The Application and Validation of High-Throughput Methods in Chinese Hamster Ovary Cell Line Development

Linas Tamošaitis

A thesis submitted to the University of Kent for the degree of Doctor of Philosophy
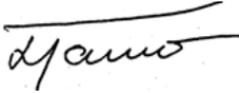
University of Kent

University of Kent, Faculty of Sciences

2019

Declaration

No part of this Thesis has been submitted in support of an application for any degree or other qualification of the University of Kent or any other University or Institute of learning.

Linas Tamošaitis   30th December, 2019

Preface and Acknowledgments

This doctoral thesis is quite a curious thing. In one way this is the culmination of life's work of preparation of growth, it is the product of a person who has spent around a decade of his life in the cloister of academia. At the same time, it is an infinitesimal contribution to the vast collective body of knowledge that is science and a significant contribution to the collective experience of one person's journey through life. It is a rite of passage stating that you are ready to join the ranks of a people who have dedicated a part of themselves to the study of the world, to illuminating the alluring darkness of the cosmos in an attempt to give it form and structure. I believe this metaphor is apt because the journey felt a lot like walking in the dark, slowly and carefully mapping unknown terrain. This work is the result of a lot of failures - of experiments not working, hunches leading down dead ends and theories being proven wrong punctuated by tantalizing moments of understanding and comprehension driving you forward. It is not a journey undertaken alone and even though my name is on the title of this body of text, it is as much a product of the people who have supported me in my personal and professional environments. This section is dedicated to those people.

The first acknowledgement is dedicated to my supervisor prof. **Christopher Mark Smales**. I definitely know at least a few people who owe getting through their PhDs to Mark's patience and support. Mark has always entertained my odd ideas and tangents, but also provided guidance and clarity in moments of need. He has never told me I can't pursue an avenue of research that I wanted or told me something can't be done in an environment that can at times be very discouraging. I will always appreciate the trust that I was given to drive my project and pursue my research independently and the input given in those moments where I felt stuck and did not see a path forward. It has made me more confident in my own abilities and more comfortable in the shoes as a researcher. I am also grateful for the invaluable opportunities in education beyond the academic that I had through you – my time in SynEARC (also huge thanks for **Chieh**, for all his help organizing this!), the course in scientific communication, the conferences and collaborations that you helped me secure. I will deeply miss our talks where I felt I could speak my mind and your insights into the intersection of academia and industry.

My industrial mentor, **Fay Saunders**, has been amazingly supportive and patient with this project. Industrial collaborations can sometimes get very complicated and PhD students sometimes fall through the cracks of beurocracy. My experience at Fujifilm Diosynth has been nothing but positive and I consider myself very lucky to have had the experience and insight of working at contract manufacturing company.

Troy Lionberger and **Phillip Elms** at Berkley Lights have also been instrumental into making a big part of this thesis happem. This collaboration happened through pure chance, however, visiting and working with you at San Francisco have been some of the best experiences of my academic career. I am immensely grateful and lucky to have worked with such passionate people and to have had access to your pioneering technology.

I would also like to thank the The Marie Skłodowska-Curie actions that operate under the EU Horizon 2020 framework for believing in this project and providing access to funding and supporting international scientific collaboration and mobility. Without this opportunity I'm not sure I would've been able to take my research career to the international stage.

The laboratory has been a very fun and relaxed environment to work in and a lot of people have left an imprint on me in that time. **Theo** and **Davide** have been my comrades in arms in this project and I was very grateful to have you at my side through this journey. I will miss **Theo's** critical insights and I will miss **Davide's** opinions on the European Union, graphic design and the chats we had on life in general. I will not miss **Davide's** minion collection, but that liquorice candy will probably be in my possession for a while. I will also miss the talks I had with **James**, as you were one of the few people in the department who would talk shop with me in the late hours of the day. Also, in all the trips we had together, we ended up talking almost until morning. I hope **Tanya** forgives me for me causing such a mess all the time around the lab – breaking printers and triggering alarms, forgetting to put on lab coats and having to worry about my tannin mugs. I'd like to thank **Emma** for her feedback and caring about the quality of my work enough to tell me that things can be done in better ways.

Rokas deserves a special thanks for being the person who showed me the advertisement for the PhD position in the first place and for being a reliable friend who I could always approach to discuss anything and share a cigarette over a break. This journey would've been much more difficult without the regular interactions we had to ground and deconstruct the experience. You were one of the few good friends I made in Canterbury and I'm sure our friendship does not end with this chapter of my life. **Darija** deserves a mention next to him as she was both his student and had to take on the mantle of my smoke-buddy when Rokas had to depart. It is incredible how much we had opened up about in such a short amount of time and I appreciate the trust you had in me.

Justas and Kamilė - It's hard to find other people who have spent so much time challenging my opinions and like you. In a way we are kindred spirits that share a passion for life, debate and celebration. You have spent a lot of time listening to me sharing my existential crises about the PhD and life in general and I know that I can always come to you for an honest and straightforward opinion. I hope we never lose this connection and that we will have each other's backs.

My other professional extended family was the **eCHO network**. We spent so much of time together trying to collectively understand the experience that is undertaking a PhD. All of you were extremely supportive and warm-hearted and it's hard to imagine that we were such a randomly assembled group of people in the beginning and by the end of the journey we felt like a family. When I started this project, I dreamt of being part of a multicultural and varied group of people and I got my wish. I doubt that any other Marie-Curie ITN will have the teambuilding experiences that we did between our communal living arrangements at Hillerød, the hike to the Schneealpenhaus our escapades in Dublin and scouting the seaside at the cliffs of Dover? Also, **Wolfgang** deserves a special mention for his invaluable and brutal lessons in public speaking.

My non-professional Canterbury extended family get a shoutout. You guys accepted me and you are the sole reason why Canterbury grew on me as a place to live. For someone who left his country and his parents' house to live on his own, building a new community from scratch and making friends was very difficult. You guys kept me sane and accepted me at my wackiest and quirkiest. **Matthew Watkins** deserves a special mention as your hands were behind a lot of the events that served as focal points for this community. Without your efforts, I'm not sure this community would exist in the form that it does now and the Canterbury music scene would be much less magical.

Lucas & Poppy, you were the best housemates I have ever had. It's so sad that this living arrangement only manifested in the last year of our time in Canterbury. So much has already been said both in person and in letter form that I don't have much else to add at this point. I hope our paths cross again. Love you guys! **Cameron and Will** – thanks for introducing me to drum meditation! **Johannes** – you were probably our most frequent guest in the house and you always brought good times with you. **Oli** – thanks for taking over as the premier community host after we left and for taking Ixcha. Will miss your conspiracy theories and the sounds of your saxophone and throat singing. **Nikita** – Our friendship was short, but your disco vibes pack a punch. A friendship forged over pickles is eternal. **Jemima** – it is unfortunate that our time together ended the way it did, but if you ever read this, know that I bear you no ill will. You were there in a critical point in my life and your support and company made all the difference. The time we spent together was very healing and magical, and I hope sometime in the future we will still keep in touch. Keep on spreading that fairy magic.

Čiompiks – these people have been with me for 10 years now and some of them for 20. They have been with me throughout my academic journey and supported me all the way through. These are the people that have seen multiple incarnations of my personality and have been there through multiple stages of my life. You guys form the foundation of my support group. Even when I left to live abroad, I discovered that friendships don't disappear just because of distance. Not everyone is so lucky to still keep in touch with

their friends after all this time, and I realize that I am blessed in this regard. Listing every one of you here would take up way too much space, but since you guys have had acknowledgements in my Bachelors and Masters theses, I'm sure you won't mind. **Kostas, Edvinas and Kamilė**, I am sorry I couldn't find a picture with you in it, but rest assured you aren't forgotten. When I left home, I said that

distance is not a factor for friendship and that promise holds true for me to this day. Thanks **Vidmantas** and **Kamilė** for the opportunity to put that theory to the test when I had buy a sudden 6 hour journey home back to Lithuania for an emergency wedding.

Simona – our story probably and the content we generated could probably fill another thesis with letters, photos and stories. The relationship we had was a formative one which has significantly contributed to me understanding myself and integrating my shadow. You were there for me more than any one person has over the course of this PhD, and I'm not sure what kind of person would I be at this point and how much harder my Canterbury journey would have been without you by my side. Also, I have you to thank for teaching me photography, which has stuck with me through this day and has become an indispensable part of how I view my life and my surroundings.

Last but not least, my family has always been there for me. We may be a family of few words but we always have supported each other through thick and thin. You have never questioned my career choices or tried to put me off this challenging path and I know that this achievement will mean the world for you as it does for me. I know that no matter what happens the is always somewhere I can come home to and be safe. I am lucky to have you.

Abstract

Chinese hamster ovary (CHO) cells are the main production platform for biotherapeutic proteins. In this thesis, two areas were identified and investigated with improving recombinant protein production from CHO cells and cell line development; transcriptomics and technologies for industrial cell line development (CLD).

Transcriptomics wise, publicly available data were identified and analysed in order to identify 'common' transcriptomic signatures of cell growth or productivity in order to devise novel cell line engineering strategies. From the literature, 19 different transcriptomic datasets were aggregated that explored the differences between high productivity and fast growth phenotypes. Here, we proceeded to analyse the data in terms of the two simplest dimensions – the frequencies of genes appearing across these data sets and the concordance (the arithmetic mean of expression values) with regard to cell growth (μ) and productivity (Qp). By mapping out the contributing genes it was possible to construct a transcriptomic 'fingerprint' of a high-performing cell line. After identifying the most common and concordant genes, those genes that had a frequency of two or more were analysed using a pathway enrichment algorithm. From this it was identified that the cell cycle and lysosome pathways are significant targets for cell line engineering. To our knowledge, this effort is the first of its kind within CHO transcriptomics.

CLD involves labour and resource intensive cloning out a genetically diverse pool of cells engineered to produce the protein of interest. We sought to analyse a new single cell analysis methodology (Berkley Lights Beacon, BLB) against an industrial ClonePix™ 2 CLD process adapted from FUJIFILM Diosynth Biotechnologies (FDB). We found that there were no statistically significant differences between cell groups generated from the BLB or ClonePix™ 2 processes. Using the Beacon® system, it was possible to predict 3 out of the top 5 producing clones for both Etanercept and Blosozumab. Within the standard ClonePix™ 2 CLD group of cell lines, predictions were most accurate from 24-well plate fed-batch and TubeSpin® batch culture ranks. Further, using the BLB, the time from recovery from transfection to cultures that were ambr® ready was reduced from 65 days to 42 days. Based on the findings of this research it is proposed that the Beacon® is an attractive and powerful new tool in industrial cell line development efforts. To the authors knowledge, this is the first in depth work validating a next-generation CLD process in such detail.

Contents

List of Abbreviations

ADCC - Antibody-dependent cell-mediated cytotoxicity

BEVS - Baculovirus expression vector system

BLI – Berkley Light Instruments

CHO – Chinese hamster ovary

CLD – Cell line development

DHFR – Dihydrofolate reductase

DNA – Deoxyribonucleic acid

DOE – Design of experiment

DTE – Difficult to express

EPO – Erythropoetin

FDA – U.S. Food and Drug Administration

FDB – FUJIFILM Diosynth Biotechnologies

FOG – Fed-batch overgrow

FRET – Fluorescence resonance energy transfer

GG – Growth group

GS – Glutamine synthetase

HEK – Human embryonic kidney

IgG – Immunoglobulin

kDa – Kilodalton

KW – Kruskal-Wallis test

LPS – Lipopolysaccharides

mAb – Monoclonal antibody

miR, miRNA – micro RNA

PG – Productivity group

PTM – Post-translational modification

PTM – Posttranslational modifications

RNA – Ribonucleic acid

siRNA – silencing RNA

TNF – Tumour necrosis factor

List of figures

List of tables

1 INTRODUCTION

1.1 THE EMERGENCE AND GROWTH OF THE MARKET FOR BIOPHARMACEUTICALS

What is a biopharmaceutical? The term describes a diverse group of biologically derived multi-amino acid or polypeptide/protein based molecules (Leader *et al.*, 2008):

1. "protein therapeutics with enzymatic or regulatory activity (e.g. replacement therapies such as insulin, growth hormone, Factor IX, b-glucocerebrosidase).

2. protein therapeutics with special targeting activity (e.g. mAbs or other binding proteins, including FcFPs, that bind specific therapeutic targets. Two examples are the anti-tumour necrosis factor (TNF)-a mAb, Remicade®, and the anti-TNF-a/b FcFP, Enbrel®).

3. protein-based prophylactic vaccines (e.g. human papillomavirus (HPV) vaccine made using virus-like particles containing HPV major capsid protein L1).

4. protein diagnostics (e.g. biomarkers such as glucagon, and imaging agents such as technetium- or indium-conjugated antibodies)".

The global biopharmaceuticals market was valued at USD $237,250.8 million in 2018 and it is estimated it will have a value of USD $388,997.3 million by 2024 according to a recent market report (Mordor Intelligence, 2018). The trends of global R&D expenditure on biopharmaceutical development are presented in Figure 1.1.1.

Figure 1.1.1. Worldwide Total Pharmaceutical R&D Spend in 2010-2024

CAGR; compound annual growth rate. Adapted from (EvaluatePharma, 2019)*.*

Since 2013, the amount of investment in the area of biopharmaceutical development has almost tripled. This increase in funding correlates with an increasing amount of new biobased drugs approved by regulatory agencies. There has been a steady incline of such drug approvals since 1989 (

). Up to 1989, only 9 such drugs obtained approval status whilst in the period 2010-2014 that number rose to 60 and to 112 in the period 2015-July 2018 (Walsh, 2018).

Products approved over the four and a half years from 2015-2018 include 68 monoclonal antibodies (mAbs), 23 hormones, 16 clotting factors, 9 enzymes, 7 vaccines, 5 nucleic acid-based products and 4 engineered cell-based products (Walsh, 2018). While this group is dominated by monoclonal antibodies, other types of recombinant biotherapeutic have been generating interest as new biotechnologies and new format protein-based biologics are developed for industrial drug production. This trend is depicted in Figure 1.1.3, showing how other, non-mAb recombinant proteins have been increasing in FDA approvals since 2013, reaching more than half the number of monoclonal biotherapeutics.

Figure 1.1.2. The numbers of biotherapeutic product approvals in the US, Canada and Europe since 1989
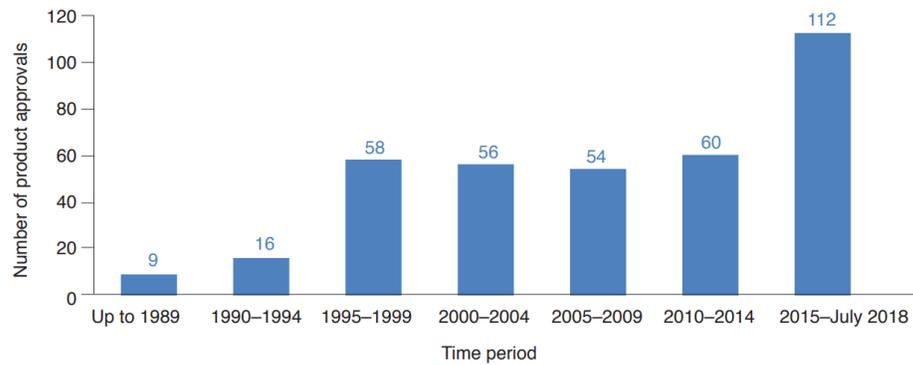
Figure reproduced from (Walsh, 2018).

Figure 1.1.3. Number of biotherapeutics approved by the FDA each year
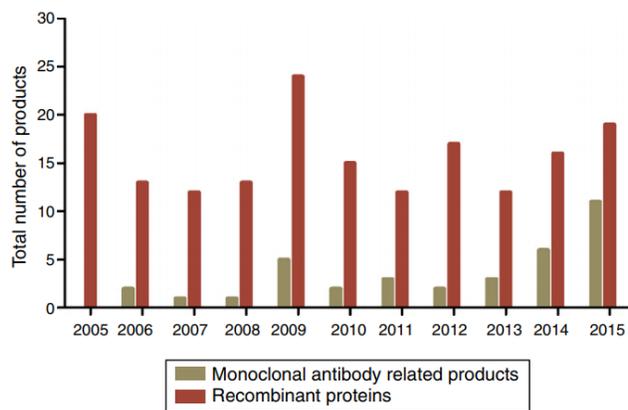
Grouped by monoclonal antibodies and other recombinant proteins. Figure reproduced from (Jozala *et al.*, 2016).

1.2 DIVERSITY OF ANTIBODY THERAPEUTICS

As described in section 1.1 above, antibody-based therapeutics currently dominate the biopharmaceutical landscape. The first monoclonal antibodies were isolated from murine B lymphocytes in the mid-1970s (Köhler and Milstein, 1975). Antibodies are proteins that immune cells secrete that have a specificity to bind a target molecule (antigen). The basic monomeric Ig (there are multiple classes of Ig molecule, IgA, IgD, IgE, IgG, IgM) assembled antibody complex is composed of two heterodimeric protein chains; heavy (50 kDa) and light (κ or λ, 25 kDa). The C-terminus of the heavy chain is known as the Fc (fragment crystalline) region, which can be recognised by other immune cells. The light chain and the N-terminus of the heavy chain comprise the Fab region which contains the variable and hypervariable protein sequences that are responsible for recognizing the target antigen. These variable sites are distinguished from the constant regions. There is one present on the light chain and three on the heavy chain. Different types of antibodies are present in the human body as outlined above, but the most commonly used Ig class for recombinant biotherapeutic application is IgG which represents approximately 75% of serum antibodies (Vidarsson *et al.*, 2014b). The IgG class has 4 different subtypes, IgG1-4, which have over 90% sequence homology between them. Most of the variation is observed at the hinge region and the N-terminal CH2 domain. Such differences mediate the effector functions of the antibody as it contains the binding site for C1q and IgG-Fc receptors that effector cells possess (Vidarsson *et al.*, 2014a). The other antibody classes (Figure 1.2.1) all have different roles to play in the immune system and there is no antigenic cross-reaction between them. IgM are pentameric antibodies held together by disulphide bonds. They are the first response antibodies produced by B cells against an antigen. Together there 10 antigen-binding sites, however they are of lower affinity then other classes. They are, however, most efficient at the agglutination, causing the antigen to clump together. IgA is the most abundant antibody found in human secretions and not much of it is present in the human serum. IgAs are dimeric and the final molecule is associated with a protein that enables it to be transported across epithelial cells into secretion fluids. IgDs are also almost absent from the human serum and they are mostly found on the surface of B cells. IgEs are similar to IgGs in structure, however they possess an extra CH domain enabling them to bind to basophils and mast cells. They are the primary antibodies involved in mediating allergic reaction responses and overabundance of IgE causes allergic hypersensitivity. They are also involved in immunogenic responses against parasitic infections.

As antibodies started gaining more interest and application in the biotherapeutic market they

have become prime targets for protein engineering. Some of the simplest antibody based biotherapeutics are Fab fragments. These antibodies have had their Fc regions deleted. This technology has already brought several biotherapeutics to market such as abciximab, an anti-gpIIb/IIIa Fab fragment for prevention of blot clots in angioplasty, and certolizumab pegol, a PEGylated anti-TNFα Fab fragment for Crohn's disease and rheumatoid arthritis. Another type of antibody fragment that has seen commercial success is the scFv (single chain variable fragment). Here, the heavy and light chain variable domains are connected with a flexible linker. Unlike Fabs, scFvs lack the constant CH1 and CL domains. Brolucizumab, an scFv against VEGF-A for neovascular AMD is currently in phase 3 clinical trials as is Otlertuzumab, the scFv targeting CD37 for chronic lymphocytic leukemia. The lack of an Fc domain has advantages and disadvantages. It simplifies the molecule to allow expression in bacterial systems and their smaller size allows steric access to cryptic epitopes. It also means that these fragments lack immune cell activation typically mediated by the Fc domain. Unfortunately, scFvs are prone to low thermostability and aggregation increasing their risk and decreasing their half-lives (Bates and Power, 2019).

Figure 1.2.1. The structural differences between Ig classes M, G and A.
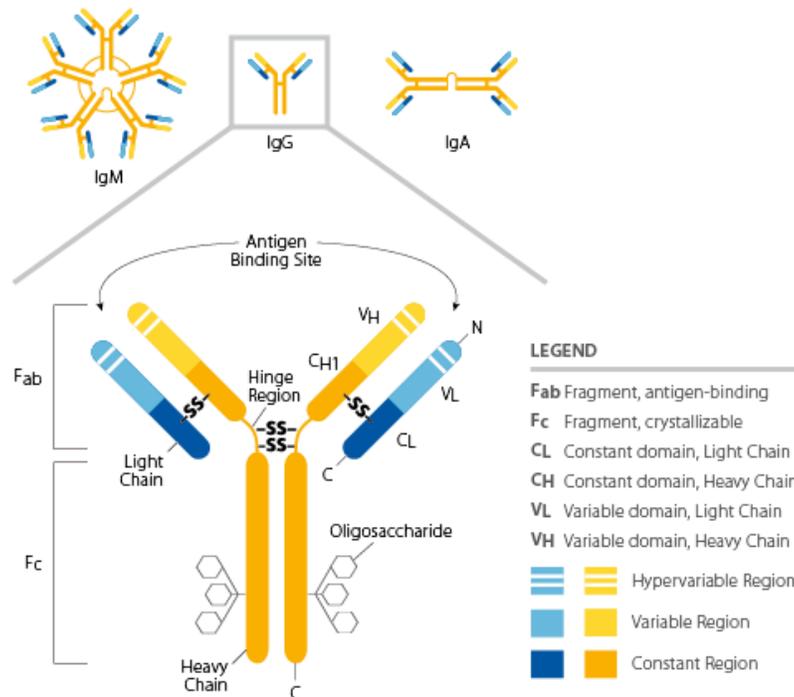
Nanobodies® are an antibody class that were created from camelid IgG heavy-chain VHH domains. They are very compact with a MW of around 12-15 kDa. Unlike their fragment counterparts they resist wide pH ranges and high temperatures while having good solubility. Although they are not of human origin, they are not particularly immunogenic due to a high similarity to the human VH3 family (Harmsen and De Haard, 2007). They can be arranged in tandem in order to create Nanobodies® with multiple binding targets. Caplacizumab, a bivalent humanised Nanobody® which targets the von Willebrand factor against aTTP is currently in phase 3 clinical trials.

Diabodies are bivalent dimers that consist of two chains each containing a VH and VL domain. The domains on each chain are connected with a G4S linker that prevents intrachain dimerization leading to interchain head-to-tail dimerization. DART®s are an improvement over traditional diabody technology as they include an interdomain disulphide bond for increased stability (Kipriyanov *et al.*, 2003). MacroGenics, the company that owns the rights to DART® technology, currently have 4 DART®s in phase I clinical trials for oncology, autoimmune disease and HIV infection (https://www.macrogenics.com/pipeline/).

Non-fragment bispecific antibodies have drawn great interest in recent years (Wang *et al.*, 2019), but their development is challenging due to their stability and difficulty to express. IgG-scFv types have VH2 and VL2 fragments fused to the Fc region. Fab-scFv-Fc type bispecifics switch one Fab arm exchanged for an scFv. Both of these variants have drug candidates in clinical trials. A broad overview of the types of novel antibody-based therapeutics in clinical development has been provided by (Sheridan, 2017).

1.3 HOSTS FOR RECOMBINANT PRODUCTION OF BIOPHARMACEUTICALS

1.3.1 Overview of biopharmaceutical production

Most biopharmaceuticals are made using a recombinant gene editing technology where an array of different methodologies and host expression cell lines or organisms are utilised as cell factories by inserting the required foreign DNA coding for the gene(s) of interest in order to produce large amounts of the target recombinant protein. Before the development of recombinant DNA technology, classical genetic studies used to rely on methods that could cause or discover spontaneous mutations. These included the use of ionizing radiation, screening large numbers of organisms and exposure to various mutagens or strong selective pressure. Whilst this allowed development of alternative cell systems, it did not allow for the introduction of exogenous genetic material such as that required to produce an exogenous recombinant protein for therapeutic purposes. The advent of DNA and the polymerase chain reaction (PCR) technologies in the 1970s changed this paradigm, however. The discovery of restriction enzymes allowed scientists to digest DNA at specific sites, transfer and rearrange this by ligation into new sites to suit the needs of mankind (Cohen *et al.*, 1973). This raised a great amount of concern both within the scientific and public communities and in 1976 the NIH published the first guidelines for using the technology "NIH Guidelines for Research Involving Recombinant DNA Molecules". One of the first real beneficial applications of this technology emerged when *E. coli* was developed into a production platform for recombinant insulin and recombinant growth hormone production which ushered in the beginning of the era of recombinant biotherapeutics.

As the collective skill in biotechnology grew more proficient, alongside the discovery/development of monoclonal antibody generation and isolation, researchers harnessed the power of our own immune systems to combat various diseases. The first monoclonal antibody to be approved for medicinal use was in 1985 by the FDA, Muromonab-CD3. It was also known by its trade name Orthoclone OKT3 (Janssen-Cilag). Interestingly, the name Muromonab was

created before the WHO established the monoclonal antibody nomenclature, so its name means <u>mur</u>ine <u>mono</u>clonal <u>anti</u>body targeting <u>CD3.</u> This antibody was made using murine hybridoma technology, where a cell line is established from a single B-cell that was merged with an immortal murine myeloma cell. While it was hugely successful in treating acute renal failure, it was not without its adverse reactions, especially the possibility of an immune response resulting from its murine origin. There was also a significant ethical concern about the use of hybridoma technology as the immunization of animals causes them measurable harm and discomfort. These drawbacks, particularly the potential for immunogenicity, resulted in a concerted effort to develop fully humanized and recombinant antibodies (Gomes *et al.*, 2016).

Humanization of an antibody is a complicated process that requires splicing parts of antibodies of different origins. It was initially an imperfect process that was fraught with uncertainty and resource costs. The burden of making such antibodies was lifted with the arrival of high-throughput screening technologies like phage display and yeast display, which allowed these respective organisms to display antibodies from their exterior into the environment allowing the screening of entire libraries of antibody sequences (Bazan *et al.*, 2012). The first fully human recombinant monoclonal antibody to obtain FDA approval was a product from phage display technologies. It was named Adalimumab or Humira (<u>hu</u>man <u>m</u>onoclonal antibody <u>in</u> <u>r</u>heumatoid <u>a</u>rthritis, Abbott Laboratories).

Nowadays nearly all biotherapeutics in development are of recombinant origin. Most are made in suspension cell lines (Walsh, 2018), which involves a technologically challenging production method that consists of inserting the gene of interest into a host cell, selecting a clone, optimizing the conditions for production and scaling up the production of the material for the desired quality. This part of the development process is known as upstream bioprocessing. Downstream bioprocessing is classified as everything that happens after the harvest of the biomass grown in the production bioreactor which includes the removal of the cells from the suspension by centrifugation or filtration, precipitation of the product or purification through chromatography, various viral contaminant removal steps and formulation of the actual drug. This process is depicted Figure 1.3.1.

Figure 1.3.1. A typical pathway for the manufacturing process of a biotherapeutic
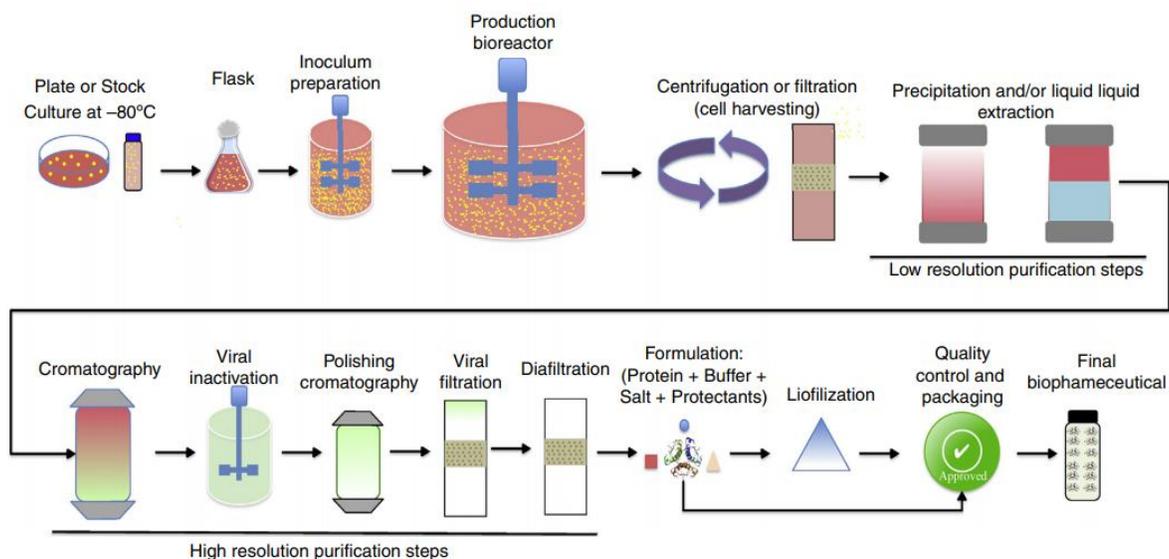
Including the upstream and downstream process. Figure reproduced from (Jozala *et al.,* 2016)*.*

The modern landscape of biotherapeutics production is varied. Because of the diverse range of molecules that are produced, the industry utilizes a variety of host systems to manufacture biotherapeutics with different hosts being more appropriate for the expression of a particular target protein. A short summary of these systems is presented in Table 1.3.1.1 and the following sections explore and describe a number of the most common systems in more detail. It is noted here that when discussing these expression systems the descriptions are with a view to using these to express, at industrial scale, clinically relevant biotherapeutic proteins, not lab-scale proteins for scientific investigation.

Table 1.3.1.1. Summary of the different host cell systems used for expression of recombinant proteins. Adapted from *(Gomes* et al.*, 2016)*.

Host system	Merits	Demerits
Escherichia coli	Easy, Quick, Cheap Rapid growth rate Continuous fermentation capacity	Cannot remove introns Susceptibility to termination signals in foreign DNA Codon bias Lack of PTM Glycosylation is extremely uncommon Inclusion bodies Degradation of proteins Accumulation of endotoxins
Bacillus subtilis	No LPS/endotoxins Cross compatibility with other bacterial plasmids Recombinant protein secretion capacity	Extracellular proteases degrade proteins Plasmid instability Reduction of heterologous proteins

Yeast	Rapid growth rate	Hyperglycosylation of proteins
	Appropriate PTMs	Codon bias
	Safe	Inefficient secretion and intracellular
	No endotoxins	retention
Filamentous fungus	High expression levels	Culturing can be complex
		Lack of knowledge of use and application
Insect cells	High expression levels	Continuous expression not possible
	Appropriate PTMs	Demanding culture conditions
	Safe	Expensive
	Very good for glycoproteins	
Mammalian cells	Proper folding	Expensive
	Appropriate PTMs	Animal virus contamination
	Human glycosylation	Demanding culture conditions
Transgenic plants	Easy, low-cost scaleup	Expression very target-dependent
	Proteins can be localized to different organs	Lack of functional assays
	High expression levels	
Transgenic animals	Proper folding	Long production process
	Appropriate PTMs	Low yield
	Human glycosylation	Most expensive scale-up

PTM – post translational modification.

1.3.2 Bacterial expression systems

Some of the most commonly used expression systems are bacterial systems, with *E. coli* being a workhorse of the biotherapeutic industry for the production of recombinant proteins, especially for proteins that do not require human-like post-translational modifications such as glycosylation. *E. coli* has a rapid growth rate (20-30 min doubling time), cheap media components and an easily manipulatable genome making this one of the easiest and most attractive solutions for recombinant protein manufacture. Further, there is a wealth of information and tools for the optimization of the expression of target proteins of interest in *E. coli*, including the ability to screen entire libraries of protein variants (Jia and Jeon, 2016). However, *E. coli* does have some intrinsic limitations that limit its usefulness as a platform for particular biotherapeutics.

Bacterial systems have much simpler cellular machinery compared to eukaryotes lacking a nucleus, endoplasmic reticulum and a Golgi complex. When expressing proteins of eukaryotic origins some optimization is usually needed because of this and high titres cannot be guaranteed. When complex PTMs or humanised glycosylation profiles are required, bacterial systems can prove inadequate. Because of a lack of eukaryotic folding machinery such as provided by the ER, many eukaryotic proteins are incorrectly folded in *E. coli* and form inclusion bodies, solid intracellular deposits of unfolded protein that accumulate and have to be recovered from the cell and then refolded. Whilst refolding can often be achieved successfully, these protocols are costly in both time and reagents. This can be mitigated by engineering the target protein in terms of

truncation, mutagenesis or fusion with other more soluble proteins. Co-expression with other proteins has proved successful for some proteins. In the case of prion protein production, solubility is greatly enhanced by the co-expression of the prion protein along with human quiescin sulfhydryl oxidase (Abskharon *et al.*, 2012). For proteins requiring disulphide bond formation, recombinant protein production is directed into the periplasmic space in order to prevent inclusion body forming and allow disulphide bond formation. It is noted that *E. coli* may maintain the initiating amino acid N-Formylmethionine, which mammalian cells do not use for initiating translation when using ribosomes in the cytosol. This can affect the immunogenicity or the intended functioning of the target recombinant protein (Jia and Jeon, 2016).

Several different industrial *E. coli* strains are used, each with their different attributes that alleviate particular deficiencies of the system (Huang *et al.*, 2012; Rosano *et al.*, 2019). An example of such an approach is the initiative to reduce acetate production during fermentation. This by-product greatly lowers the pH of the medium, inhibiting further cell growth and recombinant protein yields. There have been successful strategies of mitigating this, for example by deletion of the pstHI operon in strain GJT 001 (Wong *et al.*, 2008). Another recent development in improving *E. coli* as a biotherapeutic production host has been the utilization of the twin arginine translocation system (TAT), which is able to export fully folded proteins, plus co-factor substrates up to 150 kDa in size, into the periplasm (Castiñeiras et al., 2018).

1.3.3   Yeast expression systems

These eukaryotic host systems are partway between the simplicity of the bacterial protein production machinery and the complexity of mammalian cell culture. On one hand, yeasts have a secretory pathway and can undertake folding and assembly of complex human proteins, production is relatively cheap and simple, and some such as *Saccharomyces cerevisiae* have model organism status meaning they have a wealth of tools and knowledge available to be harnessed for protein production. On the other hand, fungi possess non-human glycosylation patterns that can critically alter recombinant protein function and induce unacceptable immunogenicity of said protein. In particular, fungi generate glycans with high terminal mannose that elicit immune responses. *S. cerevisiae* does have a safety profile that goes back for more than 25 years and is an established host for the expression of recombinant proteins, being considered one of two yeasts that are the most economically valuable for production worldwide (the other being *Picha pastoris*).

Yeast expression systems have been used to produce recombinant hepatitis B and Hantavirus

(Antoniukas *et al.,* 2006) viral particles. In 2015 there were 11 approved vaccines against hepatitis B and one against human papillomavirus in the US and Europe made in yeasts, while only two such approvals have been granted using *E. coli* and insect cells at the time. Since then, there have been many examples of yeast recombinant vaccine production in the scientific literature (Bill, 2015).

An advantage of yeasts over their prokaryotic counterparts is their cellular machinery that includes the secretory pathway giving them the ability to correctly fold, assemble and post-translationally modify target recombinant proteins before secreting the recombinant protein into the extracellular medium (Daly and Hearn, 2005). However, few products are currently produced using yeasts for biotherapeutic production due to their tendency to hypermannosylate their target proteins. There have been substantive efforts in order to humanize these glycosylation patterns to reduce immunogenicity and increase efficacy. One such effort introduced a number of mannosidase and glycosyltransferase enzymes across the tree of life coupled with the diversion of the existing glycosylation path from *Och1* which ends in hypermannosylation (Hamilton *et al.*, 2006). Meanwhile, in *S. cerevisiae*, humanized $Man_5GlcNAc_2$ N-linked oligosaccharides, an intermediate of mammalian hybrid- and complex-type oligosaccharides, have been successfully generated by introducing manosidases and disrupting internal O-linked mannosylaiton (Abe *et al.,* 2016). This shifts the cell line engineering goalposts towards making sure that the resulting humanized glycosylation is homogenous on the recombinant protein of interest. The comparisons between human *and S. cerevisiae* glycosylation systems are outlined in Figure 1.3.2. In humans, the core $Man_5GlcNAc_2$ structure undergoes mannose trimming via MnsIA, IB and IC. In the Golgi further glycosylation processing is undertaken via addition of β-1,2-N-acetylglucosamine residues that get capped with β-1,4-Galactose and α-2,3-N-acetylneuraminic acid. In yeasts, however, the $Man_5GlcNAc_2$ core structure gets polymannosylated in the Golgi. A hypothetical humanization pathway is proposed in *P. pastoris* starting from a Δalg3, Δoch1 strain that prevents mannosylation. To introduce human-like glycosylation pathways the introduction of human GnT I, GnT II, GalT and ST enzymes is proposed.

Some filamentous fungi have been drawing attention due to their ability secrete over 100 g/L of proteins into their environment. If this productive capacity could be harnessed to its full potential it would be a very efficient expression system, providing the product was of appropriate quality. There are few examples of such 'non-standard' systems being used to develop biopharmaceuticals. One such example was the utilization of *Trichoderma reesei* to secrete interferon alpha-2b with yields up to 4.5 g/L with an additional 1.8 g/L still bound to the secretion

carrier protein (Landowski *et al.*, 2016).

Figure 1.3.2. A comparison between the N-glycosylation machinery in human, baker's yeast (S. cerevisiae) and P. pastoris with a synthetic pathway for glycan humanization
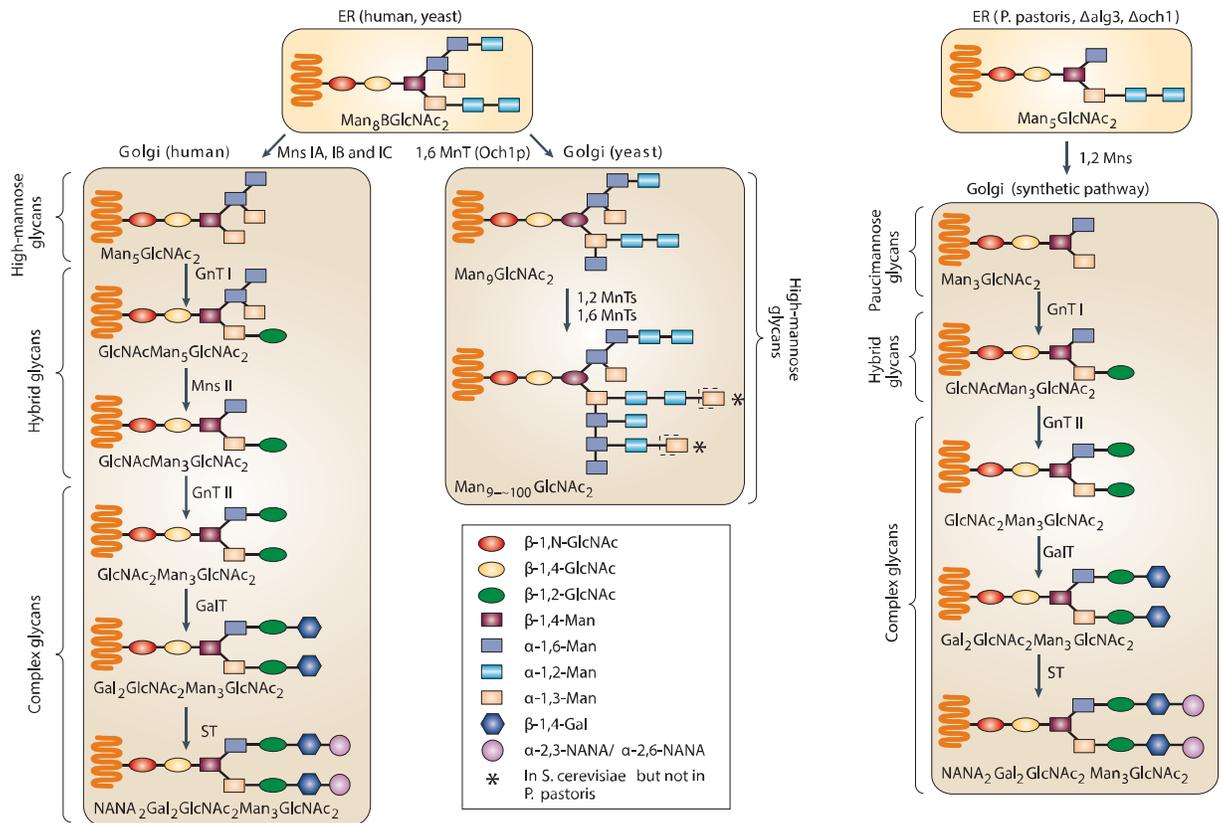
Adapted from (Wildt and Gerngross, 2005)*. ER, endoplasmic reticulum; GalT, galactosyltransferase; GlcNAc, N-acetylglucosamine; GnTI, N-acetylglucosaminyl transferase I; GnTII, N-acetylglucosaminyl transferase II; Man, mannose; MnsII, mannosidase II; MnTs, mannosyltransferase; NANA, N-acetylneuraminic acid; ST, sialyltransferase; Man, manose; Gal, galactose.*

1.3.4   Insect cells as a recombinant protein expression system

Insect cell lines were first established in 1963 and since that time over 500 cell lines have been developed for use across more than 100 insect species (Lynn, 2001). The most common insect expression systems are derived from three species, *Spodoptera frugiperda* (Sf-9 and Sf-21), *Trichoplusia ni* (Hi5), and *Drosophila melanogaster* (S_2). Insect cells are closer to mammalian cells in terms of the cellular machinery than fungi or prokaryotic expression systems. The doubling times, for example, are very similar to immortalized human and CHO cell lines; for S_2 cells the doubling time is ~15 h, Sf9 20-30 h, Hi5 18-24 h (Saarenpää *et al.*, 2015). However, their *N*-glycosylation patterns are different. Both groups share a common glycosylation precursor,

which then diverges; in insects, the terminal *N*-acetylglucosamine is removed and typically glycans contain both 1,3-linked and 1,6-linked core fucose, the former being absent in mammals. Insect cells also lack sialic acid modifications, which are common in human cells (Geisler *et al.*, 2015). These differences are outlined in Figure 1.3.3 along with a synthetic humanized glycan pathway.

Figure 1.3.3. The differences between the glycosylation patterns of insect cells and mammalian cells.
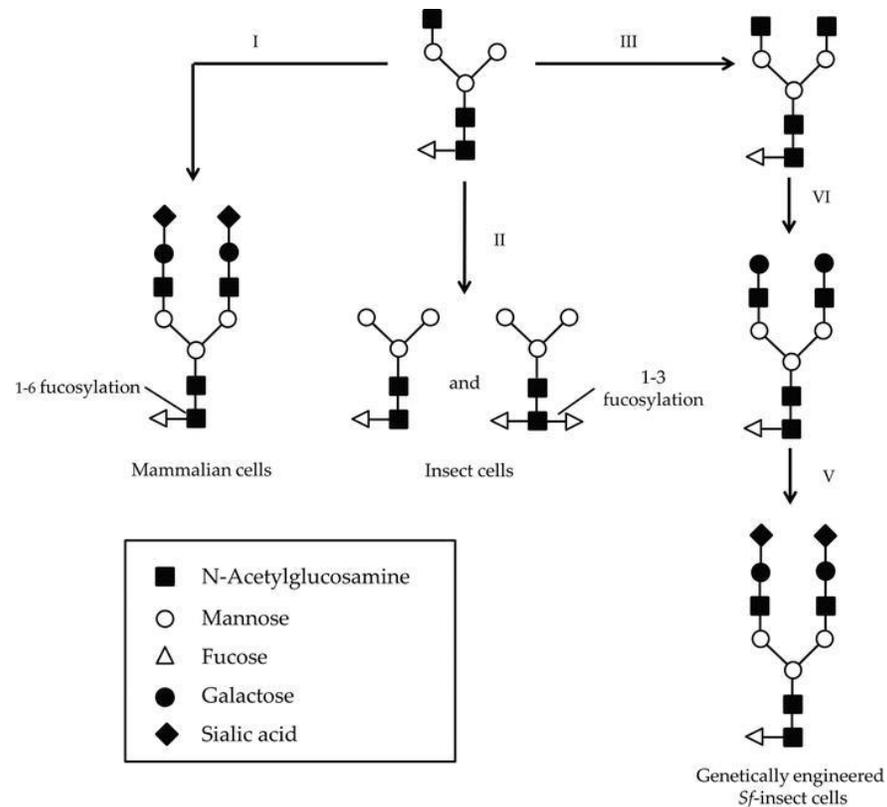
Adapted from (Zitzmann *et al.,* 2017)*.*

The development of the BEVS (baculovirus expression vector system) was one of the main advances that contributed to the growing popularity of insect cell lines as recombinant protein expression systems. Baculovirus is specific to insect cells and poses no threat to animals and is therefore considered safe. Baculoviruses have a large genome, are easy to edit and proliferate in insect cell culture. Due to the lytic characteristics of this expression system, there are short turnaround times reducing the need for the development of stable cell lines. Insect cells now have a ~1% share of all FDA/EMA approved protein therapeutics for human use. Among these is a vaccine against cervical cancer, Ceravix and an immunotherapy against prostate cancer - Provenge (Yee *et al.*, 2018).

The main difference between viral and non-viral insect expression systems is that when using stable plasmid-based expression there is no lysis of cells due to viral proliferation. Cell lysis induces spill over of cellular content into the surrounding media, potentially affecting product quality due to proteases and making purification more arduous. The lytic capacities of the virus ensure short culture times, however, there may not be enough time for protein to be properly processed and modified resulting in greater heterogeneity. Though great efforts have been made to reduce baculoviral lytic capacities, for some biotherapeutics expression in stable cell lines is preferable (Yee *et al.*, 2018). Stable transfection enables different modes of bioprocessing such as fed-batch and perfusion culturing, which can drastically increase protein yields as these approaches provide an environment that facilitates cultures achieving much higher cell concentrations. This also means that through the use of stable clonal cell lines and strict bioprocessing protocols the user can achieve a greater level of reproducibility between runs, a criterion very important to biopharmaceutical production (Zitzmann *et al.*, 2017).

1.3.5 Plants and plant cells as biopharmaceutical cell factories

While the idea of producing biopharmaceuticals in plants has been around for decades, research into plant expression systems has not yet materialized into large scale industrial uptake of the platform. It certainly has nothing to do with variety as there are both whole plant, plant tissue and suspension cell culture systems available for use with a variety of expression methods (Kermode and Jiang, 2018). There are several attractive qualities about plants that make them worth considering for recombinant protein manufacture. In essence, plant systems combine the benefits of inexpensive media costs of microbial systems with the ability to produce complex proteins like mammalian cells. Transient expression platforms based on vacuum infiltration with *Agrobacterium tumefaciens* or viral vectors enable short production time scales with a much higher scale-up ceiling compared to mammalian cells (Buyel and Fischer, 2015; Chen and Davis, 2016). Plant-specific glycans, while different to human glycans, do not usually cause adverse effects in humans, though significant progress has been made to humanize the glycosylation pathways in plant expression systems (Schoberer and Strasser, 2018).

Based on these criteria, the use of plant systems has consolidated in several distinct applications. One such application is the production of rapid response biotherapeutics such as vaccines during outbreaks, allowing rapid and small scale manufacture of individual medicines for patients where scale-up is not economically viable. Plant-based systems are also attractive for the production of high volume/low margin bulk proteins. Plant systems are the only expression

system that possess a safe oral administration route without complicated downstream processing of the plant matrix (Kermode and Jiang, 2018).

An analysis of the literature reveals a changing landscape of utilisation of plant-based expression systems to produce recombinant proteins. Between 2010 and 2016 transient systems based on *N. benthamiana* were the most favoured platforms for vaccine and antibody production while suspension-based moss and carrot systems were more focused on enzyme productions. Although there are only a few recombinant therapeutic products approved for use derived from plants, increasing numbers of reports of pre-clinical and clinical research highlight a drive towards the commercialization of these platforms (Kermode and Jiang, 2018). One such drug is recombinant glucocerebrosidase (prGCD) marketed as Elelyso (Protalix Biotherapeutics) (Zimran *et al.*, 2018) an enzyme replacement therapy approved for the treatment of Gaucher disease and produced using cultured carrot cells. In 2014 during the West Africa Ebola virus outbreak, the FDA approved ZMapp, a *N. Benthamiana* based transiently expressed cocktail of 3 chimeric monoclonal antibodies (Na *et al.*, 2015). A further plant-based protein comes from Biolex Therapeutics that has released a biobetter version of interferon α2a under the name Locteron. In the case of Locteron and Elelyso the efficacies of these proteins were improved because of the plant glycan profile. The bioavailability of Elelyso was increased due to the absence of sialic acid which increased uptake by macrophages. The equivalent CHO cell-derived biologic Imiglucerase needs to be trimmed *in vitro* to remove the sialic acid residues to achieve a similar potency, adding complexity and costs to the downstream process (Tekoah *et al.*, 2013).

Currently, there are two major disadvantages that are keeping plant-based expression systems from wider adoption in recombinant protein manufacture. The first challenge is productivity. In an optimised CHO cell production process the specific productivity can reach as much as 50-90 pg/cell/day, human secretory plasma cells have been observed to reach an astounding 200-400 pg/cell/day (Hansen *et al.*, 2017). While plant-specific productivity values are seldom found in literature, one example of a suspension cell production of a full-size antibody estimates this productivity around 8 pg/cell/day (Havenith *et al.*, 2014). A second challenge is downstream purification. The plant cellular matrix is much more robust than mammalian cell culture, requiring additional resources spent on the removal of debris and soluble host cell proteins. Plant cell proteases can also be problematic to remove resulting in degradation of the target protein of interest if not removed or inactivated. While there seems to be no reason to believe that these problems cannot be overcome through cell line engineering and downstream process optimization, plants will remain a more niche production platform in the near future until these issues are overcome (Schillberg *et al.*, 2019).

1.3.6 Mammalian cell expression platforms

Mammalian cell expression systems have evolved since their establishment to become the most widely utilized biotherapeutic production platform (Figure 1.3.4.). In the period 2010-2014, mammalian cell lines were the expression system for around 60% of all approved biotherapeutics. Hundreds of different cell lines exist, however the two mainstays of recombinant protein production are the CHO cell, a group of immortalized cell lines derived from Chinese hamster ovarian tissue (Kao and Puck, 1968), and HEK-293 cells, a human embryonic kidney derived cell line that was transformed with fragments of adenovirus type V DNA (Graham *et al.*, 1977) increasing the HEK293 cell growth rates, tolerance of reduced serum concentrations and making it easier to genetically manipulate. Both HEK and CHO based cell lines have been subjected to cell engineering and genome editing to make them compatible with either non-antibiotic based selection platforms or to aid stable cell line generation. The HEK-293T cell line has been genetically engineered to encode the SV40 Large T antigen which enables the episomal replication of plasmids that contain the SV40 origin of replication. Three dominant cell lines have emerged for academic and scientific use; CHO-K1, CHO-S and CHO- DG44. The DG44 cell line has a deletion of the dihydrofolate reductase gene *DHFR* whilst the CHO-K1 and CHO-S hosts are usually used with a glutamate synthetase (GS) selection system and an inhibitor of this (MSX) although now GS gene deletion cells have also been made. There is still debate regarding the precise lineages of these cell lines, however phylogenetic reconstructions using the diversity of SNPs have allowed scientists to establish the genetic similarities between them (Lewis *et al.*, 2013).

