with RMCE eGFP donor plasmids and bulk sorted by FACS (mCherry negative/eGFP positive cells). Copy number of eGFP was measured in recovered bulk-sorted populations using dPCR, being reported as eGFP/COSMC reads (%), with *COSMC* being used as one-copy control gene. Three bulk-sorted populations (namely A-CP, A-SV40 and T2-CP) were generated twice, independently (for more details refer to Material and Methods in the main text), here marked as Rep2 samples. The error bars represent 95% confidence intervals, averaged between the technical replicates (n = 2).



References

1. Pristovšek,N., Hansen,H.G., Sergeeva,D., Borth,N., Lee,G.M., Andersen,M.R. and Kildegaard,H.F. (2018) Using Titer and Titer Normalized to Confluence Are Complementary Strategies for Obtaining Chinese Hamster Ovary Cell Lines with High Volumetric Productivity of Etanercept. *Biotechnol. J.*, **13**, 1–10.
2. Lee,J.S., Grav,L.M., Pedersen,L.E., Lee,G.M. and Kildegaard,H.F. (2016) Accelerated homology-directed targeted integration of transgenes in Chinese hamster ovary cells via CRISPR/Cas9 and fluorescent enrichment. *Biotechnol. Bioeng.*, **113**, 2518–2523.
3. Grav,L.M., Lee,J.S., Gerling,S., Beuchert Kallehauge,T., Holmgaard Hansen,A., Kol,S., Lee,G.M., Ebdrup

Pedersen,L. and Faustrup Kildegaard,H. (2015) One-step generation of triple knockout CHO cell lines using CRISPR Cas9 and fluorescent enrichment. *Biotechnol. J.*, **10**, 1446–1456.

4. Pybus,L.P., Dean,G., West,N.R., Smith,A., Daramola,O., Field,R., Wilkinson,S.J. and James,D.C. (2014) Model-directed engineering of 'difficult-to-express' monoclonal antibody production by Chinese hamster ovary cells. *Biotechnol. Bioeng.*, **111**, 372–385.
5. Hefzi,H., Ang,K.S., Hanscho,M., Bordbar,A., Ruckerbauer,D., Lakshmanan,M., Orellana,C.A., Baycin-Hizal,D., Huang,Y., Ley,D., *et al.* (2016) A Consensus Genome-scale Reconstruction of Chinese Hamster Ovary Cell Metabolism. *Cell Syst.*, **3**, 434–443.e8.
6. Rupp,O., Macdonald,M.L., Li,S., Dhiman,H., Polson,S., Griep,S., Heffner,K., Hernandez,I., Brinkrolf,K., Jadhav,V., *et al.* (2018) A reference genome of the Chinese hamster based on a hybrid assembly strategy. *Biotechnol. Bioeng.*, 10.1002/bit.26722.
7. Singh,A., Kildegaard,H.F. and Andersen,M.R. (2018) An Online Compendium of CHO RNA-Seq Data Allows Identification of CHO Cell Line-specific Transcriptomic Signatures. *Biotechnol. J.*, **1800070**, 1800070.
8. Feichtinger,J., Hernandez,I., Fischer,C., Hanscho,M., Auer,N., Hackl,M., Jadhav,V., Baumann,M., Krempl,P.M., Schmidl,C., *et al.* (2016) Comprehensive genome and epigenome characterization of CHO cells in response to evolutionary pressures and over time. *Biotechnol. Bioeng.*, **113**, 2241–2253.
9. De Vree,P.J.P., De Wit,E., Yilmaz,M., Van De Heijning,M., Klous,P., Verstegen,M.J.A.M., Wan,Y., Teunissen,H., Krijger,P.H.L., Geeven,G., *et al.* (2014) Targeted sequencing by proximity ligation for comprehensive variant detection and local haplotyping. *Nat. Biotechnol.*, **32**, 1019–1025.
10. Lee,J.S., Kallehauge,T.B., Pedersen,L.E. and Kildegaard,H.F. (2015) Site-specific integration in CHO cells mediated by CRISPR/Cas9 and homology-directed DNA repair pathway. *Sci. Rep.*, **5**, 8572.
11. Petersen,S.D., Zhang,J., Lee,J.S., Jakočiūnas,T., Grav,L.M., Kildegaard,H.F., Keasling,J.D. and Jensen,M.K. (2018) Modular 5-UTR hexamers for context-independent tuning of protein expression in eukaryotes. *Nucleic Acids Res.*, 10.1093/nar/gky734.
12. Grav,L.M., Sergeeva,D., Lee,J.S., Marín de Mas,I., Lewis,N.E., Andersen,M.R., Nielsen,L.K., Lee,G.M. and Kildegaard,H.F. (2018) Minimizing clonal variation during mammalian cell line engineering for improved systems biology data generation. *ACS Synth. Biol.*, 10.1021/acssynbio.8b00140.