Figure 1.3.4. Number of recombinant protein products approved for use as drugs in humans, depending on the type of production platform.
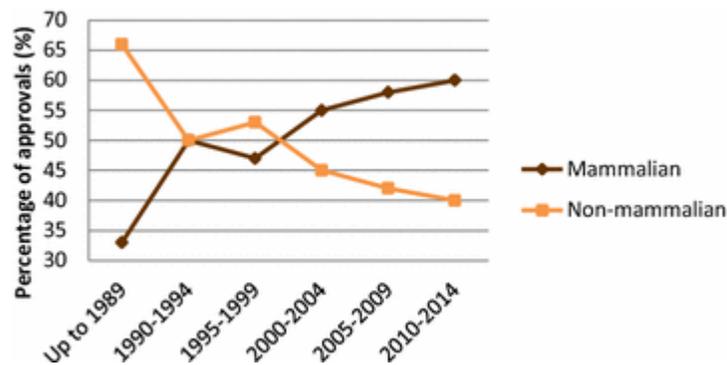
Reproduced from (Sanchez-Garcia *et al.*, 2016)*.*

The ability to utilize these mammalian host cell lines to produce ever increasing amounts of biotherapeutic protein over time has increased dramatically. This is illustrated in Figure 1.3.5 which shows a comparison between a hypothetical unpublished bioprocess in 1986 and a 10 L fed-batch bioprocess from 2004 run by the commercial contract manufacturer, Lonza Biologics. This data is now more than 15 years old and further improvements have been made such that modern in-house CHO cell lines are capable of reaching viable cell concentrations of over 10 x 10^6 cells/ml (Yongky *et al.*, 2019) with titres in excess of 13 g/L (Huang *et al.*, 2010).

Figure 1.3.5. Evolution of the bioprocess capabilities of recombinant protein therapeutics in cultivated mammalian cells from 1986 to 2004.
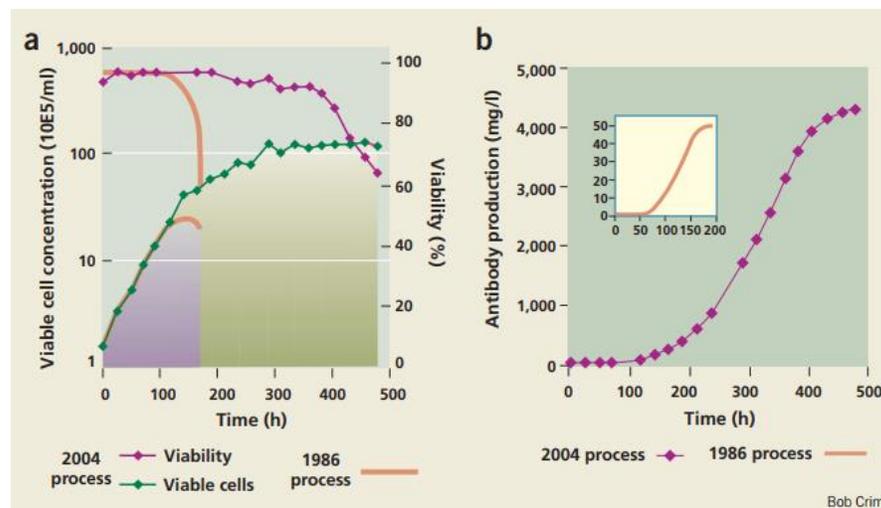
Reproduced from (Wurm, 2004)*. A – Bioreactor viable cell concentration 10^5/mL; B – Bioreactor antibody production capacity, mg/L.*

HEK-293 cells have seen more uptake in gene and cell therapy applications, clinical and basic

research as a model human cell line, while CHO cells appear to be the preferable industrial expression host for stable and high yielding expression of biopharmaceuticals. Both cell lines can be cultured in suspension and adherent format, have comparable doubling times which are less than 24 h and can grow to high cell concentrations. Both cell lines are easily transfected and handled in routine cell culture. The reason why CHO cells are preferred in the industrial settings is because currently bioproduction processes can reach productivities of up to 50 pg/cell/day in fed-batch culture, a number that HEK-293 cell lines cannot currently match. These productivities, and the quality of the protein produced, is enough to offset the fact that CHO cells can produce human-like glycoforms only, although much work has gone into engineering humanized glycan pathways in CHO cell lines (Chung *et al.*, 2012; Tian *et al.*, 2019). Nevertheless, several drugs can be found on the market that are manufactured using HEK-293 and HT-1080 cells (human fibrosarcoma), including recombinant factor VIII-Fc, Dulaglutide, Idursulfase and Velaglucerase alfa (Dumont *et al.*, 2016). A production platform for erythropoietin in HEK-293 cells was demonstrated by deleting the glutamine synthetase gene with CRISPR-Cas9 and then using the GLUL selection marker along with methionine sulfoximine (MSX) for stable cell line generation. The end result was a titre of 696 mg/L demonstrated in a 2 L stirred-tank fed batch bioreactor (Chin *et al.*, 2019) outperforming existing CHO cell based production methods in the literature and yielding EPO with human glycoforms.

Figure 1.3.6. Illustration comparing the glycoforms of recombinant factor VII in different expression systems.

Other cell lines such as mouse derived NS0 and Sp2/0 exist that are used in the production of a small number of biotherapeutics such as Synagis and Erbitux. New cell lines are constantly emerging as well as alternatives such as CAP-T cells derived from human aminocytes and PER.C6 derived from human retina. CHO cells however, have decades of safety profiling, produce higher yields of good quality protein and have well established bioprocesses which make then attractive for biomanufacturer of recombinant proteins for use in the clinic. HEK-293 cells are also proven to be safe, however human cell lines are inherently more susceptible to human pathogens such as viruses, which downstream processing screen and try to remove at great expense. Ironically, for this reason human cell lines are inherently less safe to use than CHO cells which have proven to be quite resistant to viral infections.

An in-depth comparison between the glycoforms observed in different expression systems using various types of mass spectrometry of recombinant factor VII is presented in Figure 1.3.5. The two main mammalian glycan epitopes known to cause immunogenic reactions are Gal(α1,3)Gal residues and N-glycolylneuraminic acid. These epitopes are more common in mouse derived cell lines such as NS0 than in CHO cells. HEK-293 cells and other human cell lines, however, have one distinct advantages in their glycosylation patterns that other mammalian cell lines cannot match, the ability to undertake glutamic acid γ-carboxylation and tyrosine sulfation, which are essential to certain biotherapeutics efficacy such as Drotrecogin alfa and recombinant factor IX-Fc (Dumont *et al.*, 2016). This illustrates the complex landscape of recombinant protein glycosylation and highlights that host selection is of paramount importance in biotherapeutic expression, as a protein that is incorrectly glycosylated can have adverse interactions, poor half-life and efficacy *in vivo*.

1.4 EFFORTS TO IMPROVE CHO CELLS AS A BIOTHERAPEUTIC PRODUCTION CHASSIS

1.4.1 The case for improving CHO for recombinant protein production

In the previous section, we established that the CHO cell is the dominant biotherapeutic production platform on the market. It is no surprise that significant efforts have been undertaken to improve and optimize CHO cell lines and the bioprocesses used to culture them. The need for the advancement of these platforms cannot be understated. In particular, we see that titres and specific productivities have improved by an order of magnitude (Figure 1.4.1). Treatment costs for biotherapeutics are on average 22 times more expensive than small molecule drugs (McCamish and Woollett, 2011). A breast cancer patient's average cost for Herceptin® is $37000 USD, a rheumatoid arthritis or Chrohn's disease patients for Humira® is $50000 USD, Gaucher disease patients with Cerezyme® is $200000 USD a year **for the rest of their lives.** The top 6 biologics already consume 43% of the Medicare part B drug budget. Assuming that the usage of these drugs increases, and new biotherapeutics will keep getting approved, it will put an unbearable amount of stress on the medical budget and ultimately keep these drugs out of the reach of people that need them (So and Katz, 2010).

Figure 1.4.1. Tracking the improvements of specific productivities, titres and viable cell densities (VCD) reported in the literature in the period from 2000-2010.
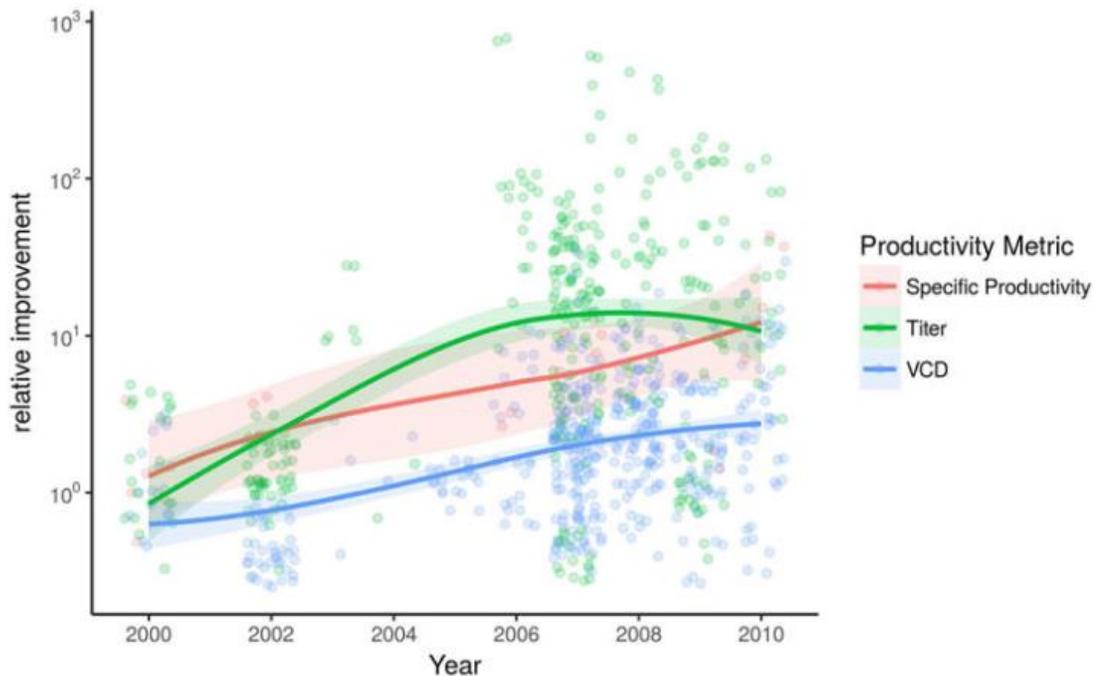
Reproduced from (Kuo *et al.,* 2018)*.*

One way of slashing the price of biopharmaceuticals is to incentivize the development of biosimilars, biotherapeutics that have gone off patent and have become public domain. It is estimated that the sale of such biosimilars will represent 40% of the anticipated global sales in 2015-2020 which amounts to just over $100 billion USD. The development costs for biosimilars are still high though, one estimation puts the development cost in the region of of $75-200 million USD for one drug, which drastically limits the places where these products can be made due to capital restrictions (McCamish and Woollett, 2011). Therefore, decreasing the costs of production for all biotherapeutics remains of paramount importance for the public interest.

1.4.2   The road to better CHO cell bioprocessing

One of the earliest methods applied to the optimizing CHO cell bioprocesses was media optimization. In the past, due to ease of use and cheaper cost of manufacturing, complex media was favoured with components such as fetal bovine serum. Eventually, due to ethical issues of sourcing and concerns around safety due to potential contaminants from animal origin, the industry started to move from complex to defined media. Modern media optimizing has now progressed to the stage where a formulation contains 50-100 compounds; among them energy

substrates, amino acids, chelators, pH buffers, vitamins, surfactants, nucleic acid derivatives, trace elements, fatty acids, salts and lipids (Mccoy *et al.*, 2015). The widespread use of chemically defined protein and serum-free cell culture media has resulted in more reproducible cell culture and simplification of downstream processing. Further, media has been developed that is cell-line and purpose-specific; for example, a media for transient transfection will not be optimal for single-cell cloning. It is now accepted that media optimisation is an essential step towards an efficient and productive bioprocess and high-throughput optimizations are common (Baktur *et al.*, 2016).

Small-scale process development used to be exclusively performed in shake or spinner flasks. These are not suited towards high-throughput process development because they do not automate easily, take up too much space with cell culture volumes of 20-60 mL and do not reproduce conditions in a bioreactor. Scaling down cultures is not easily achievable though. One of the reasons for this is described by the volumetric mass transfer coefficient (k_La) (Petříček *et al.*, 2018). For spinner flasks, this coefficient is roughly 2-3 h^{-1} (Nienow, 2006). In stirred bioreactors, the mass transfer coefficient has been observed to be as high as 10-20 h^{-1} (Zhang *et al.*, 2008). This means that scaled-down culture will perform very different compared to scaled-up cultures. Despite this, the industry has seen an uptake of various scaled-down systems such as TubeSpin® shake flasks (Pereira *et al.*, 2011), 24- and 96-deep well plates (Porter *et al.*, 2010; Mora *et al.*, 2018; Wang *et al.*, 2018). However, perhaps the biggest success in scale down cell line culturing to date was the introduction of the ambr® 15 system. This system allows for the simultaneous stirred-tank culture of 48 mini bioreactors with automated liquid handling and environmental control. This system has now industrially become the first line of reference to determine or predict how a cell line is likely to behave at scale-up and allowing more throughput for DoE, media optimization and stability studies (Bollmann *et al.*, 2019).

Large scale mammalian culture used to be almost exclusively performed in fed-batch mode within stainless steel bioreactors outfitted with the necessary machinery to sustain cell culture; aeration, stirring, pH, DO and temperature probes. Currently, there is a trend towards using disposable and single-use equipment, because of the time saved not having to do clean-up and sterilization in between runs, smaller footprints and potential to run different products in the same facility. The first successful disposable commercially was the wave-type bioreactor introduced in 1999, which uses a rocking mechanism for agitation. Disposable stirred bioreactors at larger volume scales are also available such as at 1-2000 L (Singh, 1999). Another potential innovation was the application of large-scale disposable shaken bioreactors. One estimation of the cost savings achieved from using disposable systems versus stainless steel systems was

around 30% (Guldager, 2010). Currently, stirred-tank bioreactors are aerated by sparging oxygen from the bottom of the bioreactors, while in shaken bioreactors oxygen exchange happens with surface aeration. For cylindrical shaken vessels $k_L a$ values of 2-7 h^{-1} have been observed at working volumes up to 100 L. In contrast, an orbitally shaken 2000 L bioreactor had a $k_L a$ value of 2-3 h^{-1}. This is considered sufficiently close to compete with stirred-tank aeration and potentially this number can be increased by using oxygen-enriched air within the bioreactor (Stettler *et al.*, 2007). Recently, an Influenza A virus production system was demonstrated in an orbital shake bioreactor coupled with a perfusion system with a VCD that reached 50 x 10^6 cells/mL (Coronel *et al.*, 2019).

Perfusion culture deserves a separate paragraph in this section as it has the potential to radically change the nature of industrial bioprocessing. The fundamental idea behind this initiative is that you can greatly extend the time and yield of a single culture by constantly filtering out cells and keep a steady flow-through of media through the culture, continuously harvesting the product. One estimation of the cost savings of such a setup is visualised in Figure 1.4.2. This assumes "a media cost of $10 USD/L for bulk-discounted chemically defined serum-free media, USD $1/mL of supplements, and 30 min/day of labour for tanks and AcuSyst® bioreactors. For stirred-tank systems, estimated scale-up is assumed to be 50% of the final working volume" (Wozniak and Biesecker, 2017). Sustained cell densities in these systems have been reported in the 20-130 x 10^6 cells/mL range with a steady-state duration of up to 50 days (Bielser *et al.*, 2018). This often allows for kg levels of recombinant protein product yields on a benchtop scale in a comparable amount of time to a fed-batch bioreactor. However, it has been estimated that the environmental footprint of perfusion based processes might be higher "because they consumed 35% more water, demanded 17% more energy, and emitted 17% more CO_2 than the fed-batch process" (Bunnak *et al.*, 2016).

**Figure 1.4.2. Estimated seed-train labour and material costs for stirred-tank and AcuSyst®
perfusion bioreactors assume a single production run.**
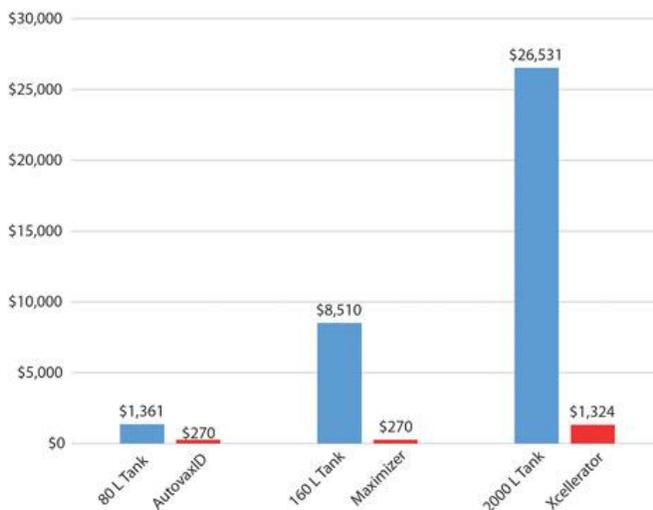
Adapted from (Wozniak and Biesecker, 2017)*.*

With the arrival of technologies that allow the manipulation of the DNA sequence of a gene,
mRNA elements and the genome directly, CHO cells have become a system of great interest for
genetic engineers. The first wave of CHO cell line engineering relied on black box unreproducible
methods. Most stable cell lines are created using random integration of foreign DNA into the
genome. This results in a very diverse population of cells that do not have the same copy numbers
of the target gene nor the same genomic insert site, which has proven to affect both the stability
of the gene and its expression levels.

More precise methods of gene insertion have now become available, although random
integration is still widely used. One such technology relies on recombination. Recently, a site-
specific stable cell line engineering system was developed which leverages the unidirectional
recombination properties of bacteriophage PhiC31 integrase (Chi *et al.*, 2019). Other
recombination-based systems have been proposed in the past that utilize the Cre/loxP system
(Kawabe *et al.*, 2015) system and Flp-Recombinase Mediated Cassette Exchange (Turan *et al.*,
2013). A commercially available cell line Flp-In exists offered by Thermo Fisher Scientific that
provides a host cell line and a vector that can be used to perform Flp-recombinase based
integration out of the box (Serpieri *et al.*, 2010).

Another cell line engineering platform that has become available recently leverages the use of
transposases, specifically the piggyBac™ transposase commercialised by Lonza under the name
GS piggyBac™. This transposase cuts out the target gene from a donor plasmid and integrates it
into TTAA sites in the genome, which are roughly about 1 in 256 base pairs in frequency (Zhao

et al., 2016). A feature of this system is it can potentially allow the integration of DNA up to 100 kb in length, allowing for the insertion of massive gene clusters enabling complex synthetic biology efforts in CHO cells. Comparatively, using this approach was found to yield better-producing clones than using the random integration of the Lonza GS system alone (Balasubramanian *et al.*, 2015; Ahmadi *et al.*, 2017; Lonza, 2019).

The discovery of the CRISPR-Cas9 system has been revolutionary in genetic engineering since it opened up the ability to target almost any sequence in the genome and introduce indels via double-strand breaks. This technology has already been used in CHO cell engineering in a variety of ways. Because of the ability of the Cas9 enzyme to targeted to different loci in the genome, scientists have been able to identify novel stable and active integration sites (Zhao *et al.*, 2018). Glycoengineering in CHO cells has also received a boost from CRISPR technology allowing tailor-made glycan profiles in cell lines. Expressing the HIV-1 vaccine in CHO cells required the viral envelope protein to have mannose-5 or earlier glycan intermediates which were achieved by deleting the *MGAT1* gene (Byrne *et al.*, 2018). Targeting 43 different glycan metabolism-related genes, differently glycosylated forms of α-galactosidase A were screened in a Fabry disease mouse model. It was found that "an α2-3 sialylated glycoform designed to eliminate uptake by the mannose 6-phosphate and mannose receptors exhibits improved circulation time and targeting to hard-to-reach organs such as heart" improving drug pharmacokinetics (Tian *et al.*, 2019).

CRISPR interference, a method of silencing mRNA levels of a target gene, has been used to improve traditional MTX based gene amplification by knocking down the DHFR gene. This approached increased the "*egfp* copy number ~3.6-fold and enhanced the eGFP expression ~3.8-fold, without impeding cell growth" when applied (Shen *et al.*, 2017). One study attempted to undertake the rewiring of amino acid catabolism using Cas9. Nine genes were targeted in 7 amino acid catabolic pathways. The disruptions to expression of *Hpd* and *Gad2* were of particular note resulting in "unchanged AA uptake rates while having growth rates increased up to 19%, and integral of viable cell density as much as 50% higher, and up to 26% decrease in specific ammonium production and to a lesser extent (up to 22%) decrease in lactate production" (Ley *et al.*, 2019). CRISPR-based targeted epigenetic editing has been used to turn on the dormant beta-galactoside alpha-2,6-sialyltransferase 1 (*St6gal1*) gene by targeting the catalytic domain (CD) of Ten-Eleven Translocation methylcytosine dioxygenase 1 (TET1) via deactivated Cas9 to its methylated promoter. Stable upregulation of this gene was achieved over a time span of more than 80 days. This process was reversible by targeted methylation with DNA methyltransferase 3A fused to a deactivated Cas9 which resulted in 5.4 fold reduction of *St6gal1* mRNA expression (Marx *et al.*, 2018).

Ever since miRNAs, small non-coding RNAs that form hairpins by folding on themselves were discovered, researchers have tried to utilize them to control gene expression levels. One method to achieve this is to constitutively express the target miRNA. One such constitutively expressed miRNA, miR-557, was found to significantly enhance cell line development (CLD) in a product independent manner. Higher producing clones were easier to discover and production cell lines expressing this miRNA exhibited significantly increased final product yields in fed-batch cultivation without compromising product quality. Further, in one instance, co-expressing this miRNA along with a difficult to express antibody increased the product titre two-fold compared to a negative control miRNA expression (Fischer et al., 2017a). In another study, miRNA-17 and miRNA-92a and cluster miRNA17-92a were found to enhance CHO cell growth and productivity. CHO cells stably expressing miRNA-17 demonstrated a 2-fold increase in productivity and a 3-fold increase in EpoFC titre (Jadhav et al., 2014).

High-throughput engineering approaches performed in CHO are far and few in between. One such study utilised a genome-wide high-content miRNA screen in a recombinant CHO-SEAP cell line. Out of 1139 miRNAs examined, 21% enhanced specific productivity, while cell proliferation was accelerated by 5% of the miRNAs. Apoptosis was found to be reduced by 13% and 4% of the total pool of miRNAs reduced necrosis. The study implicated the miRNA-30 family as a critical component of cell proliferation (Fischer et al., 2014). Another recent endeavour utilised a cross-species whole-genome siRNA library to screen CHO cells for engineering targets. It was found that in the end only two genes were validated by a second knockdown screen, *Rad21* and *Chd4*. siRNA mediated knockdown of these two genes conferred an increased productivity phenotype but it was cell line and clone specific (Klanert *et al.*, 2019). A summary of cell line engineering efforts undertaken in CHO is presented in Table 1.4.2.1 for the convenience of the reader.

Table 1.4.2.1. Summary of the cell engineering approaches undertaken within the community in order to improve bioprocess characteristics.

Cell engineering approach	Strategies involved	Result	Reference
Regulation of apoptosis	Over-expression of anti-apoptotic genes	Limit cell-apoptosis	(Arden *et al.*, 2007; Baek *et al.*, 2017; Zhang *et al.*, 2018)
	Inhibition or down-regulation of pro-apoptosis genes		(Han and Rhee, 2018)
Regulation of cell cycle progression	Inducible expression of cell-cycle factors	Cell cycle arrest	(Weinstein *et al.*, 1994; Renner *et al.*, 1995; Lee *et al.*, 1996; Fussenegger *et al.*, 1998; Mazur *et al.*, 1998; Ifandi and Al-Rubeai, 2003; Lee *et al.*, 2013; Kim *et al.*, 2014)

	Inhibition of CDK or overexpression of CDK inhibitor		(Du *et al.*, 2015)
	Use of mTOR-based engineering of mammalian cell lines	Slowed cell cycle progression	(Dreesen and Fussenegger, 2011; Dadehbeigi and Dickson, 2015)
Engineering of chaperones and foldases	Over-expression of protein disulphide isomerase	Increased disulphide bond formation	(Borth *et al.*, 2005; Mohan *et al.*, 2007)
Post-translational modifications	Knocking out the fucosyltransferase gene fut8	Enhanced ADCC activity	(Yamane-Ohnuki *et al.*, 2004; Chung *et al.*, 2012)
	High-throughput glycan engineering	Tailor-made glycan profiles for target proteins	(Tian *et al.*, 2019)
	Specific glycoengineering	Simplified glycans	(Byrne *et al.*, 2018)
Metabolic engineering	Over-expression of glutamine synthetase	Reduction in ammonia generation as a by-product	(Zhang *et al.*, 2006)
	Over-expression of ornithine transcarbamylase, carbamoyl phosphate synthetase I		(Park *et al.*, 2000)
	Over-expression of pyruvate carboxylase	Reduction in lactic acid accumulation as a by-product	(Kim and Lee, 2007b; Toussaint *et al.*, 2016; Gupta *et al.*, 2017)
	Down-regulation of pyruvate dehydrogenase kinases/lactate dehydrogenase A		(Kim and Lee, 2007a; Zhou *et al.*, 2011; Wilkens *et al.*, 2019)
	Growth factor engineering	Quicker adaptation to suspension	(Lee *et al.*, 2016)
	Amino acid catabolism	Improved cell growth and fewer by-products	(Ley *et al.*, 2019)
Engineering for hypothermic growth	Stable over-expression of cold stress genes, such as cold-inducible RNA-binding protein	Improvement in the productivity and yields of recombinant protein	(Tan *et al.*, 2008)
miRNA engineering	Overexpression of miRNA or miRNA sponge	Various impacts on VCD, productivity, titre and bioprocessing	(Barron *et al.*, 2011; Jadhav *et al.*, 2012; Druz *et al.*, 2013; Strotbek *et al.*, 2013; Jadhav *et al.*, 2014; Loh *et al.*, 2014; Fischer *et al.*, 2014; Fischer *et al.*, 2015; Patel *et al.*, 2016; Fischer *et al.*, 2017b)
High throughput screens	Genome-scale siRNA screen	Increased productivity	(Klanert *et al.*, 2019)

VCD – viable cell density; siRNA – silencing RNA; miRNA – micro RNA; ADCC - Antibody-Dependent Cell Cytotoxicity; CDK – cyclin dependent kinase.

As systems biology tools have developed, the CHO cell and biopharmaceutical community has assembled large amounts of 'omics data. The first CHO cell line to have it's genome sequenced was the CHO-K1 (Lewis *et al.*, 2013). It was reported by (Wurm and Hacker, 2011) that there is much CHO cell instability and a variety of different CHO cell line lineages that are used today in the industry. The CHO-K1 cell host was generated from the original CHO cell line

isolated by Puck in 1957, however, the DHFR⁻ phenotype was achieved via exposing these cells to random mutagenesis. Two of the most common industrially used cell lines, CHO-DG44 and DXB11 (DUKX), are derived from CHO-K1, however it was unclear how different these cell lines are, given that the karyotype is unstable. It was understood that more sequencing efforts needed to be undertaken to fully understand the landscape of CHO cell 'omics. To this end, the CHOgenome database was created to aggregate and disseminate these datasets for the community (Hammond *et al.*, 2012).

Another study in this area published the draft genome of a female Chinese hamster, *Crisetulus griseus* and sequenced the genomes of 6 CHO cell lines from the CHO-K1, DG44 and CHO-S lineages (Lewis *et al.*, 2013). Subsequently, a reference genome of the Chinese hamster became available using a hybrid strategy using single-molecule real-time sequencing and merged this with Illumina-based assemblies. This reduced the number of scaffolds by >28-fold, with 90% of the sequence in the 122 longest scaffolds (Rupp *et al.*, 2018). Eventually, the mitochondrial genome was sequenced as well revealing the widespread heteroplasmy, "89% of the heteroplasmic mutations identified were cell line-specific with the majority of shared heteroplasmic SNPs and INDELs detected in clones from 2 cell line development projects originating from the same host cell line" (Kelly *et al.*, 2017).

The metabolome has also received attention from researchers as it reveals the pathways of metabolism and energy exchange opening the door to rewire CHO cells to be more metabolically efficient. A consensus genome-scale model of CHO cell metabolism was constructed with 1766 genes and 6663 reactions describing metabolism and protein production (Hefzi *et al.*, 2016). This has now been used to help understand how metabolic engineering of the cell might impact cell metabolism and phenotype.

CHO cell omics have progressed to the point where reserarchers are exploring the dark matter of gene expression, long non-coding RNA transcripts and relating their abundance to production (Vito and Smales, 2018; Vito *et al.*, 2019). Among these, *Adapt15* (linked to ER stress), *GAS5*, (mTOR signalling/growth arrest), and *PVT1* (Myc expression) were highlighted as being correlated with productivity and growth.

1.4.3 CLD technologies

The process of expressing recombinant proteins for industrial manufacture involves the transfection of the cells with the DNA coding for the recombinant gene(s) of interest such that these become stably integrated into the genome allowing for them to produce the protein of

interest. Selecting the 'best' expressing cells from the resulting population can be a labour and time intensive process. Once a so-called clonal cell line is isolated and allowed to divide, the cells within the population will, in theory, be genetically identical with the selected parent cell. This process is necessary because it is currently accepted that monoclonal cell lines are more consistent in their production attributes and, in turn, the product quality (International Conference on Harmonisation (ICH), Guideline Q5D, 1997). It is often the case that in a panel of monoclonal cell lines derived from the same parent pool the growth rates, specific productivities, and post-translational modifications will vary significantly across the isolated clones. Depending on the bioprocess that a company uses for manufacturing, and the product's specific quality and activity profile, the developer can then make an informed decision as to which cell line will be the best production platform for the product in that instance. Cell line development (CLD) strategies are not consistent between different companies. For monoclonal cell line isolation, a number of technologies can be used, but the three most popular methods are:

1) Fluorescence activated cell sorting (FACS) with or without cell labelling.
2) Limiting dilution cloning where single cells are isolated by progressively diluting the sample in multi-well plates.
3) ClonePix™ 2 systems with automated colony picking from semi-solid medium.

Regardless of the technology used the timelines for isolating single cell clones, scaling them up and then validating their production characteristics can take up to 6 months (Cell Line Development (CLD) 2.0, 2017). It is also the case that regulatory agencies demand strict evidence from manufactures that the cell lines they have are monoclonal, placing additional emphasis on this part of the CLD pipeline.

Further, the CLD process usually involves screening thousands of clones at different scales for product titre and other attributes as the cell lines are progressed and gradually honed down to a single cell line; 96 well plate, 6 well plate, 24 well plate, T75 flask, 50 mL shaking tube, E125 flask and, finally, bioreactor scales. This is necessary because as the cell environment scales up the cells can behave differently due to differences in the maturation and stability of cell lines, gas exchange, medium composition, shear stress and feeding regimes.

A number of issues that arise during CLD can be addressed using picodroplet microfluidic technology. It allows for high-throughput single cell labelling and sorting and these picodroplets can undergo individual measurements as well as encapsulation and maintenance of single cells. New analyses have been developed that allow the detection of secreted protein within the droplet and, therefore, sorting based on production output. One example of this technology, the Sphere Fluidics's Cyto-Mine platform, has successfully been applied in industrial CLD processes in order

to replace previous generation methods (Josephides *et al.*, 2020).

Another microfluidic technology that has arrived on the scene recently is optofluidics. This utilizes the manipulation of cells within a microfluidic environment using light and does not use droplets to encapsulate cells. The Berkeley Lights Beacon® is one such system that has seen a rapid uptake within industry. This device allows for long term cell cultivation on a microchip with pens designed to house cell populations in order to monitor them in terms of growth, productivity and monoclonality (Le *et al.*, 2020).

1.5 AIMS OF THE WORK DESCRIBED IN THE THESIS

The aims and objectives of the work described in this thesis were therefore to;

1) Leverage existing public transcriptomic datasets in order to recommend targets for CHO cell engineering to improve recombinant protein production.

2) To design an industrial-like next-generation CHO cell line development workflow using high-throughput single-cell technologies.

3) To characterize and validate these technologies by comparing them against existing representative cell line development workflows in terms of resource overhead or efficiency.

2 MATERIALS AND METHODS

2.1 STATISTICAL ANALYSIS

All sample populations were initially checked with a D'Agostino and Pearson omnibus normality test to determine whether the populations were drawn from a normal distribution. If the p-value was below 0.05 then it was assumed that the distributions were not normal and nonparametric statistical methods were applied, mainly the Kruskal-Wallis test to check for differences between the population means. In parametric testing, standard t-tests were used for comparing means. For ANOVA testing, the p-value threshold for multiple comparisons was adjusted. Most statistics were carried out within GraphPad Prism 6 unless stated otherwise. The p-value thresholds for the asterisks in all figures highlight the decimal point of the value (*<0.05, **<0.01, ***<0.001, ****<0.0001)

Growth rates of cell lines were calculated using the *growthrates* package in R using the *fit_easylinear* function.

2.1.1 DNA concentration measurements

DNA concentration measurements were carried out using a NanoDrop 2000 spectrophotometer (Thermo Fisher, UK). 2 µL of the sample was routinely used to assay plasmid and PCR product concentrations.

2.1.2 DNA visualization software

For visualizing and planning experiments, the freeware software from Snapgene (GSL Biotech, USA) and Benchling (Benchling, USA) were used to simulate restriction digest, primer annealing, PCR products and ligations.

2.1.3 DNA sequencing

All plasmid and gene Sanger sequencing was outsourced to GENEWIZ. Vials of 1 mL sample in water were sent at 10 ng/µL concentrations. Primers used for sequencing covering region from promoter to poly A sequence:

Fwd: TCAAGCCTCAGACAGTGGTTC

Rev: TAGAAGGCACAGTCGAGG

2.1.4 Agarose gel electrophoresis analysis of DNA

DNA agarose gels were prepared with TAE buffer (50x stock consisting of 2 M Tris base, 50 mM EDTA and 20 mM acetic acid). Agarose 1% (w/v) was used in the final gel unless specified otherwise. After addition of agarose in buffer, the solution was microwaved until the agarose had fully melted. Ethidium bromide (Invitrogen, USA) was added to the gel and to the buffer solution to visualize DNA bands. To estimate DNA band sizes, a 1 kb DNA ladder was used (Promega, UK). Gels were typically run at 7 V/cm for 1 hour. DNA fragments were visualized by exposing the agarose gel to UV light in a G Box (Syngene, UK).

2.1.5 Polymerase chain reaction amplification of DNA

Phusion High Fidelity DNA Polymerase was used to carry out PCR experiments (Thermo Fisher Scientific, UK). PCR was performed according to typical manufacturers guidelines; in brief 4 µL of 5x Phusion HF buffer, 0.8 µL of forward and reverse primers, 0.2 µL of Phusion DNA Polymerase, 0.4 µL of 10 mM dNTPs, 1 ng of template and nuclease free water (Promega, UK) were added to achieve a final volume of 20 µL. A 2 step PCR protocol was used that consisted of an initial denaturation temperature of 98°C for 10 s and an extension step of 72° for 45 s. A final extension time was set at the end of 30 cycles at 72°C for 10 min.

Primers used to clone in Etanercept fragment into pAVE 1062 backbone were as follows;

Et_Fwd; ATAGAATTcgccgccaccatggctcctgtggctgtt

Et_Rev; ATAGGATcctcacttgccaggggacagag

where those nucleotides in capitals represent overlap regions containing sites recognised by the specific restriction enzymes.

Template DNA from with Etanercept sequence was amplified with reading frame underlined and primer sites highlighted: Further information on plasmid sequence is not available due to the plasmid being proprietary and under IP protection.

gGAGACCCAAGCTGGCTAGCGCCGCCACCATGGCTCCTGTGGCTGTTTGGGCTGCTCTGG
CTGTTGGACTGGAACTGTGGGCTGCTGCTCATGCTCTGCCTGCTCAGGTGGCCTTCACACCTT
ATGCTCCAGAGCCTGGCTCTACCTGCAGACTGAGAGAGTACTACGACCAGACCGCTCAGATG
TGCTGCTCCAAGTGTTCTCCTGGCCAGCACGCCAAGGTGTTCTGCACCAAGACCTCCGATAC
CGTGTGCGACTCCTGCGAGGACTCCACCTATACTCAGCTGTGGAACTGGGTGCCCGAGTGCC
TGTCTTGTGGCAGCAGATGCTCCTCCGACCAGGTGGAAACCCAGGCCTGTACCAGAGAGCAG
AACCGGATCTGCACCTGTAGACCCGGCTGGTACTGTGCCCTGTCTAAGCAAGAGGGCTGTAG
ACTGTGCGCCCCTCTGAGAAAGTGCAGACCTGGCTTCGGAGTGGCTAGACCTGGCACCGAGA

CATCTGACGTCGTGTGCAAGCCTTGTGCTCCCGGCACCTTCTCCAACACCACCTCCTCTACCG
ACATCTGCAGACCCCACCAAATCTGCAACGTGGTGGCTATCCCTGGCAACGCCTCTATGGAT
GCCGTGTGCACCTCTACCTCTCCAACTCGGTCTATGGCTCCCGGCGCTGTTCATCTGCCTCAG
CCTGTGTCTACCAGAAGCCAGCACACCCAGCCTACACCTGAGCCTTCTACCGCTCCTTCCACC
AGCTTTCTGCTGCCCATGGGACCATCTCCACCAGCCGAAGGATCTACAGGCGACGAGCCTAA
GTCCTGCGACAAGACCCATACCTGTCCTCCATGTCCTGCACCTGAGCTGCTCGGAGGCCCTTC
CGTGTTTCTGTTCCCTCCAAAGCCTAAGGACACCCTGATGATCTCTCGGACCCCTGAAGTGAC
CTGCGTGGTGGTGGATGTGTCTCACGAGGACCCAGAAGTGAAGTTCAATTGGTACGTGGACG
GCGTGGAAGTGCACAACGCCAAGACCAAGCCTAGAGAGGAACAGTACAACAGCACCTACAG
AGTGGTGTCCGTGCTGACCGTGCTGCACCAGGATTGGCTGAACGGCAAAGAGTACAAGTGCA
AGGTGTCCAACAAGGCCCTGCCAGCTCCTATCGAAAAGACCATCAGCAAGGCCAAGGGCCAG
CCTAGGGAACCCCAGGTTTACACCTTGCCTCCAAGCCGGGAAGAGATGACCAAGAACCAGGT
GTCCCTGACCTGCCTGGTCAAGGGCTTCTACCCTTCCGACATTGCCGTGGAATGGGAGAGCA
ATGGCCAGCCTGAGAACAACTACAAGACCACACCTCCTGTGCTGGACTCCGACGGCTCATTC
TTCCTGTACTCCAAGCTGACAGTGGACAAGTCCAGATGGCAGCAGGGCAACGTGTTCAGCTG
CTCCGTGATGCACGAGGCCCTGCACAATCACTACACACAGAAGTCCCTGTCTCTGTCCCCTG
GCAAGTGAGGCGCGCCCCAAGCT

2.1.6   DNA restriction enzyme digests

The restriction enzymes used were sourced from the FastDigest catalogue of Thermo Fisher Scientific, UK, unless otherwise specified. Digestion times used were according to manufacturer's recommendations; 1 µL of enzyme to digest 1 µg of DNA in a 20 µL volume containing 2 µL of FastDigest buffer for 30 min at 37°C was used as a typical reaction setup.

2.1.7   DNA purification

DNA excised from agarose gels and PCR products were purified using the Wizard® SV Gel and PCR Clean-Up System (Promega, UK) according to manufacturer's instructions. Incubation time was extended by 5 minutes for gel slices to ensure complete melting of the agarose. Elution was performed by using preheated 60°C nuclease-free water (Promega, UK) twice, using the elute for the second time on the column.

2.1.8   Ligation of DNA fragments

Only sticky end ligations were performed throughout this work and T4 ligase (Promega, UK)

was used following manufacturer's specifications. An insert to backbone ratio of 1:3 was consistently used. Where possible, 100 ng of backbone was digested with appropriate restriction enzymes for ligations making sure the total amount of DNA in the 10 µL mix did not exceed 200 ng. Ligase (1 µL) was used to shorten the ligation time to 15 mins at room temperature. Ligation calculations were carried out using the online BioCalculator (Promega, UK).

2.1.9 Preparation of DH5α *Eschericia coli* competent cells

A colony of DH5α cells (provided from the laboratory at Kent) was picked and streaked across a Lysogeny Broth (LB) agar plate (1.5% agar, 1% NaCl, 1% peptone, 0.5% yeast extract) and grown overnight at 37°C. A single colony from a fresh plate was then used to inoculate 5 mL of LB media which was left overnight at 37°C and shaking at 150 rpm. This starter culture was then used to inoculated two 250 mL flasks containing 50 mL of Super Optimal Broth (SOB) (2% w/v of peptone, 0.5% w/v of yeast extract, 0.1% w/v of NaCl, 2.5 mM KCl, 10 mM $MgCl_2$). Cultures were grown at 37°C and 150 rpm until an absorbance at 600 nm of 0.4-0.6 was reached. Cultures were then immediately chilled on ice and harvested by centrifugation at 3000 g for 15 min at 4°C. The media was aspirated and pellets were resuspended in 10 mL of 100 mM ice-cold $CaCl_2$ which had been filter sterilized. After resuspension the cells were incubated on ice for 30 min and then pelleted by centrifuging at 4000 g for 15 min. The pellets were then resuspended with 2 mL of ice cold $CaCl_2$ and the two cultures were pooled together. 1 mL of 80% glycerol (Sigma, USA) was added to the cells and cells were aliquoted (100 µL) into cryovials and snap frozen on dry ice. Cells were stored in -80°C until required for experiments.

2.1.10 Plasmid DNA amplification and purification

Based on the quantity of material the appropriate QIAgen DNAprep kit was used for purification (QIAprep® Spin Miniprep/HiSpeed® Plasmid Maxi/HiSpeed® Plasmid Giga, Qiagen). Based on the manufacturer's instructions, starter cultures were inoculated overnight from a freshly streaked colony which were then incubated overnight at 37˚C, 150 rpm. The next day the cells were harvested by centrifugation at 3000 *g* for 20 min and the pellets were processed using the appropriate kit.

2.1.11 Transformation of DNA into *E. coli* competent cells

An aliquot of competent DH5α cells was thawed on ice. 3 µL of plasmid or 10 µL of ligation mix was then added to cells and incubated for 30 min on ice. A heat shock at 42˚C for 90 sec was

then applied to transform the cells. After a 2 min recovery period on ice, 900 μL of LB media was added to the competent cells and the cells were incubated for an hour at 37°C, 180 rpm. At the end of the incubation step, 100 μL of competent cells was spread on a LB media agar plate containing ampicillin at 100 μg/mL. The plates were then placed at 37°C overnight.

2.1.12 Linearization of plasmid DNA

To linearize plasmid DNA, 200 μg of plasmid, 200 μL of NheI-HF restriction enzyme (NEB, UK), and 40 μL of 10x HF buffer was used in a reaction volume of 400 μL to set up overnight linearization reactions at 37°C. 1 μL of reaction mix was then analysed on an agarose gel to confirm complete digestion. If multiple bands were observed, then linearization was extended for another 2-3 hours. After confirming the presence of a single band on an agarose gel, plasmid DNA was ethanol precipitated. Contents of the digest were transferred to a sterile 2 mL tube. 0.1 volumes of filter sterilized sodium acetate (3 M, pH 5.2, Sigma Aldrich, USA) was added and mixed by inversion. Precipitation was then initiated by adding 2.5 volumes of ice cold 95% v/v ethanol. The tube was mixed by inversion until a white precipitate was observed. The tube was then incubated at -80°C for 1 hour and centrifuged at 13200 rpm for 30 min at 4°C. The supernatant was then decanted and 1 mL of 70% ice cold ethanol added. After a final centrifugation at 13200 rpm for 10 min, the supernatant was removed and the pellet allowed to air dry for 10 min or until all ethanol had evaporated. 105 μL of sterile culture grade water (Sigma Aldrich, USA) was then added to resuspend the DNA pellet and incubated overnight at 37°C . The DNA concentration was then confirmed using a NanoDrop 2000 spectrophotometer and the DNA was then diluted to a final working concentration of 1 μg/uL.

2.2 TISSUE CULTURE

2.2.1 Fed-batch overgrow (FOG) cultures

For fed-batch overgrow (FOG) experiments, 250 mL Erlenmeyer flasks with vented caps were inoculated at 0.2 x 10^6 viable cells/mL in 60 mL FDB-MAP media containing 8 mM L-glutamine. The flasks were incubated at 37°C, 125 rpm, 80% humidity and 5% CO_2. Monitored parameters throughout the FOG culture include cell concentration, culture viability, concentration of glucose, glutamine, glutamate, lactate and ammonium ions. Cell concentration and culture viability were monitored daily from day 2 using a Vi-CELL cell counter. Metabolite levels were evaluated daily from day 3 using a YSI 2950 metabolic analyser (YSI, USA). Every other day from day 4, 1 mL of culture was sampled, centrifuged for 5 min at 1400 g and both pellet and supernatant were kept at -20°C. Regarding the feed regime, the cultures was fed from day 2 with 2% (v/v) HyClone™ Cell Boost 7A (GE healthcare, USA) and 0.2% (v/v) HyClone™ Cell Boost 7B (GE healthcare, USA), 0.5% (v/v) and 200 mM L-glutamine when the glutamine concentration dropped below 0.22 g/L. Glucose (Sigma Aldrich, USA) was added when the concentration dropped under 3 g/L. The volume of glucose added to the culture was calculated using the following formula:

$$Glc\ feed\ volume\ (ml) = \frac{4 - current\ GLC\ conc.\ \text{g/L}}{3\ g/L} \times starting\ volume\ (mL)$$

2.2.2 FUJIFILM Diosynth Biotechnologies (FDB) CHO cell cryopreservation

After verification of the viability of a culture (>95%) using a Vi-CELL instrument, 1 x 10^7 viable cells were centrifuged at 200 g for 5 min. The cells were then resuspended in 1 mL of FDB-MAP media containing 8 mM L-glutamine and 10% (v/v) DMSO (Sigma Aldrich, USA) and transferred to a cryovial. Vials were kept overnight at -80°C in Nalgene Mr. Frosty containers before being transferred into liquid nitrogen storage. For banks made from TubeSpin® flasks before 24-well plate screening, 0.5 mL of DMSO was added to a 5 mL sample of cell culture and deposited into a 5 mL cryovial and kept at -80°C.

2.2.3 Cell line generation via electroporation of DNA into host cells

The required SF buffer was brought up to room temperature for at least 30 minutes prior to use. Cells and DNA were prepared up to a final cell concentration of 5 x 10^7 cells/mL and 80

µg/mL of DNA (this equates to 5×10^6 cells per cuvette and 8 µg of DNA per cuvette). Before electroporation, the calculated volume of cell suspension was centrifuged for 10 min at 180 g. Post-centrifugation, the supernatant was aspirated and the pellet was suspended in SF buffer. The DNA was then added into the cell suspension and mixed by gently flicking the tube. 100 µL was then transferred into the Nucleocuvette™ making sure that there was full contact with the surface and no bubbles were present. Nucleofection™ was carried out using the Amaxa Nucleofector™ on program setting FF-137. Post Nucleofection™, 400 µL of prewarmed FDB-MAP media was added to the Nucleocuvette™ and incubated at 37°C for 10 minutes. After incubation, the entire contents of the Nucleocuvette™ were transferred to the T75 flasks containing 20 mL of pre-warmed FDB-MAP medium and kept in a static humidified incubator at 37°C and 7.5% CO_2. One Nucleocuvette™ was used to seed one pool in a T75 flask at a seeding density of 5×10^6 cells/flask (200 µL of transfected cells per flask).

2.2.4 FDB cell revival protocol

A cryovial was thawed gently in a water bath at 37°C . The content of the vial was then resuspended in 20 mL of pre warmed CD-DG44 media containing 8 mM L-glutamine and 0.18% pluronic F-68 added drop by drop. A cell concentration and culture viability count was then performed. Cells were then diluted to a concentration of 0.2×10^6 viable cells/mL and incubated at 37°C, 125 rpm, 5% CO_2 and 80% humidity.

2.2.5 FDB stable cell line transfection

After a culture viability check, 5×10^6 viable cells were centrifuged at 180 g for 10 min and resuspended in 100 µL of SF buffer provided in the SF cell line 4D-Nucleofector™ X Kit L (Lonza, CH). The cells were placed into the Nucleocuvettes™ and 8 µg of linearized DNA was added. The cells were electroporated using program FF-137 (parameters not communicated) of the 4D-Nucleofector™ System (Lonza, CH). Post-nucleofection, 400 µL of pre-warmed FDB-MAP containing 1 x HT supplement (Thermo Fisher Scientific, USA) and 8 mM L-glutamine were added into the cuvettes and then incubated at 37°C for 15 min. After incubation, 20 mL of pre-warmed FDB-MAP containing 1 x HT supplement and 8 mM L-glutamine in a T75 flask were inoculated with the transfected cells. The flasks were incubated at 37°C, 10% CO_2 overnight. The next day, cells were centrifuged for 10 min at 180 g and the media was replaced by FDB-MAP media containing 8 mM L-glutamine and the selection pressure agent (175 nM MTX or 7.5 µg/mL puromycin). Flasks incubated for a recovery period of 14 days, and progressed if the cell count

was sufficient to seed into E125 Erlenmeyer flasks at 0.2 x 10^6 cells/mL in a 15 mL volume.

2.2.6   FDB host cell line maintenance

Clone 27 is a CHO-DG44 derived cell line used as a host at FDB. Clone 27 was routinely seeded at 0.2 x 10^6 viable cells/mL in 30 mL of CD-DG44 media (Life Technologies, USA) containing 8 mM L-glutamine (Thermo Fisher Scientific, USA)) and 0.18% (v/v) pluronic F-68 (Thermo Fisher Scientific, USA) in a 125 mL Erlenmeyer flask with vented caps (Corning, USA). Flasks were incubated in a Certomat® CT Plus incubator (Sartorius, DE) at 37°C, 125 rpm, 5% CO_2 and 80% humidity. Cells were passaged every 3 days. Stably transfected Clone 27 cells were cultivated in FDB-MAP (SAFC, USA) containing 8 mM L-glutamine and a selection pressure, methotrexate (MTX) at 175 nM (Sigma-Aldrich, USA) or puromycin at 10 µg/mL (Sigma-Aldrich, USA). Stably transfected Clone 27 cell lines were maintained every 3-4 days. TubeSpin® vessels were also used with a working volume of 10 mL in 50 mL spin tube bioreactors with vented caps (TPP, CH) and incubated in a Certomat® CT Plus incubator (Sartorius, DE) at 37°C, 200 rpm, 5% CO_2 and 80% humidity.

2.2.7   General FDB cell culture maintenance

Culture viability and cell concentration were determined by processing a sample (0.6 mL) of the culture with a Vi-CELL cell counter (Beckman Coulter, USA). The cell suspension was diluted to the desired level in pre-warmed media and returned as soon as possible into the incubator under previous culture conditions.

2.2.8   Clone picking of recombinant Clone 27 derived stable cell lines

On the day before seeding, the CloneMedia (K8712, Molecular Devices) was placed at 4°C to thaw overnight. One hour before seeding, the CloneMedia was brought up to room temperature as per the manufacturer's instructions. After a 3 day subculture, 5 mL of transfectant pool suspension was taken by mixing the cells thoroughly by pipetting up and down using an air displacement pipette. The mixed cell suspension was passed through a cell strainer (Steriflip®, SCNY00020, Merck Millipore) to reduce clumping before proceeding to the next steps. Two separate 0.6 mL samples from the filtered suspension were counted using a Vi-Cell instrument. If the counts were not within a 10% margin of error of each other, then the process was repeated again. A dilution was then made until a cell concentration of 0.25 x 10^6 in 5 mL of media was achieved. The cells were then counted again in duplicate ensuring that the counts were within a

10% margin of error. Afterwards, the suspension was thoroughly mixed 20 times using an air displacement pipette and a 20 µL sample examined under a haemocytometer. The suspension of cells was deemed suitable for cloning if there were no visible clumps and haemocytometer counts had < 1% doublets. Once this was done, another serial dilution was made to give a final cell concentration of 0.125 x 10^5 cells/mL. The dilutions are outlined in Table 2.2.8.1.

Table 2.2.8.1. Serial dilutions undertaken of transfectant pools before plating with semi-solid medium on 6-well plates.

Dilution Number	Dilution factor	Volume of cell suspension (mL)	Volume of medium (mL)	Pre-dilution cell concentration (cells/mL)	Post-dilution cell concentration (cells/mL)
1	Calculated based on first count			Determined from Vi-Cell count	$0.25*10^6$
2	1/20	1	19	$0.25*10^6$	$0.125*10^5$
3	1/1000	0.1	100	$0.125*10^5$	12.5

Next, 8.9 mL of sterile water (W3500, Sigma) was added to a bottle of overnight thawed semi-solid CloneMedia (K8712, Molecular Devices) and shaken vigorously for 30 s. Following this, the bottle was allowed to settle for 10 min and 1 mL of CloneDetect (K8295, Molecular Devices) was then added ensuring the lights in the laminar air flow hood were off to minimize photo bleaching. The bottle was mixed by inversion 20 times. Before plating, 100 µL of the 0.125 x 10^5 cells/mL suspension was added to the bottle below the surface of the medium and mixed by inverting 20 times. This gave a final cell concentration of 12.5 cells/mL or 25 cells per well of a 6 well plate if 2 mL of semi solid medium is dispensed. The reservoirs of the 6-well plates were filled with a total of 5 mL PBS solution (D8537, Sigma) to minimize evaporation. After semi solid medium was dispensed, the plates were rotated to ensure that the medium covered the entirety of every well and any large bubbles were then manually removed with a 200 µL air displacement pipette. One bottle of medium gives enough to dispense 8 x 6-well plates. Plates were then incubated at 36.5°C, 7.5% CO_2 for 13 days. After 13 days, colonies on the semi solid medium were screened on the ClonePix™ 2 instrument (Molecular Devices) applying the criteria outlined in Table 2.2.8.2 for selection of colonies. The 600 ranked highest colonies by FITC 1000 exterior mean intensity were picked into 96-well plates in 200 µL per well of FDB-MAP supplemented with 8 mM L-glutamine (G7513, Sigma), 2x GlutaMax (35050, Life Technologies) and 175 nM methotrexate. The outer wells of the plates were excluded from picking due to evaporation. After picking, clonal cells were set up to be expanded into 96 well plates.

Table 2.2.8.2. The settings used on the ClonePix™ 2 2 to pick colonies into 96-well plates.

Criteria	Excluded colonies
Edge excluded	TRUE
Too big	Area > 0.15 mm2
Too small*	Area < 0.01 mm2
Irregularity 1	Compactness < 0.6
Irregularity 2	Axis ratio < 0.6
Proximity	Proximity < 1.00 mm
FITC exposure time	1000 ms
Aspirate volume	5 µL
Dispense volume	7 µL
Dispersal volume	20 µL
Dispersal cycles	5
'Bath cycles'	3
'Purge cycles'	3

2.2.9   Clone expansion in 96-well plates

About a week post-picking, the 96-well plates were monitored for growth on a Cell Metric®
CLD cell imager (Solentim) every 2 to 4 days for up to 4 weeks following picking. Up to 500
colonies were screened for productivity (recombinant protein concentrations) and colonies were
allowed to reach a confluence of at least 50% before sampling. A 50 µL sample of supernatant
was transferred directly into a black 96-well plate for Octet measurements taking great care not
to disturb the cells at the bottom of the well. When removing a larger sample volume than 50
µL, an equivalent volume of growth medium (FDB-MAP + 8 mM glutamine) was added to each
sampled well. The 50 µL sample was diluted to 200 µL for Octet measurement.

2.2.10 Clone expansion from 96-well plates to T-25 flasks

Once 500 cell lines had been screened, the top 120 were identified by Octet titer and
transferred to flat bottomed 24-well plates (Greiner-Bio). 1 mL of medium was added to each
well and warmed for ≥30 minutes in an incubator at 37°C. Cells were dislodged from the 96-well
plate with a P200 pipette (Thermo Fisher Scientific) and the entire well contents transferred into
a well in a 24-well plate containing warmed (37°C) FDB-MAP medium. The 96 well plates were
re-feed with ~0.2 mL suspension from the 24-well plate as a 'back-up'. 24-well plates were
incubated in a static incubator at 37°C, 7.5% CO_2, 80% humidity. The 24-well plates were
progressed to T-25 (Nunc, Thermo Fisher Scientific) flasks 4 days later. Cells were dislodged from
the 24-well plate with a P1000 pipette and transferred to a T-25 flask. Wells were re-fed in the
24-well plate with ~1 mL suspension from the T-25 flask as a 'back-up'. T-25 flasks were incubated
in a static incubator at 37°C, 10% CO_2, 80% humidity.

2.2.11 Clone expansion into shake culture

5 days after being transferred into a T-25 flask, VCD was determined using a Vi-Cell instrument. Cells were seeded at 0.2×10^6 cells/mL in a total volume of 10 mL in TubeSpin® shake flasks (TPP, Sigma). In cases where VCD was not high enough to progress a total culture volume of 5 mL was used instead. Spin tubes were then incubated at 37°C, 5% CO_2, 80% humidity, 200 rpm (25 mm throw) in a Certomat® CT Plus incubator (Sartorius, DE). Tube spins were sub-cultured on a 3 to 4 day regime (avoiding the weekend) and cells were passaged at 0.2×10^6 cells/mL in a total volume of 10 mL until characterised in terms of product titre using a fed-batch MWP screen.

2.2.12 24-deep well plate fed-batch screening

TubeSpin® cultures were counted with a Vi-Cell instrument and the wells of a 24-deep well plate seeded with a starting concentration of 0.5×10^6 cells/mL with a total volume of 2.5 mL. 24-deep well plates were then transferred to a Certomat® CTplus shaking incubator at 200 rpm, 37°C, 10% CO_2, 80% humidity. The shaking incubator was custom fitted with a Duetz-sandwich cover system to keep the plates in place during high shaking speeds, and to maintain sterility while allowing gas exchange. The feeding regime used is outlined in Table 2.2.12.1 below. After the end of the 14 day fed batch culture, the supernatants were measured for titre on an Octet system.

Figure 2.2.1. The configuration of the 24-well plate Duetz-sandwich cover system.

Table 2.2.12.1. A typical schedule of feeding 24-deep well plate fed-batch cultures

Day	Feed A, 50 μL	Feed B, 20 μL	L-glutamine, 25 μL	Glucose, 30 μL
0				
1				
2				
3	X	X	X	
4	X			
5	X		X	
6	X			
7	X		X	
8	X	X		X

9	X		X
10	X		
11	X		X
12	X	X	
13	X		X
14	Harvest		

2.2.13 ambr® 15 culturing

Materials used for ambr® 15 studies were as follows; Feed 7A (GE: SH31026.01), Feed 7B (GE: SH31027.01), 50% glucose (Sigma: G8769), 200 mM glutamine (Sigma: G7513), 1 M sodium carbonate (Sigma, S7795), 10% ACDF antifoam (GE, 50897.01) diluted 1 in 2 in water for use, PBS (Sigma, D8537).

The day before inoculation each vessel was filled with 10.06 ml of FDB-MAP medium without selective pressure followed by 10 μL of antifoam. The medium was then incubated overnight to allow pH and DO to stabilise at 37°C. After stabilization, a 60 μL sample of medium was removed to check the initial pH offline. On day 0, all vessels were seeded with 5.0 mL of cell culture at a concentration of 0.5×10^6 cells/mL, which had been counted in duplicate with a Vi-Cell instrument with no more than 10% error. The cells were cultured for at least two 4-day passages before seeding for them to adapt to shake flask environment. Fed-batch culture in the ambr® 15 were kept running for a total of 14 days. The feeding regime is described in Table 2.2.13.1 below.

Table 2.2.13.1. Ambr feeding, sampling and maintenance. All columns specify additions to the bioreactor with the exception of metabolites, titre, offline pH and cell count which were subtracted from the bioreactor volume.

Day	Feed A, 2% vol.	Feed B, 0.2% vol.	Metabolites, 200 µL	Antifoam, 30 µL	Titre, 200 µL	Offline pH, 60 µL	Cell count	Glutamine, 75 µL
1						X		
2	X	X	X	X		X	X	X
3	X	X	X	X		X		X
4	X	X	X	X		X	X	X
5	X	X	X	X		X		X
6	X	X	X	X	X	X	X	X
7	X	X	X	X		X		X
8	X	X	X	X	X	X	X	X
9	X	X	X	X		X		X
10	X	X	X	X	X	X	X	X
11	X	X	X	X		X		X
12	X	X	X	X	X	X	X	X
13	X	X	X	X		X		X
14					X	X	X	

2.2.14 Beacon® CLD workflow

The Beacon® cell line development workflow consisted of a number of distinct steps as follows: cleaning of the system, wetting of the chips, priming of the system with media, collecting calibration images, loading single cells with OptoElectro Positioning (OEP), culturing the cells, assaying for protein secretion, selecting clones for export, exporting cells.

The Beacon® system was initially made aseptic by cleaning the system. To achieve this, the system was loaded with plastic flush chips in each nest and 10% bleach (Clorox® Concentrated Germicidal Bleach, CLO31009CT) was flushed through the system and incubated for 1 hour. The system was then extensively flushed in two stages with autoclaved DI water to remove any residual bleach.

Next. the system was loaded with chips and the surface prepared by wetting the chip. The fluidic lines were purged with 100% CO_2 followed by loading of an OptoSelect Chip on to each nest. BLI Wetting Solution (SSRL22 (BLI PN 520-00009)) was then flushed onto the chip and incubated at 50°C to drive the chemical conjugation with the proprietary surface coating. The chip was then flushed with a Wetting Additive (1x Dulbecco's Phosphate-Buffered Saline (Gibco PN 14190136) with 0.1% Pluronic™ F-127, 0.2 µm filtered (Thermo Fisher Scientific PN P6866). Finally, the system was primed with assay media by filling the syringe pump with 1 mL of media and flushing to waste prior to flushing 1 mL of media across the chip. Once media was

on the chip and the system was idle, it was considered in a culture state of slow perfusion.

Three calibration images were then collected prior to loading cells: a background image, a fluorescence reference image, and a diffusion reference image. All images were collected using the Texas Red filter with an exposure time of 1000 ms. Images were collected sequentially for each field of view (FOV) with each chip divided into 22 FOVs. The background images were collected with only media on the chip. All reagents imported or exported through the well plate incubator were prepared in a 96 well plate (Falcon® 96 Well Clear Round Bottom Not Treated Microplate, with Lid, Individually Wrapped, Corning PN 351177). The fluorescence reference images were collected after importing the Spotlight Hu3 reagent (1x in assay media, BLI PN 520-00024) from the well plate incubator and allowing it to equilibrate across the chip and into the pens for 45 minutes. The diffusion reference images were collected after flushing the chip for 12 minutes after equilibration.

To load cells onto the system, the media on the system was changed to load media and the chip was primed with the media. All media on the system was buffered with 5% CO_2 to maintain the pH of the media. The cells were prepared in 96 well plate in load media to a concentration of 2.5 million cells per mL. The well plate was then loaded into the Beacon® Well Plate incubator set to a culture state (36°C with humidity). The nest was set to a temperature of 36°C and was maintained throughout the remaining workflow. Cells were then imported onto the chip at 1 μL/sec. The cells were then identified with the BLI CNN watershed algorithm and single cells were loaded into individual pens with OEP. For the pool screen (1st Beacon® workflow) the Two Stage penning algorithm was used to pen single cells and reduce the number of multiples loaded. For the clone characterization workflow (2nd Beacon® workflow) the Targeted penning algorithm was used to target 200-300 singles loaded per cell line per chip. The OEP settings were 4 V and 1.3 MHz with a cage speed of 8 μm/sec. All other settings were the default. After the load, a brightfield image was collected and analysed with the CNN watershed algorithm to report the number of cells per pen. After loading, the media was changed to culture media and the system was re-primed. Following this the system was set to a culture state. This perfuses media across the chip at 0.01 μL/sec for 72 μL (for 2 hours) with a faster periodic flush of 2 μL/sec of 128 μL.

After three days of culture, the media was exchanged with assay media and the system was re-primed. The system was then left to culture for 1 h prior to the start of the assay. The Spotlight Hu3 reagent was aliquoted into a 96 well plate (200 μL per well per nest) at a final concentration of 1X in assay media. Prior to import, a bright field image was collected at 23% illumination for 7.5 ms to allow for cell counting with the CNN watershed algorithm and a per pen cell count was

generated. The CLD diffusion gradient assay was run by importing 85 μL into the fluidic line from the well plate at 1 μL/sec. The reagent was then flushed at 0.014 μL/sec to allow the reagent to equilibrate across the chip for 45 minutes. Next, the chip was flushed with assay media at 0.167 μL/sec for 12 minutes. While maintaining the flush rate, the chip was imaged with the Texas Red filter for 1000 ms. Secretion assays were collected once a day on day 3, 4, and 5, depending on cell growth.

For analysis of secretion assay images, the images were first corrected for optical roll-off (flat fielding) using the background ($B_{x,y}$), fluorescence reference ($R_{x,y}$) images to normalize the assay images ($A_{x,y}$) by:

$$N_{x,y} = \frac{\left(A_{x,y} - B_{x,y}\right)}{\left(R_{x,y} - B_{x,y}\right)}$$

This correction was also applied to the diffusion reference image. The diffusion reference is used to partially correct for the additional loss of signal due to diffusion as a function of imaging time as later images will have had more diffusion out of the pens and therefore less signal. The average pen score (slope of the intensity in the specified region of the pen) of the normalized diffusion reference image for each FOV was calculated and subsequent assay images were flat file corrected and then normalized to the first FOV by a linear regression of the diffusion reference images.

To select clones for export for the pool screen, the clonality, final cell count (or doubling time), final secretion score, and final relative specific productivity (rQp) were calculated for each pen. Doubling time was calculated by dividing the time elapsed in hours between the post load image by the $\log_2$ of the final cell count. The 96 top clones (i.e. single cell loaded into a pen) were selected for export (48 by score and 48 by rQp) with a minimal threshold of 6 cells in the final cell count.

To prepare the cells for export, the Optoselect chips were temporarily removed from the system and stored upright at 37°C. The Beacon® was then loaded with flush chips and flushed with 10 cycles of autoclaved DI water and then re-primed with export media. The Optoselect chips were then reloaded on to the system and re-primed to remove any air bubbles in the system. The chips were then treated with 10x trypsin for 30 minutes to make the cells less sticky for unpenning. Next, each chip was pruned using OEP to remove any cells in the top third of each pen to reduce the chances of clonal contamination during export. For each pen selected for export, the cells were unpenned by OEP with a minimum threshold of 2 cells and a maximum of 6 unpenning attempts. If less than 2 cells were detected in the channel, the chip channel was

flushed to waste and the system went to the next pen. If enough cells were detected in the channel above the pen, a leading volume (10 to 13 µL, variable depending on pen position on chip) was flushed to waste prior to pushing a 5 µL package volume containing the cells for export into a specific well in the well plate incubator with 195 µL of export culture media. After each successful export, a blank export (same fluidic actions but with no OEP) was performed with a 5 µL package volume being collected in the well plate to determine the rate of clonal contamination post-export.

Post-export, the well plate was removed from the well plate incubator and stored in a standard incubator at 37°C , 5% CO_2 for 10 days. At this point, growth was measured and viable colonies were scaled up for future evaluation.

2.2.15 Conditioned medium

Cells were seeded at 3×10^5 viable cells/mL in a 50 mL volume of a 250 mL shake flask in FDB-MAP medium and cultured for 2 days. After culturing, the cells were centrifuged for 10 min at 300 g and 4°C. The supernatant was then filtered through a 0.2 µM (Sartorius, DE) filter and the filtered medium was then stored for up to one week at 4°C. Prior to use on the Beacon® or in the export well plates, the conditioned medium was diluted 1:5 in fresh medium to give a final concentration of conditioned medium of 20% (v/v).

2.3 OCTET MEASUREMENTS TO DETERMINE PRODUCT CONCENTRATIONS

Protein containing samples were centrifuged at 1400 g for 5 min and 200 µL of sample was placed into a 96-well plate. The final columns 11 and 12 were filled with FDB-MAP media and regeneration buffer (10 mM glycine, pH 1, Sigma, G8790) respectively. The media and buffer is used in the Octet to wash the tips and remove bound ligand in between measuring different rows. Before measurement, Dip and Read™ Protein A biosensors (PALL FortéBio, USA) were incubated in media for 10 minutes. Titre measurements were done using an Octet® QKe system at 0.6 Hz frequency, 30°C, 400 rpm and the results analysed using the Octet QKe system analysis (version 9.0) with a 5PL unweighted curve fitting curve. For concentration standards, IgG1-kappa and IgG4-lambda was used for Etanercept and BlosozumAb respectively. For preparation of standards a concentration range of 125 µg/mL, 100 µg/mL, 75 µg/mL, 50 µg/mL, 25 µg/mL, 10 µg/mL, 5 µg/mL and 1 µg/mL was used.

3 META-ANALYSIS OF PUBLICLY AVAILABLE CHINESE HAMSTER OVARY (CHO) CELL TRANSCRIPTOMIC DATASETS FOR IDENTIFYING ENGINEERING TARGETS TO ENHANCE RECOMBINANT PROTEIN YIELDS

Note: The work described in this chapter has been published in the peer reviewed journal Biotechnology Journal as following;

3.1 ABSTRACT

Transcriptomics has been extensively applied to the investigation of the CHO cell platform for the production of recombinant biotherapeutic proteins to identify transcripts whose expression is regulated and correlated to (non)desirable CHO cell attributes. However, there have been few attempts to analyse the findings across these studies to identify conserved changes and generic targets for CHO cell platform engineering. Here we have undertaken a meta-analysis of CHO cell transcriptomic data and report on those genes most frequently identified as differentially expressed with regard to cell growth (μ) and productivity (Qp). By aggregating differentially expressed genes from publicly available transcriptomic datasets associated with μ and Qp, using a pathway enrichment analysis and combining it with the concordance (the arithmetic mean of expression values), we have identified a refined target gene and pathway list whilst determining the overlap across CHO transcriptomic studies. We find that only the cell cycle and lysosome pathways show good concordance. By mapping out the contributing genes we have constructed a transcriptomic 'fingerprint' of a high-performing cell line. This study provides a starting resource for researchers who want to navigate the complex landscape of CHO transcriptomics and identify targets to undertake cell engineering for improved recombinant protein output.

3.2 INTRODUCTION

The most widely industrially utilised mammalian cell expression system for the manufacturing of biotherapeutic proteins is the Chinese hamster ovary (CHO) cell. The CHO cell expression system has now been used for the manufacture of a number of classes of biotherapeutic proteins, notably monoclonal antibodies (mAbs) (Walsh, 2014), however there remains the potential to further optimise this system, particularly for the expression of novel format and difficult to express (DTE) molecules. The appeal of the CHO cell for the manufacture of biopharmaceuticals is explained by several factors as outlined in detail in the introduction chapter of this thesis. In brief, first, CHO cells have been in use as protein expression 'factories' for several decades, meaning there is an established precedent to using this system and a track record of approval from regulatory agencies. Secondly, CHO cells have appropriate specific productivity, can grow in suspension in chemically defined, serum-free media (Dumont *et al.*, 2016). CHO cells can now deliver high recombinant product yields, with reports of recombinant antibody yields of >10 g/L compared to other systems such as HEK 293 where yields of approximately 1 g/L have been reported (Huang *et al.*, 2010; Steger *et al.*, 2015). They also have the ability to produce human like glycosylation patterns that are bio-compatible with the human immune systems (Kim *et al.*, 2012). However, current CHO cell research is still driven by a need to reduce development times (and costs), increase recombinant protein yields/quality, enhance cell growth and express novel molecules.

CHO cell research is presently experiencing a paradigm shift in terms of how the cell factory is understood due to the availability of a variety of omics data. The Chinese hamster, CHO-K1 (Xu *et al.*, 2011) and various other cell line genomes have been sequenced and published along with a library of proteomic, transcriptomic and metabolomic data (Kildegaard *et al.*, 2013; Lewis *et al.*, 2016; Hefzi *et al.*, 2016). These studies and databases provide the community with a wealth of information around the CHO cell platform and allow for the rational and precise fine-tuning of the CHO recombinant protein expression platform. However, in order to identify pathways and targets for CHO cell engineering, the investigator needs to know what genes are being expressed under which conditions and how this affects phenotype. Investigations into the CHO transcriptome have been underway since 2006 (Wong *et al.*, 2006) using in house CHO cDNA microarrays and cross-species microarrays. More recently, RNAseq as a technique has been applied to CHO transcriptomics, with the first reports in 2010 (Birzele *et al.*, 2010). According to the CHO bibliome (Golabgir *et al.*, 2016) up to 2015, 52 CHO gene expression and transcriptomic publications had been identified with datasets being generated for panels of CHO cell lines with

different growth and production characteristics (Clarke *et al.*, 2011a; Doolan *et al.*, 2013), under cold shock (Yee *et al.*, 2009), butyrate treatment (Kantardjieff *et al.*, 2010), adaptation to suspension (Shridhar *et al.*, 2017) and other culture conditions (Becker *et al.*, 2014; Fomina-Yadlin *et al.*, 2015). Here we describe a meta-analysis of different CHO transcriptome datasets to identify common pathways and genes identified as underpinning CHO cell growth and product yield. These genes and pathways represent priority targets for cell engineering and manipulation to further enhance the CHO platform for manufacturing of biotherapeutic proteins.

3.3 METHODS

3.3.1 Identification of Publicly Available CHO Transcriptomic Datasets for Analysis

The CHO bibliome (Golabgir *et al.*, 2016) was used to identify CHO based transcriptomics publications up to 2015. Additional datasets sourced from those published 2015 – 2017 were also included in the analysis. The list of final genes and their datasets of origin are provided in Table 3.4.1.1 and Supplementary file S1. Transcriptomic studies that used a cross-species microarray approach were omitted since the accuracy of cross-species microarray data is still under debate. From these datasets, we extracted lists of differentially expressed genes and assigned them to one of two groups based on their association with either specific productivity (Qp) or growth (μ). This was undertaken in order to accurately discern the impact of genes to a specific phenotype as it has been shown that Qp can come at the cost of μ and vice versa (Du *et al.*, 2015). An expression value (+1 or -1) was assigned to all genes and corresponds to the upregulation (+1) or downregulation (-1) of the gene. This did not consider the absolute fold change in the datasets, only the direction in which expression changes were observed. Comparing fold change values across datasets without having access to the raw data of the omics experiment would not be meaningful and, unfortunately, such data is not available from most of the datasets included in this study.

After the assembly of an aggregate gene list, two parameters were calculated for unique gene entries in the Qp and growth categories:

Frequency; the number of times a gene appears across selected datasets.

Concordance; the arithmetic mean of expression values (from the assigned -1 or +1 expression value assigned as described above). A concordance threshold of -0.2 and 0.2 was established to differentiate which genes show an agreement in expression data. This corresponds to a minimum of three fifths (0.6) of the gene entries in the group having an agreement of the expression value.

These two parameters form the cornerstone of our analysis.

3.3.2   Ensuring Consistent Gene Annotation for Analysis

Most of the available publications have annotated the gene sets as mouse, rat or human gene ID's or by using official gene symbols. To compare the different gene lists all datasets had to be re-annotated to a single format so that these could be compared and analysed. Re-annotation was performed using the Mouse Genome Information database batch gene lookup tool (http://www.informatics.jax.org/batch) into an Entrez ID format. This format is preferable to an official gene names-based annotation because gene name designations tend to change with time and may cause duplications of genes under synonym entries. ID's identified as pseudogenes and non-coding genes were discarded. Entrez ID's given in publications were not changed. The annotated master gene list is provided in Supplementary file S2.

3.3.3   Pathway Enrichment Analysis

For pathway enrichment, entrez ID's of genes with a frequency of 1 in the growth and productivity groups (GG, PG) were rejected and these genes account for roughly half of the master gene list. Entrez ID's of genes that had a frequency of ≥ 2 were submitted to DAVID Knowledgebase 6.8 (https://david.ncifcrf.gov/) for functional annotation analysis with the option to chart KEGG pathway enrichment as we wished to identify conserved differentially expressed genes across CHO cell lines and conditions. KEGG was used as the functional annotation database because the use of KEGG in pathway enrichment is widespread for interpreting the biological meaning of transcriptomic datasets and is well curated (Kantardjieff *et al.*, 2010; Harreither *et al.*, 2015; Zhang *et al.*, 2017). Default functional annotation parameters were used (Threshold count = 2 and EASE value of 0.1). Pathway charts were generated in DAVID using the KEGG database. We then included an overlay of concordance values for each gene present in the meta-analysis and in the pathway enrichment to visualise the dynamics of pathway expression. Once the gene list was submitted to DAVID, the number of viable targets was reduced due to insufficient coverage in the database. At the time of undertaking this study, 7720 genes were present in the KEGG pathways for *Mus musculus*.

3.4 RESULTS

3.4.1 Datasets used in this study

We wanted to screen the publicly available CHO transcriptomic data to aggregate and analyse patterns of changes at the transcript level relating to high specific productivity (Qp) and growth rate (μ). The working datasets used in this study consisted of publicly available species-specific transcriptomic data that was generated using CHO cell lines expressing recombinant proteins under various conditions. The reported transcriptomic experiments were set up using a number of different approaches. Some experiments compared a panel of cell lines with a range of parameter values, while in others cells were exposed to known productivity or phenotype changing treatments such as cold shock or sodium butyrate to enhance their recombinant protein yields or change cell growth. The selected publications for data mining are presented in Table 3.4.1.1. Out of the 19 datasets, only 4 used RNAseq while 2 compared the use of RNAseq to a microarray in the same experiment. Affymetrix based custom microarrays were the most often used across the datasets. In the Qp group, 2 studies used copper to reduce lactate levels while 4 studies used butyrate to enhance Qp. One dataset was generated under high osmotic stress and 4 induced cold shock in the culture. Six of the studies directly investigated the differences in transcriptomic gene expression amounts between cell lines with different Qp. In total, we assigned 16 lists to the Qp category and 6 to growth. Growth datasets included in this study compared a panel of cell lines with different growth characteristics; no growth enhancing processes were used in any of the sources. Lists from 3 sources are present in both groups because they contained data that was partitioned for these phenotypes separately. Genes present in these lists were then assigned values for their frequency and concordance as outlined in section 3.4.1. The top most frequent genes across the datasets (≥ 5) are listed in Table 3.4.1.2 along with their individual concordance values. We note that definition of Qp as 'high' differs between studies and is a subjective judgement made by the investigators of each study.

Taking into account the clonal variation of the cell lines used in the datasets is also important. These are included in the supplementary table 1. We can see that the dominant cell line was CHO-DXB11, which was used in 8 studies. These cells are DHFR deficient so that MTX can be used as a selection tool. DHFR deficient cells were used in 11 of the 19 studies. Unfortunately, 7 studies failed to self-report the type of CHO cell line they were using.

Table 3.4.1.1. List of publications selected for transcriptomic meta-analysis in this study.

DATABASE ENTRY	TITLE	TYPE	AUTHOR/ DATE
	Predicting cell-specific productivity from CHO gene expression	Microarray - Wye2aHamster CHO K1, CHO DUX Fed-Batch shakeflask mAB and fc-fusion product.	(Clarke *et al.*, 2011a)
E-GEOD-30321	Gene expression profiling of Chinese Hamster Ovary production cell lines	Microarray - Wye2aHamster CHO DUX, CHO K1 Bioreactor, shakeflask mAB, fusion protein, growth factor, coagualation factor	(Clarke *et al.*, 2011b)
E-GEOD-37251	Transcriptomic analysis of clonal growth rate variation during CHO cell line development	Microarray - Wye3aHamster CHO K1 Batch shakeflask	(Doolan *et al.*, 2013)
	Microarray and proteomics expression profiling identifies several candidates, including the valosin-containing protein (VCP), involved in regulating high cellular growth rate in production CHO cell lines	Microarray - Wye2aHamster; proteomics CHO DUKX Batch mAB	(Doolan *et al.*, 2010)
	Transcriptome and proteome analysis of Chinese hamster ovary cells under low temperature and butyrate treatment	Microarray - Custom-made Affymetrix® CHO mAB CHO-S Shakeflask	(Kantardjieff *et al.*, 2010)
	Translatome analysis of CHO cells to identify key growth genes	Microarray - Niblegen 13k CHO DG44 mAB Batch	(Courtes *et al.*, 2013)
	Transcriptome and proteome profiling to understanding the biology of high productivity CHO cells	Microarray - 15 K CHO cDNA, proteomics Batch ATCC CRL-9096 GFP	(Nissom *et al.*, 2006)
Bioproject 79563	Into the unknown: expression profiling without genome sequence information in CHO by next generation sequencing.	Microarray - CHO affymetrix; RNAseq mAB Fed batch Bioreactor	(Birzele *et al.*, 2010)
	CHO Gene Expression Profiling in Biopharmaceutical Process Analysis and Design	Microarray - CHO Affymetrix Fed batch mAB Bioreactor	(Schaub *et al.*, 2010)
	Genomic and proteomic exploration of	Microarray - CHO cDNA	Yee et al(Yee *et*

		library, proteomics T-flask	*al.*, 2008), 2008
	Comparative transcriptome analysis to unveil genes affecting recombinant protein productivity in mammalian cells	Microarray - CHO cDNA library mAB Shake flask Batch	(Yee *et al.*, 2009)
	Cell Line Profiling to Improve Monoclonal Antibody Production	Microarray - Custom-made Affymetrix® CHO, proteomics Spin-tube Fed batch mAB	(Kang *et al.*, 2014)
	Microarray expression profiling identifies genes regulating sustained cell specific productivity (S-Qp) in CHO K1 production cell lines	Microarray - CHO wye2a CHO K1 Bioreactor Fed batch mAB	(Doolan *et al.*, 2012)
	Transcriptome analysis of a CHO cell line expressing a recombinant therapeutic protein treated with inducers of protein expression	RNAseq RANK-Fc fusion protein Shake flask Batch	(Fomina-Yadlin *et al.*, 2015)
	Effect of Temperature Downshift on the Transcriptomic Responses of Chinese Hamster Ovary Cells Using Recombinant Human Tissue Plasminogen Activator Production Culture	RNAseq CHO TF70R rh-tPA protein Shake flasks Batch	(Bedoya-López *et al.*, 2016)
	Cell culture and gene transcription effects of copper sulphate on Chinese hamster ovary cells	Microarray - Custom-made Affymetrix® CHO IgG-fusion protein B0 Fed Batch Shake flask	(Qian *et al.*, 2011)
	Transcriptomic responses to sodium chloride-induced osmotic stress: A study of industrial fed-batch CHO cell cultures	Microarray - Custom-made Affymetrix® CHO Fc-fusion protein Batch, fed batch Shake flasks, bioreactor	(Shen *et al.*, 2010)
	Effects of Copper on CHO Cells: Insights from Gene Expression Analyses	Microarray - Custom-made Affymetrix® v3 CHO; RNAseq DUXB11 host derived mAB Shake flask Fed Batch	(Yuk *et al.*, 2014)
	CHO gene coexpression database	Microarray - WyeHamster2a	www.cgcdb.org

Table 3.4.1.2. Frequency analysis results from datasets relating to high growth rate (µ) and specific productivity (Qp) phenotypes as described in the text.

		Frequency			Concordance		
Gene	Name	Sum	Growth	Qp	All	Growth	Qp
Cd36	CD36 molecule	9	2	7	-0.50	-1.00	-0.33
Ctsl	cathepsin L	8	4	4	-0.43	-0.50	-0.33
App	amyloid beta (A4) precursor protein	7	5	2	-0.67	-0.60	-1.00
Eif6	eukaryotic translation initiation factor	7	2	5	0.33	0.00	0.50

	6						
Nedd4	neural precursor cell expressed, developmentally down-regulated 4	7	2	5	0.00	1.00	-0.50
Hnrnpk	heterogeneous nuclear ribonucleoprotein K	6	4	2	0.60	1.00	-1.00
Lamp1	lysosomal-associated membrane protein 1	6	4	2	-0.60	-0.50	-1.00
Hdgf	hepatoma-derived growth factor	6	3	3	0.33	0.33	0.33
Mcm5	minichromosome maintenance complex component 5	6	3	3	0.33	1.00	-0.33
Rab10	RAB10, member RAS oncogene family	6	3	3	-0.67	-0.33	-1.00
Slc25a20	solute carrier family 25 (mitochondrial carnitine/acylcarnitine translocase), member 20	6	3	3	-1.00	-1.00	-1.00
Eif5a	eukaryotic translation initiation factor 5A	6	3	3	0.20	0.33	0.00
Ldha	lactate dehydrogenase A	6	2	4	-0.33	0.00	-0.50
Atp6ap2	ATPase, H+ transporting, lysosomal accessory protein 2	6	2	4	-1.00	-1.00	-1.00
Acaa2	acetyl-Coenzyme A acyltransferase 2 (mitochondrial 3-oxoacyl-Coenzyme A thiolase)	6	0	6	0.33	N/A	0.33
Glul	glutamate-ammonia ligase (glutamine synthetase)	5	4	1	-0.60	-1.00	1.00
Cbx5	chromobox 5	5	3	2	1.00	1.00	1.00
Cct3	chaperonin containing Tcp1, subunit 3 (gamma)	5	3	2	0.20	1.00	-1.00
Hspa8	heat shock protein 8	5	3	2	0.00	0.33	-1.00
Kpnb1	karyopherin (importin) beta 1	5	3	2	0.60	1.00	0.00
Lamp2	lysosomal-associated membrane protein 2	5	3	2	-1.00	-1.00	-1.00
Mcm7	minichromosome maintenance complex component 7	5	3	2	0.60	0.33	1.00
Rsu1	Ras suppressor protein 1	5	3	2	-0.20	-0.33	0.00
Tuba1b	tubulin, alpha 1B	5	3	2	0.50	0.33	1.00
Retsat	retinol saturase (all trans retinol 13,14 reductase)	5	3	2	-1.00	-1.00	-1.00
Mrpl14	mitochondrial ribosomal protein L14	5	3	2	-0.20	-0.33	0.00
Atic	5-aminoimidazole-4-carboxamide ribonucleotide formyltransferase/IMP cyclohydrolase	5	3	2	0.60	1.00	0.00
Mthfd1	methylenetetrahydrofolate dehydrogenase (NADP+ dependent),	5	3	2	0.50	1.00	-1.00
Bsg	basigin	5	2	3	-0.20	0.00	-0.33
Ccnb2	cyclin B2	5	2	3	0.50	0.00	1.00
Itgb1	integrin beta 1 (fibronectin receptor beta)	5	2	3	-0.50	0.00	-1.00
Npc1	NPC intracellular cholesterol transporter 1	5	2	3	-0.50	-1.00	0.00
Ccl2	chemokine (C-C motif) ligand 2	5	2	3	-0.20	-1.00	0.33
Cdc20	cell division cycle 20	5	2	3	1.00	1.00	1.00
Hadhb	hydroxyacyl-Coenzyme A	5	2	3	-0.60	-1.00	-0.33

	dehydrogenase beta subunit						
Anxa2	annexin A2	5	1	4	-0.50	1.00	-1.00
Serpinh1	serine (or cysteine) peptidase inhibitor, clade H, member 1	5	1	4	0.60	1.00	0.50
Grb2	growth factor receptor bound protein 2	5	1	4	-0.20	1.00	-0.50
Kpna4	karyopherin (importin) alpha 4	5	1	4	-0.20	-1.00	0.00
Vim	vimentin	5	0	5	0.00	N/A	0.00

3.4.2 Pathway enrichment analysis

Figure 3.4.1. Diagram showing workflow of the analaysis of the 19 CHO transcriptomic datasets

CHO transcriptomic datasets of high growth and specific productivity cell lines

Aggregation of gene expression direction & frequency ->
-> Pathway enrichment analysis ->
Concordance analysis

High producer transcriptomic fingerprint and gene frequency distributions
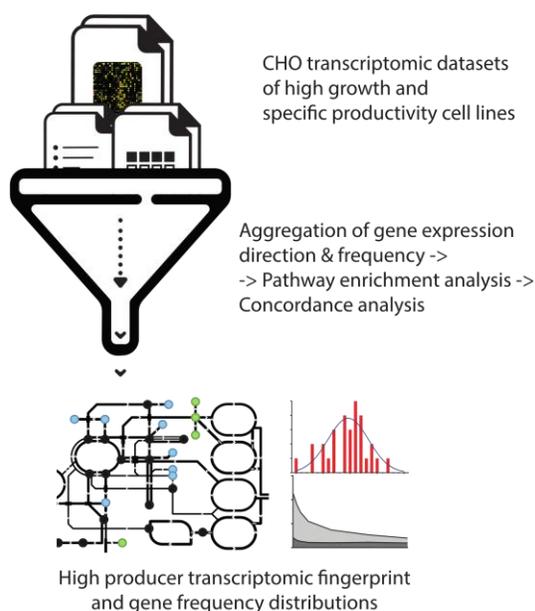
We set out to determine whether particular pathways were enriched within the lists that we extracted from the datasets. Enrichment analysis tries to identify whether the genes in your dataset tend to overrepresent certain pathways than if assembled together by chance. It has been suggested that single gene overexpression or knock-down alone is unlikely to govern complex changes underpinning phenotypes such as growth or recombinant protein yield (Boyle *et al.*, 2017), except in cases where a cell line has a specific bottleneck or a product specific requirement. On-the-other-hand, groups of genes (or pathways) can be co-expressed together with moderate fold change values (Clarke *et al.*, 2011b), where the cumulative contribution effect results in an improvement in the phenotype required (e.g. growth, productivity). Thus, in terms of developing a cell line engineering strategy, changes at the transcriptomic level that reflect (a) high value single gene targets, (b) global transcriptomic analysis of groups of genes that are co-expressed, and (c) entire pathways that are enriched within the expression data, should be considered.

To analyse the results from the selected transcriptomic studies, the differentially expressed gene lists from these sources were aggregated and analysed for frequency and concordance of expression direction. In total, 4783 unique differentially expressed genes were identified (4044 *Qp* and 1406 growth associated as visualised Figure 3.4.2). Between these groups, an overlap of 667 genes was established. The frequency distributions for these groups are reported in Table 3.4.2.1. A detailed annotation master list reporting on the frequency, direction of expression and concordance of discovered genes across the datasets analysed here is provided in Supplementary file S2. The results from the pathway enrichment analysis using KEGG pathways data are presented in Table 3.4.2.2 and are more extensively described and reported in the Supplementary tables S3 & S4. We have integrated these enrichment results onto pathway maps, which enables a more integrative look at the interactions between the genes identified.

Table 3.4.2.1. Frequency distribution of unique genes found in the literature relating to transcriptomic changes associated with productivity and growth rate.

Frequency	Sum	Growth (μ)	Qp
1	3461	1166	3269
2	918	186	636
3	283	49	118
4	81	4	16
5	25	1	3
6	10	0	1
7	3	0	1
8	1	0	0
9	1	0	0

From the pathway enrichment analysis, a number of what might be considered 'unusual' pathways were identified including biosynthesis of antibiotics and Epstein-Barr (EB) virus infection. This can be explained by the fact that these pathways share a broad overlap with other major pathways. In the case of the EB virus infection pathway, half of the genes assigned are present in the cell cycle, while almost all hits in the biosynthesis of antibiotics pathway term are present in the general cell metabolism pathway. Therefore, we deemed these pathways as being non-specific and they were excluded from further considerations for identification of potential cell engineering targets. We have kept these non-specific pathways in the list to reflect a typical enrichment result and for reference, should anyone try to replicate or use our work in the future.

For those genes associated with the growth group, we observed that only a small number of relevant pathways were found to be enriched; the cell cycle, phagosome and lysosome (Benjamini-Hochberg adj. p-value <0.05). The cell cycle (0.42) and lysosome (-0.73) pathways

had high concordance within the data sets, while there was little concordance in the phagosome (-0.02) pathway for the genes being up- or down-regulated. In comparison, the only pathway that showed concordance in the *Qp* group was the lysosome (-0.36). The overlap between genes in these two pathways (cell cycle and lysosome) for both groups is shown in Figure 1B & 1C. In both cases, there were more genes in the *Qp* group for both pathways; 22 and 25 respectively for lysosome and the cell cycle. This is most likely a result of the fact that the *Qp* group is larger, therefore has more coverage of the pathways. We have used the pathway enrichments to explain changes in cellular mechanisms that could lead to fast growth or high specific productivity phenotypes and also compared genes identified in the study with engineering strategies that others have applied to engineer increased yields in recombinant CHO cell lines. The pathways are presented in more detail in the following sections.

Figure 3.4.2. A Venn diagram showing the number of unique genes in both Qp and growth (μ) categories
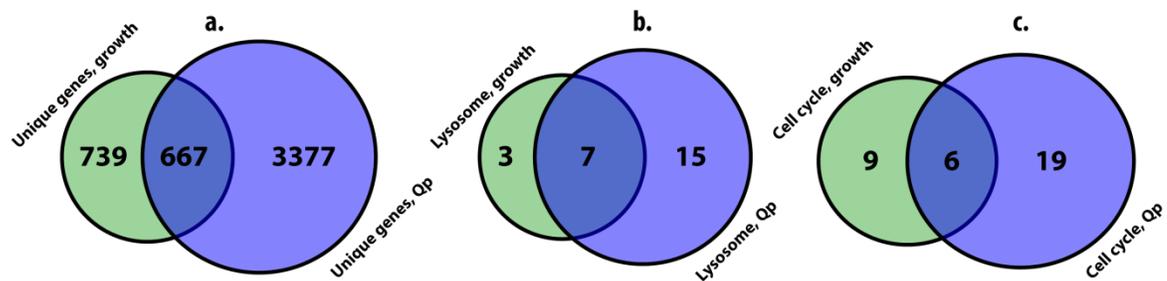
(A) and the lysosome (B) and cell cycle (C) pathway enrichments.

Table 3.4.2.2. Pathway enrichment results from datasets relating to high growth rate (μ) and specific productivity (Qp) phenotypes as described in the text

Pathway	Count	P-Value	FE	BH p-value	FDR	Concordance
Growth						
Cell cycle	15	1.60E-07	6.00	3.00E-05	1.90E-04	0.42
Phagosome	15	9.40E-06	4.30	9.10E-04	1.20E-02	-0.02
***Epstein-Barr virus infection**	14	3.70E-04	3.20	2.40E-02	4.60E-01	0.55
Lysosome	10	7.40E-04	4.10	3.50E-02	9.20E-01	-0.73
***Biosynthesis of antibiotics**	13	1.20E-03	3.00	4.60E-02	1.50E+00	0.46
Specific productivity (*Qp*)						
Cell cycle	25	5.00E-09	4.1	1.30E-06	6.50E-06	0.15
***Biosynthesis of antibiotics**	32	4.20E-08	3	5.30E-06	5.50E-05	-0.13
Lysosome	22	3.80E-07	3.7	3.20E-05	4.90E-04	-0.36
FoxO signaling pathway	20	2.50E-05	3	1.60E-03	3.30E-02	0.18
Steroid biosynthesis	7	2.10E-04	7.5	8.80E-03	2.70E-01	0.14

MicroRNAs in cancer	29	1.90E-04	2.1	9.40E-03	2.40E-01	0.11
Metabolic pathways	89	3.40E-04	1.4	1.20E-02	4.40E-01	-0.12
Fatty acid degradation	10	5.30E-04	4.1	1.70E-02	6.90E-01	-0.07
Fatty acid metabolism	10	7.20E-04	4	2.00E-02	9.30E-01	-0.07

*Pathways marked with * are non-specfic and are only included as a representation of a general enrichment result. High concordance values are marked in bold. PH – total genes in KEGG pathway. FDR – false discovery rate. FE – Fold enrichment, BH – Benjamini-Hochberg.*

3.4.3   Cell cycle pathway

The cell cycle map that depicts the concordance values of differentially expressed genes are presented in more detail in Figure 3.4.3. There are several functional clusters of genes in the KEGG cell cycle pathway that are present in the enrichment data. One such group is clustered around *P53*, one of the most studied genes in the scientific literature, due to its status as the "guardian of the genome" and P53's role in controlling the DNA damage checkpoint (Bieging *et al.*, 2014). MDM2 directly binds to P53 preventing it's mechanism of action; *MDM2* shows a strong downregulation concordance in the growth group (GG) and no concordance in the productivity group (PG), while *P53* is upregulated. P53 is known to be mutated in CHO-K1 cells and facilitates DNA repair but not UV-induced G2/M arrest or apoptosis (Chang *et al.*, 2008). It is unclear how expression of *P53* helps promote cell growth. Interestingly, the transcripts that code for proteins that lead to growth arrest as a response of p53 upregulation (*GADD45A* and *P21* (*CDKN1A*)) both show downregulation with good concordance. GADD45A and P21 can interact with PCNA to initiate DNA damage repair response and inhibit transition into S-phase (Chen *et al.*, 1995; Strzalka and Ziemienowicz, 2011). P130(RBL2) is known to interact with proteins of the EF2 family as part of a UV-induced DNA damage repair pathway to cause cell cycle arrest (Genovese *et al.*, 2006) and was strongly downregulated in the PG. On-the-other-hand, *CREBBP (EP300)* is upregulated in the PG even though it is a tumour supressing gene because of its ability to activate P53 through acetylation (Goodman and Smolik, 2000). Based on this it seems that the mechanisms associated with DNA repair growth arrest are inhibited in the GG while PCNA is upregulated due to its role in DNA synthesis as a processivity factor. The MCM genes are upregulated with strong concordance in the GG as well. MCMs together form a hexamer that acts as a helicase essential for the function of the replication fork in DNA synthesis (Lei, 2005). *MCM7* is also found upregulated in the PG. However, *MCM5* and *MCM3* are downregulated. It has been observed that overexpression of *MCM3* leads to inhibition of the G1/S checkpoint, while knockdown does not affect the entry or progression of said checkpoint (Li *et al.*, 2011). *MCM5* knockdown leads to S-phase arrest in CHO cells and overexpression was shown

to prevent over-duplication of centrosomes (Ferguson and Maller, 2008). Based on available data it is not clear how downregulation of these two genes would contribute to an increased *Qp* phenotype. DNA-PK(PRKDC) is known to be an upstream activator of p53 and the knockdown phenotype is known to be sensitive to UV irradiation and is downregulated in the PG group as well (Woo *et al.*, 1998). *MYC* was found to be upregulated in the GG, which is not surprising as it is a characterised oncogene that promotes DNA synthesis and has been implicated in DHFR/MTX associated gene amplification (Denis *et al.*, 1991).

Another cluster of genes appears to be involved in the entry/exit of the mitotic stage of the cell cycle. Cyclin B1 signals the irreversible start of cell division and CDC20 is responsible for activating the APC complex which degrades G2/M cyclins and signals start of anaphase, while MAD2 stalls the separation of the chromosomes until they are properly aligned (Castro *et al.*, 2005). All three of these genes showed upregulation with strong concordance in the GG as well as *YWHAE/14-3-3 ε*, which binds CDC25 proteins based on their phosphorylation state preventing a premature entry into mitosis before replication of the genome (Chen *et al.*, 2003; Sur and Agrawal, 2016). While the Cyclin B2 gene was found to be upregulated in the PG, *CDK1* was downregulated. Typically, *CDK1* downregulation is associated with a prolonged G2/M phase and it has been proposed that CDK1 can have an inhibitory effect on the secretory pathway which would decrease *Qp* (Enserink and Kolodner, 2010; Yeong, 2013). *PLK1* is upregulated in the PG which activates the CyclinB/CDK1 complex and the APC. This is supported by upregulation of *CDC20* in both groups. *BUB1B* is downregulated in *Qp* and inhibits the APC and PLK1 (Castro *et al.*, 2005; Bolanos-Garcia and Blundell, 2011). However, *CDC27* which is a core subunit of the APC and responsible for ubiquitin mediated degradation of B-Cyclins and degradation of CDC20 (Prinz *et al.*, 1998), is downregulated in the PG. Our meta-analysis therefore suggests that the cell cycle in CHO cells can be rewired in three major ways related to increased growth; upregulation of proteins that facilitate the passing of the G1/S checkpoint, upregulation of DNA synthesis and those that assure proper separation of chromosomes in anaphase.

A number of cell cycle based engineering strategies have been attempted in CHO cells that provides further evidence that this pathway has potential for engineering to improve desirable phenotypes. *MDM2* was overexpressed in batch cultures increasing viable cell concentration two times over control cells in spent media conditions (Arden *et al.*, 2007). *GADD45A* was used to arrest the cell cycle via inducible expression controlled by doxycyclin in CHO-TREx, showing a 110% increase in yields of Fc fusion protein Valpha (Kim *et al.*, 2014). Overexpression of *CDC20* in CHOd cells led to a 4-fold increase in the VCD of cells growing on plates by day 14 compared to cells transfected with antisense *CDC20* cDNA. The antisense cells also grew larger and had

more DNA per cell as shown by flow cytometry (Weinstein *et al.*, 1994). A small molecule inhibitor of CDK4/CDK6 was able to induce sustained G1/S checkpoint arrest for up to 4 days without causing cell death or decrease of product quality. As a result Qp was increased ~2 fold across a panel of cell lines (Du *et al.*, 2015). One of the most obvious candidates to induce cell cycle arrest are the cyclin dependant kinase inhibitor proteins. Fussenegger et al. have successfully overexpressed *P21* along with CCAAT/enhancer-binding protein α by tetracycline enhancing the yields of SEAP by 10-15 times (Fussenegger *et al.*, 1998). The overexpression of *BCL-XL* with *P27* was found to significantly increase SEAP yields in the same study. A similar method was applied to overexpression of *CDKN1B* with comparable results to *P21* overexpression induced cell cycle arrest (Mazur *et al.*, 1998). *E2F-1* was overexpressed in CHO-K1 cells leading to elevated cyclin A levels and bypassing the need for serum in the growth media (Lee *et al.*, 1996). Similar effects have been observed in CHO-K1 by overexpression of cyclin E (Renner *et al.*, 1995). Overexpression of *CDC25A* and *CDC25B* has successfully been used to increase recombinant protein yields as well, however cell lines displayed an increased incidence of chromosomal aberration (Lee *et al.*, 2013). Finally, *MYC* has been stably overexpressed in both suspension and adherent cells resulting in increased growth rate and VCD (Ifandi and Al-Rubeai, 2003).