Transient Co-Expression of Novel Effector Genes and Different IgG1 Molecules Increases Specific Productivity of CHO-S Cells

Supplementary data

Nuša Pristovšek[1], Henning Gram Hansen[1], Anne Mathilde Lund[2], Johan Rojek[1], Claes Nymand Nilsson[1], Gyun Min Lee[1,3], Mikael Rørdam Andersen[2], Helene Faustrup Kildegaard[1]

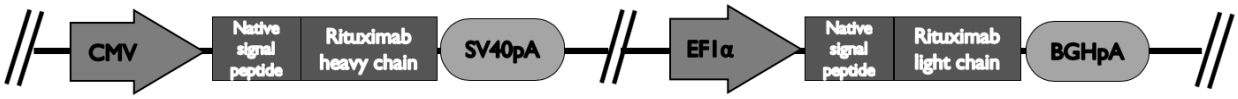
[1]The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Kemitorvet 220, 2800 Kgs. Lyngby, Denmark
[2]Department of Biotechnology and Biomedicine, Technical University of Denmark, Søltofts Plads 223, 2800 Kgs. Lyngby, Denmark
[3]Department of Biological Sciences, KAIST, 291 Daehak-ro, Yuseong-gu, Daejeon 305-701, Republic of Korea

Supplementary Table S1: Cloning of effector genes, used in this study

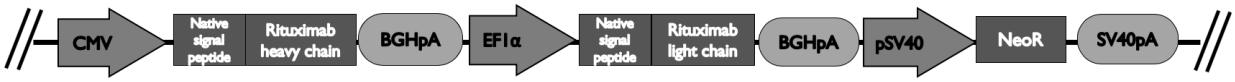
Gene	Refseq ID	Uniprot ID	template type	template name/ID	amplicon size
Mm_Xbp1s	NM_001271730.1	O35426-2	single stranded oligo	PR00398 XBP1s_part1_ss oligo	136 bp
Mm_Xbp1s	NM_001271730.1	O35426-2	gBlock	G0001 Xbp1S_part2	1 kb
Mm_p97	NM_009503.4	Q01853-1	IMAGE clone	G0017 Vcp/p97	2.4 kb
Mm_BiP	NM_001163434.1	P20029-1	IMAGE clone	G0019 Hspa5/BiP	2.5 kb
Mm_Gosr2	NM_019650.3	O35166-1	gBlock	G0022 Gosr2_part1	657 bp
Mm_Tbp	NM_013684.3	P29037-1	gBlock	G0023 Tbp_part1	969 bp
Mm_Sec11c	NM_025468.2	Q9D8V7-1	gBlock	G0024 Sec11_part1	597 bp
Mm_Mafb	NM_010658.3	P54841-1	gBlock	G0025 Mafb_part1	518 bp
Mm_Mafb	NM_010658.3	P54841-1	gBlock	G0026 Mafb_part2	450 bp
Mm_Eif4e3	NM_025829.4	Q9DBB5-1	gBlock	G0027 Eif4e3_part1	623 bp
Mm_PDI	NM_011032.2	P09103-1	cDNA	amsbio cDNA (#C4334566)	1530 bp
Mm_Ero1a	NM_015774.3	Q8R2E9-1	cDNA	amsbio cDNA (#C4334566)	1395 bp
Mm_Ero1b	NM_026184.2	Q8R2E9-1	cDNA	amsbio cDNA (#C4334566)	1404 bp
Mm_CREB5	NM_172728.2	Q8K1L0	gBlock	G0015 CREB5_part1	0.5 kb
Mm_CREB5	NM_172728.2	Q8K1L0	gBlock	G0016 CREB5_part2	0.6 kb
Mm_Sorcs3	NM_025696.3	Q8VI51-1	cDNA	amsbio cDNA (#C4334566)	3660 bp
Mm_Epas1	NM_010137.3	P97481-1	gBlock	G0028 and G0029	
Mm_map3k5	NM_008580.4	O35099-1	cDNA	amsbio cDNA (#C4334566)	4143 bp
Mm_Cebpd	NM_007679.4	Q00322	cDNA	amsbio cDNA (#C4334566)	807 bp
Mm_Casq1	NM_009813.2	O09165-1	cDNA	amsbio cDNA (#C4334566)	1875 bp
Mm_Tcp1	NM_013686.4	P11983	cDNA	amsbio cDNA (#C4334566)	1671 bp
Mm_Scd1	NM_009127.4	P13516	cDNA	amsbio cDNA (#C4334566)	1068 bp
Mm_Cryab	NM_009964.3	P23927	cDNA	amsbio cDNA (#C4334566)	528 bp
Mm_Nfe2	NM_008685.2	Q07279	cDNA	amsbio cDNA (#C4334566)	1122 bp
Mm_Creb3l2	NM_178661.4	Q8BH52	cDNA	amsbio cDNA (#C4334566)	1566 bp
Mm_Bicd1	NM_009753.4	Q8BR07	cDNA	amsbio cDNA (#C4334566)	2508 bp
Mm_Ppp2r2d	NM_026391.2	Q925E7	cDNA	amsbio cDNA (#C4334566)	1362 bp
Mm_CHOP	NM_001290183.1 and NM_007837.4 (make nested primers so both variants can be amplified)	P35639-1	cDNA	amsbio cDNA (#C4334566)	
Mm_SRP14	NM_009273.4	P16254-1	cDNA	amsbio cDNA (#C4334566)	
Mm_SNAP-23	NM_001177792.1, NM_001177793.1, NM_009222.3.	O09044-1	cDNA	amsbio cDNA (#C4334566)	
Mm_VAMP8	NM_016794.3	O70404-1	cDNA	amsbio cDNA (#C4334566)	
Mm_cdc42	NM_009861.3	P60766-2	cDNA	amsbio cDNA (#C4334566)	
Mm_cdc42F28L	N/A	N/A	cDNA	amsbio cDNA (#C4334566)	
Mm_Sly1	NM_029825.2	Q8BRF7-1	cDNA	amsbio cDNA (#C4334566)	
Mm_Munc18	NM_011504.1	Q60770-1	cDNA	amsbio cDNA (#C4334566)	
Mm_CERT	N/A	Q9EQG9-2	cDNA	amsbio cDNA (#C4334566)	
Mm_CERTS132A	N/A	N/A	cDNA	amsbio cDNA (#C4334566)	
Mm_ATF6c	NM_001081304.1	F6VAN0-1 (aa 1-373)	cDNA	amsbio cDNA (#C4334566)	
Mm_mTOR	N/A	Q9JLN9-1 (R628C natural variant)	cDNA	amsbio cDNA (#C4334566)	

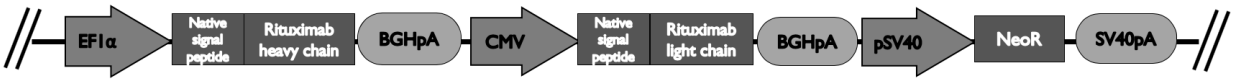
Supplementary Figure S1: Depictions of plasmids, used in this study