Figure 3.4.3. Pathway enrichment map for the cell cycle pathway
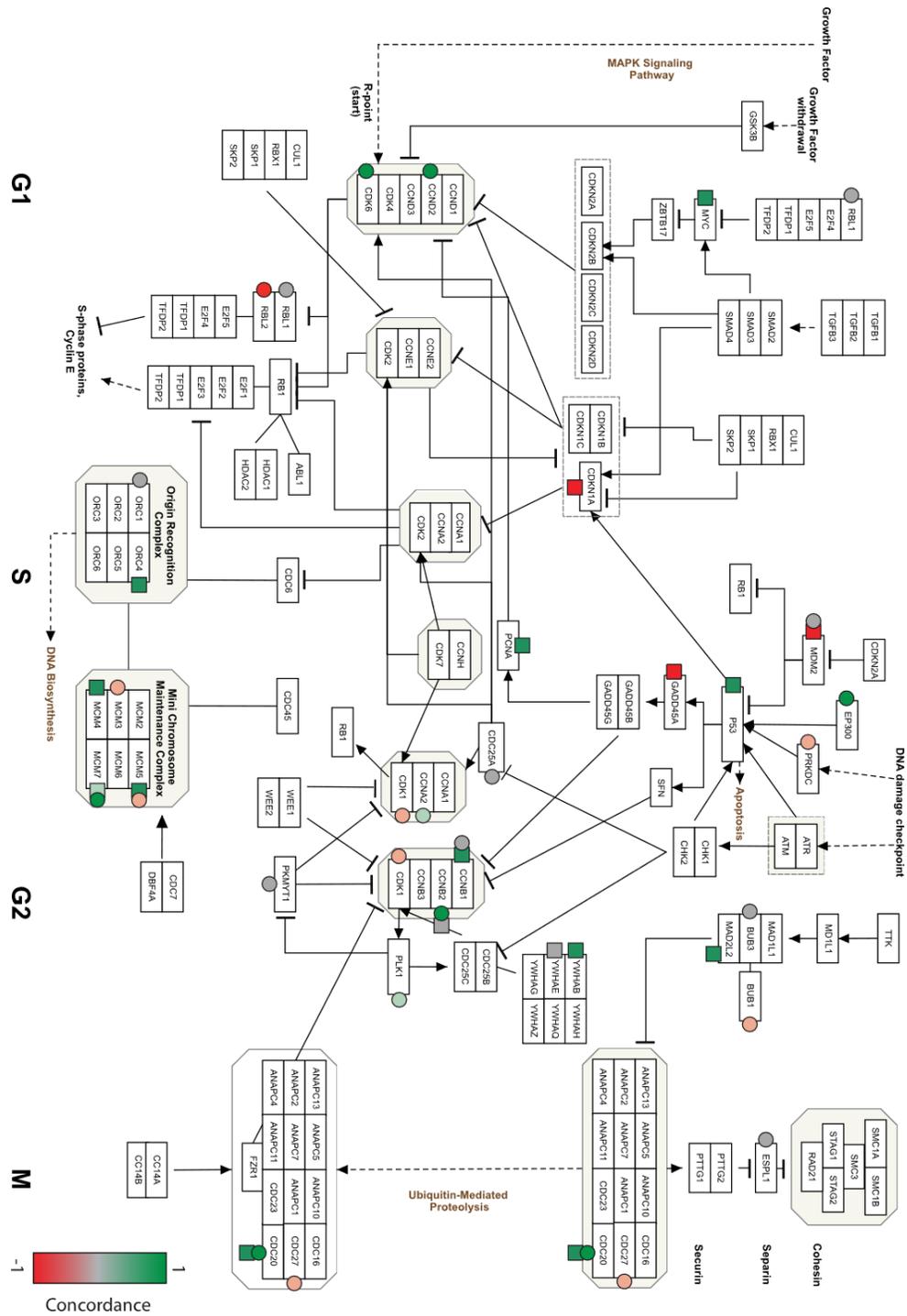
Hits for the growth group are shown in squares (□) and the productivity group is represented as circles (○). Overlap between shapes (⊡) indicates a hit in the same gene, while adjacent but non-overlapping shapes (○□) convey hits in the same gene family. Concordance values for each hit are shown as a colour value as visualised in the concordance bar. G1, S, G2 and M are cell division stages.

3.4.4 Lysosome pathway analysis

The KEGG lysosomal pathway graphic (Figure 3.4.4) provides an overview of the progression of endosome maturation and genes belonging to the pathway are roughly classified based on their functions. Cathepsins are some of the most pivotal proteins in the degradation and recycling machinery of the lysosome. Of these, *CTSL* (GG, PG) and *CTSA* (GG) were found to be downregulated. *CTSL* knockout mice have been shown to have hyperproliferation of hair follicle epithelial cells and basal epidermal keratinocytes (Roth *et al.*, 2000). Cathepsins have also been implicated in mAb degradation during production from CHO cells via proteomic analysis (Park *et al.*, 2017). Glycosylceramidase gene *GBA* was found to be downregulated and a number of sphingolipid metabolism genes can be found within the *Qp* group; ceramide synthase (*CERS2*) and sphingosine-1-phosphate (*SP1*), lyase-1 (*SGPL1*), and alkaline ceramidase 3 (*ACER3*) were downregulated, while *SGPHK1* was upregulated. This suggests an overall trend towards downregulation of ceramide levels and an increase in sphingosine-1-phosphate. Ceramide has been implicated in promotion of apoptosis, while S1P induces proliferation in HEK293 cells (Oskouian *et al.*, 2006). Yusufi et al. reported an increase in the levels of ceramide and it's derivatives in a high producing SH-87 cell line when compared to the host cell (Yusufi *et al.*, 2017). Another two genes involved in sphingolipid metabolism coding sphingolipid activator proteins (SAP's) were downregulated in the PG; prosaposin (*PSAP*) and GM2 ganglioside activator (*GM2A*). These genes are responsible for degrading lysosomal membrane bound glucocerebrosides. Accumulation of these lipids can lead to Gauche disease and are linked to mutations in *PSAP* and *GBA*, while GM2A deficiency is implicated in GM2 gangliosidosis (Xu *et al.*, 2016). These genes are mainly studied in neuronal context and their role in CHO cell metabolism is not clear.

The major lysosomal genes *LAMP1* and *LAMP2* were downregulated in both groups and represent some of the most frequent hits across the meta-study; 6 and 5 respectively. Lysosomal content has been shown to be negatively correlated with *Qp* in a tissue plasminogen producing CHO cell line along with *LAMP2* mRNA levels. The study also reported that glutamine depletion on its own is enough to increase levels of autophagy (Jardon *et al.*, 2012). The Niemann-Pick type C1 *NPC1* gene was downregulated in the GG; CHO cells lacking NPC1 have been observed to have impaired lipid recycling, accumulating in late endosomes. However, no data was given on any impact on cell growth (Pipalia *et al.*, 2007). *LAPTM4A* was found to be downregulated in both groups. Little is known about this protein, except that it is a transmembrane protein localized to the lysosome and possibly facilitates transport across the membrane. It has been shown to co-

precipitate with NEDD4, which was upregulated in growth and downregulated in the PG with a cumulative frequency of 7 across both groups. NEDD4 deficient mice seem to divert LAPTM4 from the lysosome towards the plasma membrane (Milkereit and Rotin, 2011). *CLN5* was downregulated in the PG, but it's exact function is not well understood. Depletion of CLN5 has been shown to degrade lysosomal sortilin receptors and cation-independent mannose 6-phosphate receptors (CI-MPR) (Mamo *et al.*, 2012). *CLN5* null human fibroblast cells were observed to have decreased levels of ceramide, sphingomyelin and glycosphingolipids along with increased growth and apoptosis. Based on these findings it was proposed that CLN5 has a function in the *de novo* synthesis of sphingolipids (Mamo *et al.*, 2012).

Clathrin light chain A (*CLTA*) was found to be upregulated in growth but downregulated in the productivity group. Clathrin is a key protein in vesicle formation and has an essential role in endocytotic trafficking and protein secretion (McMahon and Boucrot, 2011). It has been shown that MAD2B is co-localized with CLTA at the mitotic spindle for stabilization of kinetochores. *MAD2A* was also found to be upregulated in the growth group as part of the cell cycle pathway suggesting a possible explanation for inclusion of *CLTA* in the GG, but not the PG (Medendorp *et al.*, 2010). The GGA family genes were implicated in both PG and GG; *GGA2* was downregulated in both and *GGA3* upregulated in the GG. GGA depletion has been shown to have a mis-sorting effect on mannose-6-phosphate receptors, cathepsin D and APP secretory inhibition (Ghosh *et al.*, 2003; Von Einem *et al.*, 2015), which was one of the top hits in our master gene list. In HeLa cells it was found that overexpression of GGAs increases fragmentation and vacuolization of the *trans*-Golgi network implying that these proteins have a role in maintaining Golgi integrity (Takatsu *et al.*, 2000). Genes coding for the δ and μ subunits of *AP-3* were found to be downregulated in the PG. AP-3 has been shown to regulate LAMP1 and LAMP2 sorting into late endosomes/lysosomes and knockdown of AP-3 led to an increase in LAMP proteins in tubular endosomes and on the cell surface (Peden *et al.*, 2004). In HEK293 cells depletion of AP-3 was shown to have an impact on lysosomal distribution, causing them to accumulate at the end of microtubules in the peripheral cytoplasm (Ivan *et al.*, 2012).

Both the regulatory profiles of the PG and GG point towards a clear pattern of downregulation of lysosomal activity by disrupting trafficking and recycling of lysosomal proteins and structural lipids and impairing lysosomal processing. None of these proteins have been engineered in recombinant CHO cells, however strategies to induce autophagocytic and supress lysosomal pathways have been implemented before using inhibitors as described in Kim et al. with a subsequent up to 30% increase in recombinant mAb yields (Kim *et al.*, 2013).

Figure 3.4.4. Pathway enrichment map for the lysosome pathway
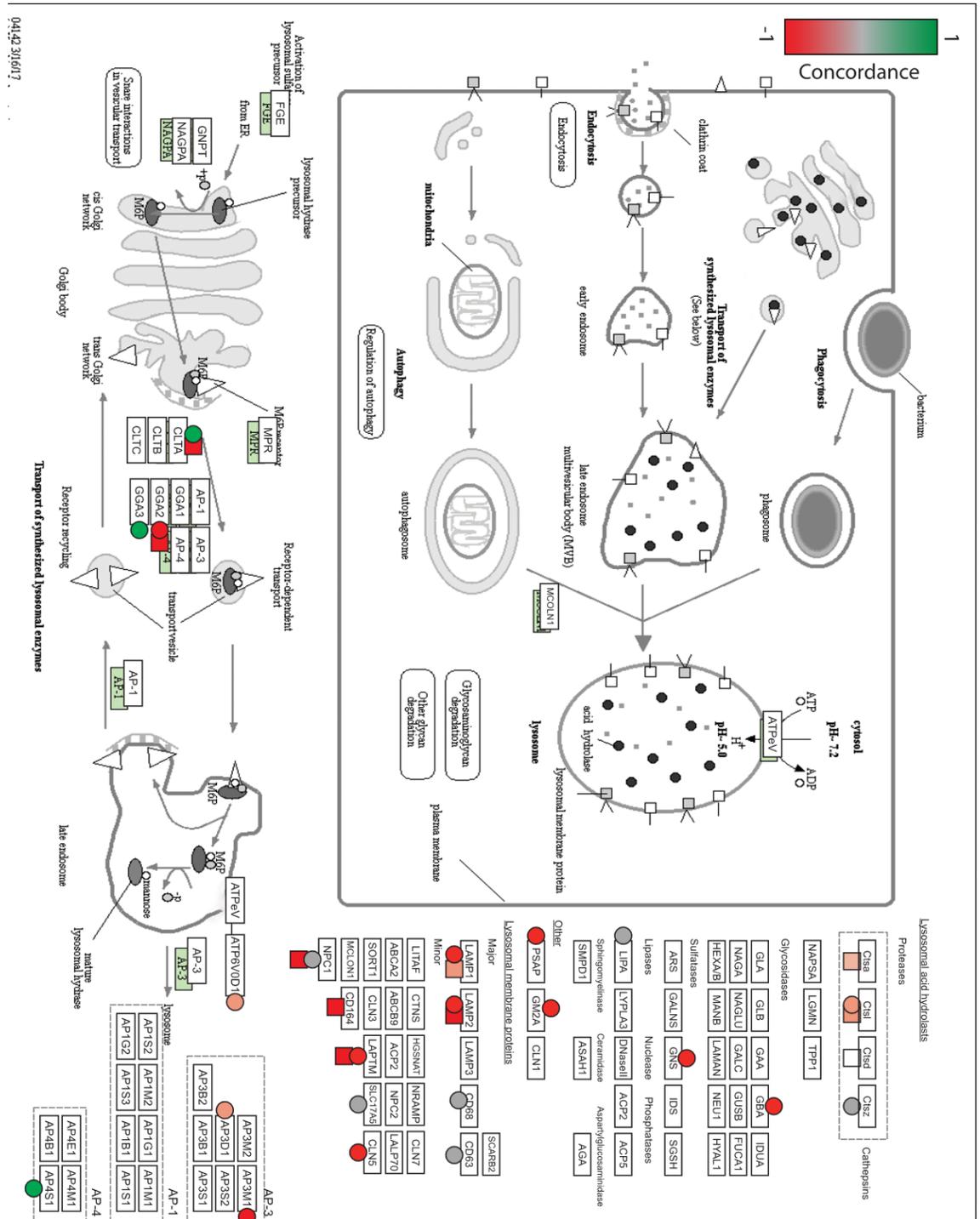
Hits for the growth group are shown in squares (□) and the productivity group is represented as circles (○).
Overlap between shapes (⬚) indicates a hit in the same gene, while adjacent but non-overlapping shapes (○□
) convey hits in the same gene family. Concordance values for each hit are shown as a colour value as
visualised in the legend.

3.5 DISCUSSION

3.5.1 Challenges in the evaluation of publicly available datasets

Of the data investigated, only two data sets/publications report on the application of RNAseq to investigate transcriptomic changes associated with *Qp* and growth rate. Studies comparing RNAseq and microarray approaches suggest that the two techniques can complement each other. Birzele et al. reported expression data for 10428 genes in a microarray group and 13375 genes in an RNAseq group (Birzele *et al.*, 2010). Between these approaches there was an overlap of 8404 genes with 2024 and 4971 unique genes in the microarray and RNAseq groups respectively (Birzele *et al.*, 2010)**.** On-the-other-hand, Yuk et al. reported that there was almost no overlap between differentially expressed genes identified by microarray and RNAseq (Yuk *et al.*, 2014). In this study, samples were taken at different times through culture at 4 and 48 h, and the subsequent microarray and RNAseq data sets had only 1 differentially expressed gene in common. This is surprising as it has been shown that RNA-seq and microarrays can have a high degree of concordance in the same biological system (Wang *et al.*, 2014). Whilst microarrays can give a good indication of relative expression levels of genes in a given experiment, these studies cast doubt on the ability of single transcriptomic analysis platforms to provide us with a representative snapshot of the transcriptome and hence a wider surveying and compiling of multiple studies may provide a better insight into those cellular processes important during CHO cell bioprocessing.

Combining omics approaches is a potentially powerful approach for constructing multi-dimensional and comprehensive models of CHO cell biology (Hefzi *et al.*, 2016). However, to date undertaking such an approach has not been widely applied in comparative cell line analysis to investigate the underlying changes in cellular machinery. This is likely because of the costs of running such experiments combined with challenges in data interpretation. The work reported by Yusufi et al. (Yusufi *et al.*, 2017) is one such noteworthy attempt to compare a parental CHO-K1 cell line with an antibody producing derivative. In this work, not only are changes in mRNA levels, but also copy number variant changes, reported and analysed. Using DAVID enrichment,

they identified groups of genes enriched after differential expression analysis. Among these were genes involved in DNA damage repair, mRNA processing and transport, vesicle transport and mitochondrial metabolism. Some of the genes singled out in this report (Yusufi *et al.*, 2017) were also identified in the meta-analysis undertaken and reported here including *Mmp14*, *Tm9sf2*, *Slc1a4, cers2, lpin1, rps2, Hnrnpa1, Nsmce2, Erccl* and *Eps8*. We also note that when comparing transcriptomic datasets some overlap can be missed and our study does not account for this nuance. This emphasises the need for enrichment analysis as different sets of stochastic transcriptomic changes can identify similar changes at a pathway level.

3.5.2   Limitations of the meta-analysis

Using aggregation methods and pathway enrichments, we present a meta-analysis of CHO high *Qp* and growth transcriptomics. We noted that the ability of a meta-analysis to identify common features and differentially expressed genes is highly dependent on the quality of the data available. In the case of the data that have been investigated here, there are several limitations for a meta-analysis. The most obvious limitation was the lack of accessibility to the transcriptomic platform expression data e.g. probe intensities for microarrays and raw RNAseq data (Zhao *et al.*, 2017). Out of the 4 available published RNAseq datasets, only 1 has made the raw RNAseq data available, and only 2 of the microarray-based transcriptomic studies have deposited their raw microarray data in public databases. This is out of step with generally accepted good practice for accessibility of 'omic' type data whereby the scientific community can only use and review/judge such reports if the raw data (as opposed to analysed data) is made available. This situation is exacerbated in the CHO cell field as the majority of the microarrays used in the experiments published are listed as proprietary and their probe sets are not disclosed. Further, the unavailability of the raw transcriptomic data prevents reanalysis of the data by others in the field, integration with other datasets or the reader reproducing any of the analysis or statistical outputs reported. Differential gene expression fold changes and listed p-values cannot be meaningfully compared between different studies due to experimental and biological variation. In our master gene list, around half of the genes appear only once across the 19 transcriptomic datasets as differentially expressed, which is indicative of a highly heterogeneous dataset to begin with.

To complicate meta-analyses further, there is a high degree of variance between the experimental methods of transcriptomic analyses performed. Further, the datasets reported in the literature around CHO cell biology are analysed using dramatically different workflows

ranging from partial least squares regression (Clarke *et al.*, 2011a), to co-expression clustering (Clarke *et al.*, 2011b) and gene set enrichment analysis (Doolan *et al.*, 2013). Naturally, these methods tend to produce gene lists that are derived from different methods of analysis and format, making it difficult to aggregate and interpret results across datasets. In human and mouse, a wealth of easily accessible and comparable transcriptomic data is available in data repositories like the Gene Expression Omnibus (GEO, see https://www.ncbi.nlm.nih.gov/geo/), which requires depositing MIAME (Minimum Information About a Microarray Experiment) compliant information transcriptomic datasets from investigators (including raw data file for each hybridization, processed data, annotation information, experimental design, gene identifiers and other annotations, data processing protocols) and facilitates target identification under specific conditions for further research.

Pathway enrichment as a method has its own set of limitations. To use pathway enrichment, you must have a background list of genes in the genome and a database of pathways. Both of these tend to change over time and variate depending on the reference genome and the pathway curator. Therefore, running the analysis on different databases may not always produce the same results and if significant changes occur within the reference genome or the way the pathways are rearranged, the analysis would need to be updated as well as it only makes sense within the context of the currently available data.

Similarly, with concordance, these values might change with the advent of new datasets that will affect the concordance values of these genes. In general, the more datasets you have the more accurate the concordance value will be. With smaller datasets, the concordance values are less precise, because the addition of one new member to an arithmetic mean calculation has more weight with regards to the outcome.

In the CHO cell field, while there are now a number of transcriptomic data sets generated and publicly available, very few studies actually follow up on their results and validate transcriptomic findings. In one of the few instances where such work has been undertaken, out of 21 potential targets from a transcriptomic and proteomic analysis of a CHO K1 cell line, 5 targets were selected for further validation (Doolan *et al.*, 2010). Only one of these 5, VCP, had a substantial effect on CHO cell growth. This is not unexpected as it is well known that transcriptomic data does not always correlate to abundance of protein (Courtes *et al.*, 2013) making validation a cumbersome ordeal. However, in order to build more comprehensive multi-omic models the CHO cell community should strive towards not only the generation of high-quality omics data, but more high-throughput rigorous validation, so that a comprehensive understanding of the cell and potential engineering strategies can be developed. This study here will help provide a framework

for researchers looking to interpret the currently available transcriptomic datasets as a 'whole' and want to apply the findings for improving the CHO cell platform. The pathways and genes identified as high frequency differentially expressed genes await validation by others as potential targets for achieving enhanced cell growth and/or productivity of recombinant biotherapeutics from cultured CHO cell expression systems.

3.6 SUMMARY

1) A 'first of its kind 'CHO meta-transcriptomic analysis of 19 datasets was performed along with concordance, frequency and pathway enrichment analysis.
2) A list of candidate genes and pathways for gene engineering was produced.
3) Methodological issues within existing CHO transcriptomic datasets were identified and guidelines were suggested to improve future work.

3.7 CONCLUSIONS

In this study, currently available CHO transcriptomic datasets were analysed to identify enriched pathways and genes differentially regulated with respect to cell growth or productivity. While individual studies have suggested these pathways as relevant for CHO cell recombinant protein expression, we have established and examined the landscape of transcriptomic variability between CHO cell specific studies. The datasets isolated from these studies were aggregated and processed to yield a reduced and manageable number of target genes and relevant pathways. This work should prove most useful for those wishing to undertake validation studies or trying to mine transcriptomic data from existing CHO cell literature as most of the data is not in the same format and not conveniently indexable. As a result of undertaking this analysis, we have also discovered and highlighted deficiencies in currently published transcriptomic studies and suggest improvement to these practices. Disclosing the raw data from transcriptomic experiments and using open, non-proprietary platforms are key to experiment reproducibility and producing data that is of use to the whole community. While platforms for depositing and analysing data exist such as NCBI's Biosample and Gene Expression Omnibus, they are not widely adopted in bioprocess transcriptomics providing unnecessary barriers for transparency of research and utilisation of the data. There is also a significant need for an indexed CHO bioprocess omics resource for target selection and gene cross-referencing. Projects including the CHO genome project (http://www.chogenome.org/) and the CHO co-expression database have

already taken the first steps toward this goal, however they will rely on the community to provide the required data in appropriate depth and format to capture the scope of the CHO cell omics landscape. While new CHO cell transcriptomic data is regularly being generated using increasingly more sophisticated tools and analysis, the curation of data must not be neglected and researchers should look to validate results.

Without presuming lysosomal or cell cycle involvement *a priori*, through the use of an aggregation and frequency based meta-analysis of publicly available transcriptomic data we were able to deduce the involvement of these pathways based on the concordance of transcriptomic data. Some of the identified targets have already been investigated in engineering recombinant CHO cells and validate our meta-study as having predictive value. We have yet to see many CHO cell engineering projects in the literature that have been informed by transcriptomic studies and this work should prove useful in that regard.

4 Design and validation of a novel configuration of the Beacon optofluidic platform for CHO CLD to 15 ml scale

4.1 Introduction

In this chapter we describe the application of a next-generation optofluidic platform, the Berkley Lights Beacon® to improve existing industrial CLD processes. We compare CLD processes of cell lines generated using this technology expressing either BlosozumAb, an IgG4 antibody against sclerostin used in treating osteoporosis, or Etanercept, an IgG1-receptor fusion protein that binds TNF-α and is clinically relevant in treating arthritis. Both of these products were run through two CLD processes to compare the performance using a scaled down version of a ClonePix™ 2 GMP certified process and the Berkley Lights Beacon® CLD workflow. We describe the limitations and advantages of each method as they are compared head to head and make recommendations for the academic and industrial community about how to improve these CLD platforms

4.2 An overview of the CLD processes

The ClonePix™ 2 CLD process we have adapted from FDB GMP protocols consists of transfecting 4 pools for each construct (Blosozumab and Etanercept) using an in-house developed DG44 DHFR- cell line. After stable transfection, the cells were then frozen and banked for future use (Figure 4.2.1). Post revival, cells were then taken through the ClonePix™ 2 colony picker. Up to 5000 colonies were screened per construct and the top 600 (120 per pool) were deposited into 96-well plates. After plate outgrowth, the titre was measured with an Octet® for each well that had a confluence of over 50% as measured by the Solentim® and the top 60 cells were progressed onwards to 24-well plate growth. 4 days later the cells were progressed into T25 flasks and after another 4 days into 50 mL TubeSpin® culture. These cultures were then used to seed 24-deep well plates for a 14-day fed-batch evaluation of the cultures. At the end of the 14-day fed-batch, supernatants were collected and measured with the Octet® system for titre. The best 12 cell lines per construct were then selected to be screened on the ambr® 15 system.

Figure 4.2.1. The ClonePix™ 2 and the Beacon® CLD processes
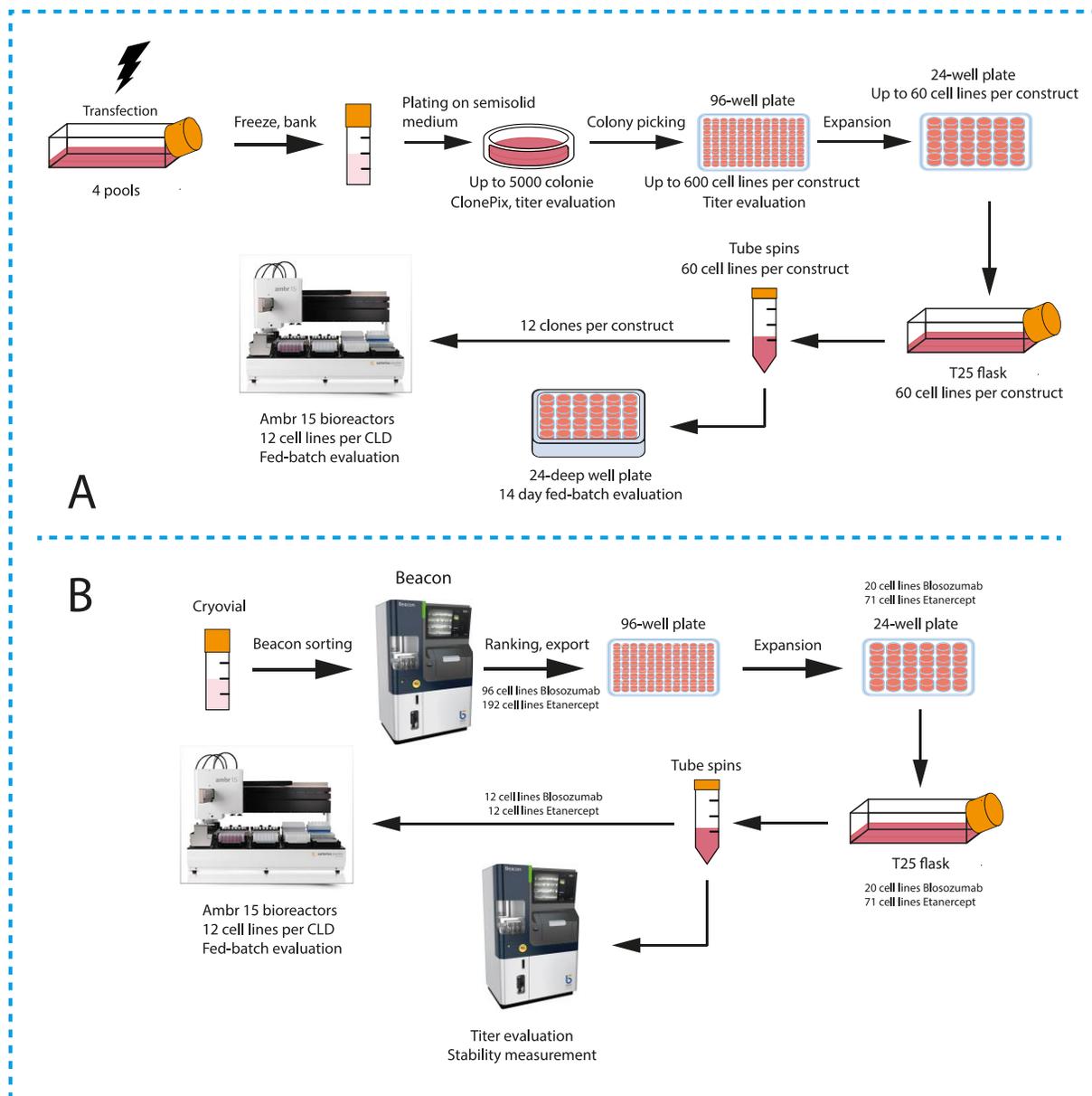
A – the ClonePix™ 2 process. Showing transfection, outgrowth and cryopreservation of expressing pools. These are then plated onto semisolid medium and up to 5000 colonies are then evaluated on the ClonePix™ 2. Up to 600 cell lines are exported into 96-well plates and expanded up to shake culture. Cells are then evaluated in a 24-well plates fed-batch format and 12 cell lines are then selected for ambr® 15 bioreactor screening based on titre. B – The Beacon® CLD process depicting recovery of cells from cryovial and deposition onto the Beacon® device. The cells are then ranked based on their expression of the recombinant protein and exported into 96 well plates following expansion into shake flask culture. These cell lines are then subject to a novel population measurement on the Beacon® to evaluate productivity and stability. 12 cell lines are then selected to be screened on the ambr® 15 platform.

With the Beacon® system we perform a similar CLD workflow, where instead of the ClonePix™

2 we ranked the cells from the transfected pools on an optofluidic microchip and then proceeded to expansion. A typical process would then progress the top cells straight into the ambr® 15 platform. However, in this work we add a novel step of reintroducing populations from the expanded cell lines back into the Beacon® and then make our predictions based on population averages from multiple pens rather than from single pen measurements.

4.3 CLONEPIX™ 2 CLD WORK

4.3.1 Transfection of FDB DG44 host cell line to express Etanercept and BlosozumAb

To construct the plasmids for stable expression of BlosozumAb, an in-house FDB expression system known as pAVE was used. To generate the construct, the heavy and light chain genes of the antibody were initially cloned into two different vectors which were then digested and ligated together to make a single double cassette vector capable of expressing the full antibody. The double gene vector coding for BlosozumAb was obtained from FDB, while the Etanercept gene was cloned into the pAVE light chain vector which also contains the DHFR selection gene. The Etanercept gene was obtained by PCR amplification from a pcDNA3.1/Hygro(+) plasmid used to express Etanercept in the group at the University of Kent. The primers contained the terminal flanking restriction sites that matched the restriction sites in the pAVE light chain vector cassette. Cloning was carried out using simple digestion of PCR product and pAVE vector with the corresponding restriction sites and subsequent ligation with a 3:1 insert to vector ratio.

To generate CHO cell lines that produce the target recombinant proteins of interest, an electroporation-based transfection method to randomly integrate the linearized plasmid DNA into the cell genome was used. Schematic representation of constructs used is represented in **Error! Reference source not found.**. After electroporation, cells were placed under selective pressure by the addition of MTX into the cell medium. After allowing 14 days for the cells to recover, the transfected pools were checked for titre and viability with a total of 10 pools being generated by transfection for each of the two DNA constructs used (Blosozumab and Etanercept). The titre and growth characteristics of the pools were then assessed.

Figure 4.3.1.1. Schematic representation of constructs used in electroporation
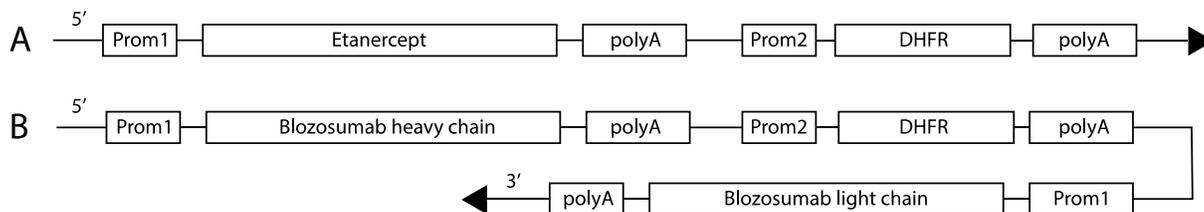
Due to plasmids being proprietary full sequences and names of promoters cannot be disclosed. A – pAVE vector containing Etanercept gene, expressed via promoter 1 (Prom1) and terminated with a polyA sequence. The dihydrofolate reductase gene is expressed in a separate cassette from promoter 2 (Prom2). Both sequences are terminated with a polyA sequence. B – pAVE double cassette pAVE vector expressing Blozosumab heavy and light chains using protomer 1 (Prom1) and terminated with a polyA sequence. The dihydrofolate reductase gene is expressed from promoter 2 (Prom2). 5' and 3' denote the orientation of the plasmid DNA sequence.

The means of all the variables measured between the two product or recombinant protein groups were found to be statistically significant with an α of < 0.05 using a two-tailed unpaired t-test (Figure 4.3.1.2). For titre, the means were 20.47 µg/mL and 17.48 µg/mL for BlosozumAb and Etanercept respectively. For BlosozumAb, titres were within a range of 17.25 to 24.80 µg/mL, while the Etanercept pools had less spread and the titres were between 15.25 to 19.65 µg/mL. In contrast, both the mean of the viable cell concentration (0.59×10^6 cells/mL and 0.93×10^6 cell/mL) and the culture viabilities (75.29% and 79.66%) were higher for Etanercept. Overall, pools generated from both DNA constructs had recovered successfully and pools that were above the mean were selected for further cell line development work.

Figure 4.3.1.2. Evaluation of Clone 27 transfection pools two weeks after selection
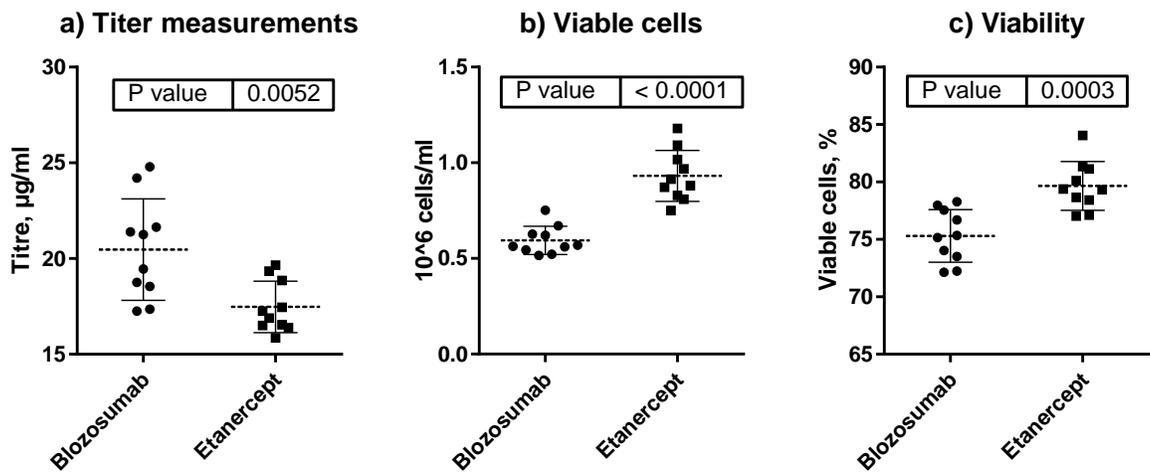
Clone 27 pools had been transfected with constructs for the expression of either Blosozumab or Etanercept. a) Depicts titre measurements as determined on an Octet® instrument. b) and c) show viable cell concentrations and percentage of viable cells as measured with a Vi-Cell. Error bars show the mean and standard deviation values. P-values for a unpaired two-tailed t-test are shown.

4.3.2 Isolation of clonal cell lines from transfected pools using the ClonePix™ 2 technology and instrumentation

As discussed previously in this chapter, while the use of transfected pools for recombinant protein production is not uncommon, mammalian derived therapeutics destined for use in the clinic are invariably produced from stably expressing monoclonal cell lines. In the cell line development of monoclonal cell lines, a panel of producers with different bioprocess characteristics (growth rate, productivity, PTMs) are generated that allows the isolation and selection of cell lines with the required product quality attributes, titre and growth characteristics. These are isolated by screening technologies, with the best screens in theory allowing the isolation of the best performers from the original population. To generate a panel of monoclonal cell lines here, a colony picking approach using the ClonePix™ 2 device was used, which is standard approach used in the industrial setting. This process relies on seeding a dilute cell suspension on a semi-solid medium so that colonies can grow out from single cells. Once these colonies have grown out and are visible by the naked eye, the colonies can be screened for titre with the aid of a fluorescein isothiocyanate-tagged Fc biding dye. A typical image showing the colonies fluorescing under image acquisition in the ClonePix™ 2 is shown in Figure 4.3.2.3 below.

Figure 4.3.2.3. Image of a colony expressing mAb from the ClonePix™ 2 camera
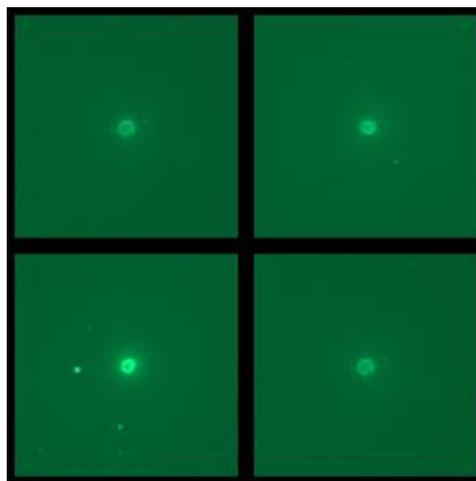
Image aquired using the FITC1000 image acquisition mode. The 4 stitched together images show colonies that had already been selected to be picked.

In total, 4 pools per construct were taken through the process of colony picking after being grown out on semisolid medium. Pool number 4 was seeded twice to allow screening of one pool

in more depth compared to the others. This approach is advantageous if the distribution of the pool population in terms of productivity is very bottom heavy and the highest producers are very rare within the population. Up to 5000 colonies were screened for each construct. The attribute selected by which to screen the colonies was by measuring the average intensity of fluorescence around the perimeter of the colony denoted by the value FITC 1000 exterior mean intensity. In Figure 4.3.2.4 and Figure 4.3.2.5 the picking histograms for the pools are presented. These show that based on the picking results, the general shapes of the distribution of the FITC 1000 exterior mean intensity scores remain the same for both pool types and we were able to observe outgrowth in all pools screened. While undesirable events such as close proximity, irregular shapes or colonies that were too big were present in all cases, this did not impact the ability to pick 120 colonies per pool. These events were judged to be within expectations for a typical picking run and do not show abnormalities with the outgrowth in semi-solid medium.

For the Etanercept pools, overall, the number of clones that were present in the viable groups (FITC 1000 + Ungated) was approximately 240 colonies. Meanwhile, in the BlosozumAb group, picking runs from pools B3 and B4 had a lower clone count (201, 162 and 217) compared to picking runs from pools B1 and B2 (283 and 265). The differences between the means of the clones selected for picking in this group were also significant ($\alpha < 0.05$) compared to clones selected from pools B1 and B2 as established by a Kruskal-Wallis (KW) test with Dunn's multiple comparisons correction. Within the Blosozum*A*b group, the fold difference between the mean of the lowest picking run and the mean of the highest picking run was 1.55 (

) with the differences being statistically significant as established by a KW test. In the Etanercept group, however, the maximum fold difference between the highest and the lowest picking run means was 1.36 which was also found to be statistically significant via the KW test. Comparing the average means of the groups we can see that the BlosozumAb group had a lower fluorescence signal on average 2254.4 vs 527.5 with a ratio of 4.3.

Figure 4.3.2.4. The results of the colony picking procedure for BlosozumAb pools performed with the ClonePix™ 2 instrument
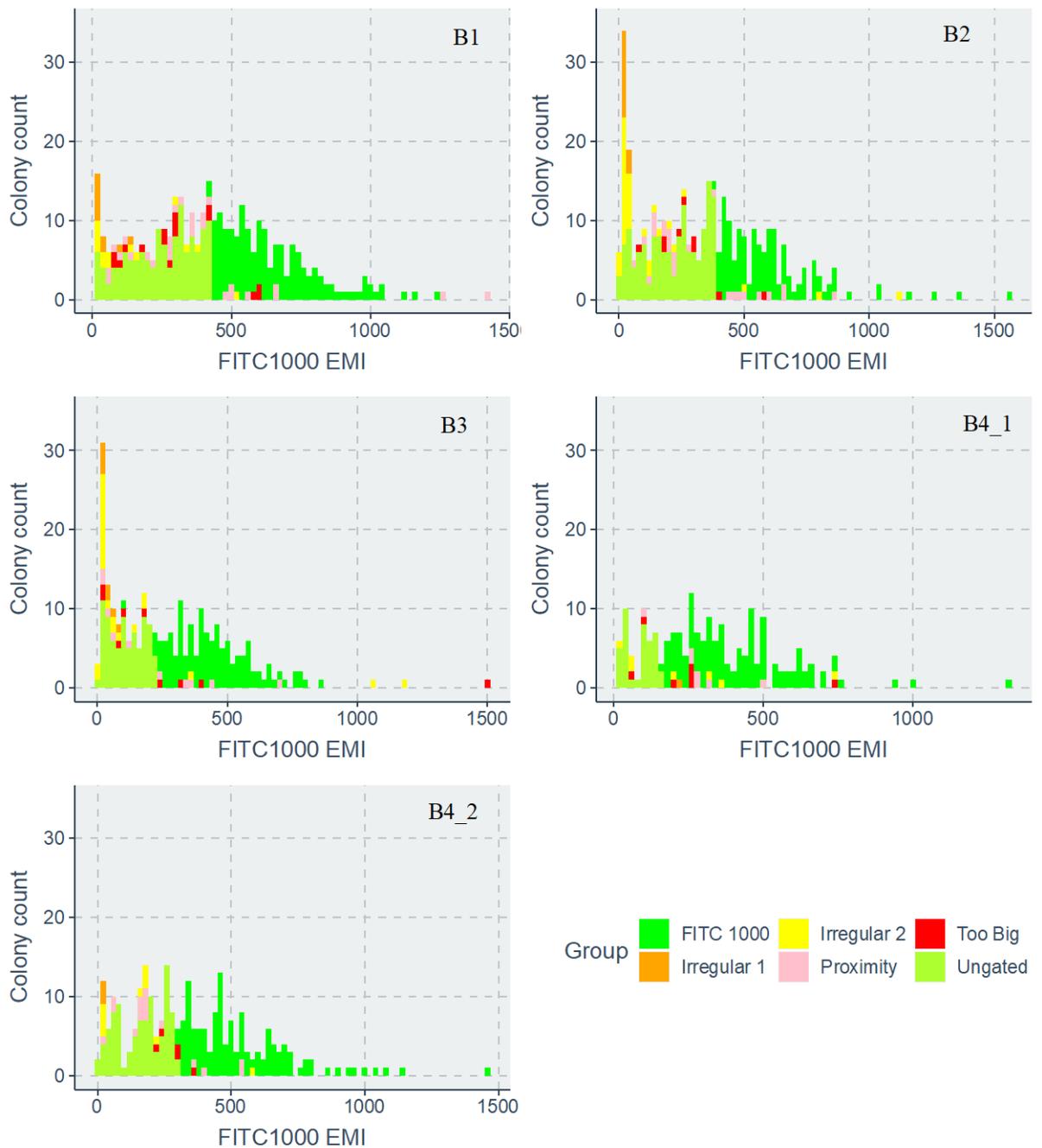
The group labelled FITC 1000 are the clones that were selected for export. FITC 1000 EMI signifies the fluorescence intensity of the fluorescein-tagged Fc region binding dye after 1000 ms exposure.

Figure 4.3.2.5. The results of the colony picking procedure for BlosozumAb pools performed with the ClonePix™ 2 instrument
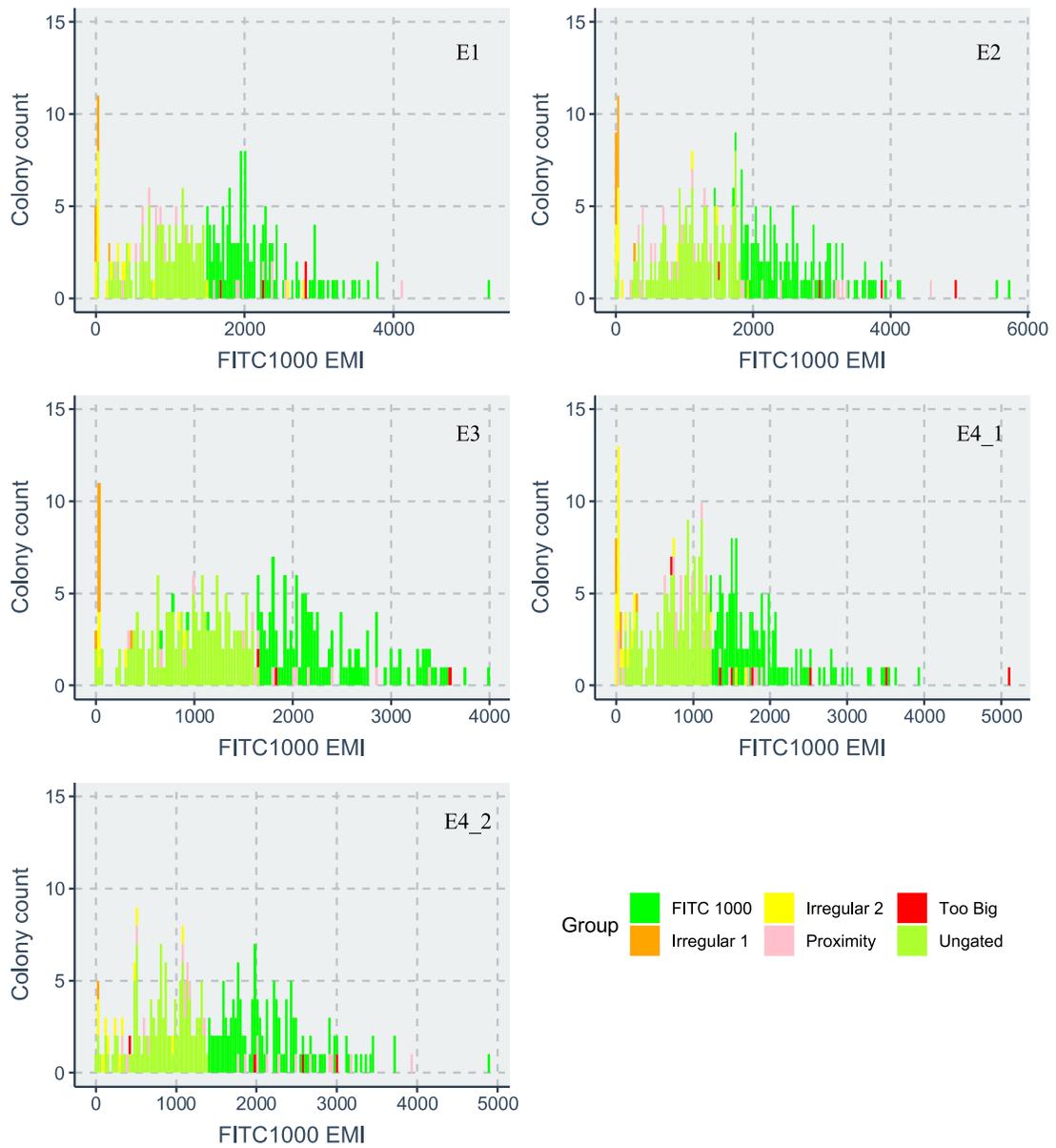
The group labelled FITC 1000 are the clones that were selected for export. FITC 1000 EMI signifies the fluorescence intensity of the fluorescein-tagged Fc region binding dye after 1000 ms exposure.

Figure 4.3.2.6. Graphs depicting the Tukey box-plots of the FITC1000 exterior mean intensity score of the picked colonies for the BlosozumAb and Etanercept pools

Pools	E1	E2	E3	E4_1	E4_2
Total colonies	230	265	265	263	238
Mean	2210	2601	2367	1901	2193
Std. Deviation	624.1	716.4	566.1	588.1	607.8
Std. Error of Mean	56.74	65.13	51.9	53.24	55.25

Pools	B1	B2	B3	B4_1	B4_2
Total colonies	283	265	201	162	217
Mean	643.3	604.1	441.4	414.2	534.4
Std. Deviation	171	196.6	158.1	185.3	196.8
Std. Error of Mean	13.82	17.8	14.49	16.78	17.74
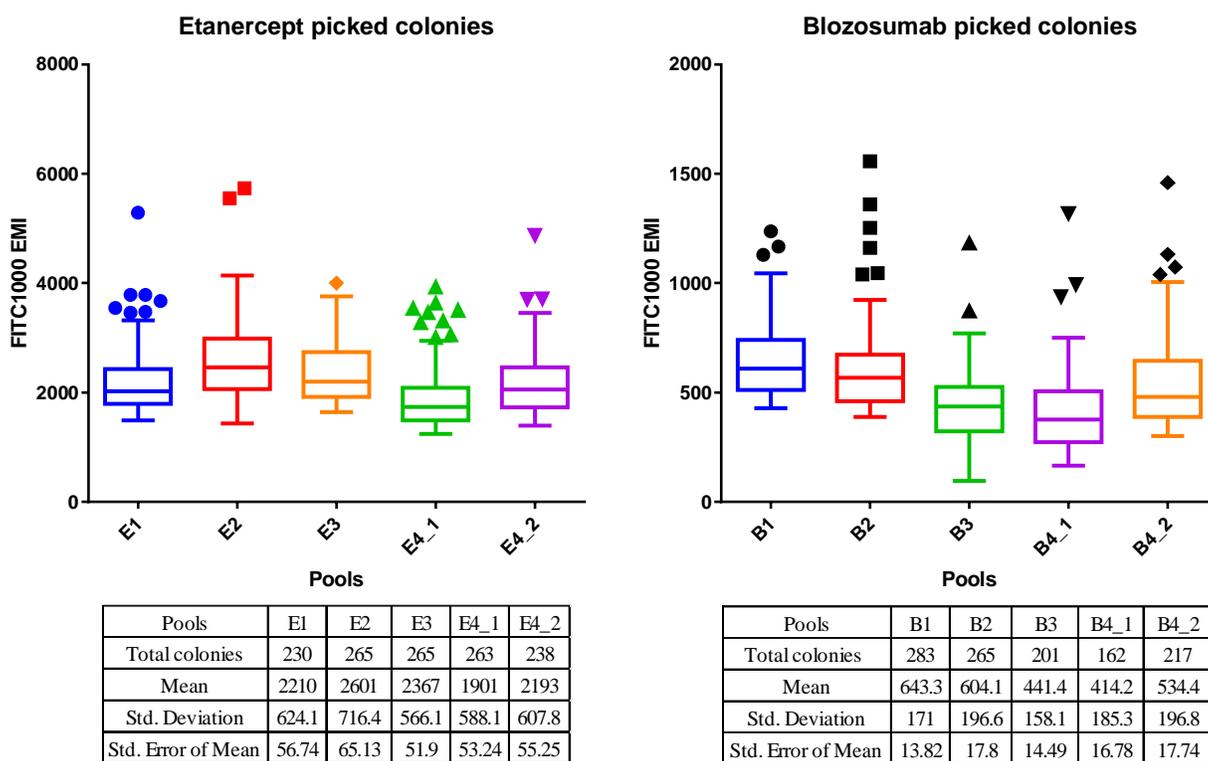
p-values	E1	E2	E3	E4_1	E4_2
E1		< 0.0001	0.1335	< 0.0001	> 0.9999
E2	< 0.0001		0.1821	< 0.0001	< 0.0001
E3	0.1335	0.1821		< 0.0001	0.0965
E4_1	< 0.0001	< 0.0001	< 0.0001		0.0001
E4_2	> 0.9999	< 0.0001	0.0965	0.0001	
p-values	B1	B2	B3	B4_1	B4_2
B1		0.3232	< 0.0001	< 0.0001	< 0.0001
B2	0.3232		< 0.0001	< 0.0001	0.0152
B3	< 0.0001	< 0.0001		> 0.9999	0.0027
B4_1	< 0.0001	< 0.0001	> 0.9999		< 0.0001
B4_2	< 0.0001	0.0152	0.0027	< 0.0001	

The table below displays the p-values of a Kruskal-Wallis test with Dunn's multiple comparisons correction applied to the means of the FITC 1000 EMI. The highlighted cells in italic are p-values below the multiple corrections adjusted 0.05 threshold.

4.3.3 Outgrowth of ClonePix™ 2 selected clones in 96-well plates

After colony picking with the ClonePix™ 2, the cells were cultured in 96-well plates at a scale of ~120 clones for pools 1, 2, 3 and 240 for pool 4. These cells were then cultured in a static incubator for a week and then measured for confluence with a Solentim® every 3 days. When cell confluence reached 50%, the well titre was measured with an Octet™ instrument using a protein

A probe. The screening continued until a minimum of 250 clones had been screened for each construct; 261 for BlosozumAb and 366 for Etanercept. For selection towards further CLD we wanted to maintain pool diversity at this stage so we progressed the top 12 clones from each pool and 24 from pool 4 into 24 well plates for expansion. The clone titre data is presented in Figure 4.3.3.7. For BlosozumAb clones, the means of the titres between the pools was fairly consistent; 21.66, 22.39, 19.40 µg/mL for B1, B2, B4 and 16.23 µg/mL for B3. The means between B3-B2, B3-B1 were found to be statistically significant with a KW test. In the Etanercept group, the means for the pool titres were consistent as well (12.03, 11.36, 12.18 µg/mL) with E2 having the highest mean titre 14.10 µg/mL. The E2 titre mean was also found to be significantly different from the others with a KW test. At this stage we can see that the Etanercept colonies appeared to be producing less than the BlosozumAb ones; the BlosozumAb means were 60% higher on average with a statistically significant difference (P = 0.0073, two tailed t-test) between the groups of the means.

Figure 4.3.3.7. Titres obtained from outgrowth of clones selected from the ClonePix™ 2 in 96-well plates
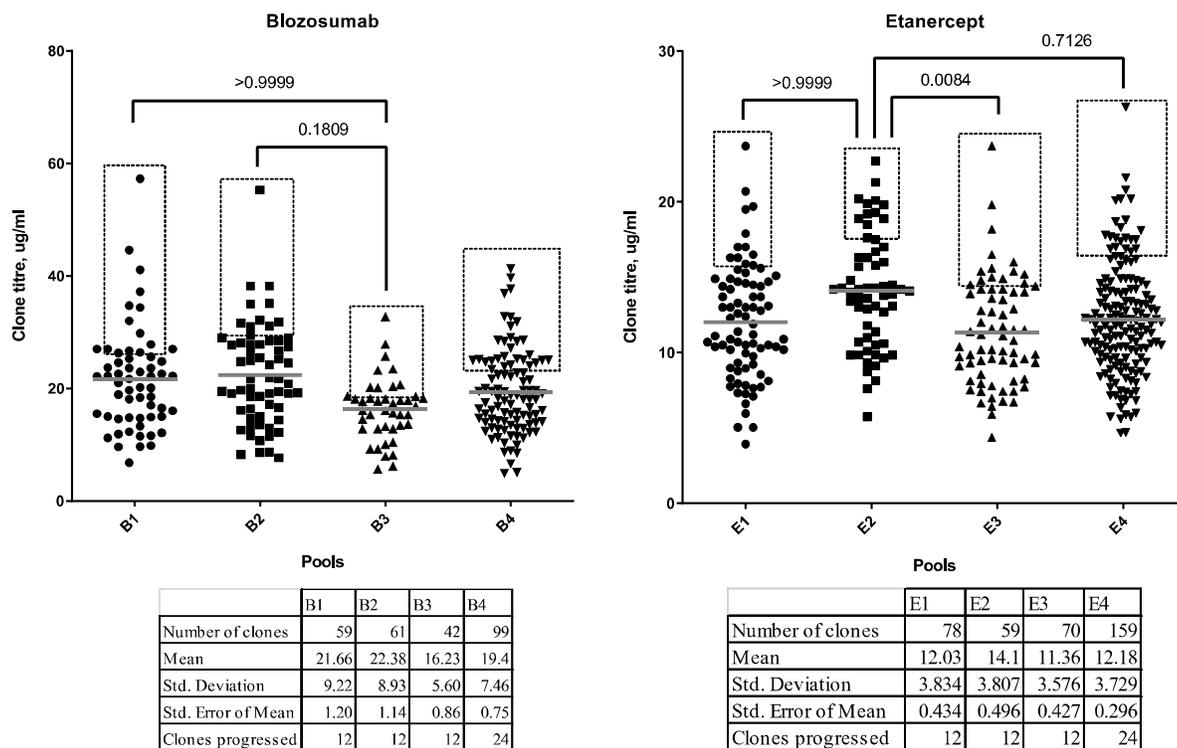
The lines above the distributions highlight statistical significance between the groups as calculated with a Kruskal-Wallis test with Dunn's multiple comparisons correction. The grey line signifies the mean. The tables below provide the summaries for the pool titre distributions.

	B1	B2	B3	B4
Number of clones	59	61	42	99
Mean	21.66	22.38	16.23	19.4
Std. Deviation	9.22	8.93	5.60	7.46
Std. Error of Mean	1.20	1.14	0.86	0.75
Clones progressed	12	12	12	24

	E1	E2	E3	E4
Number of clones	78	59	70	159
Mean	12.03	14.1	11.36	12.18
Std. Deviation	3.834	3.807	3.576	3.729
Std. Error of Mean	0.434	0.496	0.427	0.296
Clones progressed	12	12	12	24

4.3.4 Screening of colonies in 24-deep well plates

After screening the ClonePix™ 2 clones after outgrowth in 96-well plates, the cells were progressively expanded into 24-well plates 4 days later, allowed to grow out in the 24-well plates and then similarly transferred to T75 flasks after 4 days of expansion. Four days post expansion into T75 flasks, the cells were transferred into TubeSpin® flasks marking the transition from stationary culture into shaking culture and kept on a split regime of 3-4 days avoiding the weekends. At this stage there were 60 cell lines per construct with a 120 cell lines in total. These 120 cell lines were then screened again in order to select the final 12 cell lines per construct that would be assessed in the ambr® 15 miniature bioreactor system. At this scale, we used fed-batch fermentation to mimic the conditions within the ambr® 15 system. The feeding regimes were kept as similar as possible to the ambr® 15 protocol within the FDB Apollo CLD framework and are further expanded upon in section 2.2.12.

At TubeSpin® scale, undertaking 60 fed-batch cultures without any high-throughput support was deemed to be too laborious so the selection was carried out within a 24-deep well plate (Sartorius, A-0038) format at the 2.5 mL scale. This is because the feeding regime can then be administered with multi-channel pipettes ensuring speed and accuracy at the cost of being able to measure cell concentration and culture viability due to the scale of the experiment. At the end of the 14-day feeding regime the titres were measured for each well. The results are presented within Figure 4.3.4.8.

Figure 4.3.4.8. Dot plots showing the titres for clones screened in 24-well plates in a 14 day fed batch format
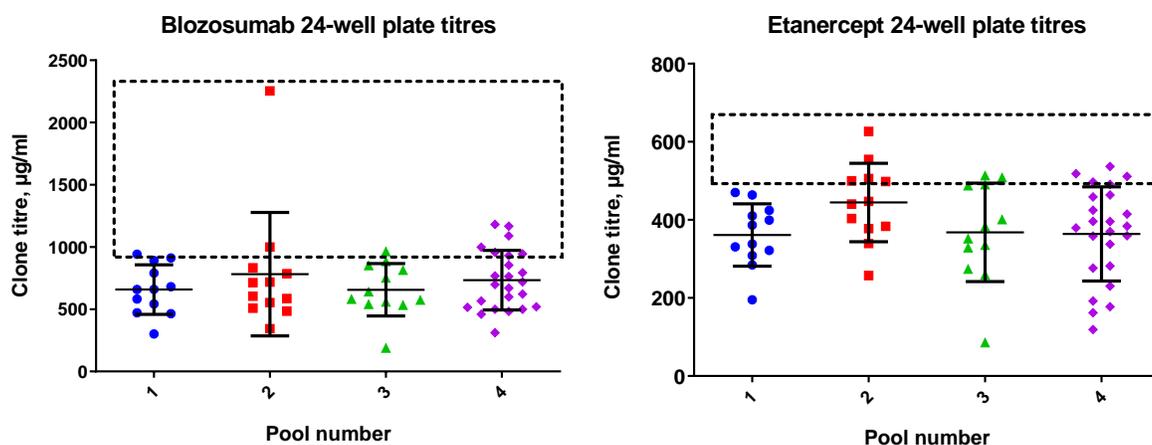
The error bars denote the mean of the distribution along with the standard deviation. The dotted lines show the cut-off for cell line progression towards the ambr® 15 screening.

No significant differences between any of the means within the pool groups for these samples using a KW test with Dunn's multiple testing corrections were observed. Similarly, to previous stages of selection, there was consistency between the means of the pools; 657.9, 781.9, 655.9, 733.4 μg/mL for the BlosozumAb group with a collective mean of 707.3 μg/mL. For the Etanercept groups the means were lower; 361.1, 444.6, 367.9, 364.0 μg/mL. The mean for this group was 384.4 μg/mL and the difference between the group means was found to be statistically significant as determined via an unpaired two tailed t-test with Welch's correction (p=0.0003). The relative difference between the groups of the means was found to be 1.84.

4.4 BEACON® CLD WORK

4.4.1 Overview of the Beacon® system

Because the Berkley Lights Beacon® system is relatively new and has only seen adoption by companies in the past few years, this subchapter is dedicated to familiarize the reader with the system. The system utilizes microfluidic chips and opto-electronic positioning (OEP) to sort cells within a microchip into pens. The visual result of sorting is depicted in Figure 4.4.1.1. Within the image it can be seen that the system is able to successfully identify individual cells within the flow channel of the microfluidic chip and then slowly guide selected cells with OEP into designated pens with the system validating how many cells are present in each pen via brightfield imaging and cell-recognition software.

Figure 4.4.1.1. Brightfield image of a row of pens within a Beacon® OptoSelect™ chip before and immediately after the penning algorithm
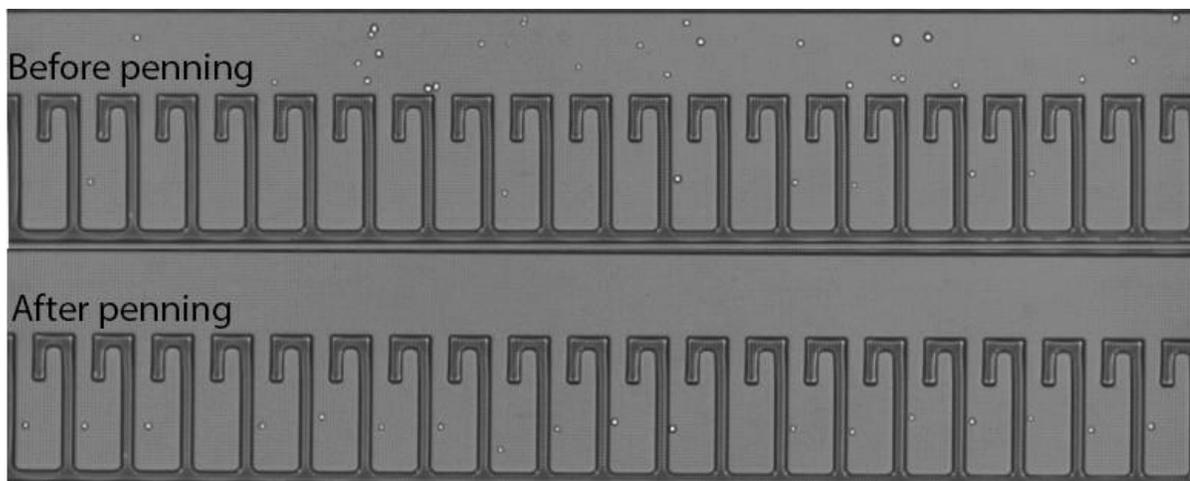
The cells can then be cultured on said microchip and exported from any pen into a 96-well plate or similar format. Assays can also be performed on the microchip using fluorescence as the reporter signal by introducing media containing fluorescent dyes or beads that have been specially tagged. The Beacon® system is equipped with several excitation emission combinations that enable it to image based on commonly used fluorescence reagent characteristics for example FITC or Texas Red.

In collaboration with the internal staff at Berkley Lights, a CLD pipeline for work with the Beacon® was therefore developed and is reported as a schematic in Figure 4.4.1.2. The pools that were generated in section 4.3.1 were cryopreserved and shipped to the Berkley Lights facilities in California, US. There, the cells were revived and pre-screened to check whether growth was

observed on-chip. At that time, Berkley Lights suggested the use of 20% conditioned media, an in-house growth supplement and avoid the use of selective pressure. We were still able to use FDB-MAP medium as the base which was used in all the ClonePix™ 2 CLD work. A more detailed description of operation and media preparation is outlined in sections 2.2.14 and 2.2.15.

After establishing the appropriate culturing conditions on chip we could progress to performing the cell line development workflow. For reference, we have included a brightfield image of cell growth on chip that is depicted in Figure 4.4.1.4. The cells shown in the figure had been in culture for 5 days and a broad range of pens with different numbers of cells could be observed clearly showing clonal growth rate differentiation.

Figure 4.4.1.2. Schematic showing the typical workflow for a Beacon® based CLD
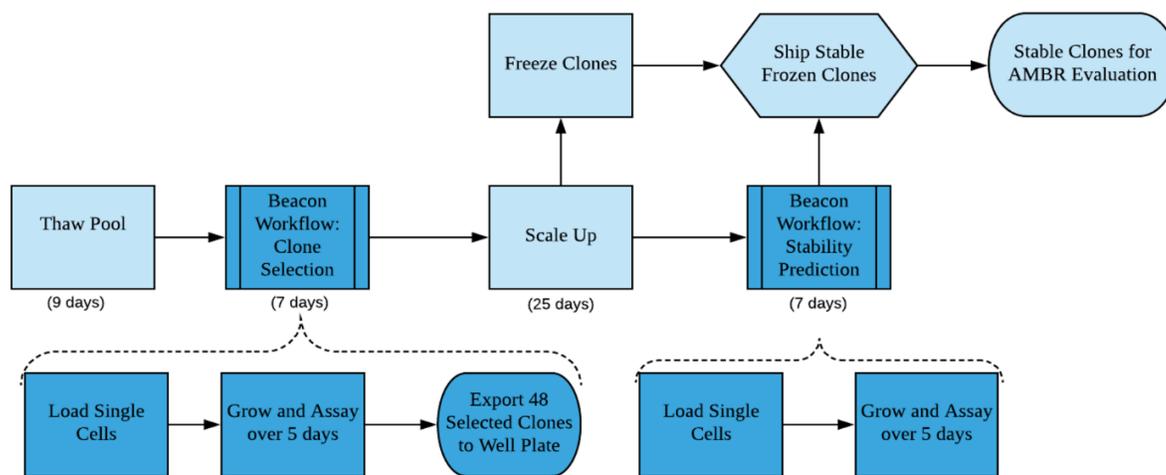
Every 2-3 days the cells were imaged with a fluorescent dye that binds to the Fc region of the target molecule and the cell number and assay scores based on pen fluorescence intensity measured as reported in Figure 4.4.1.3. For the assay, the dye is first introduced into the medium and allowed to equilibrate throughout the chip after which the chip is flushed with media lacking the dye. As the fluorescent tagged recombinant protein starts diffusing back into the flow channel, the relative concentrations of the recombinant protein can be measured based on the total diffusion time which is dependent on concentration and the mass of the diffusing molecule. This way, cells can be ranked based on desired bioprocess characteristics and target the clones that seem to perform the best on-chip. After the selection had been made, the cells were exported from the desired pens into 96-well plates.

Figure 4.4.1.3. Colour image of the fluorescence intensities observed during the secretion assay on the Beacon® OptoSelect™ chip
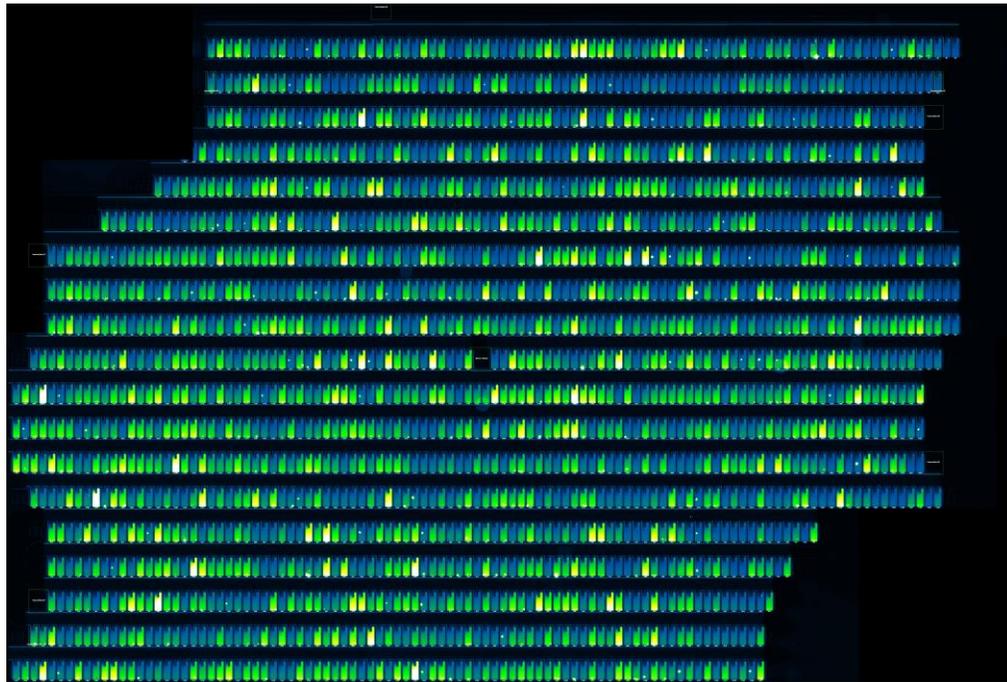
The surviving clones were then scaled up and reintroduced into the chip to measure their stability scores based on the growth rates of population members with different fluorescence assay scores. This allows the avoidance of clones that might have their productivity decline quickly. These clones were then shipped back to FDB in the UK to be run on the ambr® 15 system alongside the clones produced according to the ClonePix™ 2 CLD guidelines. The different Beacon® CLD steps are expanded upon in the following sections.

Figure 4.4.1.4. Brightfield image of a section of a Beacon® OptoSelect™ chip after 5 days of culturing CHO cells
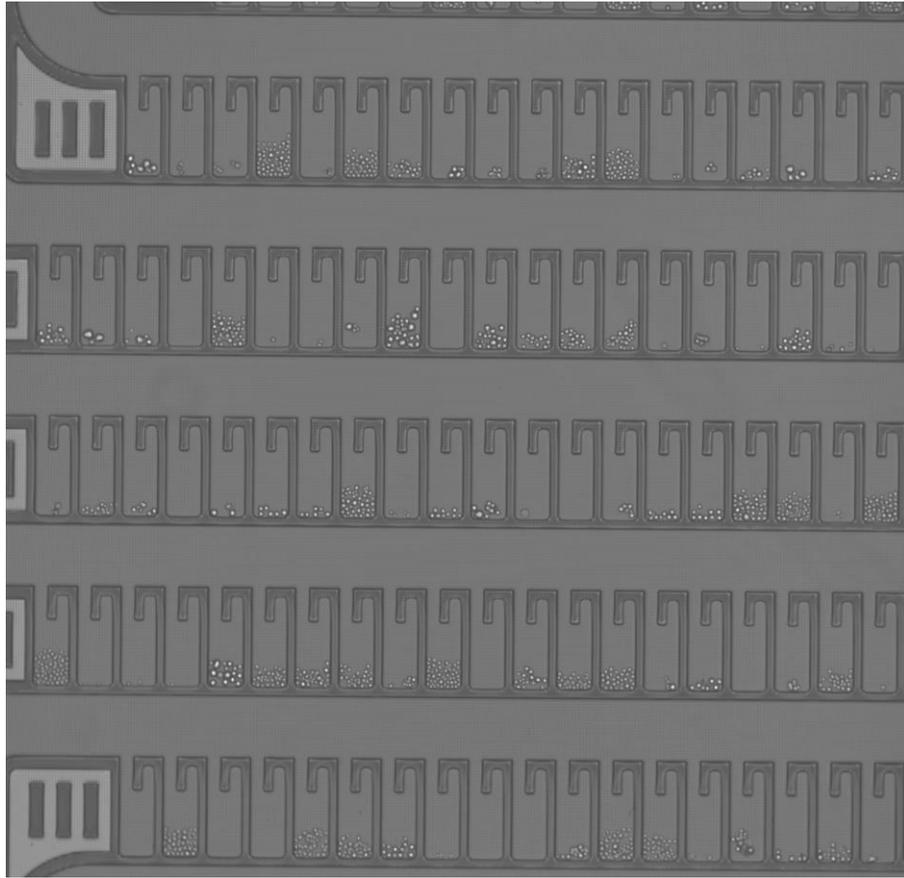
The bars on the left-hand side of the image are calibration marks for the optical system alignment.

Figure 4.4.1.5. Images showing the status of the pen and cells during different stages of the export process
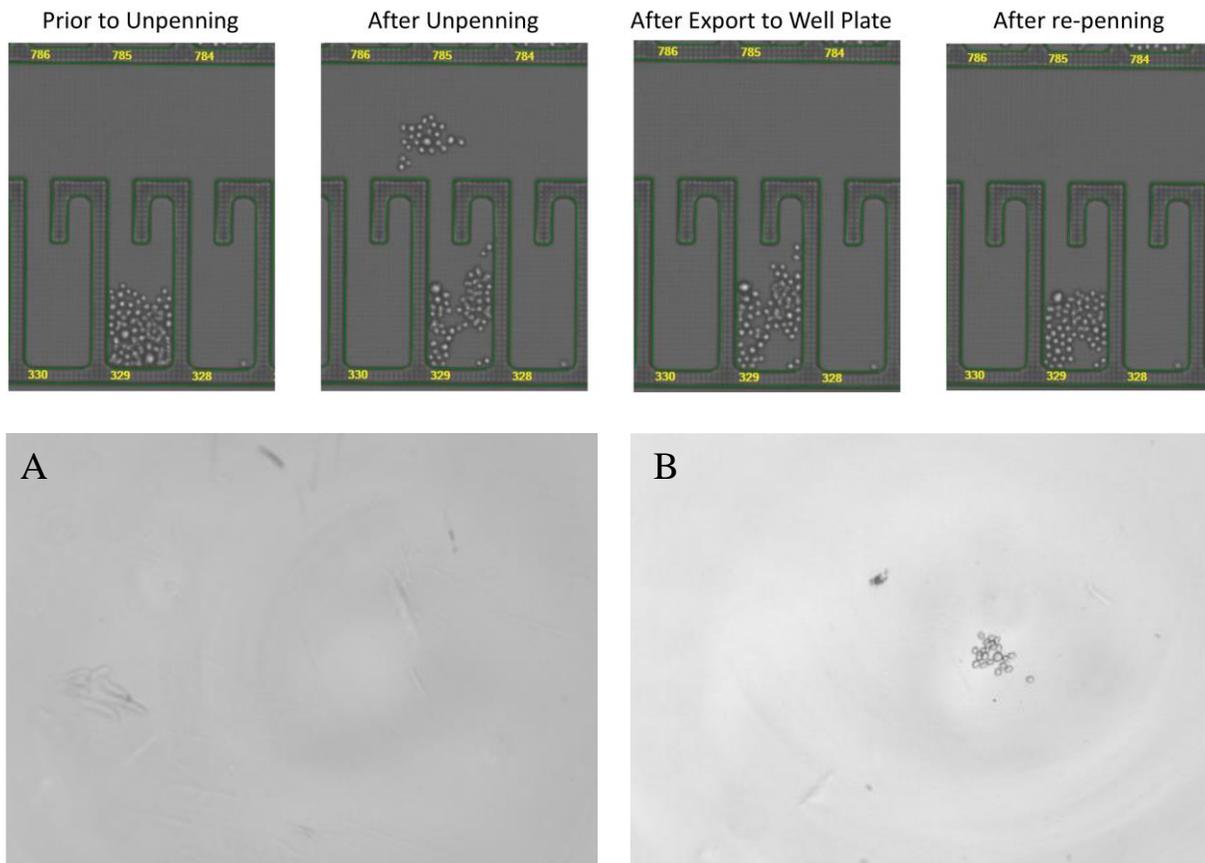
The numbers represent the internal labels of each pen (top pictures). Brightfield images taken from an export 96-well plate: (A) a blank export from the chip flow channel, (B) export after unpenning pen 329.

4.4.2   Loading of the cells onto the microfluidic chip

The cell suspension was deposited on chip and cultured for 7 days, taking images every few days to avoid running experiments on the weekend. The results of cell loading are shown in Figure 4.4.2.6. As in the ClonePix™ 2 CLD, pool 4 for both groups/molecules was screened at double capacity in case the high producers are very rare events which might be overlooked otherwise. For the BlosozumAb group, the number of single cells that were confirmed to be loaded was 1237, 1148, 1223 and 2394 for pools 1, 2, 3 and 4 respectively. Similarly, for the Etanercept group, we were able to load 1232, 1220, 1239 and 2304 cells. The target amounts to be loaded for each pool was 1200 and 2400 for pool 4. We were able to be very precise in introducing the desired number of cells in this regard and the number can further be adjusted with a manual operation of the OEP system or further rounds of the visual recognition software mediated algorithm. A total of 6002 cells were loaded for the BlosozumAb group and 5995 cells for the

Etanercept group, with a relative deviation of 0.00033% and 0.00083% percent from the desired target of 6000 cells.

Figure 4.4.2.6. Bar chart showing the number of pens that were confirmed to be loaded with cells on chip for the BlosozumAb and Etanercept pools
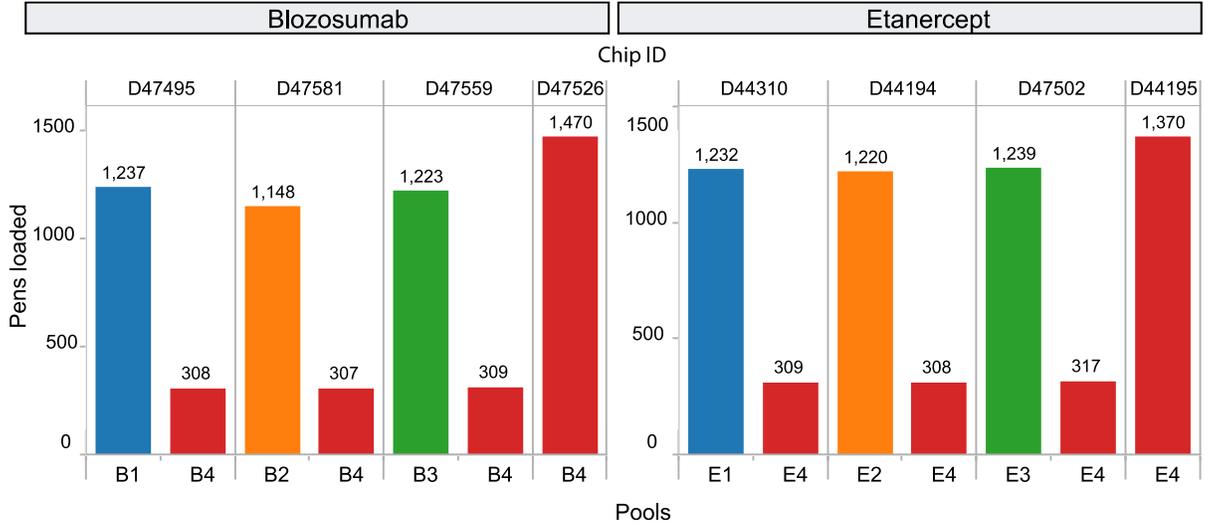
Device ID denotes which chip was used to load the cells. Different colours represent the subpools for each product.

The growth of the cells derived from the pools was observed to be higher than the lower limit of 0.3. This limit describes the proportion of occupied pens that did not grow out to 6 cells or more at the time of measurement. The cell growth was visualised and is reported in Figure 4.4.2.7. Pools B4 and E4 were distributed between the chips as a control to see whether the chips had an impact on cell growth. In the case of the BlosozumAb pools, the average ratio of grown pens was 0.64 and within the figure it can be seen that the ratio for each chip hovers around the average. The same trend was observed in the Etanercept group, across all chips the growth ratio hovered close to the average value of 0.72. The averages are not statistically different as determined via a two tailed unpaired t-test, showing that on the Beacon® system the cells behaved the same between the BlosozumAb and Etanercept groups in terms of on-chip clonal expansion.

Figure 4.4.2.7. On-chip viability for the cells derived from different pools
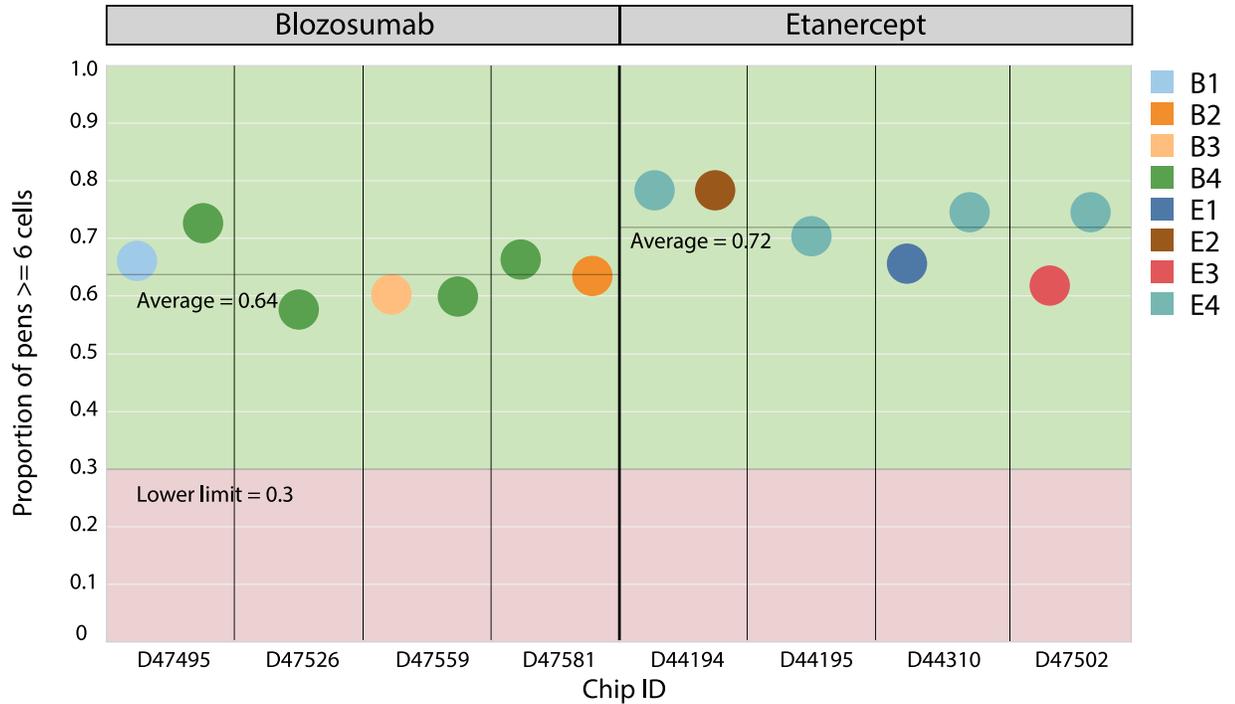
Viability is represented as the proportions of cells that grew to 6 or more cells after 5 days.

4.4.3 Comparing the bioprocess parameters on-chip and ranking clones for export

After confirming that clones are able to grow within the system, we then proceeded to characterize the cells to determine which to select for scale up. For our initial selection criteria, we tried to make as few assumptions as possible and limit the criteria to pens that had >6 cells and sorting by their assay scores. The distributions of the final cell counts before export, assay scores and specific productivities are presented in Figure 4.4.3.8. Again, we see good reproducibility between the different pools within the groups: the means for the BlosozumAb pools for the cell counts within the pens were 10.99, 10.71, 10.21, 9.99 and 0.01603, 0.01644, 0.0164, 0.01572 units for the titre scores. The Etanercept group follows the same trend with the means for the cell count being 18.88, 18.96, 18.84, 17.20 and 0.0619, 0.059, 0.0579, 0.0549 units for the titre scores. Subgroups B3 and B4 were statistically significant in their titre score means (p = 0.031). When comparing cell counts statistical significance was observed between subgroups B1/B4, B2/B4, E1/E4, E2/E4. All comparisons between subgroups E and B were significant with a p-value <0.0001 (Table 4.4.3.1).

Figure 4.4.3.8. Final cell count and intensity score measurement for different pens on the Beacon® instrument contains either BlozosumAb or Etanercept expressing cells
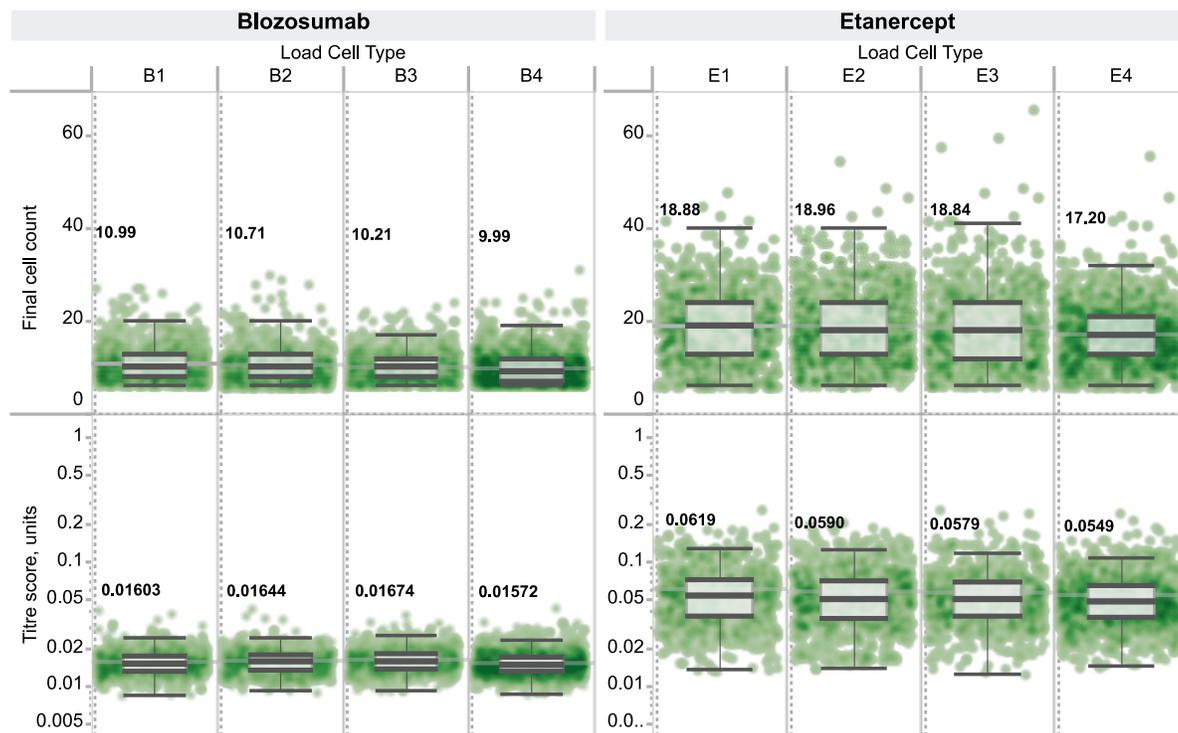
The boxplots represent the median and 0-1.5 of the interquartile range. The grey line and numeric value represent the mean. B stands for Blozosumab pool and E stands for Etanercept pool.

Table 4.4.3.1. Pool intragroup statistical comparisons as determined by one-way non-parametric Dunn's multiple comparisons corrected Kruskal-Wallis test

Titre score		Final cell count	
Comparison	p-value	Comparison	p-value
B3 vs. B4	0.031	B1 vs. B4	<0.0001
		B2 vs. B4	0.0374
		E1 vs. E4	0.008
		E2 vs. E4	0.0022

All intergroup comparisons between members of groups B and E were significant with a p-value of <0.0001. B stands for Blozosumab pool and E stands for Etanercept pool.

The reason that the Etanercept cell counts were higher than the BlosozumAb is likely due to the fact that the Etanercept clones were exported a day later than the Etanercept. This discrepancy disappears if we instead look at the doubling times between the two groups (Figure 4.4.3.9). The average growth rates for the BlosozumAb and Etanercept groups were quite similar; 29.01 and 31.08 h respectively, with a relative difference of 1.07. The statistical differences between the pools are presented in Table 4.4.3.2. Because the differences in the means between

the B and E groups were <10%, we concluded that the cells grew equally well on chip and that this would not impact on the evaluation of the cells using the system.

Figure 4.4.3.9. Density box-plot depicting the doubling times of the clones in the BlosozumAb and the Etanercept groups
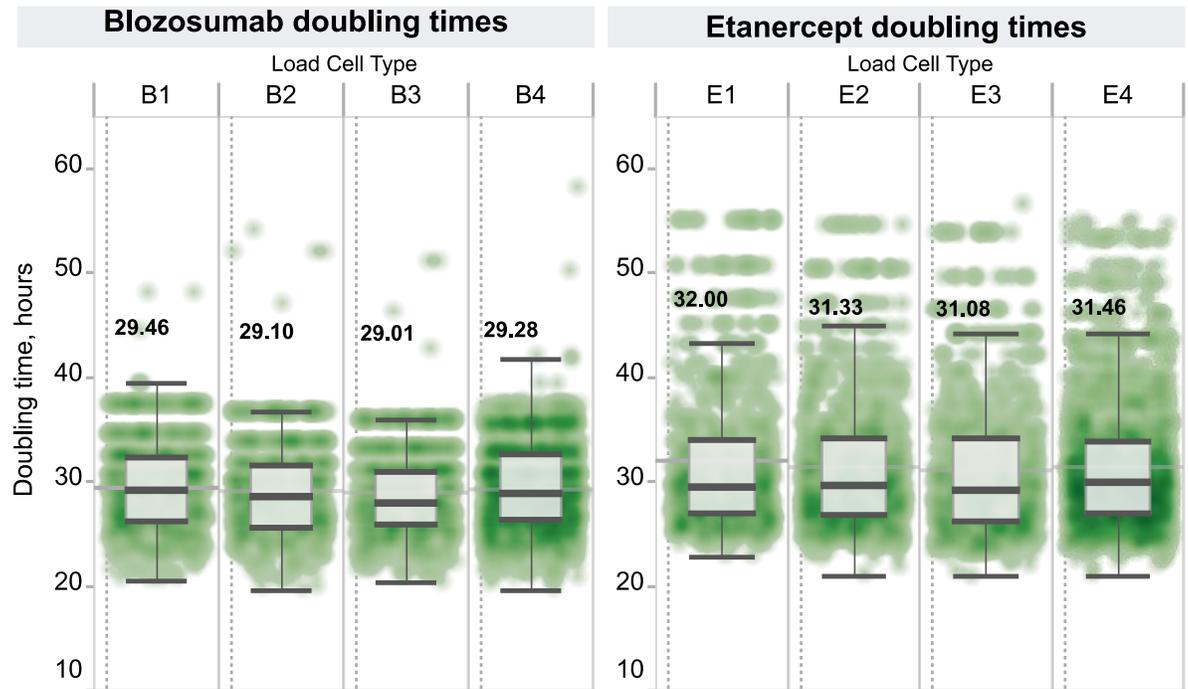
The boxplots represent the median and 0-1.5 of the interquartile range. The grey lines signify the means and their values are presented adjacent to the box-plots. B stands for Blozosumab pool and E stands for Etanercept pool.

Table 4.4.3.2. Pool doubling time intragroup statistical comparisons as determined by one-way non-parametric Dunn's multiple comparisons corrected Kruskal-Wallis test

Comparison	p-value
B1 vs. E2	0.0004
B1 vs. E4	<0.0001
B2 vs. E1	<0.0001
B2 vs. E2	<0.0001
B2 vs. E3	0.0059
B2 vs. E4	<0.0001
B3 vs. E1	<0.0001
B3 vs. E2	<0.0001
B3 vs. E3	0.0017
B3 vs. E4	<0.0001
B4 vs. E1	<0.0001
B4 vs. E2	<0.0001
B4 vs. E3	0.0473
B4 vs. E4	<0.0001

All intergroup comparisons between members of groups B and E were significant with a p-value of <0.0001. B stands for Blozosumab pool and E stands for Etanercept pool.

The relative titre differences between the groups, however, were larger than the differences in the final cell counts. This is to be expected as the mAb is considered an 'easy' to express molecule whilst the Etanercept molecule is considered difficult to express. The titre score means for the BlosozumAb and Etanercept groups were 0.0162 and 0.0584 with a ratio of 3.6 between them, which would not be expected if Etanercept is more difficult to express, as confirmed in the ClonePix™ 2 CLD. Such a difference is unlikely to be accounted for by a shift in export time as the number of cells had not even doubled compared to the BlosozumAb group. One hypothesis is that this discrepancy has occurred due to the difference of the background fluorescence from the conditioned media which was freshly prepared between each of the two runs. To determine whether this was the case, we looked at the average background fluorescence intensity levels of the empty pens in the two runs which are presented in Figure 4.4.3.10. The differences between the means of the groups were found to be significant using a two-tailed t-test with a p-value of < 0.0001 with a relative difference of 1.48. From the figure we can observe that the fluorescence intensity mean variance between the chips was small compared to the variance across the runs. The standard deviations within the groups were 4.6×10^{-4} and 5.5×10^{-4} for the BlosozumAb and Etanercept, and they were not found to be significant using an F-test. This means that the base intensity signal was consistent across the chips within the group. This led us to believe that it is difficult to compare fluorescence intensities in a meaningful way across different batches of conditioned media.

Figure 4.4.3.10. Density box-plots depicting the fluorescence intensities of empty pens across all chips for the BlosozumAb and Etanercept pool assays
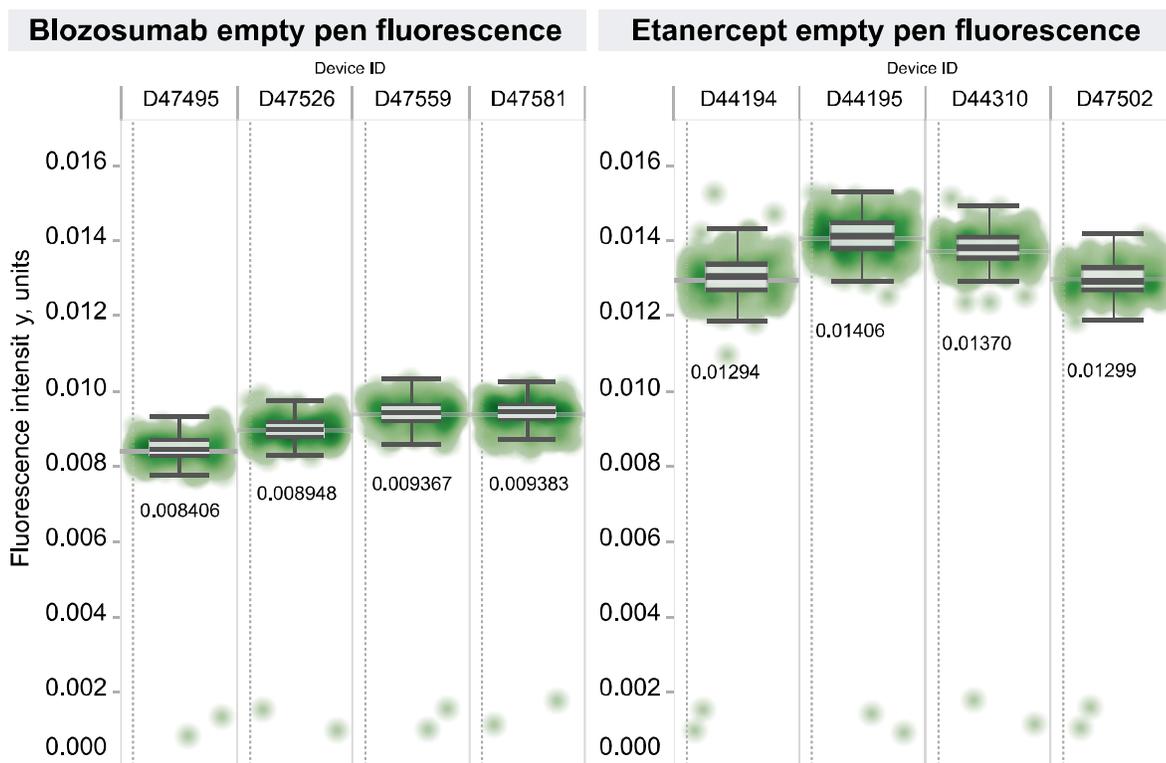
The boxplots represent the median and 0-1.5 of the interquartile range. The grey lines signify the means and their values are presented below the box plots.

4.4.4   Export and scale up of BlosozumAb and Etanercept clones

After ranking the clones from the pool screen by their fluorescence assay scores and filtering by whether the pens had >=6 cells, we selected the top 96 cell lines and exported them into 96-well plates. These clones then had to then be scaled up to 24-well plate, T75 and E125 flask stages. A histogram of the ranked clones is presented in Figure 4.4.4.1 which shows the overall fluorescence assay scores of the clones and the top 100 zoomed in. The clones that survived scale up to shake flasks are highlighted in red. As described in the previous sections, the average assay scores for the Etanercept group were higher than the BlosozumAb group; 0.0578 compared to 0.01613 with a relative difference of 3.6. We can see a difference within the overall titre score distributions for these groups as well. The means with the standard deviations for these distributions were 0.0161 ± 0.0038 (Blozosumab) and 0.0578 ± 0.0323 (Etanercept) with a relative difference of 8.52.

The number of clones that were scaled up and progressed through the different stages of scale up are reported in Table 4.4.4.1. It can be seen that attrition of clones was most significant at the 96-well plate stage. A total of 96 clones were exported for the BlosozumAb group and 192 clones for the Etanercept group. A total of 24 and 71 clones survived export for BlosozumAb and Etanercept respectively, which corresponds to a survival rate of 25% and 36.9%.

Table 4.4.4.1. The progression and survival of clones through the Beacon® CLD workflow

Group	Clones screened			
	96 well plates	24-well plates	T25	Shake flask
BlosozumAb	24	20	20	17
Etanercept	71	71	71	50

Figure 4.4.4.1. Histograms showing the fluorescence intensity distributions of the cultured BlosozumAb and Etanercept cell lines
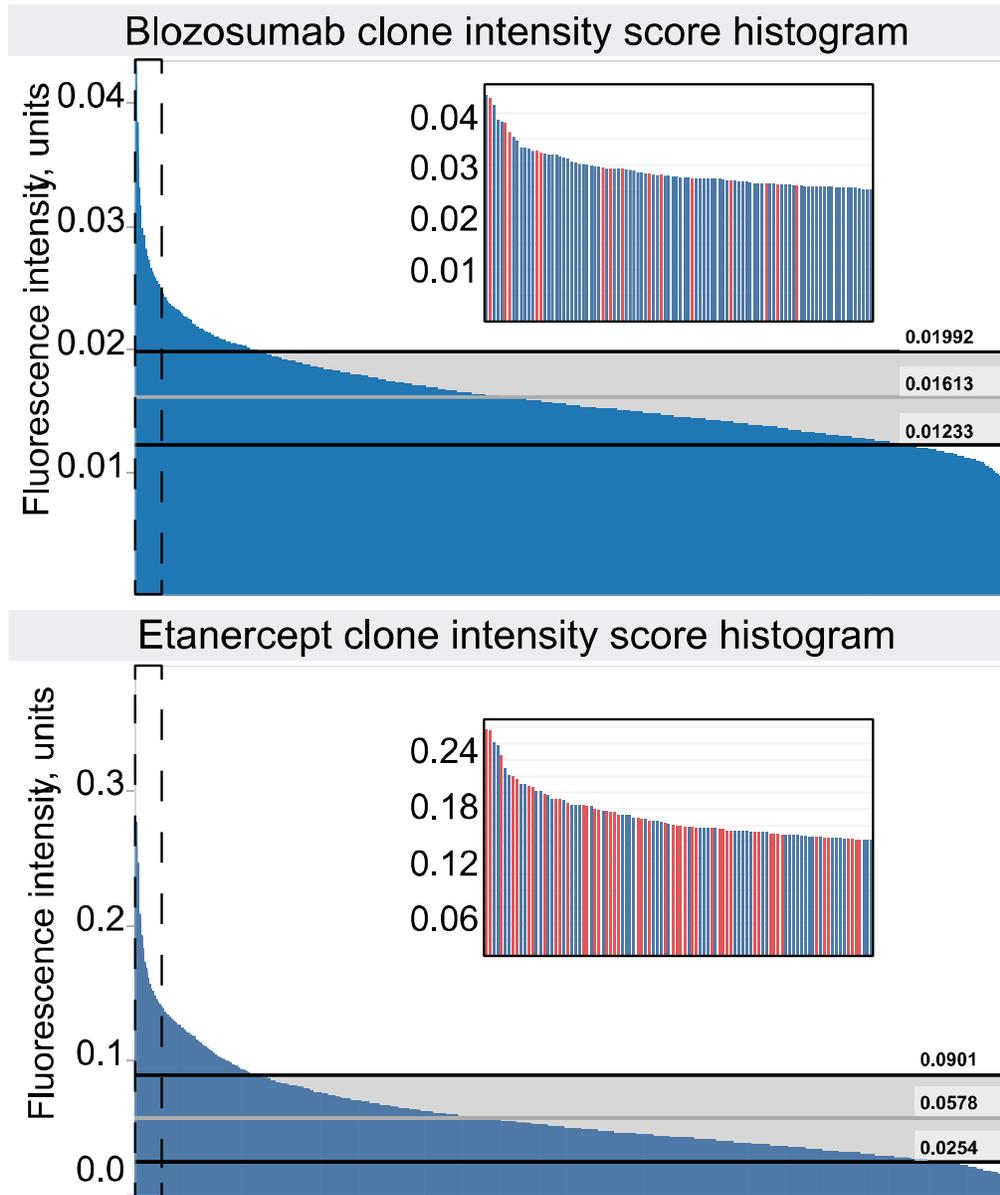
The embedded histograms are zoomed in on the dotted regions containing the top 100 clones and show which clones survived export and scale-up in red. The grey lines signify the mean fluorescence intensity values for the distributions and the black lines above and below are the values for +1 and -1 standard deviations from the mean.

4.4.5   Population analysis of exported and scaled up clones

In the literature, it has been established that single-cell measurements are hindered by high biological noise of the systems underpinning cellular function (Arriaga, 2009; Barron and Li,

2016; Kim *et al.*, 2019). To counteract this, we performed a secondary stage of screening in which we loaded a minimum of 200 clones that were derived from each scaled-up cell line generated in the previous section. This in-depth measurement allowed us to counteract the biological noise and make more robust predictions about cellular behaviour off-chip. The average specific productivities obtained from these population measurements is reported in Figure 4.4.5.1. These figures show data only from pens that had a cell growth of >=6 cells at the time the assay was performed.