A – Rituximab A

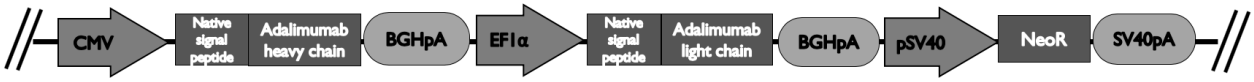


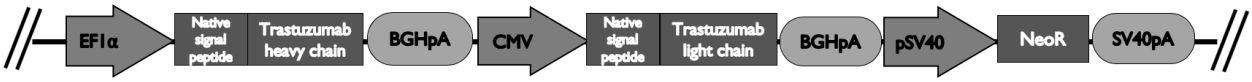
B – Rituximab B and B swap





C – Adalimumab and Trastuzumab





D – Etanercept

E – Effector genes and the empty vector

Supplementary Table S2: TaqMan assays (primers and probes) for qPCR

Gene	Forward Primer	Reverse Primer	TaqMan Probe	Dye	Primer Efficiency
Rituximab HC	CCTGACATGCCTCGTGAAA	GGGTGGTCTTGTAGTTGTTCTC	ATATCGCCGTGGAATGGGAGTCC	FAM-MGB	94.5%
Rituximab LC	TGTTCATCTTCCCACCTTCC	CACCTTCCACTGCACCTT	TCGTGTGCCTGCTGAACAACTTCT	ABY-QSY	92.0%
BiP	ACCACCTATTCCTGCGTTGG	AGACCGTGTTCTCGGGATTG	TCATAGCCAACGATCAGGGCAACC	ABY-QSY	119.4%
HERP	TGTCGGGTGGTTTCCATTCA	CTGTGCCTTCCTGGAGGTTA	ACCATCGTCCGGGAAGTTCTGAAC	FAM-MGB	103.1%
ACTB	Commercially available TaqMan assay (Thermo Fisher Scientific): Cg04424026_gH			VIC-MGB	93.7%
GAPDH	Commercially available TaqMan assay (Thermo Fisher Scientific): Cg04424038_gH			VIC-MGB	94.2%

Supplementary Table S3: 'Positive control' effector gene panel

Genes with published positive effects on antibody qp/titer														
Gene	**Antibody**	**Fold change**	**Qp after (pcd)**	**Host cell line**	**Expression (IgG)**	**Media**	**Amplified**	**Serum**	**Expression (effector gene)**	**Inducible (effector gene)**	**Function**	**Gene origin**	**Note**	**Reference**
PDI	HIV-1	1.37	22.6	CHO2F5 MCB		DMEM/Ham's F12	DHFR	5%	Stable	No	Folding/disulfide	?		N. Borth, D. Mattanovich, R. Kunert, and H. Katinger, "Effect of increased expression of protein disulfide isomerase and heavy chain binding protein on antibody secretion in a recombinant CHO cell line," *Biotechnol. Prog.* , vol. 21, no. 1, pp. 106–111, 2005.
PDI	?	1.15 - 1.27	3.86	DG44		IMDM	DHFR	10%	Stable	Yes	Folding/disulfide	CHO		C. Mohan, S. H. Park, J. Y. Chung, and G. M. Lee, "Effect of doxycycline-regulated protein disulfide isomerase expression on the specific productivity of recombinant CHO cells: Thrombopoietin and antibody," *Biotechnol. Bioeng.* , vol. 98, no. 3, pp. 611–615, Oct. 2007.
ERO1-Lα	?	1.37	7.9	DG44		IMDM	DHFR	10%	Transient	No	Disulfide	Human	Does not work with stable overexpression	C. Mohan and G. M. Lee, "Effect of inducible co-overexpression of protein disulfide isomerase and endoplasmic reticulum oxidoreductase on the specific antibody productivity of recombinant Chinese hamster ovary cells.," *Biotechnol. Bioeng.* , vol. 107, no. 2, pp. 337–46, Oct. 2010.
ERO1-Lα + PDI	?	1.55	8.95	DG44		IMDM	DHFR	10%	Transient	No	Folding/disulfide	PDI (CHO), ERO1L (Human)	Does not work with stable overexpression	C. Mohan and G. M. Lee, "Effect of inducible co-overexpression of protein disulfide isomerase and endoplasmic reticulum oxidoreductase on the specific antibody productivity of recombinant Chinese hamster ovary cells.," *Biotechnol. Bioeng.* , vol. 107, no. 2, pp. 337–46, Oct. 2010.
Xbp1s	?	1.2	N.D.	DG44		BI	?	No	Stable	Yes	UPR	Human	Lowers growth	E. Becker, L. Florin, K. Pfizenmaier, and H. Kaufmann, "Evaluation of a combinatorial cell engineering approach to overcome apoptotic effects in XBP-1(s) expressing cells.," *J. Biotechnol.* , vol. 146, no. 4, pp. 198–206, Apr. 2010.
Xbp1s + XIAP	?	1.6	N.D.	DG44		BI	?	No	Stable	Yes	UPR/survival	Human		E. Becker, L. Florin, K. Pfizenmaier, and H. Kaufmann, "Evaluation of a combinatorial cell engineering approach to overcome apoptotic effects in XBP-1(s) expressing cells.," *J. Biotechnol.* , vol. 146, no. 4, pp. 198–206, Apr. 2010.
Xbp1s + ERO1-Lα	N.D.	Yield (1.7 - 6.3)	N.D.	CHO-S	Transient	CD CHO + 2mM Glutamax	No	No	Stable	No	UPR/disulfide	Human		K. Cain, S. Peters, H. Hailu, B. Sweeney, P. Stephens, J. Heads, K. Sarkar, A. Ventom, C. Page, and A. Dickson, "A CHO cell line engineered to express XBP1 and ERO1-Lα has increased levels of transient protein expression.," *Biotechnol. Prog.* , vol. 29, no. 3, pp. 697–706, 2013.