Figure 4.4.5.1. The ranked mean specific productivities of the BlosozumAb and Etanercept clone population screenings
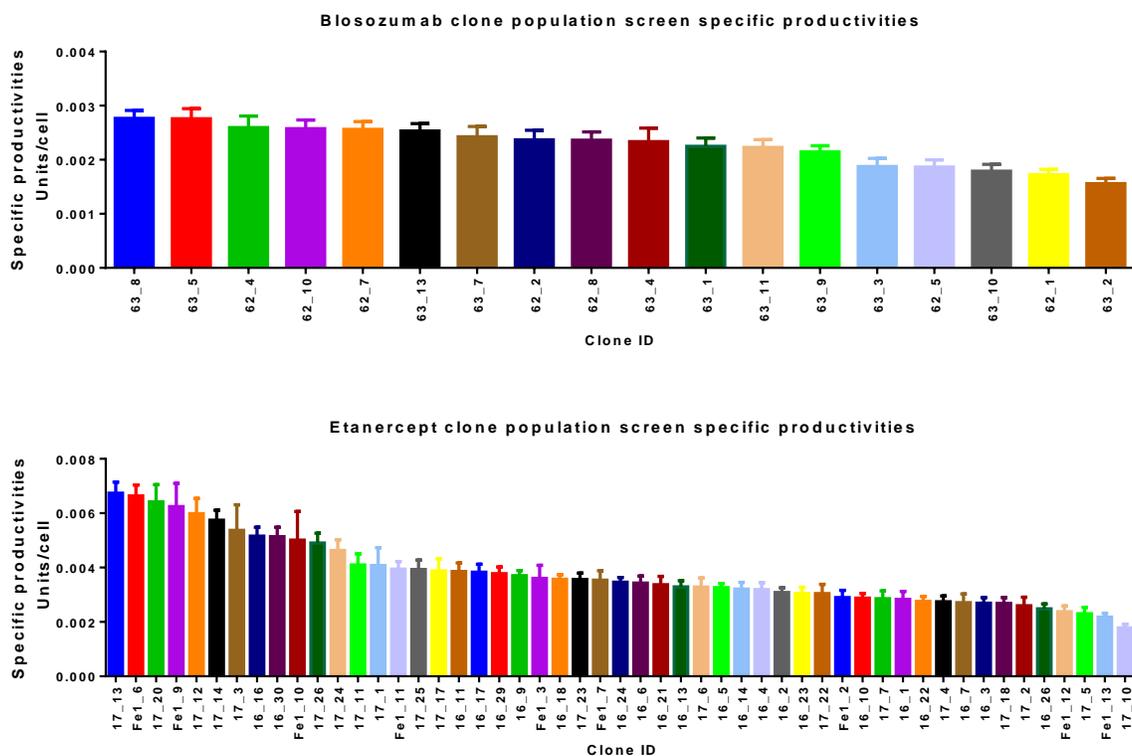
The error bars in both graphs show the 95% confidence interval of the means. Only data from pens that had >=6 cells at the time of measurement are shown.

For the BlosozumAb group, the top performing clone was 63_8 with a Qp of 0.00276 units/cell and the lowest was 63_2 (Qp 0.00172 units/cell). In the Etanercept group, however, the highest value belonged to clone 17_13 (0.00673 units/cell) and the lowest value to clone 17_10 (0.00178 units/cell). The means of the specific productivities of these groups were 0.00226 units/cell and 0.00384 units/cell for BlosozumAb and Etanercept respectively, with a ratio of 1.70. The coefficient of variance for these two groups were 16.11% and 35.10%, meaning that in

the Etanercept group there was greater variance between the mean specific productivities of the population of different clones.

This contrast is highlighted in Figure 4.4.5.2 which reports the ability of the clone population analysis to statistically discriminate between the different clones. To this end we used a Kruskal-Wallis test with Dunn's multiple comparisons correction to test all of the means of the specific productivities against each other. The colour in the graph visualises the p-value obtained from the test; from red (not statistically significant) to green (significant). In the BlosozumAb group, we could discriminate between two different groups among the clone populations (1-13 and 14-18) which seem to be statistically different from each other (p-value < 0.05). In the Etanercept group, however, we could isolate 4 groups; 1-12, 13-36, 27-47, 37-51. In each case we progressed the top 12 clones to go through the ambr® 15 screening stage and the cells were banked down and shipped back to the FDB site at Billingham, UK.

Figure 4.4.5.2. One-way non-parametric Dunn's multiple comparisons corrected Kruskal-Wallis test heatmap for the mean specific productivities of clones obtained after population analysis on the Beacon®

ID		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
	Rank	63_8	63_5	62_4	62_10	62_7	63_13	63_7	62_2	62_8	63_4	63_1	63_11	63_9	63_3	62_5	63_10	62_1	63_2
1	63_8		1.00	1.00	1.00	1.00	1.00	1.00	0.07	0.03	0.17	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2	63_5	1.00		1.00	1.00	1.00	1.00	1.00	0.33	0.17	0.50	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00
3	62_4	1.00	1.00		1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.38	0.00	0.00	0.00	0.00	0.00
4	62_10	1.00	1.00	1.00		1.00	1.00	1.00	1.00	1.00	1.00	0.37	0.79	0.02	0.00	0.00	0.00	0.00	0.00
5	62_7	1.00	1.00	1.00	1.00		1.00	1.00	1.00	1.00	1.00	0.23	0.48	0.01	0.00	0.00	0.00	0.00	0.00
6	63_13	1.00	1.00	1.00	1.00	1.00		1.00	1.00	1.00	1.00	0.64	1.00	0.03	0.00	0.00	0.00	0.00	0.00
7	63_7	1.00	1.00	1.00	1.00	1.00	1.00		1.00	1.00	1.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00
8	62_2	0.07	0.33	1.00	1.00	1.00	1.00	1.00		1.00	1.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00
9	62_8	0.03	0.17	1.00	1.00	1.00	1.00	1.00	1.00		1.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00
10	63_4	0.17	0.50	1.00	1.00	1.00	1.00	1.00	1.00	1.00		1.00	1.00	1.00	0.08	0.06	0.07	0.00	0.00
11	63_1	0.00	0.00	1.00	0.37	0.23	0.64	1.00	1.00	1.00	1.00		1.00	1.00	0.04	0.04	0.06	0.00	0.00
12	63_11	0.00	0.01	1.00	0.79	0.48	1.00	1.00	1.00	1.00	1.00	1.00		1.00	0.02	0.02	0.03	0.00	0.00
13	63_9	0.00	0.00	0.38	0.02	0.01	0.03	1.00	1.00	1.00	1.00	1.00	1.00		0.26	0.20	0.30	0.00	0.00
14	63_3	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.08	0.04	0.02	0.26		1.00	1.00	1.00	0.15
15	62_5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.06	0.04	0.02	0.20	1.00		1.00	1.00	0.47
16	63_10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.07	0.06	0.03	0.30	1.00	1.00		1.00	1.00
17	62_1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00		1.00
18	63_2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.15	0.47	1.00	1.00	
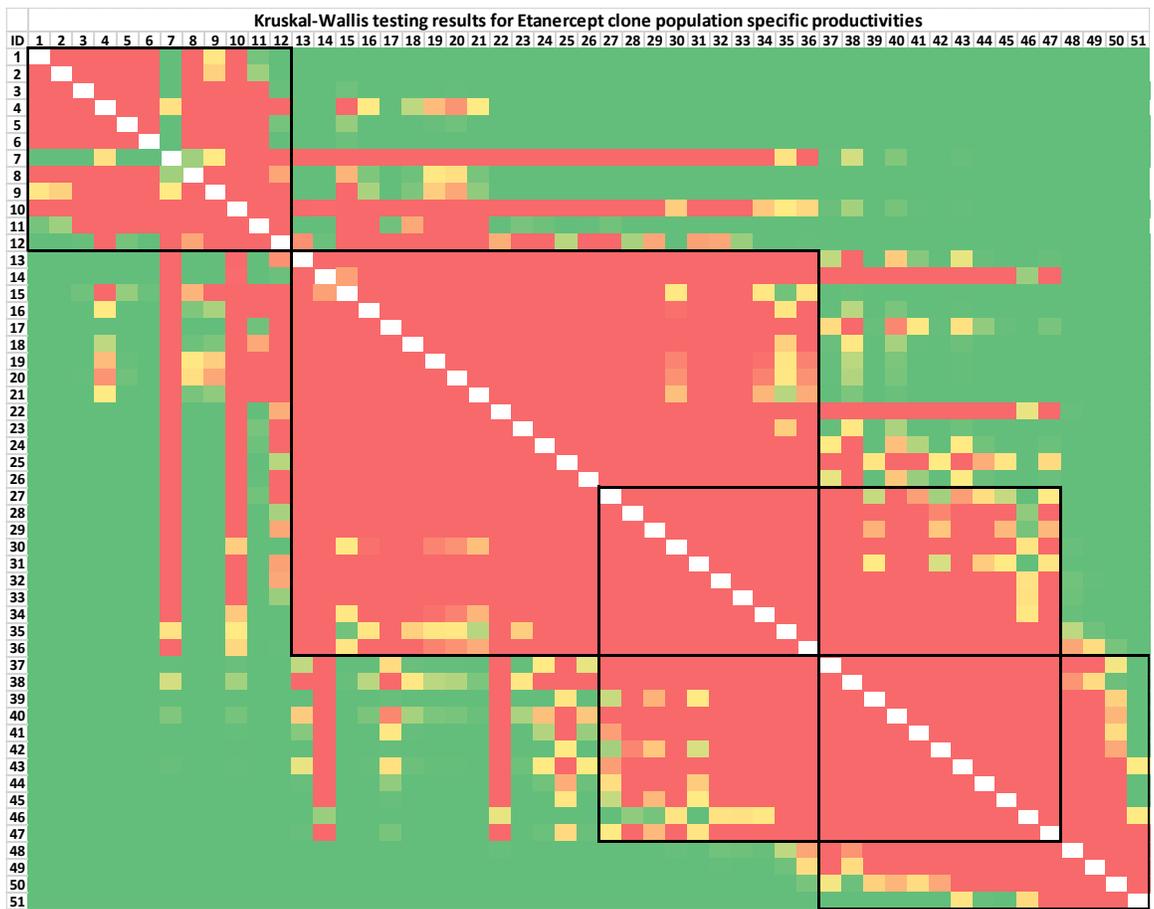
The transition from red into green is set at p=0.05 with green areas being statistically significant. The bold squares highlight the different groups within the table.

4.5 ANALYSIS OF VARIANCE IN INTRACLONAL AND INTERCLONAL POPULATIONS

Biological noise is one of the biggest issues in working with mammalian cell lines for manufacturing biopharmaceuticals. Regulatory agencies require companies to control for heterogeneity of product quality by developing monoclonal cell lines with CLD workflows as described in this thesis. Single cell data investigations into CHO heterogeneity are rare in the literature so we decided to investigate the variance of the titres and cell counts of cell populations in the Beacon® when they are derived from pools and from clonal cell lines (Figure 4.5.1). The means of the coefficients of variance for the BlosozumAb pool and clonal population titre measurements were 23.53% and 17.92% and these two groups were statistically significantly different from each other. The means of the measurements for the coefficients of variance of pool and clonal population cell counts were 34.48% and 35.30% with no statistical significance being observed between these two groups.

Figure 4.5.1. Comparisons of the coefficients of variance in titres and cell counts between clonal population and pool measurements.
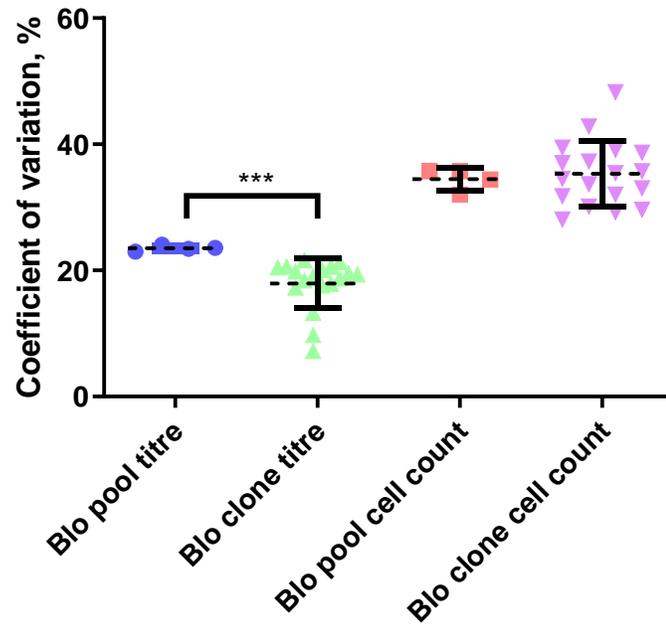
The stars denote the p-value as calculated by a two-tailed Mann-Whitney test. The error bars signify the means and the standard deviations of the populations.

4.6 DISCUSSION

4.6.1 Rationale for exploring CLD of two recombinant molecules using two different CLD approaches

We carried out the CLD process using two different production molecules, BlosozumAb and Etanercept. BlosozumAb is an anti-sclerostin IgG4 antibody that is produced by Eli Lily (Recker *et al.*, 2015). At the time of writing the drug had progressed through stage 2 clinical trials for the treatment of osteoporosis and based on an initial estimate, the projected sales of this drug could be as much as $500 million dollars per year if approved in the US (Taylor, 2019). While this is not a drug established in the market with a history of industrial manufacture, the reason this particular molecule was chosen is because it is commonly used in internal development at FDB and was classified as 'easy to express'. Therefore, our experiments would have some continuity with the body of cell line development work at the company and we would be able to contextualize our results within that experience.

Etanercept, however, is not an antibody, but an Fc fusion molecule. It was made by fusing the soluble part of the TNF receptor 2 protein and the Fc region of a human IgG1 antibody. The resulting fusion protein when dimerized has a similar molecular weight as a fully formed antibody would, 150 kDa. It is made to target the protein Tumor Necrosis Factor alpha (TNF-α), which is commonly known as a master regulator of the inflammation response. Unsurprisingly, for such a high-value target, there are several biotherapeutics on the market that directly target the downregulation of TNF-α (Mcdermott, 2016). The most well-known trade name of Etanercept is Enbrel®; in 2017 Enbrel® was one of the bestselling biotherapeutics by value in the world (Walsh, 2018). As a class, broadly, fusion proteins are considered to be more difficult to express than standard mAbs (Kiss *et al.*, 2018). One reason for this is that CHO cells have been optimised throughout the years as an expression platform to make mAbs. This includes efforts to formulate media, select cell lines based on titres and growth for mAb production and developing bioprocesses historically around mAb production. The other reason is that fusion proteins are synthetic constructs and thus not optimised for expression by nature in general, leading to potentially suboptimal yields.

As future biotherapeutics in development are considered to likely include more non-standard drug formats (Sauna *et al.*, 2017), we chose to include both types of recombinant protein to show that the CLD processes evaluated can be applied to study and optimize expression of different types of molecules, as well as to highlight potential opportunities for further investigation.

4.6.2   Analysis and limitations of the ClonePix™ 2 CLD

Electroporation was performed as the transfection method of choice as it works on a variety of cell types and in a variety of cell media. In the case of FDB, electroporation is the standard applied in CLD, whilst some lipid-based transfection reagents may not be as favourable to use with the FDB-MAP medium. A second caveat of the experimental methodology is that we scaled down the CLD process for each recombinant molecule group by a factor of 2 compared to what a commercial project would run to reduce costs and complexity. This should not compromise the results or validity of the study however, as we compared two different CLD methods head to head. Any differences between their capacities to isolate rare high producing clones should become apparent regardless of the reduction in overall screened clones.

The use of the ClonePix™ 2 technology in industrial mammalian cell line development is well documented (Dharshanan *et al.*, 2011; Hou *et al.*, 2014; Mangalampalli *et al.*, 2015). While the strengths of the ClonePix™ 2 technology are many; the high-throughput automatization of cell selection, versatility to use this platform in library screening and drug discovery, both brightfield and fluorescence information acquisition, it does not come without limitations. For example, the system uses semi-solid medium as a growth matrix (Figure 4.6.2.1) to allow the separation, monitoring and growth of cells. Such an environment is quite different from suspension bioreactor culture. When cells grow together, they produce colony microenvironments that have spatial metabolic and secreted protein gradients which can affect cell growth and production.

Second, the system uses a fluorescence signal to track levels of recombinant protein production, whether it is secretion into the media or intracellular. The fluorescence reagent is added to the medium before plating and takes a minimum of 10 days for the cells to grow out into colonies. The cells need to be stored in a static incubator for that duration which can cause problems with bleaching from the incubator being opened and closed throughout this time. The exposure of light might also be dependent on the positioning of the plates within the incubator, plates that are closer to the incubator door are more susceptible to bleaching than plates stored at the back. This makes comparing plates between picking runs complicated, but it's usually not a significant problem as the picking process in the CLD is used to select the cells which perform the best within the picking run. Any bleaching effects in the CLD are not impactful because we pick a set number of clones for each run. However, it makes comparing absolute secretion between the Etanercept and BlosozumAb clones difficult. Fluorescence intensity can also vary between different lots of fluorescent antibody however the same lot was used in this experiment.

Figure 4.6.2.1 Diagram depicting the growth of a single cell into a cell colony within a semi-solid medium complex.
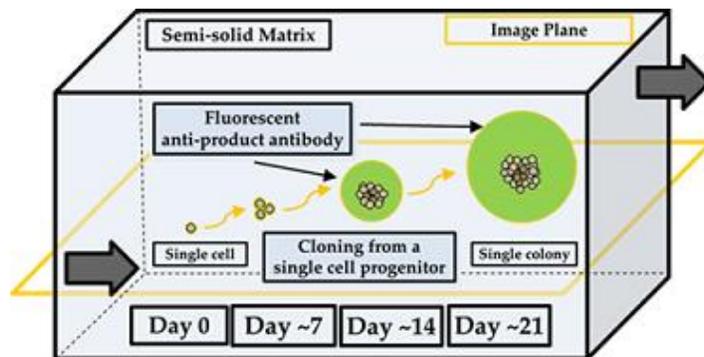
The medium contains a fluorescence anti-product antibody that produces visibility of the produced recombinant protein of interest. Taken from Acyte website http://www.acyte.com.au/technologies/axyte-cells-media.

Finally, the semi-solid medium is formulated differently from the FDB-MAP medium that FDB uses as the default medium for CLD. This is not unusual, because at the single cell stage, growth factors and additives are introduced to increase cell survivability. However, all of these differences need to be taken into account when considering predictability of clone ranking at this stage. In section 4.3.2 some statistically significant differences between the pools in both recombinant protein groups were observed at this stage. Based on the ClonePix™ 2 data alone, it's difficult to explain whether this difference arises through inconsistencies in the incubation process or intrinsic biological differences between the pools.

At the 96-well plate stage, some of the pool comparisons were similar titre wise to the ClonePix™ 2 stage. B3 was statistically significantly lower than B1 or B2, but B4, B1 and B2 were statistically indistinguishable. In the Etanercept group, E2 retained the statistical difference from the rest of the pools, but the differences between E1, E3 and E4 disappeared. These discrepancies can be explained by considering the details of the process of colony picking. The number of cells that are picked is never defined in the process, however on average bigger colonies will result in more cells being picked per well than smaller colonies. Because we compare these cells based on titres, if the colonies grew bigger in some plates more than others, we would see that pool advantage transferred at the 96-well plate stage.

Until the 24-well plate stage, a statistically significant comparison between both the Etanercept and BlosozumAb groups at the ClonePix™ 2 and the 96-well plate stages in terms of titre, was observed. Indeed, we would expect for different molecules to see a difference in titre. At the 24-well plate stage, the number of cells seeded was more controlled by ViCell counting

the TubeSpin® cultures before seeding the 24-well plates. At this stage no statistically significant differences between the titres of clones obtained from the pools for both the Etanercept and BlosozumAb groups were observed.

It is interesting to note that at the ClonePix™ 2 stage we observed that the ratio of fluorescence intensity means between the BlosozumAb and Etanercept groups was inverted compared to the 96-well plate stage and the 24- well plate stage. At the ClonePix™ 2 stage the ratio was as high as 4.3 in favour of the Etanercept, however, at and 96-well and 24-well plate this relationship becomes inverted with the ratio's being 1.6 and 1.8 in the favour of BlosozumAb. At all stages the two groups behaved differently from one another with a statistically significant difference between them being observed in all instances. Due to the fact that Etanercept and BlosozumAb are quite different recombinant proteins in nature, it is expected that they would have different expression profiles and effects on their host. At present, there is no data in the literature for us to cross-examine in order to compare whether the differences present here are due to host expression dynamics or stark differences between bleaching of the different sets of 6-well plates. These two sets were stored in two different incubators and were prepared a week apart, so it's entirely possible that this discrepancy arises purely due to different exposures to ambient light during routine incubator opening/closing over a 10-day period.

A similar type of CLD has been reported where the transfection and recovery steps are combined with the ClonePix™ 2 selection (Mangalampalli *et al.*, 2015). This allowed the group to reduce the CLD timelines to <3 months of work. Upon investigation of cell growth and morphology the authors did not find any differences between cells seeded onto semi-solid medium 48 h post transfection compared to cells seeded from a regular workflow. This highlights potential space for optimization of the CLD workflow used in this study, however we did not have the resources to validate this approach independently within FDB.

While the 24-well plate stage had control for initial cell concentrations and was subject to agitation, due to the 2.5 mL culture volume we were not able to keep track of the cell concentration during the fed batch process and thus unable to select based on specific productivity but rather just product titre, which limits the versatility of the overall process. The feeding regimes also did not consider the metabolite concentrations for determining whether a certain metabolite needed supplementation and there was no addition of base for pH control or anti-foaming reagents. All of these variables are usually taken into consideration at larger scale fed-batch culture. Last but not least, the overall V-shape of the well coupled with shaking speed of 200 rpm means that the cells will be subject to different shear stresses compared to a rotor-based agitation on the ambr® 15 scale.

4.6.3 Analysis and limitations of the Beacon® CLD process

The Berkley Lights Beacon® instrument is a relatively new platform on the market and as such there are only a few instances of it being used in the literature for cell line development (Le *et al.*, 2018; Mocciaro *et al.*, 2018; Jorgolli *et al.*, 2019; Winters *et al.*, 2019). In each of these cases the publishing group focused more on the results obtained using the system rather than investigating the performance of the system in greater detail. One of the goals of the work reported in this thesis was to provide a more thorough and transparent view of working with this system in a CLD workflow and based on our experience recommend improvement or guidelines to aspiring researchers in academia and industry. As such, we have described in great detail the entire process of preparing the Beacon® for use, cell culturing on chip and implementing fluorescence-based titre assays for selecting and exporting best performing cells for further validation.

We note the accuracy of the cell loading process compared to statistically random cell screening on the ClonePix™ 2. During colony screening we calculated that there should be 10000 cells deposited onto the semisolid medium, however that resulted in only roughly 1000 colonies screened per pool, a tenth of the original number. While it is possible that this number could be further optimized, this result is disappointing given the fact that we used the ClonePix™ 2 commercial semi-solid medium focused for CHO cells. With the Beacon® however, we have demonstrated that there can be tighter control over how many cells an operator wishes to screen and that >70% outgrowth of these cells on chip can be achieved.

It is interesting to note that in the population analysis of the clones it was difficult to make any predictions about the statistical significance between different clonal populations in terms of specific productivity. As noted in the results section, we do seem to be able to identify clusters of cell populations that seem to be significant from the others. In the case of both BlosozumAb and Etanercept the top most cluster seems to be around 12 clones, which was sufficient for our scale as that was how many we wished to carry through to the ambr® 15 screenings. However, if we wanted more resolution and discernibility between the clonal populations, extra research into the number of clones that needs to be screened to achieve optimum statistical power would be required. Finally, standardizing the media is essential to remove potential variability between the runs that might distort the way the cells behave. Further work should investigate the relationship between the time of culturing the cells on chip and the ability to compare them. It is possible that with time the statistical power becomes lower as more time is given for cells that performed worse on average to 'catch up' to the first starters. Conversely, more time may allow

a better discrimination of poor from good producing cells.

4.6.4 Cell growth and viability during both CLD processes

During all stages of the CLD process, the conditions for cell growth are 'optimized' for the different formats/stages, however we have to take into account that these three stages are very different systems of culturing cells. The limitations of using the ClonePix™ 2 are discussed above, however there are severe limitations (as well as advantages) to using 96-well plate and 24-well plate scales as well. When cells are grown statically the gas exchange is mediated by diffusion only, as the cells are not subjected to any agitation. At the 96-well plate stage of the CLD we did not control for the initial cell concentration upon seeding or track the confluence of the wells in order to determine specific productivity. Such an approach has been determined to be valid in an unbiased comparison to a titre only-based selection (Pristovšek *et al.*, 2018). Using such a combined approach on Etanercept producing CHO cells it was possible to enrich the final selection of clones with cell lines capable of reaching high IVC and high specific productivity contributing to a more versatile final selection panel. The number of cells screened at the 96-well stage was similar to our proposed scale, 852 surviving clones were screened compared to the 600 colonies we picked from each group.

One of the initial findings we found working with the Beacon® platform is that without supplementation and using regular FDB-MAP medium + 8 mM L-glutamine + 175 nM MTX we could not get sufficient growth on chip of the cells. In Figure 4.4.2.7, we introduce the concept of the outgrowth threshold as a measure of cell proliferation on-chip. This metric describes the ratio of occupied pens that have ≥6 cells at the time of measurement. Before optimizing the medium for our cells, we could not achieve growth that would surpass the 30% mark over a 3-day period of incubation. To achieve this criterion, we had to utilize a proprietary Berkley Lights medium additive that contains various growth factors along with using 20% conditioned medium. This is not particularly novel as a lot of culturing systems at the single cell level use some sort of supplementation (Chen and Pruett-Miller, 2018). In the case of another published Beacon® study (Le *et al.*, 2018), it was specified that for their FACS based cloning and Beacon® culturing they use proprietary growth medium, however it is not clear whether the same medium works across both platforms or they needed distinct formulations for both CLD workflows. We have found that the FBD-MAP medium works well as a base cloning medium and requires only minor supplementation in order to get appropriate outgrowth on chip. For the export of the cells into 96 well plates we used the regular FDB-MAP + 8 mM L-glutamine + MTX formulation and

were able to get 24/96 (25%) recovery rates for BlosozumAb export into the 96-well plate stage. The survival ratio was a little better for the Etanercept group, survival rates for CHO cells have been reported exceeding 50% in Berkley Lights internal development experiments which matches survival rates obtained through limiting dilution cloning and FACS based cloning (Lim *et al.*, 2013), however obtaining this would require further optimization of the FDB-MAP medium and determining the optimum number of cells that need to be unpenned for export.

The reason that there were more clones recovered for the Etanercept group is that the processes were staggered and after observing poor recovery statistics for the BlosozumAb group, we decided to export after an extra day of on-chip culturing. This resulted in an increase in clones exported and recovery rate. We exported a total of 96 BlosozumAb clones and a total of 24 survived the 96-well plate stage (25%) and, for Etanercept, 192 and 71 survived (31%). The second major loss of clone at scale up was at the shake flask stage, 20 to 17 from BlosozumAb and 71 to 50 for Etanercept. This makes sense considering the fact that this stage of the scale up process involves a dramatic change of culture conditions for the cells, from static to shaking culture. The only other major shift in culturing conditions was during the export stage as the cells transition from being cultured in microscale pens with a perfusion system to a much bigger milli-scale volume with a static batch culture. There is no data to our knowledge regarding the shear stress that cell undergo upon export as the literature has shown that "*Cells can be exposed to very high hydrodynamic forces when flowing through channels and nozzle in the sorting process. Results indicate that not only are cells damaged in a flow cytometer, but that this damage can vary from cell line to cell line as well as from specific conditions/type of flow cytometer and flow conditions*" (Berdugo, 2010). It has been shown that such shear stress forces can damage CHO cells (Hu *et al.*, 2011).

In short, both systems seem perfectly capable of maintaining cell culture viability and growth on a single cell level and both have their own nuances in optimization in order to ensure that precious high producing clones are not lost. The ClonePix® 2 system has had years of semi-solid media optimization and therefore is easier to work with out of the box, however, we show that with minimal optimization we can get good cell growth and viability on the Beacon® as well.

4.6.5 Comparing the Beacon® CLD workflow with market alternatives

In this section how the Beacon® CLD workflows compare to those of other devices on the market is explored alongside a discussion of the perceived advantages and disadvantages of these systems. One of the most common approaches by which CLD is performed in an industrial

setting is via FACS based sorting. There are many different FACS systems with different configurations and it is not in the scope of this work to explore them all. Rather, here is considered systems that are either particularly suited for CLD or widely adopted in industry.

FACS based sorting works by using the different optical properties of cells as they pass through a narrow, rapidly flowing stream of liquid. The cells can then be subjected to different light sources and their light scattering and fluorescence characteristics can then be analysed to establish 'cut-off' or threshold gates for cell selection. FACS is probably the most widely utilised system within industry to conduct single-cell cloning. The advantages of FACS are two-fold. Firstly, it is relatively easy to set up and the cost of operation is inexpensive (although the instruments themselves may require an initial outlay), even when considering the cost of reagents for fluorescent cell labelling. Second, it is a very versatile platform supporting a large knowledge base of fluorescence assays for a wide range of cell characteristics. The main disadvantage of a FACS system are that it is an open system and needs to be in an enclosure (the usual case for industrial CLD) or use antibiotics, which are not always viable in an industrial CLD context; to maintain sterility. There is also no ability to directly observe the cells as they go through the device, although this has been ameliorated in recent years with the addition of new FACS machines that have microscopy capacity like the Amnis® imaging flow cytometer. A further potential issue is that since the cells are screened as they go through a sheath fluid, they cannot be retained within the device or tracked and thus the measurements are more of a snapshot of the cell in time rather than being a continuous observation. Thus, there is no way to measure growth or productivity over time within a FACS device.

The main methodology used to discriminate high and low producing cells in FACS is via cold-capture labelling where the cells are subjected to a treatment with a fluorescent antibody that targets the recombinant protein of interest. The antibody can then bind to the recombinant protein that is mid-secretion or stuck in the membrane. However, this is not a direct measurement of the secretory capacity of the cell. During such staining on the Beacon® we actually observe a lot of cells that seem to glow brightly, but do not seem to have high secretion and cells that are not bright, but have high secretion capacity (Figure 4.6.5.1). This suggests that cell surface staining is not the best tool for identifying high producing clones. It is unknown however whether this is a phenomenon of cell surface staining or cellular internalization of the fluorescent dye. It is clear that in this regard the Beacon® enables a whole new area of cytological analysis previously inaccessible with FACS.

Figure 4.6.5.1. Close up fluorescence image of cells during secretion assay
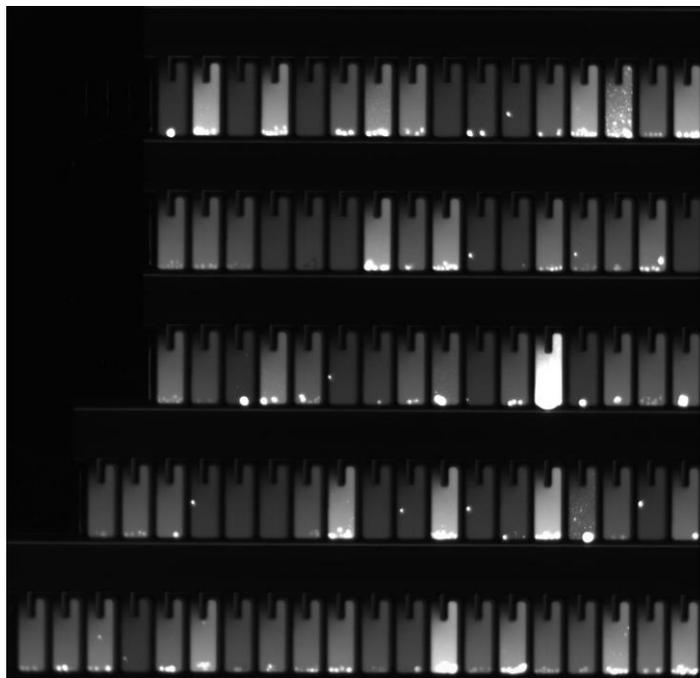
Other products on the market exist that are more specifically designed with CLD in mind, for example the Solentim® Verified *In Situ* Plate Seeding system (VIPS). The VIPS system is a cell-seeder and plate imager 2-in-1 device. The advantage over the FACS system is that it does not require a dedicated specialist to run and operate, it gives a very high probability of clonality and uses a much lower nozzle pressure than a flow cytometer decreasing cell shear stress and increasing cell survival. It does not have the ability to run assays on the cells however, though it does support fluorescent 96-well plate imaging. Thus, while convenient, this method doesn't offer a lot of power to discriminate between clones.

A further product on the market for CLD is the Cyto-Mine® from Sphere Fluidics. The technology uses microfluidics to encapsulate cells in a droplet of surfactant. The main advantage of this is that it is possible to perform secretion assays on the cells using a fluorescent reporter signal. This system works by utilizing FRET to give a fluorescent signal if there is only unassembled antibody bound in the droplet. However, when the donor-acceptor pair is in close proximity with each other in the assembled molecule a shift in the fluorescence signal towards a longer wavelength is observed. As far as the author is aware, this is the only other commercial system that allows for direct measurement of secretion apart from the Beacon®. The system also allows direct observation of the microdroplets removing the need for additional monoclonality monitoring systems. On the manufacturer's website, it has been reported that the outgrowth in 96-well plates matches that of limited dilution, assuring cells do not undergo any stress in their

encapsulated forms or during the FRET assay. Another potential advantage of this method is that encapsulated cells have a micro environment which can be potentially analysed downstream via metabolomic methods such as mass spectrometry, whereas with the Beacon® that is traded off with an open pen creating a steady-state environment which may allow a better monitoring of growth and dynamic manipulation of the environment of the cells. To the authors knowledge, at the time of writing, there are no peer-reviewed work with the Cyto-Mine® in the literature, however a report (Kelly *et al.*, 2018) by Janssen Pharmaceutical mentions the successful application of the technology in their CLD workflows. They highlight that Cyto-Mine® sorted cell titres were better at the 96-well plate stage compared to semisolid agar screening and that the automated visual recognition system was conservative and tended to overestimate the number of cells in the droplets. With the Cyto-Mine® they reduced their need for subcloning and shortened their CLD timelines by 4-6 weeks to roughly two months. There was no scale-up past the 96-well plate stage so whether this data translates to industrially relevant bioreactor conditions is currently unknown.

Figure 4.6.5.2. Schematic of a picodroplet in the Cyto-Mine® containing a single cell and visualising the FRET secretion assay
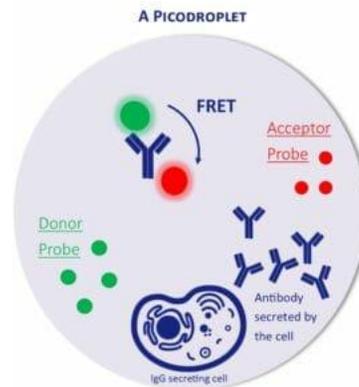
Sourced from the Cyto-Mine webpage.

Last but not least, Molecular Devices has released the CloneSelect™ Single-Cell Printer Series. The essence of this single-cell printing technology is based on a disposable cartridge combined with a high-resolution camera. The sample with a cell suspension is deposited into the cartridge where a piezo-based actuator is used to deposit the droplets out of the nozzle onto the target well on a plate. An example of this is shown in Figure 4.6.5.3 below. This allows for eyewitness proof of monoclonality and captures a sequence of events to determine the cell number in each droplet. The platform supports both brightfield label free and fluorescence-based cell sorting. The technology is different from FACS in that it does not offer light scatter data and is less

flexible, however, it is more enclosed ensuring sterility, and the sterile cartridges do not come into contact with the instrument itself.

Figure 4.6.5.3. Images showing the sequence of steps at the nozzle of the CloneSelect™ Single-Cell Printer during a single cell detection event. 1-3 cell approaching the nozzle

4 – cell detected and verified. 5 – cell ejected. Taken from CloneSelect™ manufacturer's website.
https://www.moleculardevices.com/products/biologics/clone-screening/mammalian-screening/cloneselect-single-cell-printer-series.

4.6.6 Insights gained from investigating inter and intraclonal variability on the Beacon®

What are the average standard errors of the mean in the pool screens and the population screens for both groups based on the three parameters investigated here (Qp, titre, IVC)? From the literature we know that CHO cell lines do not have a stable karyotype and that clonal stability has no correlation with genomic stability at the clonal level. In fact we can find reported cases of master cell banks that have had post-clonal mutation events resulting in distinctly genetically different populations that had no impact on cell performance (Scarcelli *et al.*, 2018). There are also reports where researchers make measurements of the heterogeneity of intraclonal populations and transgenic pools finding them to be similar in terms of titre, around 70% of the mean (Pilbrough *et al.*, 2009). The International Conference on Harmonisation (ICH), Guideline Q5D, 1997 makes a refutation of this idea "*In any circumstance, and under most efficient processing of cells from a MCB to a 1000 litre production vessel (typically 30 days) only a small part (estimated as less than 1%) of the theoretical biomass of cells emerging from a vial of frozen cells in a Master Cell Bank will be composed of a cell population structurally highly related and with transgene DNA identical to cells of the original frozen MCB vial*". To sum up, the scientific community has known for years that the idea of monoclonal homogeneity is not supported by data.

The current dogma of therapeutic recombinant drug production regulation has established the monoclonal cell line as the gold standard for a reproducible and safe manufacture process that minimizes heterogeneity. Most regulatory agencies require strict evidence of the

monoclonality of cell lines to ensure patient safety and drug quality. However, these guidelines typically rely on the assumption that monoclonally derived cell lines are more stable. So far all attempts in the literature to find either a phenotypical determinant of the stability of bioprocess characteristics have been unsuccessful with data refuting the idea of homogeneity. The measurements we have taken of the heterogeneity of the titre scores in the pool screens and the clonal population screens in the Beacon® point towards the same conclusion.

The advent of the use of next generation single cell technologies in CLD brings up this conundrum again. As we gain more capacity to characterize these cell lines at the single cell level an increasing amount of intraclonal variation will in all likelihood be observed. This brings about the question, should monoclonality be a strict guideline if a company is able to provide a production process based on cell populations that are of non-monoclonal lineages? Relaxation of such restrictions could potentially reduce the resources spent on cell line development and, if the data demonstrates appropriate stability of a transgenic pool, potentially drastically reduce development time and costs, an aim that is at the heart of this thesis albeit through the use of novel technologies such as the Beacon®.

4.7 SUMMARY

1) We found that both CLD processes are capable of maintenance of cells, clone selection and export allowed. The Beacon® offers an exceptional amount of precision in terms of cell screening (0.00083% loading precision and 70% viability on chip) whereas the ClonePix® 2 allowed us to screen approximately 1200 clones out of a hypothetical number of 5000.

2) The Beacon® platform, due to being capable of single cell real time monitoring allows us to offer precise measurements about the behaviour of isolated populations of cells and demonstrates the counterintuitive similarities between intra and interclonal variability.

3) The Beacon® allowed us to reduce CLD timelines from 54 to 39 days of work.

4) The novel population analysis step on Beacon® allowed us to make statistical comparisons between clone bioparameters in order to inform export decisions.

4.8 CONCLUSIONS

In this chapter, the ability and validation of the Berkley Lights Beacon® platform for CHO cell line development, as compared to a CLD workflow derived from GMP protocols at FDB using a colony picking approach, has been shown. The two platforms have been compared and the advantages and disadvantages of both systems discussed. With the Beacon® system it is possible to reduce the development time from approximately 54 days to 39 days comparing both of the workflows. However, the full industrial manufacture workflow includes an extra step of subcloning to ensure monoclonality extending the CLD to 147 days which would be a 70% reduction in development time. This assumes that monoclonality is accepted from the Beacon® instrument and no further work is required to confirm this. The work also shows that the Beacon® gives additional control and information on the cell populations being interrogated than the traditional method which means more support for data driven CLD and drug discovery. However, the Beacon® also has disadvantages as outlined in the chapter. Furthermore, the initial cost of the instrument is high (in the millions USD $) and chips themselves are expensive. Thus, although the Beacon® offers some advantages over other systems, the choice of CLD approach will depend on a number of factors that each industrial applicator needs to consider in ultimately deciding which approach to adopt.

5 VALIDATION OF A NOVEL OPERATIONAL CONFIGURATION OF THE BEACON® OPTOFLUIDIC PLATFORM FOR CHO CLD TO 15 ML SCALE IN AN AMBR® 15 BIOREACTOR

5.1 INTRODUCTION

In the previous Chapter, two different cell line development workflows were undertaken and reported, a colony picking-based approach adapted from GMP manufacture protocols carried out at FDB and a novel optofluidic based approach using the Berkley Lights Beacon® technology. The CLD process was carried through into a TubeSpin® scale, however, the clones need to be validated under bioreactor like conditions that more accurately predict their performance in the process environment. In this section ambr® 15 fed-batch cultures of the clones obtained through CLD were undertaken to determine whether there were differences in the cell lines selected for production of the target recombinant proteins, Etanercept and BlosozumAb. The ambr® 15 system allows 48 simultaneous fed-batch cultures with bioreactor like control over the culture conditions including stirring, gassing and automated liquid handling to be undertaken, reducing operator error in the experiment and more closely mimicking the environment at the production scale.

Thus, in this Chapter we explore the ability of these workflows to make predictions around how the cells behave in a bioreactor environment. The gold standards by which a CLD workflow is judged is the ability to discover high-producing, stable cell lines within a polyclonal cell pool, the ability to predict how such cell lines will behave at scale up and the reproducibility of the process. While the reproducibility of these workflows was not assessed as part of this work, it was possible to compare their discovery and predictive powers at different scales.

5.2 AMBR® 15 SCREEN OF CLONES GENERATED BY CLONEPIX™ 2 AND BEACON® CLD WORKFLOWS

A full layout of the ambr® 15 fed-batch culture is detailed within the Methods section 2.2.13. Fed-batch cultures were undertaken in the ambr® 15 with 48 cell lines split into two groups; 24 ClonePix™ 2 cell lines (12 BlosozumAb (Blo) and 12 Etanercept (Et)) obtained at the 24-well plate screening stage of the CLD process; 24 Beacon® cell lines obtained through the culturing of the cells on chip and przedicting their rankings based on a fluorescence assay score. These cell lines are presented below in Table 5.2.1.

Table 5.2.1. The clones selected for ambr® 15 fed-batch screening from the CLD workflows undertaken with the ClonePix™ 2 (Fuji) and with the Berkley Lights Beacon® platform for BlosozumAb and Etanercept recombinant protein production.

Blo Beacon	Blo Fuji	Et Beacon	Et Fuji
63_8	C1	17_12	C13
63_9	C2	17_17	C14
62_7	C3	Fe1_6	C15
63_13	C4	17_20	C16
63_11	C5	17_26	C17
62_10	C6	16_30	C18
62_8	C7	Fe1_9	C19
62_4	C8	17_3	C20
62_2	C9	17_24	C21
63_1	C10	17_25	C22
63_7	C11	17_14	C23
63_5	C12	17_13	C24

The growth characteristics of these cells in the bioreactor environment are initially evaluated. The cell count data along with the culture viability is reported in Figure 5.2.1. From the figures it is clear that the cells behave in a similar fashion across the board, viability remains stable until day 6 and starts declining around day 8 steadily until the end of culture. Viable cell concentration peaks at day 8 around 1.5×10^7 viable cells/mL for all groups and as the culture overgrows the cell lines in the groups start to die off. At the end of the process the culture viability was between 40-80%. One notable difference was a cell line from the Etanercept Beacon® group that crashed at day 8, however this seems to be an isolated event rather than a reflection of the general performance of the clone group.

Figure 5.2.1. Line and Tukey box plots of the cell count data of ambr® 15 fed batch cultures over a 14 day period.

Figure 5.2.2. The fitting of linear models to the period of exponential growth (red dots) using the ``growth rates made easy method'' of *(Hall* et al*., 2014)*.

Growth curve modelling of ambr® 15 cell lines
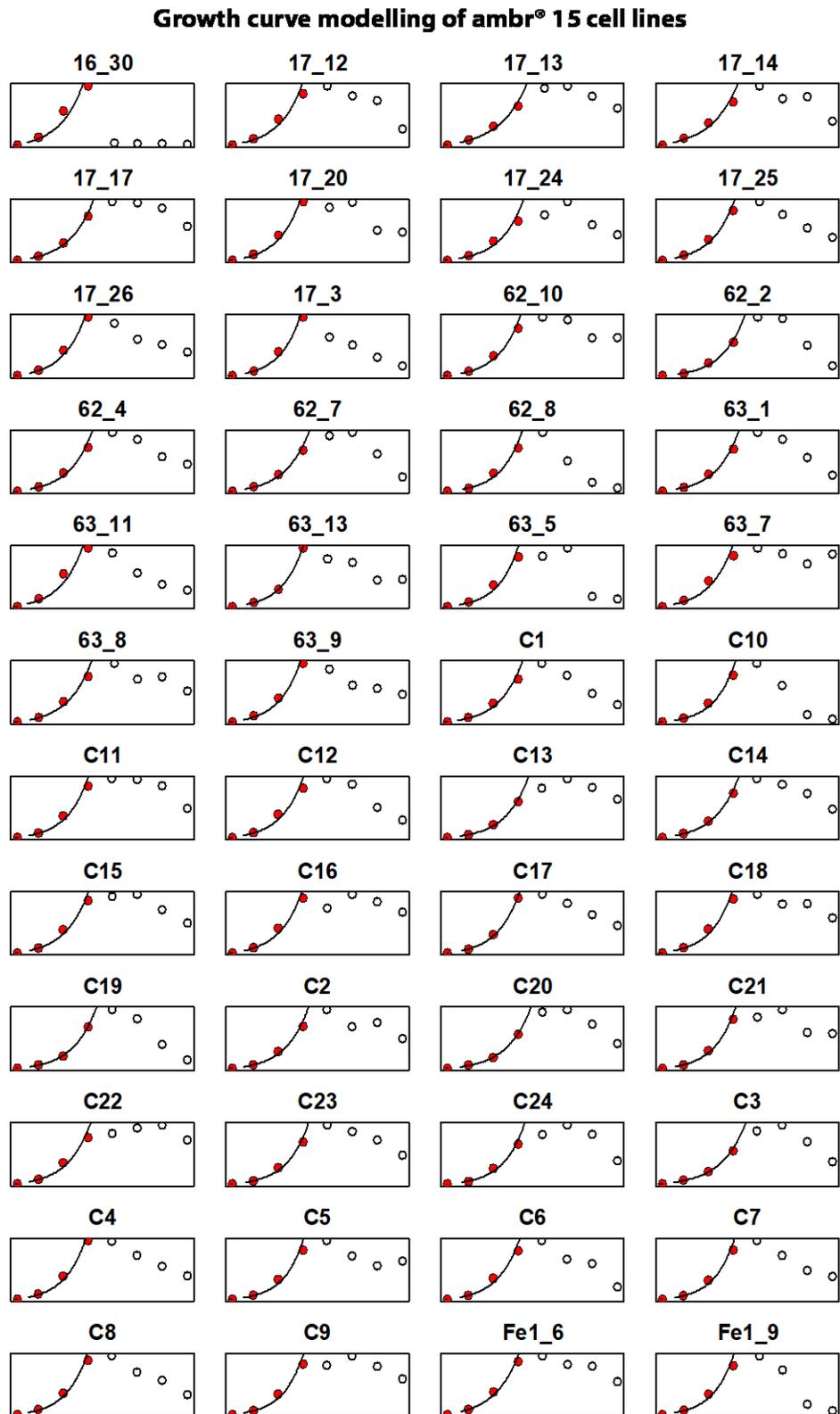
Comparing and evaluating clones based on culture viability and viable cell density, however, does not represent the entire view of cell growth dynamics. Two more comprehensive parameters to analyse these 4 groups of clones, maximum growth rate (μ) and the integral of viable cells (IVC) which is an indicator of cumulative cell hours where cells can be performing recombinant protein production, were therefore also undertaken. The growth rates were calculated using a linear fitting model which took into account the first 6 days of growth data. The plotted exponential functions have been visualised in Figure 5.2.2 and the results of the model presented in Table 5.2.2. From these data we can see that the exponential cell growth model was a good fit for the data overall, the coefficient of determination R^2 was in the high 0.97-0.99 values.

Table 5.2.2. The fitting parameters and results from Figure 5.2.2.

Cell line	Group	y_0	$y_0$lm	μ_{max}	t_{lag}	R^2	T_d
62_10	Blo Beacon	620172	338211	0.0232	26.13	0.99	29.87
62_2	Blo Beacon	715408	350528	0.0231	30.87	0.99	29.99
62_4	Blo Beacon	769911	424965	0.0227	26.18	0.99	30.54
62_7	Blo Beacon	587685	341986	0.0220	24.66	0.99	31.57
62_8	Blo Beacon	643814	314067	0.0248	28.98	0.99	27.99
63_1	Blo Beacon	548892	292020	0.0232	27.18	0.99	29.85
63_11	Blo Beacon	574864	364329	0.0234	19.48	0.96	29.61
63_13	Blo Beacon	620160	302289	0.0250	28.79	1.00	27.77
63_5	Blo Beacon	495550	266504	0.0258	24.07	0.98	26.90
63_7	Blo Beacon	588112	352787	0.0230	22.25	0.97	30.18
63_8	Blo Beacon	592790	336412	0.0221	25.67	0.99	31.41
63_9	Blo Beacon	598300	346501	0.0242	22.54	0.98	28.60
C1	Blo Fuji	614812	385527	0.0228	20.44	0.99	30.36
C10	Blo Fuji	704880	396471	0.0251	22.97	0.99	27.67
C11	Blo Fuji	601832	344292	0.0243	22.98	0.99	28.53
C12	Blo Fuji	612070	367850	0.0238	21.41	0.98	29.14
C2	Blo Fuji	616284	362802	0.0235	22.59	0.99	29.55
C3	Blo Fuji	588132	369177	0.0214	21.72	0.99	32.33
C4	Blo Fuji	629104	367409	0.0248	21.70	0.99	27.96
C5	Blo Fuji	598950	348031	0.0264	20.56	0.98	26.25
C6	Blo Fuji	644844	361407	0.0240	24.17	0.99	28.94
C7	Blo Fuji	646140	349918	0.0257	23.88	0.99	26.99
C8	Blo Fuji	623070	373098	0.0245	20.96	0.99	28.33
C9	Blo Fuji	637650	360108	0.0257	22.27	0.99	27.01
16_30	Et Beacon	665720	421680	0.0229	19.96	0.96	30.30
17_12	Et Beacon	571298	359702	0.0227	20.41	0.97	30.58
17_13	Et Beacon	692160	435339	0.0188	24.67	0.99	36.88
17_14	Et Beacon	593769	393865	0.0196	20.95	0.97	35.37

17_17	Et Beacon	561792	317395	0.0216	26.43	1.00	32.09
17_20	Et Beacon	603508	351273	0.0237	22.83	0.99	29.23
17_24	Et Beacon	521352	314354	0.0228	22.22	0.98	30.45
17_25	Et Beacon	684454	393433	0.0228	24.30	0.99	30.42
17_26	Et Beacon	659600	372645	0.0233	24.50	0.99	29.74
17_3	Et Beacon	648720	342546	0.0250	25.54	0.99	27.72
Fe1_6	Et Beacon	658240	381730	0.0208	26.15	0.99	33.27
Fe1_9	Et Beacon	619520	342797	0.0251	23.61	0.98	27.66
C13	Et Fuji	530334	295292	0.0228	25.73	1.00	30.46
C14	Et Fuji	467280	269411	0.0222	24.81	1.00	31.23
C15	Et Fuji	606594	354782	0.0221	24.23	0.99	31.31
C16	Et Fuji	570764	331346	0.0235	23.15	0.99	29.51
C17	Et Fuji	644182	327458	0.0237	28.57	1.00	29.27
C18	Et Fuji	560798	329735	0.0242	21.92	0.98	28.62
C19	Et Fuji	591085	329631	0.0227	25.75	1.00	30.57
C20	Et Fuji	569008	327348	0.0220	25.15	1.00	31.53
C21	Et Fuji	609724	325250	0.0247	25.41	1.00	28.02
C22	Et Fuji	541940	316278	0.0228	23.61	0.99	30.39
C23	Et Fuji	578976	301131	0.0245	26.69	0.99	28.30
C24	Et Fuji	505305	247652	0.0250	28.55	0.99	27.75

R^2 – coefficient of determination; T_d – doubling time, h; μ_{max} – maximum growth rate, h^{-1}; t_{lag} – the time it takes to go into exponential phase, h; y_0 and $y_0 lm$ – fitting constants for the linear model. Doubling times were calculated using the $T_d = \frac{ln2}{\mu_{max}}$ relationship.

Alongside the growth data, we also compared the titre and specific productivities of the clones based on their day 14 final recombinant protein concentrations. This aggregate data is shown in and Figure 5.2.4. In the titre category there was a clear difference between the BlosozumAb and the Etanercept group, as expected; the means for BlosozumAb titres were 1240 mg/L and 1306 mg/L compared to 539.4 mg/L and 671.7 mg/L for Etanercept clones. This is around a 2-fold difference and the statistical method we employed were able to differentiate between the BlosozumAb and Etanercept groups, however not between whether groups of clones came from the ClonePix™ 2 CLD or Beacon® workflows.

A similar situation was observed when considering the specific productivity data for the cells. The BlosozumAb group means were very similar, 9.19 pg/cell/day and 9.06 pg/cell/day for the Beacon® and Fuji clones and in the Etanercept group the means were 4.87 pg/cell/day and 5.02 pg/cell/day. In this case the ANOVA could still differentiate between the BlosozumAb and Etanercept groups, but not by the CLD subgroups. In terms of protein output on both titre and specific productivity metrics the Beacon® therefore does not appear to isolate different groups of clones in our comparisons.

Figure 5.2.3. Dot plots showing the day 14 titres, specific productivities and the maximum doubling times of the ambr® 15 fed-batch cultures of the clones obtained from the ClonePix™ 2 (Fuji) and Beacon® CLD workflows.
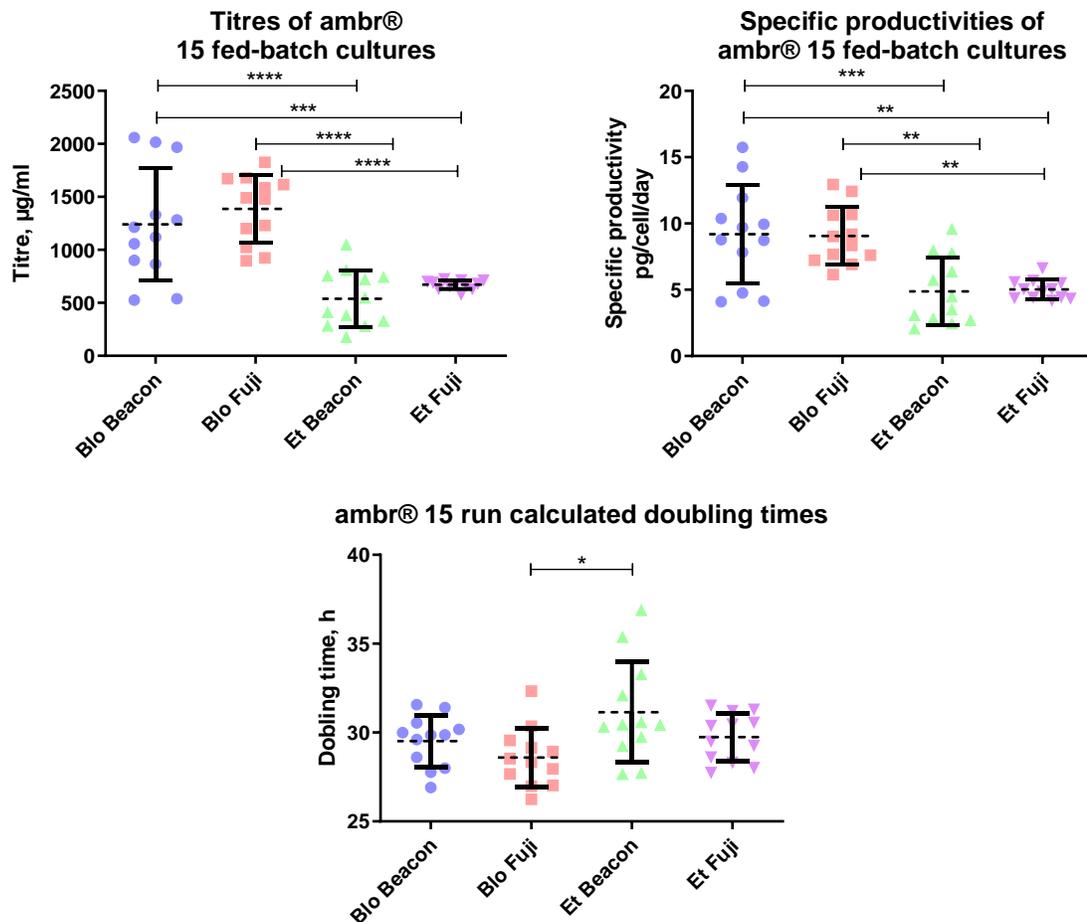
Error marks depict the mean (dotted line) and the standard deviation. Statistics were performed using a one-way Welch Geisser-Greenhouse corrected one-way ANOVA with Games-Howell's multiple comparisons correction.

When doubling times were considered, there were no statistically significant different comparisons between any of the clone groups here using ANOVA. This suggests that the recombinant protein itself is not having an effect on the maximum growth rates of the clones at the ambr® 15 scale.

It is worth noting that within the titre and specific productivity categories we observe very small standard deviations for the ClonePix™ 2 Etanercept group, 0.7340 pg/cell/day and 42.37 mg/mL. The standard deviations for the ClonePix™ 2 BlosozumAb group were 318.2 mg/mL and 2.18 pg/cell/day while the Beacon® Etanercept group standard deviations were 2.54 pg/cell/day

139

and 270.5 mg/mL. This discrepancy suggests that the deviation is not explained by either the CLD or the phenotype of Etanercept producing cells.

When the integrals of viable cell concentrations were considered (Figure 5.2.4) over time, the cells in the bioreactors were accumulating similar amounts of cell work hours, 3257 cells/h/mL and 3699 cells/h/mL for BlosozumAb and 2709 cells/h/mL and 3268 cell/h/mL for Etanercept. The means for the Beacon® groups we were able to discern differences between some of the subgroups. From day 6 and onwards there was a statistically significant difference between the BlosozumAb Fuji and the Etanercept Fuji groups, but all the statistical significance was observed between BlosozumAb and Etanercept groups and not between the type of CLD that was used to generate the cell lines.

Figure 5.2.4. Tukey box plot showing the IVCs of the ambr® 15 cultures calculated over the period of the 14-day fed-batch culture.
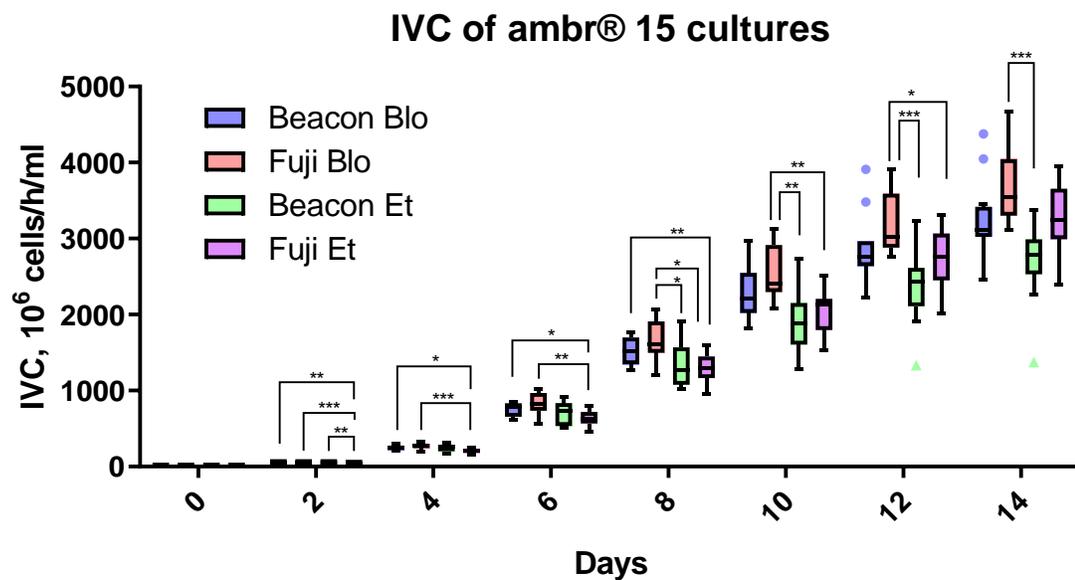
Statistical analysis was performed using a two-way repeated measures Geisser-Greenhouse corrected ANOVA with Tukey's multiple comparisons correction.

A further important aspect for the comparison of the two CLD processes was to look at the ranks of the titres and specific productivities of the clone distributions for BlosozumAb and Etanercept. In Figure 5.2.5 the clones have been sorted by rank and colour coded by CLD. For BlosozumAb, the top 3 titre cell lines and top 2 Qp came from the Beacon® workflow. In the Etanercept group the top clones came from the Beacon® CLD as well, the top 4 titre clones and top 3 Qp were from the Beacon® CLD.

Cell bioprocess characteristics are not limited to growth and titre. To add another layer of dimensionality to the dataset we can also examine whether any significant metabolic differences in the glucose and lactate trends exist in the clone groups. This data is presented in Figure 5.2.6. This data shows that all clone groups follow a general trend of an increase in glucose concentration until day 4 and then consumption overtaking the glucose fed amounts within the media. Thus, on days 4-8 the trend changed to glucose concentrations decreasing and then stabilising at days 9-14. The major outlier observed in the Beacon® Etanercept group was the cell line that crashed on day 8 and we observe this outlier in the lactate plot as well.

The general trend for lactate seems to be a steady increase until day 5 and then the concentration of lactate stabilising until day 9 after which the concentration drops off as the cell presumably consume lactate, a phenomenon that has previously been reported in CHO cells. These trends correspond to major phases within the growth curve. Cells should be dividing the fastest around day 5 and beyond this they are no longer in the exponential growth phase. After day 9 the cell cultures have reached maximum viable cell concentration and are beginning to decline. Statistical analysis of metabolite data suggests no significant systemic difference between glucose and lactate concentration patterns throughout cell culture for the cell derived from Beacon® and ClonePix™ 2 CLD workflows.

Figure 5.2.5. Bar plots showing the rank order of titres and specific productivities for Etanercept and BlosozumAb clones obtained through the Beacon® (RED) and ClonePix™ 2 (BLUE) workflows.
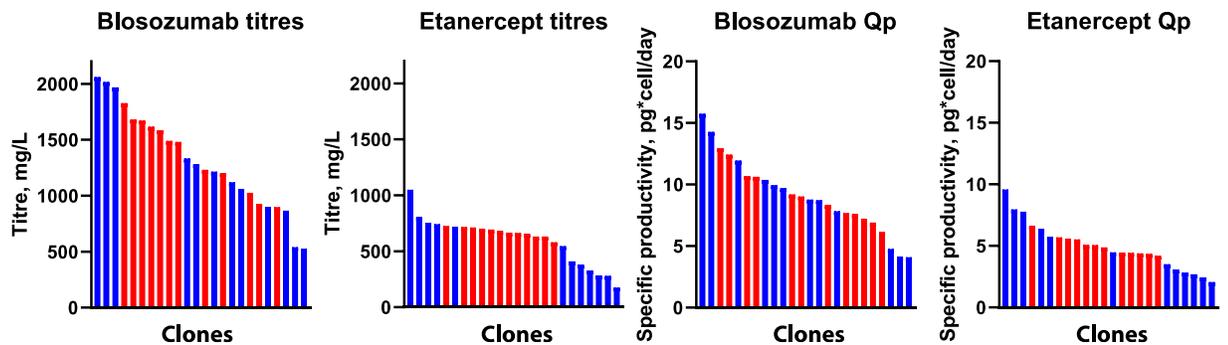
Figure 5.2.6. Box-plots showing the metabolite measurements for glucose and lactate within ambr® 15 fed-batch cultures.
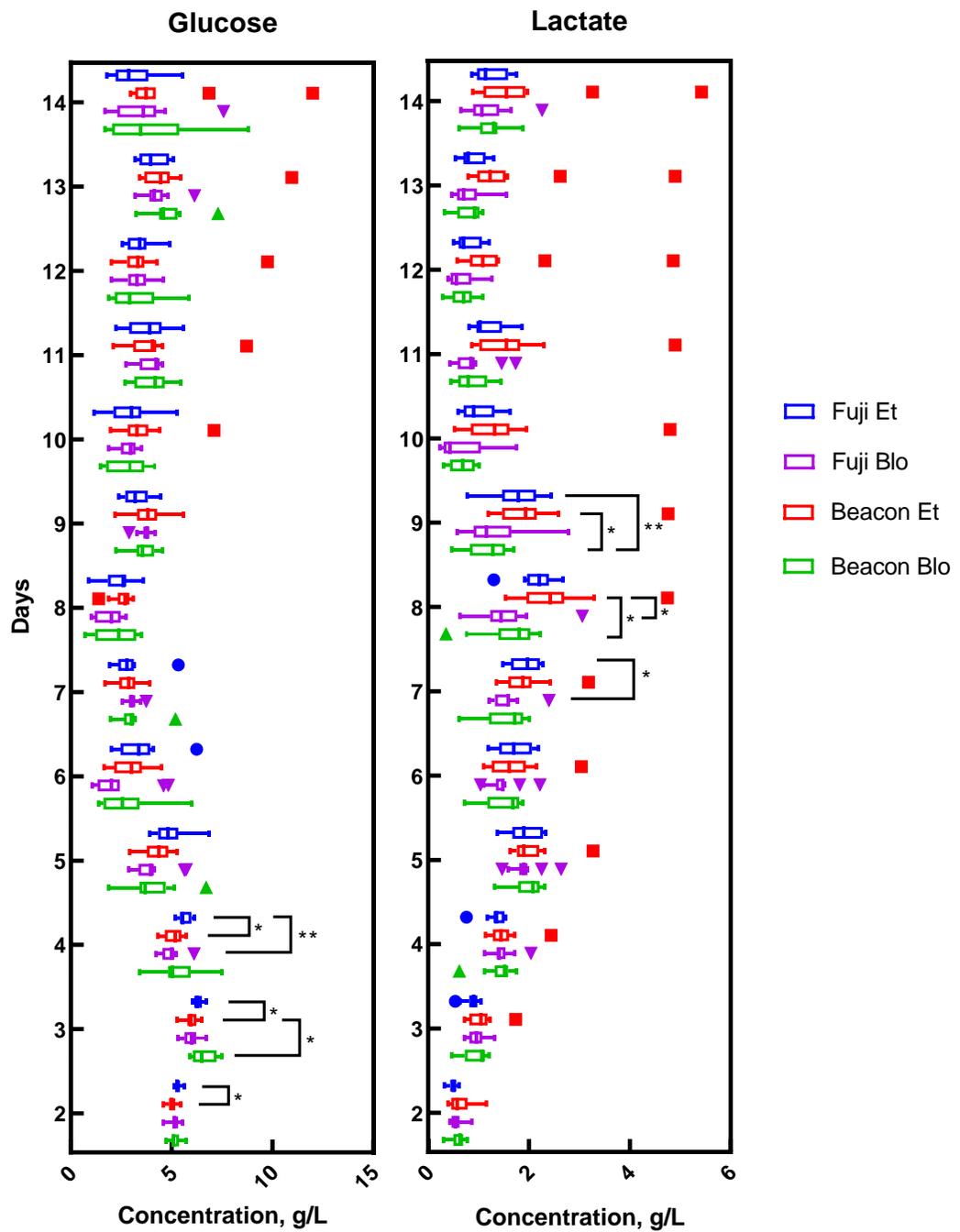
Statistical analysis was performed using a two-way mixed effects Geisser-Greenhouse corrected ANOVA with Tukey's multiple comparisons correction.

5.3 ANALYSIS OF PREDICTION OF CELL PRODUCTION CHARACTERISTICS ACROSS CLD STAGES

5.3.1 Correlations between different CLD stages

With the fed-batch process described and a comparison of the clones undertaken, the next analysis was to consider the cell line development process as a whole and see how good the predictability of the steps at different stages were. Figure 5.3.1.1 shows the Pearson correlation matrices of the cell lines at different stages of the cell line development process to determine how the different stages changed the behaviour of the population as a group. Starting with the Beacon® derived cells in the case of BlosozumAb, we see good correlation between the ambr® Qp and ambr® titres (0.94), and between the ambr® titres and the 24-well plate (0.61) and TubeSpin® (0.72) stages. Surprisingly, the Beacon® productivities and titres were found not to be significantly correlated to any of the scale-up stages even though the clonal population were found to be quite similar. No significant correlations were observed at the initial Beacon® pool screening in this group as well.

The Beacon® Etanercept group followed this general trend, high correlation between ambr® Qp and ambr® titres (0.96), and significant correlations of the ambr® titres and the TubeSpin® titres (0.79). Here, the correlation with the Beacon® titres and Qp was statistically significant. In the ClonePix™ 2 Etanercept group, however, there were not any correlations between the stages except for the 96-well plate titres and the ambr® Qps (0.79). In fact, there was a statistically significant negative correlation between the ambr® stage and the 24-well plate stages.

In the ClonePix™ 2 BlosozumAb group the correlations were higher than the Etanercept group with the TubeSpin® and 24-well stages correlating to ambr® Qp (0.63, 0.65). Only the 96-well plate screen titres did not correlate will the ambr® Qps.

Figure 5.3.1.1. Pearson correlation coefficient matrices showing how clones behaved as a group across different stages of the cell line selection process when generated using the ClonePix™ 2 or Beacon® CLD process.
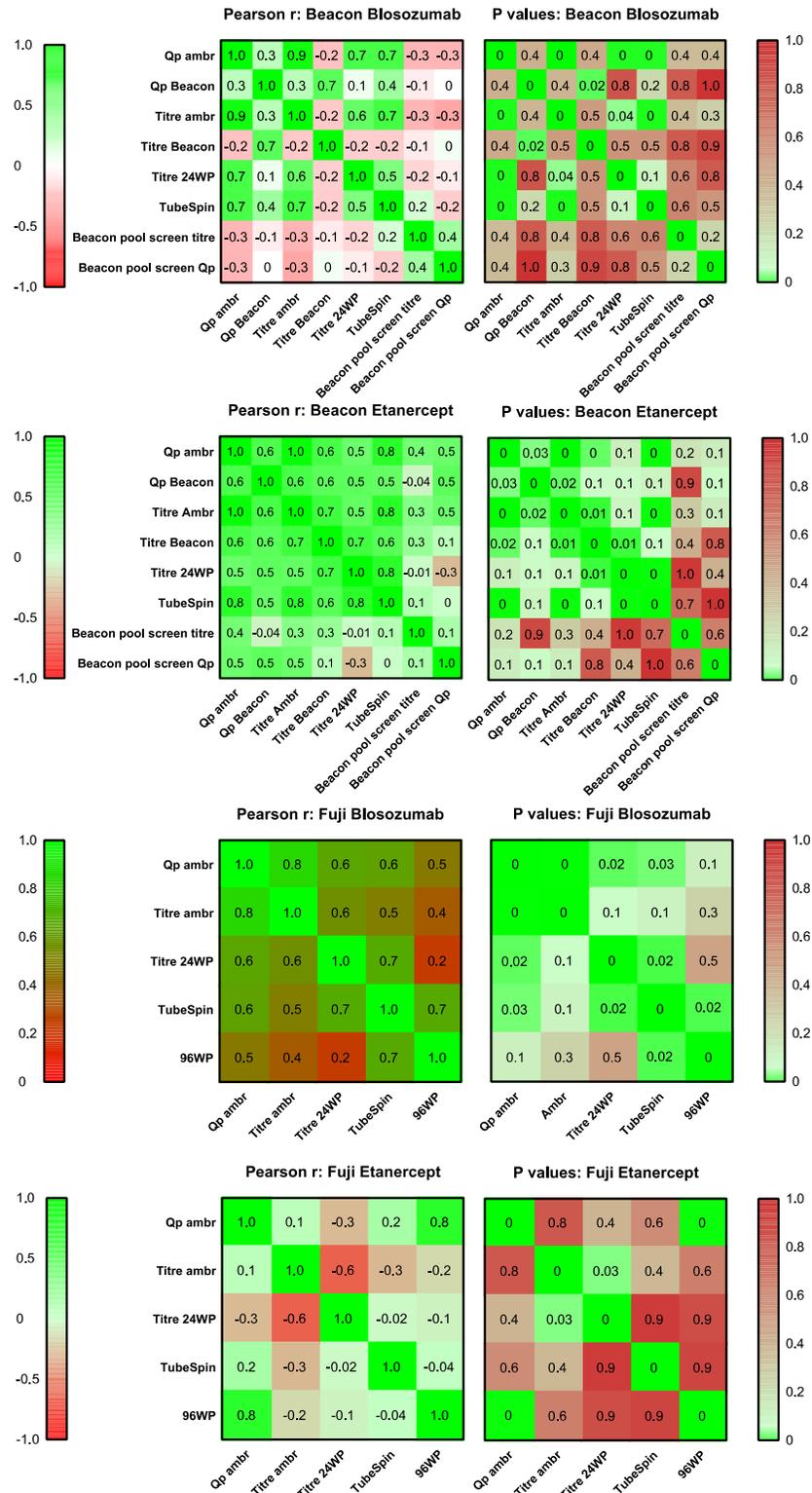
WP- well plate; Qp – specific productivity; r – correlation coefficient

5.3.2 Comparing clone ranks between different CLD stages

Correlations are not the only approach to investigate the relationship between two data sets. Because the varying conditions between different stages of a CLD process can impact how cells behave, investigating their 'ordinary' behaviour can account for some of the variance. To this end we ranked the performance of the clone production capacities at all CLD stages. This data is presented Figure 5.3.2.1) and shows that as cells transition from the 96WP stage to ambr® stage, the final cell lines tend to be selected more from the middle of the TubeSpin® and 96WP populations in the Fuji group. While most of the cell lines came from the top 50% of the previous groups, at the TubesSpin® stage we observed some cells from the bottom 50% (25% BlosozumAb, 16% Etanercept) and for the 96WP stage 4/12 for BlosozumAb and for 5/12 Etanercept. If we were to consider the predictions of just the 96WP data, we would have obtained 2 clones from the top 5 ambr® for the BlosozumAb and 1 for Etanercept. The TubeSpin® predictions were as accurate as the 24WP predictions for the top 5 in the BlosozumAb group while in the Etanercept group the TubeSpin® stage was not able to predict any of the top 5 clones and the 24WP group predicted only 1. Some cell lines such as C2, C3 and C6 have very good agreement between all the columns. Such cell lines are present in all groups except for the Fuji Etanercept group. This group is also unique in the sense that the ambr® titre ranks do not agree with the ambr® Qp ranks.

The predictions from the Beacon® were as, if not more, accurate as the traditional CLD methods. When the population measurement clones were ranked by their productivities, 3/5 and 4/5 clones in the top 5 group for BlosozumAb and Etanercept respectively were predicted while the 24WP measurements predicted 3/5 for both Etanercept and BlosozumAb groups. The TubeSpin® measurements also predicted 3/5 of the top 5 as well.

By comparing the variances of the mean ranks for the processes we can we can see the rank predictive capacity of the Fuji process at the 96 well plate level is much lower than the Beacon® (variances 132.1 and 4.5 respectively). This tells us that the ranks for each clone on average tend to diverge more across the process than for the Beacon® CLD.

Figure 5.3.2.1. Comparing the ranks of the clonal cell lines by bioparameter at different stages of cell line development.

	Clone ID	Ambr titre	Ambr QP	24WP titre	TubeSpin titre	96WP titre	Variance
Fuji Blosozumab	C1	12	11	9	27	15	48.8
	C2	4	4	6	4	2	2.0
	C3	1	1	1	1	13	27.0
	C4	10	10	12	32	54	317.0
	C5	6	8	10	15	58	420.7
	C6	2	2	3	3	4	0.5
	C7	5	5	2	9	11	12.2
	C8	9	7	5	14	17	24.2
	C9	7	9	4	37	21	161.7
	C10	11	12	7	30	29	103.3
	C11	3	3	8	10	44	263.2
	C12	8	6	11	35	51	332.7
Fuji Etanercept	C13	4	7	12	15	38	141.5
	C14	8	1	8	22	4	64.7
	C15	11	3	9	9	12	10.7
	C16	10	5	2	16	19	51.3
	C17	7	6	10	20	21	41.2
	C18	12	12	1	31	32	171.5
	C19	5	4	7	6	27	86.5
	C20	9	10	5	29	54	370.3
	C21	2	9	6	23	23	61.2
	C22	1	2	3	33	28	199.3
	C23	3	11	4	14	32	106.7
	C24	6	8	11	10	38	151.7

132.1 Mean

	Clone ID	Ambr Titre	Ambr Qp	24WP titre	TubeSpin titre	Beacon Qp	Beacon titre	Variance
Beacon Blosozumab	63_8	6	4	6	9	1	2	8.5
	63_9	7	8	5	10	12	4	6.7
	62_7	3	1	1	2	5	11	2.7
	63_13	8	9	4	12	6	3	9.2
	63_11	9	7	10	7	10	14	2.3
	62_10	1	2	2	1	4	7	1.2
	62_8	5	5	3	3	8	10	4.2
	62_4	2	3	8	4	3	5	4.3
	62_2	10	10	7	8	9	9	1.3
	63_1	12	12	11	11	11	8	0.2
	63_7	4	6	9	6	7	13	1.5
	63_5	11	11	12	5	2	1	17.3
Beacon Etanercept	17_12	4	4	1	2	5	1	2.5
	17_17	7	7	12	8	12	12	5.2
	Fe1_6	2	3	3	1	2	5	0.7
	17_20	5	5	2	4	4	4	1.2
	17_26	10	11	10	11	9	6	0.7
	16_30	12	8	4	7	8	16	2.7
	Fe1_9	8	10	8	10	2	5	10.8
	17_3	9	9	11	12	7	9	3.7
	17_24	6	6	9	5	10	8	4.3
	17_25	11	12	5	9	11	13	7.2
	17_14	3	2	6	6	6	2	3.0
	17_13	1	1	7	3	1	3	6.0

4.5 Mean

The colour scale is based on percentile values. The population variance is computed using the formula $\frac{\sum(x-\bar{x})^2}{n}$ *excluding the Ambr titre column on the Fuji group, and excluding the Beacon titre and Ambr titre column on the Beacon groups.*

5.3.3 In-depth analysis of Beacon® clone population attributes

To investigate why the ClonePix™ 2 BlosozumAb group displayed a low correlation with the final ambr® titres, we went back to the population statistics data to try to explain this discrepancy. In subchapter 4.4.5 we showed that the population productivity measurements of the Beacon® BlosozumAb clones were similar, so similar in fact that we could only identify two statistically significant clusters in their ranks. This means that the predictions in assigning the clone ranks were not accurate, even though the initial screening allowed us to obtain a group containing the top 3 clones. From the raw data of the population analysis experiment (

Figure 5.3.3.1) we see that these clones were not just similar in terms of productivity, but in their overall assay score and cell count characteristics. Looking through the detailed protocol of the experiment, the same lot of conditioned media was used for the BlozosumAb experiment and some of the Etanercept clones designated as group G1 in Figure 5.3.3.2. Knowing this, we could analyse these two groups of Etanercept clones to see if the media might have affected the ability to perform accurate rankings within the clone population stage.

The differences between Etanercept groups G1 and G2 are highlighted in Figure 5.3.3.3. We compared these two groups in terms of the average of the titres, productivities and numbers of cells per pen. While there seems to be no statistical significance between these groups in terms of productivity, most of the top 12 clones fall into group G2. When comparing these two groups a Kolmogorov-Smirnov test was applied to determine whether these two populations were sampled from the same distribution. The p-value for that test is 0.0212, highlighting that they are in fact different. The cell counts were found to be significantly different between the two group with means of 8.82 and 13.46 with a p-value of <0.0001. That is a relative difference of 1.53 between the means. A significant difference was also observed between the means of the average fluorescence intensities of the clones within the two groups with means of 0.0284 and 0.0467 and a p-value of <0.0001. After evaluating these groups, it became clear that the conditioned media had a profound effect on the way the cells behave within the Beacon® system and likely contributed to the ability to be able to make predictions from that population due to these effects.

Figure 5.3.3.1. Graphs showing the results from scaled up Blosozumab clone stability screening.
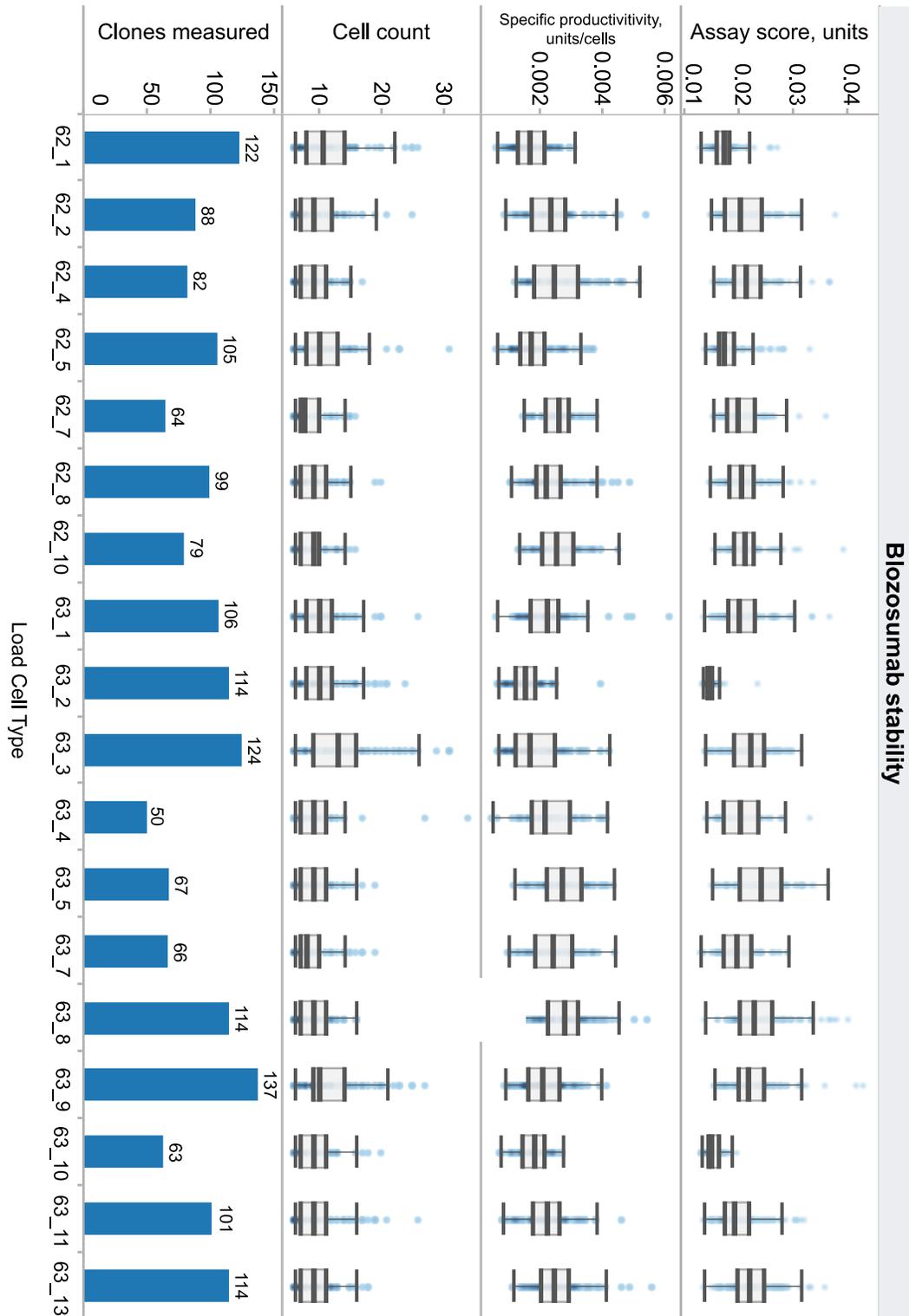
The boxplots represent the median and 0-1.5 of the interquartile range.

Figure 5.3.3.2. Graphs showing the results from the scaled-up Etanercept clone stability screening.
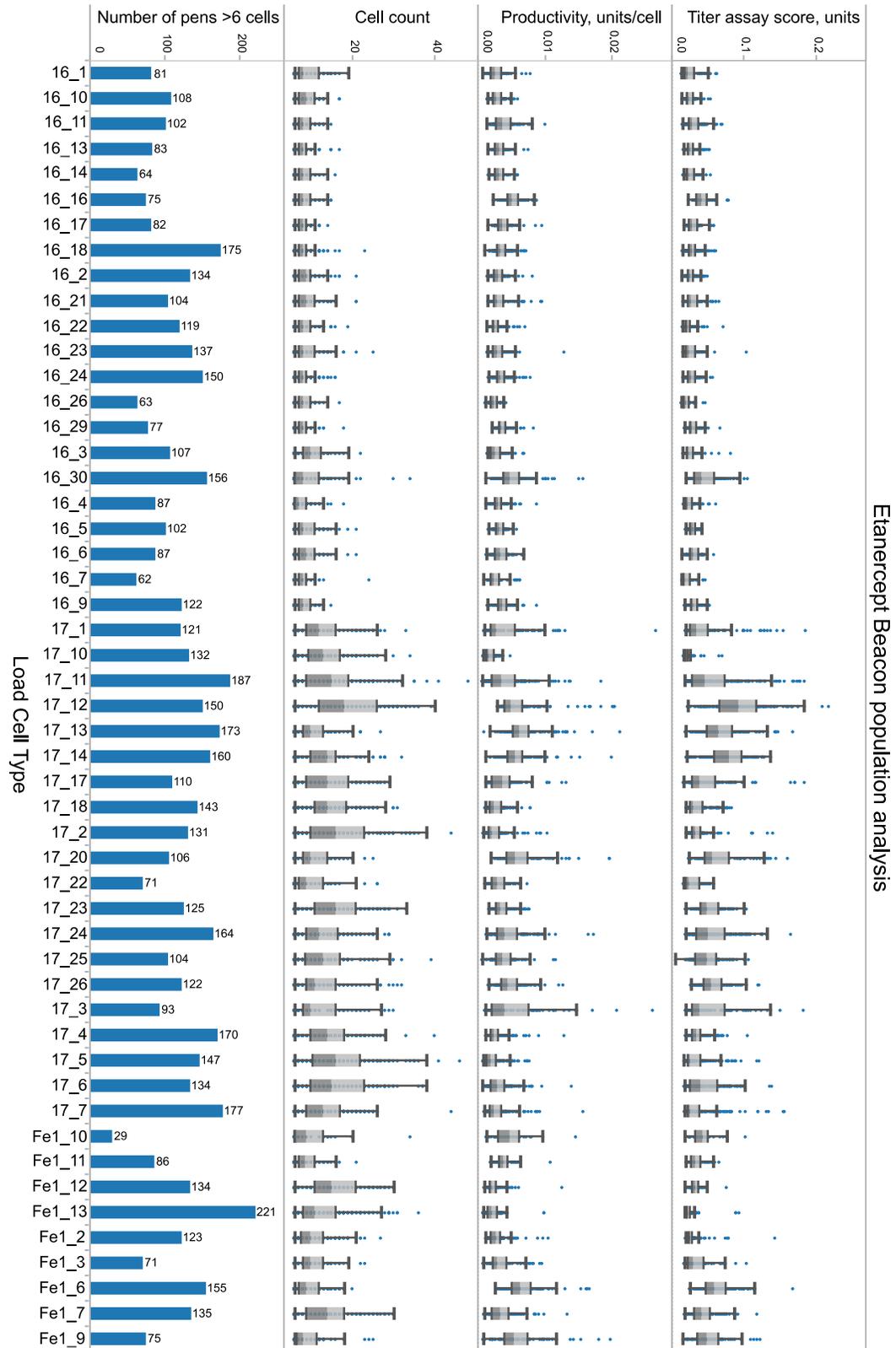
The boxplots represent the median and 0-1.5 of the interquartile range.

Figure 5.3.3.3. Dot-plots showing the differences between the fluorescence intensity, cell count, and specific productivity means between groups G1 and G2 sampled from the Etanercept pools.
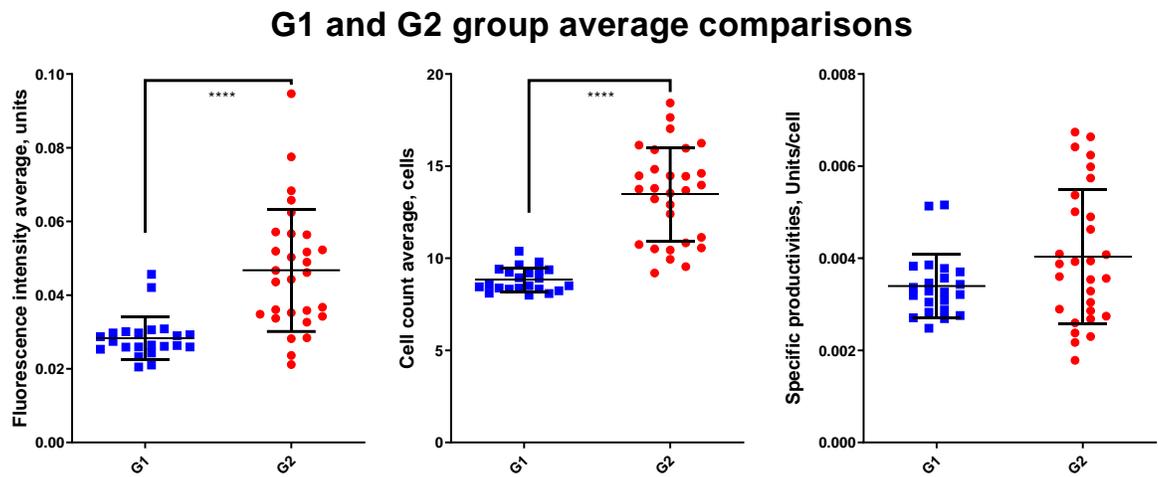
G1 and G2 group average comparisons

The stars denote the p-value as calculated by a two-tailed Mann-Whitney test. The error bars signify the means and the standard deviations of the populations.

5.4 EVALUATION OF CLONE STABILITY USING THE BEACON® PREDICTIVE ALGORITHM

One potential model of clonal stability assumes that the reason for phenotypical drift in a clonal population happens because within a population cells that grow faster but produce less, overtake cells that produce more at the expense of growth. In the previous section it was reported that the clonally derived populations can be quite diverse in their titer, productivity and growth. Examples of two such hypothetical populations are presented in Figure 5.4.1.

Figure 5.4.1 Model depicting assumptions of clonal stability mechanism.
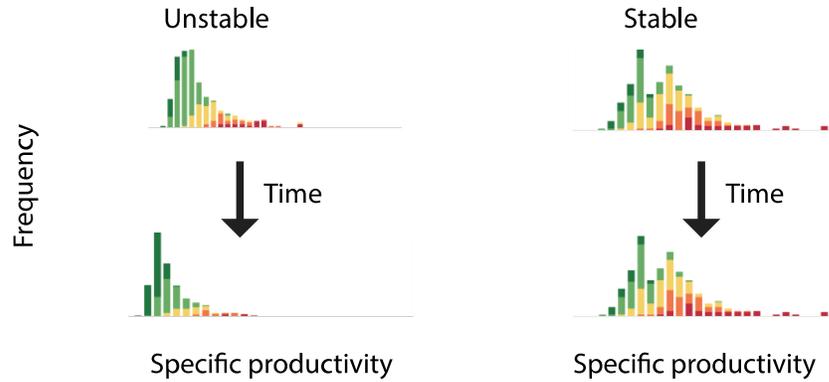
The histograms show growth rates of the different producing cells as measured in pens on-chip. Growth rate is indicated by colour (Red-Green, low-high).

Figure 5.4.2. Graph depicting the stability and specific productivity of clones that were assayed.
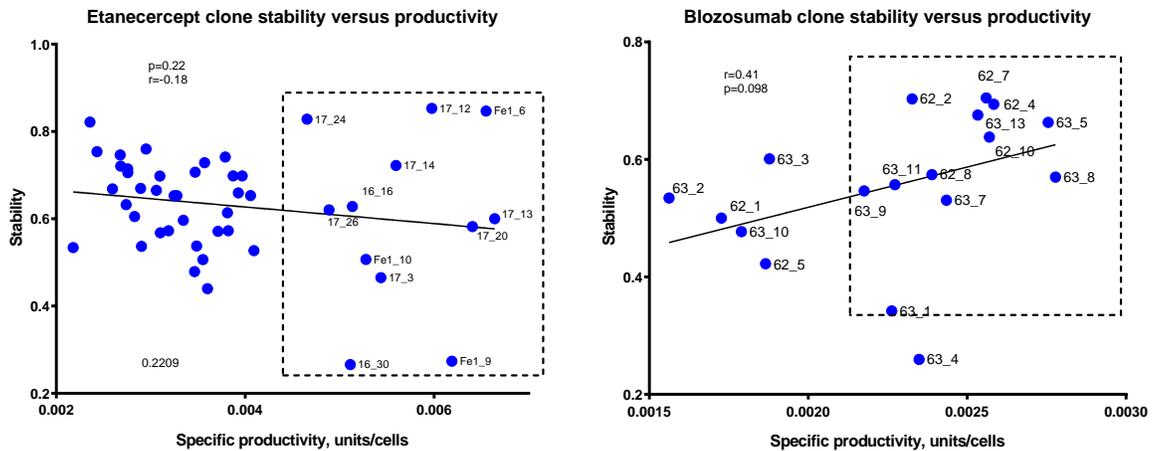
Dotted lines show the area of selection for clones to be progressed to screening on the ambr® 15 system.

Using Beacon® data from the previous section of the chapter, we applied a proprietary model built as an in-development feature for the Beacon® to give a stability score for exported clones based on the growth rates of populations in pens that have high or low productivity. This data is presented in Figure 5.4.2. The stability score here is a metric which estimates the proportional productivity of the cell line after 12 weeks of culturing. The dotted lines in the figure represent the population that were selected for ambr® 15 screening. Within the population group for Etanercept we identified a group of clones that all displayed high productivity scores but drastically different stabilities; 17_13 (0.60), Fe1_6 (0.85), 17_20 (0.58), Fe1_9 (0.27). These cell lines correspond to the top 4 predicted clones by productivity. Based on this, we decided to try

and validate this predictive model on these 4 cell lines.

To do this, we passaged these cell lines for 60 generations as would be undertaken during routine cell culture. At the end of the passaging period a comparative fed-batch culture analysis of clones before and after long term culture was performed. The viable cell number and culture viability information from these fed-batch cultures are presented in Figure 5.4.3 and Figure 5.4.4.

For all cultures a general growth trend was observed during the fed-batches in E250 shake flasks. Exponential growth was observed until day 7 after which the viable cell concentrations slowly started to decline. The culture viability actually started decreasing a day before that and cultures were terminated after they were determined to have a viability below or equal to 70%. Cell lines 17_13 and Fe1_6 demonstrated increased levels of total viable cells throughout the fed-batch process whilst this observation was not as pronounced for cell line Fe1_9.

Figure 5.4.3. Viable cell growth curve of the stability of selected cell lines by FOGs.
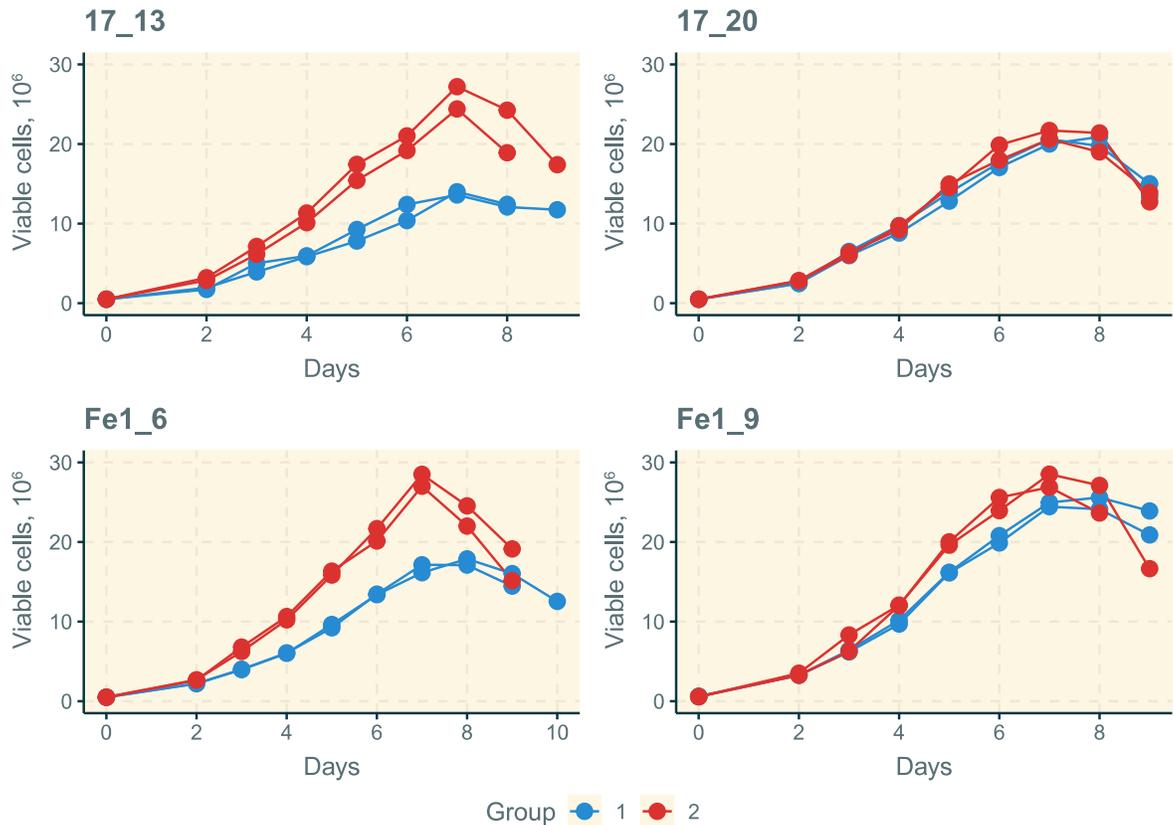
Two biological replicates were measured for each cell line. Group 1 represents cell culture before extended passaging and group 2 measurements were done after the 60 passage period.

Figure 5.4.4. Culture viability data during assessment of the stability of selected cell lines by FOGs.
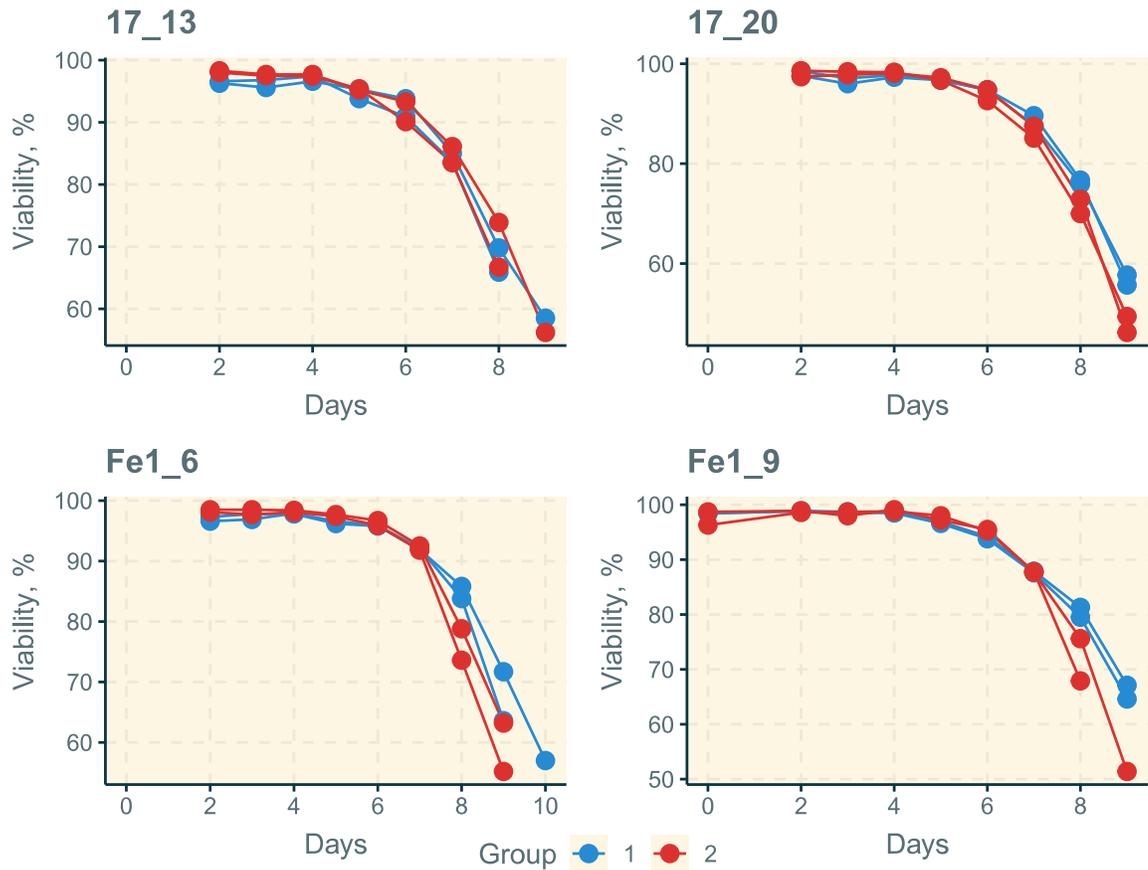
Two biological replicates were measured for each cell line. Group 1 represents cell culture before extended passaging and group 2 measurements were done after the 60 passage period.