CHOP	humanized (hTRA-8)	Yield (1.8)	N.D.	CHO-S	Transient	α-MEM	No	10%	Transient	No	UPR/folding/disulfide	CHO or Human	No known effect on qp	D. Nishimiya, T. Mano, K. Miyadai, H. Yoshida, and T. Takahashi, "Overexpression of CHOP alone and in combination with chaperones is effective in improving antibody production in mammalian cells," *Appl. Microbiol. Biotechnol.* , vol. 97, no. 6, pp. 2531–2539, 2013.
SRP14	Infliximab	1.35 - 2.4+	60	CHO-K1	Stable	SFM4CHO Hyclone + L-glutamine+HT supplement	No	No	Stable (piggyBac)	No	Translocation	Human		V. Le Fourn, P.-A. Girod, M. Buceta, A. Regamey, and N. Mermod, "CHO cell engineering to prevent polypeptide aggregation and improve therapeutic protein secretion.," *Metab. Eng.* , vol. 21, pp. 91–102, Jan. 2014.
SNAP-23, VAMP8	Anti-CD18	2.5	25	CHO-B13-24	Stable	CHOMaster	?	5%	Stable	No	Vesicle transport	Rat (his-tagged)		R. W. Peng, E. Abellan, and M. Fussenegger, "Differential effect of exocytic SNAREs on the production of recombinant proteins in mammalian cells," *Biotechnol. Bioeng.* , vol. 108, no. 3, pp. 611–620, 2011.
cdc42F28L, E1B	Multiple	2	30	Multiple	Stable	Multiple	Multiple	Multiple	Stable	No	Cell cycle/apoptosis-inhibition	Mouse		T. Rose, V. Sandig, H. H. VON, and K. Winkler, "Enhancement of protein production yield mediated by a fast shuttling cdc42 gtpase," WO2012152945 A1, 15-Nov-2012
Sly1 and Munc18 + Xbp1s	Rituximab	10 - 19	38	CHO-K1	Stable	?	?	?	Stable	?	Vesicle transport	?		H. Kaufmann, E. Becker, L. Florin, M. Fussenegger, R.-W. Peng, and J. M. Studts, "Sm-protein based secretion engineering," WO2009080299 A1, Jul-2009.
CERT, CERT S132A	N.D.	1.3	65	DG44	Stable	BI-proprietary CD SF media	N.D.	N.D.	Stable	No	Lipid transfer	Human		L. Florin, A. Pegel, E. Becker, A. Hausser, M. A. Olayioye, and H. Kaufmann, "Heterologous expression of the lipid transfer protein CERT increases therapeutic protein productivity of mammalian cells.," *J. Biotechnol.* , vol. 141, no. 1–2, pp. 84–90, Apr. 2009.
ATF6c	N.D.	>2.0	N.D.	EB27 (CHO-K1)	Transient(EBNA1)	CD CHO+6mM L-glutamine +400ug/mL Hygromycin B	No	No	Transient	No	UPR	Human	titrated	L. P. Pybus, G. Dean, N. R. West, A. Smith, O. Daramola, R. Field, S. J. Wilkinson, and D. C. James, "Model-directed engineering of 'difficult-to-express' monoclonal antibody production by Chinese hamster ovary cells.," *Biotechnol. Bioeng.* , vol. 111, no. 2, pp. 372–85, Feb. 2014.
BiP	N.D.	>1.5	N.D.	EB27 (CHO-K1)	Transient(EBNA1)	CD CHO+6mM L-glutamine + 400ug/mL Hygromycin B	No	No	Transient	No	Chaperone	Human	titrated	L. P. Pybus, G. Dean, N. R. West, A. Smith, O. Daramola, R. Field, S. J. Wilkinson, and D. C. James, "Model-directed engineering of 'difficult-to-express' monoclonal antibody production by Chinese hamster ovary cells.," *Biotechnol. Bioeng.* , vol. 111, no. 2, pp. 372–85, Feb. 2014.
mTor	anti-integrin	4	50	DUXB11	Stable	ChoMaster HTS + 10% KOSR	?	No	Stable	No	Metabolic sensor	Human		I. A. J. Dreesen and M. Fussenegger, "Ectopic expression of human mTOR increases viability, robustness, cell size, proliferation, and antibody production of chinese hamster ovary cells," *Biotechnol. Bioeng.* , vol. 108, pp. 853–866, 2011.

Supplementary Table S4: Candidate effector gene panel

Gene name MM	alias	Gene name CHO	Uniprot MM ID	Function	LogFC None vs IgG	logCPM	LR	PValue	FDR	MM Gene ID	bp	Seq ID
Casq1	Calsequestrin-1	LOC100762901	O09165	Calcium-binding protein	2.98	1.60	46.99	7.14E-12	1.29E-10	12372	1875	NM_009813.2
Map3k5	Ask1	Map3k5	O35099	Kinase, component of the MAP kinase signal transduction pathway	7.14	3.94	366.72	9.69E-82	1.49E-79	26408	4143	NM_008580.4
Gosr2	Gosr2	Gosr2	O35166	Protein transport, TGN	-1.42	4.54	84.77	3.36E-20		56494	639	NM_019650.3
Tcp1	TCP1	Tcp1	P11983	Molecular chaperone	-1.61	8.48	117.34	2.42E-27		21454	1671	NM_013686.4 /NM_001290712.1
Scd1	SCD1	LOC100770337	P13516	Lipid and fatty acid biosynthesis	1.48	10.31	34.73	3.79E-09	8.51E-05	20249	1068	NM_009127.4
Cryab	Cryab, Crya2	LOC100772623	P23927	Chaperone, protein folding, prevent aggregation	3.03	2.28	78.39	8.47E-19	1.92E-17			NM_009964.3/ NM_001289785.1/ NM_001289784.1/ NM_001289782.1
Tbp	TBP	LOC100751579	P29037	General transcription factor	-1.21	5.49	75.46	3.74E-18	8.12E-17	21374	951	NM_013684.3
Mafb	Mafb, Maf1	LOC100751374	P54841	Acts as a transcriptional activator or repressor	5.27	-1.33	12.92	3.25E-04	1.46E-03	16658	972	NM_010658.3
Epas1	HRF	Epas1	P97481	Transcription factor involved in the induction of oxygen regulated genes	-7.21	1.87	262.59	4.69E-59	1.33E-55	13819	2625	NM_010137.3
Cebpd	CEBPγ,Crp3	LOC100771036	Q00322	Transcriptional activator	1.60	4.17	70.83	3.90E-17	7.92E-16	12609	807	NM_007679.4
Nfe2	NFE1-p45	Nfe2	Q07279	Transcription factor	5.84	0.98	61.89	3.63E-15	6.41E-14	18022	1122	NM_008685.2
Creb3l2	CREB3L2, Bbf2h7	Creb3l2	Q8BH52	Transcription factor, response to unfolded protein and ER stress	4.08	-1.99	6.24	1.25E-02	3.39E-02	208647	1566	NM_178661.4
Creb5	CREB5A, crebpa	LOC100759947	Q8K1L0	cAMP response element and activates transcription	9.23	2.02	147.48	6.17E-34	8.74E-101	231991	1074	NM_172728.2
Ero1lb	ERO1-Lβ	Ero1lb	Q8R2E9	Oxidoreductase	1.29	5.13	55.18	1.10E-13	1.71E-12	67475	1404	NM_026184.2
Sorcs3	Sorcs3	Sorcs3	Q8VI51		7.99	0.73	59.64	1.14E-14	1.94E-13	66673	3660	NM_025696.3
Ppp2r2d	Ppp2r2d	Ppp2r2d	Q925E7	PP2 phosphatase	-1.68	5.67	153.47	3.02E-35	1.54E-33	52432	1362	NM_026391.2
Sec11c	Sec11c, Spc21	LOC100769695	Q9D8V7	Signal peptide processing	1.68	4.93	46.98	7.19E-12	9.48E-11	66286	579	NM_025468.2
Eif4e3	Eif4e3	Eif4e3	Q9DBB5	Translation initiation factor activity	-7.89	2.90	476.56	1.20E-105	3.74E-07	66892	624	NM_025829.4
Vcp	Vcp, p97	N/A	Q01853		N/A	N/A	N/A	N/A	N/A	269523	2421	NM_009503.4
Bicd1	BICD1	LOC100762966	Q8BR07	Transcription factor, response to unfolded protein and ER stress	2.93	3.21	108.09	2.56E-25	1.67E-23			NM_009753.4/ NM_001112796.2