To further compare the differences in growth for the different cells, the IVC as a measure of total accumulated cell time was then analysed along with the specific productivity. We chose to compare IVC up until day 8 as that is the last common day where all cultures were still in culture. Statistically significant shifts were observed in the IVC for 17_13, Fe1_6 and Fe1_9 with relative increases of 1.82, 1.58 and 1.15 respectively. With regard to specific productivity, the only statistically significant changes were observed for 17_13 (5.95 pg/cell/day and 2.71 pg/cell/day) and Fe1_6 (5.68 pg/cell/day and 3.34 pg/cell/day) with relative deceases of 2.20 and 1.7 respectively. Fe1_6 was predicted to be stable; however, stability is usually defined within FDB as a deviation of a parameter less than 20%. Fe1_6 and 17_13 do not fit this criterion. Both of these cell lines had similar drift and yet their stability scores were 0.85 and 0.6. The cell line 17_20 with a score of 0.58 did not have statistically significant changes, suggesting this was a stable cell line judged on (IVC) metric. Fe1_9 was found to have a stability score of 0.27 and was found to

be stable as well. From these data the model does not at this stage appear able to predict what cell lines will be stable and the intensity of the stability drift from the early stage Beacon® data generated and used in the model.

Figure 5.4.5. The calculated specific productivity (Qp) stability and the stability of the integral of viable (IVC) cells on day 8 of the selected different cell lines.
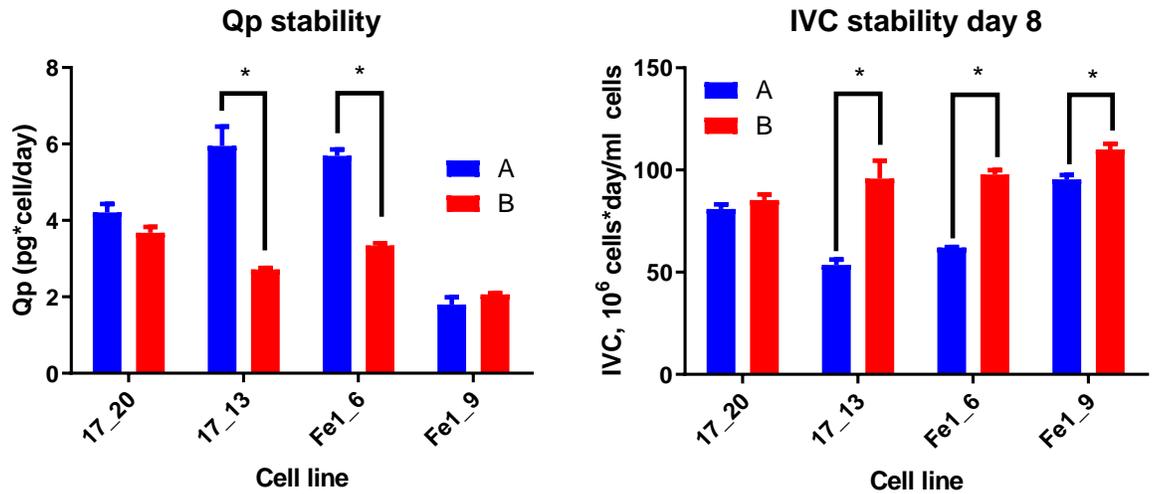
The error bars show the standard deviation of the data. A star denotes the results of a two tailed unpaired t-test with a Holm-Sidak (a = 0.05) adjusted p-value. (A) – Cell line without prolonged stability culturing; (B) – cell line after 60 generations of culturing.

5.5 DISCUSSION

5.5.1 Scale and scope limitations of the study

In this chapter we explore the predictive power of two cell line development processes and comprehensively compare their outputs. One of the most notable limitations of the process here is that we did not screen all of the clones generated. As such, we have no way of knowing whether the clones obtained at the end of the CLD process are the true top 12 clones. There is a good reason for this, at the 96 well plate stage of the ClonePix™ 2 process we had 600 clones exported. Carrying through all of those cells to ambr® 15 stage or fed batch screening would require a staggering amount of both human and financial resources, both of which were not feasible.

In a previously published study, the authors explored a CLD method where 175 clones were progressed all the way to shake flask fed-batch stage from a 96-well plate format (Porter *et al.*, 2010). Subsequent culturing steps involved a 24-well plate scale up screen, a 125 mL shake flask batch screen and a 125 mL fed-batch screen. The authors found that most of the 'real top 10' clones in the fed-batch screen came from the middle of the 96-well plate screen clone titre ranking distribution, and in the later stages almost exclusively from the top 50% of the titre distribution. These findings are similar to what we have observed in terms of how the rankings shift across different stages of CLD. In contrast, we expanded this CLD method into a more relevant ambr® 15 fed-batch measurement which more accurately represents a bioreactor environment. Seeing as the authors of the previous study knew the 'real top 10' clones out of the 175, they were able to track the predictive power of each step. They found the correlations between the steps to be 0.63 to 0.8, yet noting that the predictive power was not high between the steps. We have observed similar correlations, albeit inconsistently between the 4 clone groups. In this regard, we have a more diverse population of clones to work with and highlight that the correlations observed by the previous process might not be valid for one CLD process to the next. The authors of the previous study also applied a predictive model in order to reduce the number of clones needed to progress through CLD to obtain the true top 10 cell lines. The results were somewhat surprising, no matter which models they used they usually ended up with similar populations, although their models would sometimes result in populations with lower titre means. This highlights the intrinsic difficulty of trying to simplify CLD using conventional methodology. We have achieved better predictive results in terms of the top 5 cell lines compared to traditional CLD and with the Beacon® methodology drastically cut down on both time and resources spend on cell line development.

Another recent CLD investigation (Pristovšek *et al.*, 2018) reported on titre to confluence ratios instead of titre as the main predictive parameter in cell line development and concluded that using both methods yields clone groups with similar volumetric productivities in fed-batch cultures. The cell line development workflow was based on FACS sorting into 384-well plates -> 96-well plates -> 96-half deep well plates -> shake flask fed-batch cultures. At the 96-half deep well plate stage they had 56 clones selected based on top titre to confluence ratios and 57 clones selected based on titre alone. At the shake flask stage this number had been reduced to 12 clones per group.

Based on these two examples we believe that both our CLD workflow and our sample size successfully compares with and surpasses the examples of CLD studies in the literature. Even though we had to compromise on the number of clones that we progressed in contrast to a full blown commercial process, we believe that this does not undermine the validity of the results.

Another issue in CLD is reproducibility of the process. In this study the cell line development process was not repeated to determine how reproducible the results are. This is not typically performed in an industrial setting in a commercial CLD process. At the end stage the ambr® 15 process is run only once on 48 clonal cell lines and at this point cells are progressed through to stability studies. In a sense the way that the CLD is designed already controls for the variability of the process by taking a large number of cell lines at the end of each stage to progress to the next. While we have seen that the best clones don't typically come from the very top of the distributions, the sheer number of progressed clones allows us to reliably capture a slice of the population that is likely to contain the top clones.

5.5.2 Stability testing

Very few sources in the literature have tried to elucidate the dynamics of clonal stability. We know that CHO cells have a base rate of stability that does not change upon subcloning from a host cell line. In one study subclones showed as much chromosomal heterogeneity of cloned cell lines after transfection as before in the karyotype(s) of the host cell line for transfection (Derouazi *et al.*, 2006). The 'drift' in the stability of one cell line over 40-100 generations of continuous culturing has been previously reported (Bailey *et al.*, 2012). The subpopulations after the end of the culturing period were compared in a fed-batch screen. The study found for this cell line that statistically significant differences started manifesting after 60 generations. These changes were mainly limited to the titres of the cells while the IVC's remained similar. It was found that this discrepancy manifests at a later time in the cell culture, around day 9 out of 15.

Thus, until then the productivities remained the same. Differences in *GADD153* expression were observed, implying that later generation cells either experience increased induction of stress or decreased modulation of GADD153 expression. A number of other differences were also noted. Ribosomal profiling was employed to investigate changes in gene expression and it was discovered that with age fewer ribosomes per mRNA in later generation clones were observed as well as a shift in their glucose, lactate and alanine metabolic profiles while retaining similar levels of mRNA expression. It is noted that stability depends not just on the base genetics of the host, but also on epigenetic regulation of the recombinant DNA elements, especially increases in CpG methylation of promotors (Veith *et al.*, 2016).

We did not try and account for all of these factors in this study, focussing on more easily determined parameters to try and investigate predictions of stability based on the Beacon® data and stability prediction model. We did not focus on the underlying mechanisms of stability, but rather on the bioprocess characteristics. The generational time used to age the cultures was in accord with previous studies that suggest the generation number selected was sufficient to observe any differences that might arise in most cell lines. However, it is recognized that this is not a robust analysis of the predictions as a whole since we only investigated 4 cell lines. Taking this into account, we tried to look at cell lines that displayed a very broad range of stability scores while having similar productivities to look at the impact on the most relevant cell lines. The conclusions drawn here are that the model at this stage is unreliable in making qualitative predictions of how much productivity will decrease over 60 generations of continuous culture, classifying some cell lines as stable when they were not and vice versa. It is likely that some of these cell lines were much better producers at the beginning of the CLD process and by the time they were carried up to shake flask fed-batch culture stage they already experienced a significant productivity drop. However, we did not investigate for such a possibility. One could imagine, however, an alternate stability workflow where the Beacon® is kept running for 60 generations in a perfusion state pruning the wells of the excess cells and tracking the stability of the cells immediately after the population screening.

5.5.3   Predicting ranks of clones across CLD stages

The previous efforts of researchers in establishing correlations between different CLD stages (Porter *et al.*, 2010) have been discussed in section 5.5.1. However, in this section we focus on the patterns observed within the datasets generated in this study. Interestingly, in both cases where we observed a lack of agreement between correlations and rankings at different stages, we

encounter a population that had a very low standard error of the mean. In the ClonePix™ 2 Etanercept group the ambr® titres were clustered together very closely and so were the productivities in the Beacon® BlosozumAb population screen. A potential explanation of why we encountered this phenomenon in the Beacon® workflow in the results section based on the differences in batches in conditioned media is described here. We know from the Etanercept comparisons that the ambr® titres for the population are not an intrinsic function of the recombinant protein that these cells are producing since both CLD's used the same initial pools to pick clones from. Because the ClonePix™ 2 CLD is geared at progression through titre and the Beacon® CLD is selecting based on productivity, it is suggested that for the Etanercept recombinant molecule these two selection criteria can result in the isolation of two different populations. We know that the predictive power increases as the dynamic range of the target characteristic increases, large differences are much easier to evaluate than smaller ones as the biological noise becomes more relevant compared to the difference in the signal as the magnitude of a difference decreases. We make the argument that in the case of the ClonePix™ 2 Etanercept group the CLD is not behaving abnormally as we still observe the same predictive power at the 24WP stage to assign 3/5 clones to the top 5 group at the ambr® stage based on titres. The rank variances differ quite drastically for the processes showing that while good correlations between absolute values are hard to pin down, we can still observe differences between the processes when trying to predict ranks. Furthermore, we note that some cell lines display good predictability throughout the entire process such as C6 or 62_2, however it is not known as to why. It goes without saying that the environment of the cells gets altered quite drastically throughout CLD in both cases and we hypothesise that a cell line might have consistency phenotype that might be able to be selected for.

5.6 SUMMARY

1) We compared top clones from two CLD processes and found that as groups they behave similarly in the ambr® 15 environment.
2) The Beacon® process produced the best clones over all and was able to predict 3/5 top producers at the chip level of screening.
3) Beacon® stability predictions do not seem to hold at fed-batch shake flask level.
4) The Beacon® is able to drastically reduce hands-on cell culture time and the top 12 clones exported for both constructs contained the high producing cell lines.

5.7 CONCLUSIONS

In this chapter we have performed a thorough evaluation of the groups of clones produced by two different CLD processes for two recombinant therapeutics, BlozosumAb and Etanercept. We have found that there seems to be no statistically significant difference between groups of clones obtained through these methods when compared in an ambr® 15 fed-batch set up. We were also able to show that using the Beacon® CLD we were able to have better capacity at predicting the top 5 clones within the respective groups than from the TubeSpin® batch or 24-well plate fed-batch cultures. We were also able to cut the CLD time drastically by using the Beacon® system since there was no need to undergo the colony outgrowth on semi-solid media and 24-well plate fed-batch screens taking off at least 4 weeks from the process. Significant cost savings could have also been made in reducing the throughput at larger scales, since by exporting the 5 top clones we would have ended up with the best cell lines by ambr® 15 titre for both BlosozumAb and Etanercept. A full economic cost analysis would need to be undertaken to confirm this which was beyond the scope of the work in this thesis. In our opinion these next generation single cell technologies open the way to radically transform the CLD process by reducing development overhead, increasing throughput and allowing never before available collection of clonal bioproduction parameter data at the microscale.

6 OVERALL DISCUSSION AND FUTURE WORK

Chinese hamster ovary cell lines are the workhorse of biotherapeutic protein expression at industrial scales due to their familiarity that goes back for more than 50 years, ability to produce human-like post-translational modifications and achieve high specific productivities using scaled-up industrial bioprocesses (Kuo *et al.*, 2018). Even so, a case can be made that our current protein biotherapeutic drug production platforms need improvement; treatment costs for biotherapeutics are on average 22 times more expensive than small molecule drugs (McCamish and Woollett, 2011). As outlined in the introduction Chapter, a breast cancer patient's average cost for Herceptin® is $37000 USD, a rheumatoid arthritis or Chrohn's disease patients for Humira® is $50000 USD per year, Gaucher disease patients with Cerezyme® is $200000 USD for the rest of their life (So and Katz, 2010).

A significant effort has been made to try and understand CHO cells as a system for recombinant protein expression. With the arrival of the 'omics era in biotechnology, we can now leverage big data sets to analyse information on a number of levels; genomic, transcriptomic, proteomic, metabolomic, glycomic and lipidomic. Up until the analysis reported in this thesis, the efforts of the community have been mostly directed to gather large amounts of omics data. Chapter 3 describes our efforts to try and systemize available transcriptomic data obtained by the community to determine it's applicability and reproducibility for cell line engineering. Chapters 4 and 5 seek to investigate other bottlenecks within the cell line development process in the industry, the lack of accurate scaled-down models of cell line development and long development times from scale-down to scale-up. To this end, we employed a next-generation high-throughput optofluidic single-cell manipulation, cultivation and assay system and compared it to an existing GMP industrial cell line development workflow.

The major findings and discussion of the results are included at the end of each results section, however, here they are viewed here collectively and the opportunities for future work are considered.

6.1 Meta-analysis and application of publicly available transcriptomic data

6.1.1 Aggregating and collating the publicly available data

In Section 6 we scanned the publicly available literature for studies that utilised microarray or RNAseq analysis of CHO cell transcriptomes to analyse changes in growth or changes in productivity. In total, we found 19 viable data sets to analyse, aggregate and compare. Overall, RNAseq datasets were underrepresented, out of the 19 datasets, only 4 used RNAseq while 2 compared the use of RNAseq to a microarray in the same experiment. We also found that most studies did not conform to good 'omics publishing practice, formatting gene annotations using gene names making certain annotations vague, not publishing raw omic data which obscures the data processing to readers. Out of these 19 datasets, only 3 of them had deposited their data into omic repositories such as Bioproject or E-GEOD. To make things more complicated, the authors used a wide array of cell lines, some of which were not disclosed and cultured cells under very different conditions; reduced cultivation temperature or hypothermia (<37°C), with addition of sodium butyrate, and with divalent copper ion feeding to name a few of the variables within the culture systems reported. All of this together makes a meta-analysis very difficult to collate between these datasets.

To overcome this, we used two main criteria of analysis, the frequency of a gene to be flagged as differentially expressed and the direction of the expression (up or down). We could then analyse which genes appeared most commonly across the datasets associated with growth (μ) or productivity (Qp). To summarize the directionality of the expression we introduce a parameter dubbed concordance which is the arithmetic mean of the expression values. The concordance shows the agreement of the expression level of the gene across studies.

Another way to look at the data is to investigate pathway enrichment and determine whether certain cellular pathways are overrepresented in our dataset. We found that the only pathways that had good concordance and good enrichment were the cell cycle and lysosomal pathways. Apoptosis and cell cycle engineering have been both historically utilised and applied as successful approaches for CHO cell line engineering (Table 1.4.2.1) and it seems that the enrichment analysis generally agrees that these are conserved properties of desirable CHO cell phenotypes for biotherapeutic protein production. We made efforts to map the genes that corresponded to these pathways visually to provide guidance to future cell line engineering

efforts and to standardize the quality of omics publishing within the community going forward.

6.1.2   Future work and impact of the transcriptomic meta-analysis dataset

Originally, we conceived of this study, to aid other CHO cell research groups interested in targeted cell line engineering strategies to improve the growth or productivity of CHO host cell lines. Few groups have the funding to perform genome-wide screening on cell lines to identify targets to improve growth and productivity, however, it was very unclear how one should proceed in scaling down such efforts. Here we provided an evidence-based roadmap from which to guide people towards a consensus list of potential targets and pathways for cell line engineering purposes. Following this analysis, we designed a state-of-the-art CHO specific pooled siRNA library of 1000 genes. 500 of these genes were selected based on the frequency in the aggregate list and the other 500 were picked as whole pathways that had high fold-enrichments with good statistical accuracy. Using this siRNA screen, we plan to screen CHO cell lines producing recombinant proteins for any changes in cell concentration, growth (μ), culture viability and titre/Qp. As mentioned previously, recently a genome-wide cross-species screen has been performed using CHO cells (Klanert *et al.*, 2019) as the platform, however cross-species screening has its limitations due to siRNAs being very sensitive to deviations in the target sequence. Therefore, the screen suggested as an output of this study and the subsequent screen within these 1000 genes should provide a good comparison to the published study for future design of high-throughput screening efforts.

6.2 REFLECTIONS ON THE COMPARISON OF TWO CLD PLATFORMS

6.2.1   How does the Beacon® compare to a more traditional CLD approach?

In chapters 4 and 5 we transitioned from big-data and providing a platform from which to launch high-throughput engineering efforts to looking at another pressing issue within the CHO cell researcher community in the area of production of biotherapeutic proteins, cumbersome resource and development costs for cell line development. With current levels of single-cell manipulation and microfluidic technology, this field is poised for disruption. Currently, CLD is carried out in a very laborious and time-intensive process where transfected cells pools are carried through multiple stages of screening at increasing scales, but decreasing throughput, ultimately arriving at a final set of clones that are evaluated at pilot scale. Single-cell technologies allow more flexibility over cloning, higher throughout and numbers of cells to be evaluated early

in the CLD process, and documentation of clonality, whilst also potentially increasing the predictive power at small scale through novel secretion assays.

One such system is the Berkley Lights Beacon® platform. Herein we have conceived of a methodology in order to evaluate how this system compares to, and can improve upon, more traditional and embedded methods used in industry in order to carry out cell line development. We evaluated a scale and workflow that is compatible with previous cell line development research carried out in the literature (Porter *et al.*, 2010; Pristovšek *et al.*, 2018). The work extends beyond most other previous efforts in order to make CLD comparisons as we progressed cell lines all the way to an automated bioreactor scale, ambr® 15 scale, which gives information as to how these cultures perform under bioreactor conditions.

In terms of all the tests that we were able to perform on the two CLD platforms, we could find no statistically significant differences between the groups of cells obtained from the Beacon® and the ClonePix™ 2 CLD approaches. The methods of evaluation were comprehensive in terms of the standard parameters evaluated, determining the cell pools and lines generated in terms of viable cell concentrations and IVCs achieved, culture viabilities, titre, cell specific productivities, growth rates and basic metabolomic profiles. For a company transitioning from an existing workflow to a new one reproducibility is key. The new platform needs to be field-tested in order to know that it performs at an equivalent to the existing technology, or better. The comparison and research reported here should be considered the first step for researchers and industrial players to evaluate the new technology and gain more confidence in the platform and to set expectations of equivalence. The major gap within the work presented in this initial work, but beyond the scope of the work in this thesis is that there was no primary recovery or downstream processing analysis undertaken on the material generated from the ambr® cultures. Evaluation of the impact of the cell lines generated on subsequent downstream performance should be evaluated in further work.

While all bioprocess parameters remained statistically similar between the CLD processes evaluated, there were benefits to using the new Beacon® system that transcend the overall numbers. Most notably, the biggest benefit is likely to be around time to selection of a potential manufacturing cell line and the depth of data that can be generated compared to the standard CLD process the Beacon® was evaluated against. The traditional CLD process used in this study takes a minimum of 42 to 65 days to perform, excluding the transfection and the ambr® 15 culturing times as these were common to both processes. A month's time is a significant saving in biopharmaceutical manufacture and is potentially a key advance of using the Beacon® system. A further advantage is the ergonomics of the system. According to our estimates, a user need

only scale up a small number of cell lines selected from the initial thousands of pens or individual cells and still be able to identify and generate a top-performing cell line within the original population distribution. While more experiments need to be performed to confirm that this finding is statistically robust, the existing results suggest this is the case. The application of the system to CLD also offers the potential for more throughput in cell line development. Once most of the overhead relating to scale up is eliminated, CLD can be done on more projects at the same time or simply explored in more dimensions than was previously possible. This is also enabled by the fact that the Beacon® allows the user to perform secretion assays and precise growth measurements on cells at the microscale, which is not an option when not using microfluidic cell culturing instruments. We believe that in one form or another, these technologies have the potential to fundamentally alter cell line development. Finally, while the stability prediction experiments did not validate the current predictive nature of stability, it is noted that this was only a pilot test of the model on a very small number (4) of cell lines. Further examination is needed in order to determine whether the Beacon® can be used to more accurately predict stability of cell lines early in cell line development.

6.2.2   Future work and potential applications

While the work reported here only tested the Beacon® platform in traditional cell line development, the instrumentation has other potential applications. Due to the ability to measure the fluorescence of single cells combined with microscopy capabilities, it is possible that this platform might be able to screen cells based on organelle morphology as was recently demonstrated with the use of imaging flow microscopy (Pekle *et al.*, 2019). It is believed that in drug discovery hard to express proteins might have visible and adverse effect on organelle size due to stress caused by protein misfolding and mis-trafficking. As such, the Beacon® technology might aid in identifying drug candidates at early stages of CLD as being more or less viable for manufacturing at scale-up. Another way to test this would be to investigate aggregation on the platform. During the work reported here, it was observed that in some pens aggregation like particles could be observed, which warrants further investigation. Currently, there exist no methods to directly measure product quality at such a small scale and the ability to detect aggregation would mean extra power in the ability to discriminate between different clones and to evaluate potential drugs for development. Having a device that provides high-throughput data on cell bioprocesses characteristics and morphology is a data generating system that lends itself very well for machine learning analysis, a field that is in its infancy in bioprocessing. As such, it

is hard to evaluate the potential of this route of investigation, however, it is one of the next steps in order to supercharge our ability to predict scale-up bioprocess parameters with striking accuracy.

Since the Beacon® platform can be run indefinitely with constant pruning of the cells to prevent overgrowth, it also has the potential as a high-throughput chemostat for directed cell line evolution. Currently, it is commonplace for companies who develop their own hosts for recombinant protein production to take cells through directed evolution to achieve good growth and favourable culturing characteristics. Leveraging the Beacon® system has the potential to greatly increase discovery power and also the ability to screen large numbers of different conditions that the cells can be cultured in for evolution. This technology also works well with high-throughput cell-line engineering efforts such as CRISPR, because transfected and high Cas9 expressing cells can be easily identified within the Beacon® and their phenotypes evaluated. Very few publications currently exist that have explored the use of next-generation CLD technologies and these have not yet been cross-examined in the literature. While the Beacon® performed as well or better than current CLD technology, a logical further step would be to compare how this technology performs against competitor microfluidic technology such as the Sphere Fluidics CytoMine® system.

6.3 CONCLUSIONS

In the course of this thesis, two areas where high-throughput methodologies are used in CHO cell line development and engineering have been examined. We have identified key areas to improve in both and have made efforts to drive the field forward. In CHO transcriptomics, we have paved the path forward for researchers to utilize publicly available CH and CHO cell data in order to make educated decisions around cell line engineering projects along with guidelines for future omics research to follow a strict code of consistency in order to be useful to the community at large.

In CHO cell line development, we have applied a next-generation high-throughput single-cell culturing technology that shows promise to offer an alternative to current cell line development processes by increasing throughput, reducing resource and time costs. We have shown that the Berkley Lights Beacon® system could produce cell lines that were comparable to those obtained via the FDB traditional cell line development workflows. The predictions made by the Beacon® had similar clone rank predictive power at ambr® scale compared to the predictions made from the TubeSpin® and 24-well plate CLD stages at the ClonePix™ 2 CLD. To our knowledge, this is the first reported case of an independent academic group validating this system in a real-world industrial setting.

7 BIBLIOGRAPHY

Abe, H., Tomimoto, K., Fujita, Y., Iwaki, T., Chiba, Y., Nakayama, K., and Nakajima, Y. (2016) Development of N- and O-linked oligosaccharide engineered Saccharomyces cerevisiae strain. *Glycobiology* **26**: 1248–1256.

Abskharon, R.N., Ramboarina, S., Hassan, H. El, Gad, W., Apostol, M.I., Giachin, G., *et al.* (2012) A novel expression system for production of soluble prion proteins in E. coli. *Microb Cell Fact* **11**: 6.

Ahmadi, S., Davami, F., Davoudi, N., Nematpour, F., Ahmadi, M., Ebadat, S., *et al.* (2017) Monoclonal antibodies expression improvement in CHO cells by PiggyBac transposition regarding vectors ratios and design. *PLoS One* **12**: e0179902.

Antoniukas, L., Grammel, H., and Reichl, U. (2006) Production of hantavirus Puumala nucleocapsid protein in Saccharomyces cerevisiae for vaccine and diagnostics. *J Biotechnol* **124**: 347–362.

Arden, N., Majors, B.S., Ahn, S.H., Oyler, G., and Betenbaugh, M.J. (2007) Inhibiting the apoptosis pathway using MDM2 in mammalian cell cultures. *Biotechnol Bioeng* **97**: 601–614.

Arriaga, E.A. (2009) Determining biological noise via single cell analysis. *Anal Bioanal Chem* **393**: 73–80.

Baek, E., Noh, S.M., and Lee, G.M. (2017) Anti-Apoptosis Engineering for Improved Protein Production from CHO Cells. In *Methods in molecular biology (Clifton, N.J.)*. pp. 71–85.

Bailey, L.A., Hatton, D., Field, R., and Dickson, A.J. (2012) Determination of Chinese hamster ovary cell line stability and recombinant antibody expression during long-term culture. *Biotechnol Bioeng* **109**: 2093–2103.

Baktur, R., Elwood, A., Anderson, J., and Wight, M. (2016) Doubling the Productivity of CHO Cells. *Genet Eng Biotechnol News* **36**: 25–25.

Balasubramanian, S., Matasci, M., Kadlecova, Z., Baldi, L., Hacker, D.L., and Wurm, F.M. (2015) Rapid recombinant protein production from piggyBac transposon-mediated stable CHO cell pools. *J Biotechnol* **200**: 61–69.

Barron, M., and Li, J. (2016) Identifying and removing the cell-cycle effect from single-cell RNA-Sequencing data. *Sci Rep* **6**: 33892.

Barron, N., Kumar, N., Sanchez, N., Doolan, P., Clarke, C., Meleady, P., *et al.* (2011) Engineering CHO cell growth and recombinant protein productivity by overexpression of miR-7. *J Biotechnol* **151**: 204–211.

Bates, A., and Power, C.A. (2019) David vs. Goliath: The Structure, Function, and Clinical Prospects of Antibody Fragments. *Antibodies* **8**: 28.

Bazan, J., Całkosiński, I., and Gamian, A. (2012) Phage display—A powerful technique for immunotherapy. *Hum Vaccin Immunother* **8**: 1817–1828.

Becker, J., Timmermann, C., Rupp, O., Albaum, S.P., Brinkrolf, K., Goesmann, A., *et al.* (2014) Transcriptome analyses of CHO cells with the next-generation microarray CHO41K: Development and validation by analysing the influence of the growth stimulating substance IGF-1 substitute LongR3. *J Biotechnol* **178**: 23–31.

Bedoya-López, A., Estrada, K., Sanchez-Flores, A., Ramírez, O.T., Altamirano, C., Segovia, L., *et al.* (2016) Effect of Temperature Downshift on the Transcriptomic Responses of Chinese Hamster Ovary Cells Using Recombinant Human Tissue Plasminogen Activator Production Culture. *PLoS One* **11**: e0151529.

Berdugo, C. (2010) Cell Damage Mechanisms and Stress Response in Animal Cell Culture. 298.

Bieging, K.T., Mello, S.S., and Attardi, L.D. (2014) Unravelling mechanisms of p53-mediated tumour suppression. *Nat Rev Cancer* **14**: 359–370.

Bielser, J.-M., Wolf, M., Souquet, J., Broly, H., and Morbidelli, M. (2018) Perfusion mammalian cell culture for recombinant protein manufacturing – A critical review. *Biotechnol Adv* **36**: 1328–1340.

Bill, R.M. (2015) Recombinant protein subunit vaccine synthesis in microbes: a role for yeast? *J Pharm Pharmacol* **67**: 319–328.

Birzele, F., Schaub, J., Rust, W., Clemens, C., Baum, P., Kaufmann, H., *et al.* (2010) Into the unknown: expression profiling without genome sequence information in CHO by next generation sequencing. *Nucleic Acids Res* **38**: 3999–4010.

Bolanos-Garcia, V.M., and Blundell, T.L. (2011) BUB1 and BUBR1: Multifaceted kinases of the cell cycle. *Trends Biochem Sci* **36**: 141–150.

Bollmann, F., Riethmüller, D., Johansson, E., and Tappe, A. (2019) Optimization of HEK293T suspension cultivation with a DoE-approach in ambr®15 microbioreactor. In *Advancing Manufacture of Cell and Gene Therapies VI.* .

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.

Boyle, E.A., Li, Y.I., Pritchard, J.K., Gordon, S., Henders, A.K., Nyholt, D.R., *et al.* (2017) An Expanded View of Complex Traits: From Polygenic to Omnigenic. *Cell* **169**: 1177–1186.

Bunnak, P., Allmendinger, R., Ramasamy, S. V., Lettieri, P., and Titchener-Hooker, N.J. (2016) Life-cycle and cost of goods assessment of fed-batch and perfusion-based manufacturing processes for mAbs. *Biotechnol Prog* **32**: 1324–1335.

Buyel, J.F., and Fischer, R. (2015) A juice extractor can simplify the downstream processing of plant-derived biopharmaceutical proteins compared to blade-based homogenizers. *Process Biochem* **50**: 859–866.

Byrne, G., O'Rourke, S.M., Alexander, D.L., Yu, B., Doran, R.C., Wright, M., *et al.* (2018) CRISPR/Cas9 gene editing for the creation of an MGAT1-deficient CHO cell line to control HIV-1 vaccine glycosylation. *PLOS Biol* **16**: e2005817.

Castiñeiras, T.S., Williams, S.G., Hitchcock, A.G., and Smith, D.C. (2018) E. coli strain engineering for the production of advanced biopharmaceutical products. *FEMS Microbiol Lett* **365**.

Castro, A., Bernis, C., Vigneron, S., Labbé, J.C., and Lorca, T. (2005) The anaphase-promoting complex: A key factor in the regulation of cell cycle. *Oncogene* **24**: 314–325.

Cell Line Development (CLD) 2.0 (2017) *BioOutsource* .

Chang, Y.C., Liao, C. Bin, Hsieh, P.Y.C., Liou, M.L., and Liu, Y.C. (2008) Expression of tumor suppressor p53 facilitates DNA repair but not UV-induced G2/M arrest or apoptosis in Chinese hamster ovary CHO-K1 cells. *J Cell Biochem* **103**: 528–537.

Chen, I.T., Smith, M.L., O'Connor, P.M., and Fornace, A.J. (1995) Direct interaction of Gadd45 with PCNA and evidence for competitive interaction of Gadd45 and p21Waf1/Cip1 with PCNA. *Oncogene* **11**: 1931–7.

Chen, M.-S., Ryan, C.E., and Piwnica-Worms, H. (2003) Chk1 Kinase Negatively Regulates Mitotic Function of Cdc25A Phosphatase through 14-3-3 Binding. *Mol Cell Biol* **23**: 7488–7497.

Chen, Q., and Davis, K.R. (2016) The potential of plants as a system for the development and production of human biologics. *F1000Research* **5**: 912.

Chen, Y.-H., and Pruett-Miller, S.M. (2018) Improving single-cell cloning workflow for gene editing in human pluripotent stem cells. *Stem Cell Res* **31**: 186–192.

Chi, X., Zheng, Q., Jiang, R., Chen-Tsai, R.Y., and Kong, L.-J. (2019) A system for site-specific integration of transgenes in mammalian cells. *PLoS One* **14**: e0219842.

Chin, C.L., Goh, J.B., Srinivasan, H., Liu, K.I., Gowher, A., Shanmugam, R., *et al.* (2019) A human expression system based on HEK293 for the stable production of recombinant erythropoietin. *Sci Rep* **9**: 16768.

Chung, S., Quarmby, V., Gao, X., Ying, Y., Lin, L., Reed, C., *et al.* (2012) Quantitative evaluation of fucose reducing effects in a humanized antibody on Fcγ receptor binding and antibody-dependent cell-mediated cytotoxicity activities. *MAbs* **4**: 326–340.

Clarke, C., Doolan, P., Barron, N., Meleady, P., O'Sullivan, F., Gammell, P., *et al.* (2011a) Predicting cell-specific productivity from CHO gene expression. *J Biotechnol* **151**: 159–65.

Clarke, C., Doolan, P., Barron, N., Meleady, P., O'Sullivan, F., Gammell, P., *et al.* (2011b) Large scale microarray profiling and coexpression network analysis of CHO cells identifies transcriptional modules associated with growth and productivity. *J Biotechnol* **155**: 350–9.

Cohen, S.N., Chang, A.C.Y., Boyer, H.W., and Helling, R.B. (1973) Construction of biologically functional bacterial plasmids in vitro. *Proc Natl Acad Sci U S A* **70**: 3240–3244.

Coronel, J., Behrendt, I., Bürgin, T., Anderlei, T., Sandig, V., Reichl, U., and Genzel, Y. (2019) Influenza A virus production in a single-use orbital shaken bioreactor with ATF or TFF perfusion systems. *Vaccine* **37**: 7011–7018.

Courtes, F.C., Lin, J., Lim, H.L., Ng, S.W., Wong, N.S.C., Koh, G., *et al.* (2013) Translatome analysis of CHO cells to identify key growth genes. *J Biotechnol* **167**: 215–24.

Dadehbeigi, N., and Dickson, A.J. (2015) Chemical manipulation of the mTORC1 pathway in industrially relevant CHOK1 cells enhances production of therapeutic proteins. *Biotechnol J* **10**: 1041–1050.

Daly, R., and Hearn, M.T.W. (2005) Expression of heterologous proteins in Pichia pastoris: A useful experimental tool in protein engineenring and production. *J Mol Recognit* **18**: 119–138.

Denis, N., Kitzis, A., Kruh, J., Dautry, F., and Corcos, D. (1991) Stimulation of methotrexate resistance and dihydrofolate reductase gene amplification by c-myc. *Oncogene* **6**: 1453–1457.

Derouazi, M., Martinet, D., Besuchet Schmutz, N., Flaction, R., Wicht, M., Bertschinger, M., *et al.* (2006) Genetic characterization of CHO production host DG44 and derivative recombinant cell lines. *Biochem Biophys Res Commun* **340**: 1069–1077.

Dharshanan, S., Chong, H., Hung, C.S., Zamrod, Z., and Kamal, N. (2011) Rapid automated selection of mammalian cell line secreting high level of humanized monoclonal antibody using Clone Pix FL system and the correlation between exterior median intensity and antibody productivity. *Electron J Biotechnol* **14**: 8–8.

Doolan, P., Barron, N., Kinsella, P., Clarke, C., Meleady, P., O'Sullivan, F., *et al.* (2012) Microarray expression profiling identifies genes regulating sustained cell specific productivity (S-Qp) in CHO K1 production cell lines. *Biotechnol J* **7**: 516–26.

Doolan, P., Clarke, C., Kinsella, P., Breen, L., Meleady, P., Leonard, M., *et al.* (2013) Transcriptomic analysis of clonal growth rate variation during CHO cell line development. *J Biotechnol* **166**: 105–13.

Doolan, P., Meleady, P., Barron, N., Henry, M., Gallagher, R., Gammell, P., *et al.* (2010) Microarray and proteomics expression profiling identifies several candidates, including the valosin-containing protein (VCP), involved in regulating high cellular growth rate in production CHO cell lines. *Biotechnol Bioeng* **106**: 42–56.

Dreesen, I.A.J., and 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.

Druz, A., Son, Y.J., Betenbaugh, M., and Shiloach, J. (2013) Stable inhibition of mmu-miR-466h-5p

improves apoptosis resistance and protein production in CHO cells. *Metab Eng* **16**: 87–94.

Du, Z., Treiber, D., Mccarter, J.D., Fomina-Yadlin, D., Saleem, R.A., Mccoy, R.E., *et al.* (2015) Use of a small molecule cell cycle inhibitor to control cell growth and improve specific productivity and product quality of recombinant proteins in CHO cell cultures. *Biotechnol Bioeng* **112**: 141–155.

Dumont, J., Euwart, D., Mei, B., Estes, S., and Kshirsagar, R. (2016) Human cell lines for biopharmaceutical manufacturing: history, status, and future perspectives. *Crit Rev Biotechnol* **36**: 1110–1122.

Einem, B. Von, Wahler, A., Schips, T., Serrano-Pozo, A., Proepper, C., Boeckers, T.M., *et al.* (2015) The golgi-localized γ-ear-containing ARF-binding (GGA) proteins alter amyloid-β precursor protein (APP) processing through interaction of their GAE domain with the beta-site APP cleaving enzyme 1 (BACE1). *PLoS One* **10**: e0129047.

Enserink, J.M., and Kolodner, R.D. (2010) An overview of Cdk1-controlled targets and processes. *Cell Div* **5**: 11.

EvaluatePharma (2019) World Preview 2019, Outlook to 2024. .

Ferguson, R.L., and Maller, J.L. (2008) Cyclin E-dependent localization of MCM5 regulates centrosome duplication. *J Cell Sci* **121**: 3224–3232.

Fischer, S., Buck, T., Wagner, A., Ehrhart, C., Giancaterino, J., Mang, S., *et al.* (2014) A functional high-content miRNA screen identifies miR-30 family to boost recombinant protein production in CHO cells. *Biotechnol J* **9**: 1279–1292.

Fischer, S., Marquart, K.F., Pieper, L.A., Fieder, J., Gamer, M., Gorr, I., *et al.* (2017a) miRNA engineering of CHO cells facilitates production of difficult-to-express proteins and increases success in cell line development. *Biotechnol Bioeng* **114**: 1495–1510.

Fischer, S., Marquart, K.F., Pieper, L.A., Fieder, J., Gamer, M., Gorr, I., *et al.* (2017b) miRNA engineering of CHO cells facilitates production of difficult-to-express proteins and increases success in cell line development. *Biotechnol Bioeng* **114**: 1495–1510.

Fischer, S., Paul, A.J., Wagner, A., Mathias, S., Geiss, M., Schandock, F., *et al.* (2015) miR-2861 as novel HDAC5 inhibitor in CHO cells enhances productivity while maintaining product quality. *Biotechnol Bioeng* **112**: 2142–2153.

Fomina-Yadlin, D., Mujacic, M., Maggiora, K., Quesnell, G., Saleem, R., and McGrew, J.T. (2015) Transcriptome analysis of a CHO cell line expressing a recombinant therapeutic protein treated with inducers of protein expression. *J Biotechnol* **212**: 106–115.

Fussenegger, M., Schlatter, S., Dätwyler, D., Mazur, X., and Bailey, J.E. (1998) Controlled proliferation by multigene metabolic engineering enhances the productivity of Chinese hamster ovary cells. *Nat*

Biotechnol **16**: 468–472.

Geisler, C., Mabashi-Asazuma, H., and Jarvis, D.L. (2015) An overview and history of glyco-engineering in insect expression systems. In *Glyco-Engineering: Methods and Protocols*. Humana Press, New York, NY, pp. 131–152.

Genovese, C., Trani, D., Caputi, M., and Claudio, P.P. (2006) Cell cycle control and beyond: Emerging roles for the retinoblastoma gene family. *Oncogene* **25**: 5201–5209.

Ghosh, P., Griffith, J., Geuze, H.J., and Kornfeld, S. (2003) Mammalian GGAs act together to sort mannose 6-phosphate receptors. *J Cell Biol* **163**: 755–766.

Goh, J.B., and Ng, S.K. (2018) Impact of host cell line choice on glycan profile. *Crit Rev Biotechnol* **38**: 851–867.

Golabgir, A., Gutierrez, J.M., Hefzi, H., Li, S., Palsson, B.O., Herwig, C., and Lewis, N.E. (2016) Quantitative feature extraction from the Chinese hamster ovary bioprocess bibliome using a novel meta-analysis workflow. *Biotechnol Adv* **34**: 621–633.

Gomes, A.R., Byregowda, S.M., Veeregowda, B.M., and Balamurugan, V. (2016) An Overview of Heterologous Expression Host Systems for the Production of Recombinant Proteins. *Adv Anim Vet Sci* **4**: 346–356.

Goodman, R.H., and Smolik, S. (2000) CBP/p300 in cell growth, transformation, and development. *Genes Dev* **14**: 1553–1577.

Graham, F.L., Smiley, J., Russell, W.C., and Nairn, R. (1977) Characteristics of a human cell line transformed by DNA from human adenovirus type 5. *J Gen Virol* **36**: 59–72.

Guldager, N. (2010) Cost advantages of single use technologies. *Pharm Technol* **34**: 26–31.

Gupta, S.K., Srivastava, S.K., Sharma, A., Nalage, V.H.H., Salvi, D., Kushwaha, H., *et al.* (2017) Metabolic engineering of CHO cells for the development of a robust protein production platform. *PLoS One* **12**: e0181455.

Hall, B.G., Acar, H., Nandipati, A., and Barlow, M. (2014) Growth rates made easy. *Mol Biol Evol* **31**: 232–238.

Hamilton, S.R., Davidson, R.C., Sethuraman, N., Nett, J.H., Jiang, Y., Rios, S., *et al.* (2006) Humanization of yeast to produce complex terminally sialylated glycoproteins. *Science (80-)* **313**: 1441–1443.

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

Han, S., and Rhee, W.J. (2018) Inhibition of apoptosis using exosomes in Chinese hamster ovary cell

culture. *Biotechnol Bioeng* **115**: 1331–1339.

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.

Harmsen, M.M., and Haard, H.J. De (2007) Properties, production, and applications of camelid single-domain antibody fragments. *Appl Microbiol Biotechnol* **77**: 13–22.

Harreither, E., Hackl, M., Pichler, J., Shridhar, S., Auer, N., Łabaj, P.P., *et al.* (2015) Microarray profiling of preselected CHO host cell subclones identifies gene expression patterns associated with increased production capacity. *Biotechnol J* **10**: 1625–38.

Havenith, H., Raven, N., Fiore, S. Di, Fischer, R., and Schillberg, S. (2014) Image-based analysis of cell-specific productivity for plant cell suspension cultures. *Plant Cell Tissue Organ Cult* **117**: 393–399.

Hefzi, H., Ang, K.S., Hanscho, M., Bordbar, A., Ruckerbauer, D., Lakshmanan, M., *et al.* (2016) A Consensus Genome-scale Reconstruction of Chinese Hamster Ovary Cell Metabolism. *Cell Syst* **3**: 434-443.e8.

Hou, J.J.C., Hughes, B.S., Smede, M., Leung, K.M., Levine, K., Rigby, S., *et al.* (2014) High-throughput ClonePix FL analysis of mAb-expressing clones using the UCOE expression system. *N Biotechnol* **31**: 214–220.

Hu, W., Berdugo, C., and Chalmers, J.J. (2011) The potential of hydrodynamic damage to animal cells of industrial relevance: current understanding. *Cytotechnology* **63**: 445.

Huang, C.J., Lin, H., and Yang, X. (2012) Industrial production of recombinant therapeutics in Escherichia coli and its recent advancements. *J Ind Microbiol Biotechnol* **39**: 383–399.

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.

Ifandi, V., and Al-Rubeai, M. (2003) Stable transfection of CHO cells with the c-myc gene results in increased proliferation rates, reduces serum dependency, and induces anchorage independence. *Cytotechnology* **41**: 1–10.

Ivan, V., Martinez-Sanchez, E., Sima, L.E., Oorschot, V., Klumperman, J., Petrescu, S.M., and Sluijs, P. van der (2012) AP-3 and Rabip4' Coordinately Regulate Spatial Distribution of Lysosomes. *PLoS One* **7**: e48142.

Jadhav, V., Hackl, M., Hernandez Bort, J.A., Wieser, M., Harreither, E., Kunert, R., *et al.* (2012) A screening method to assess biological effects of microRNA overexpression in Chinese hamster ovary cells. *Biotechnol Bioeng* **109**: 1376–1385.

Jadhav, V., Hackl, M., Klanert, G., Hernandez Bort, J.A., Kunert, R., Grillari, J., and Borth, N. (2014) Stable overexpression of miR-17 enhances recombinant protein production of CHO cells. *J Biotechnol* **175**: 38–44.

Jardon, M.A., Sattha, B., Braasch, K., Leung, A.O., Côté, H.C.F., Butler, M., *et al.* (2012) Inhibition of glutamine-dependent autophagy increases t-PA production in CHO Cell fed-batch processes. *Biotechnol Bioeng* **109**: 1228–1238.

Jia, B., and Jeon, C.O. (2016) High-throughput recombinant protein expression in Escherichia coli : current status and future perspectives. *Open Biol* **6**: 160196.

Jorgolli, M., Nevill, T., Winters, A., Chen, I., Chong, S., Lin, F.F., *et al.* (2019) Nanoscale integration of single cell biologics discovery processes using optofluidic manipulation and monitoring. *Biotechnol Bioeng* **116**: 2393–2411.

Josephides, D., Davoli, S., Whitley, W., Ruis, R., Salter, R., Gokkaya, S., *et al.* (2020) Cyto-Mine: An Integrated, Picodroplet System for High-Throughput Single-Cell Analysis, Sorting, Dispensing, and Monoclonality Assurance. *SLAS Technol* **25**: 177–189.

Jozala, A.F., Geraldes, D.C., Tundisi, L.L., Feitosa, V. de A., Breyer, C.A., Cardoso, S.L., *et al.* (2016) Biopharmaceuticals from microorganisms: from production to purification. *Brazilian J Microbiol* **47**: 51–63.

Kang, S., Ren, D., Xiao, G., Daris, K., Buck, L., Enyenihi, A.A., *et al.* (2014) Cell line profiling to improve monoclonal antibody production. *Biotechnol Bioeng* **111**: 748–760.

Kantardjieff, A., Jacob, N.M., Yee, J.C., Epstein, E., Kok, Y.-J., Philp, R., *et al.* (2010) Transcriptome and proteome analysis of Chinese hamster ovary cells under low temperature and butyrate treatment. *J Biotechnol* **145**: 143–59.

Kao, F.T., and Puck, T.T. (1968) Genetics of somatic mammalian cells, VII. Induction and isolation of nutritional mutants in Chinese hamster cells. *Proc Natl Acad Sci U S A* **60**: 1275–1281.

Kawabe, Y., Inao, T., Komatsu, S., Ito, A., and Kamihira, M. (2015) Cre-mediated cellular modification for establishing producer CHO cells of recombinant scFv-Fc. *BMC Proc* **9**: P5.

Kelly, P.S., Clarke, C., Costello, A., Monger, C., Meiller, J., Dhiman, H., *et al.* (2017) Ultra-deep next generation mitochondrial genome sequencing reveals widespread heteroplasmy in Chinese hamster ovary cells. *Metab Eng* **41**: 11–22.

Kelly, T., Tuckowski, A.M., and Smith, K.D. (2018) Using FRET-Based Microfluidic Screening for Analysis, Sorting, Imaging, and Dispensing in the Rapid Generation of High-Producing Clonal Cell Lines. *Bioprocess Int* .

Kermode, A.R., and Jiang, L. (2018) *Molecular pharming: applications, challenges and emerging areas*.

John Wiley & Sons, Inc., Hoboken, NJ, USA.

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**: 1102–7.

Kim, B., Lee, E., and Kim, J.K. (2019) Analysis of technical and biological variability in single-cell RNA sequencing. In *Methods in Molecular Biology*. pp. 25–43.

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**: 917–30.

Kim, S.H., and Lee, G.M. (2007a) Down-regulation of lactate dehydrogenase-A by siRNAs for reduced lactic acid formation of Chinese hamster ovary cells producing thrombopoietin. *Appl Microbiol Biotechnol* **74**: 152–159.

Kim, S.H., and Lee, G.M. (2007b) Functional expression of human pyruvate carboxylase for reduced lactic acid formation of Chinese hamster ovary cells (DG44). *Appl Microbiol Biotechnol* **76**: 659–665.

Kim, W.H., Kim, Y.J., and Lee, G.M. (2014) Gadd45-induced cell cycle G2/M arrest for improved transient gene expression in Chinese hamster ovary cells. *Biotechnol Bioprocess Eng* **19**: 386–393.

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

Kipriyanov, S.M., Moldenhauer, G., Braunagel, M., Reusch, U., Cochlovius, B., Gall, F. Le, *et al.* (2003) Effect of domain order on the activity of bacterially produced bispecific single-chain Fv antibodies. *J Mol Biol* **330**: 99–111.

Kiss, B., Gottschalk, U., and Pohlscheidt, M. (2018) *New Bioprocessing Strategies: Development and Manufacturing of Recombinant Antibodies and Proteins*. Springer International Publishing, Cham.

Klanert, G., Fernandez, D.J., Weinguny, M., Eisenhut, P., Bühler, E., Melcher, M., *et al.* (2019) A cross-species whole genome siRNA screen in suspension-cultured Chinese hamster ovary cells identifies novel engineering targets. *Sci Rep* **9**: 8689.

Köhler, G., and Milstein, C. (1975) Continuous cultures of fused cells secreting antibody of predefined specificity. *Nature* **256**: 495–497.

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.

Landowski, C.P., Mustalahti, E., Wahl, R., Croute, L., Sivasiddarthan, D., Westerholm-Parvinen, A., *et al.* (2016) Enabling low cost biopharmaceuticals: high level interferon alpha-2b production in Trichoderma reesei. *Microb Cell Fact* **15**: 104.

Le, K., Tan, C., Gupta, S., Guhan, T., Barkhordarian, H., Lull, J., *et al.* (2018) A novel mammalian cell line

development platform utilizing nanofluidics and optoelectro positioning technology. *Biotechnol Prog* **34**: 1438–1446.

Le, K., Tan, C., Le, H., Tat, J., Zasadzinska, E., Diep, J., *et al.* (2020) Assuring Clonality on the Beacon Digital Cell Line Development Platform. *Biotechnol J* **15**: 1900247.

Leader, B., Baca, Q.J., and Golan, D.E. (2008) Protein therapeutics: a summary and pharmacological classification. *Nat Rev Drug Discov* **7**: 21–39.

Lee, K.H., Sburlati, A., Renner, W.A., and Bailey, J.E. (1996) Deregulated expression of cloned transcription factor E2F-1 in Chinese hamster ovary cells shifts protein patterns and activates growth in protein-free medium. *Biotechnol Bioeng* **50**: 273–279.

Lee, K.H., Tsutsui, T., Honda, K., Ohtake, H., and Omasa, T. (2013) Overexpression of mutant cell division cycle 25 homolog B (CDC25B) enhances the efficiency of selection in Chinese hamster ovary cells. *Cytotechnology* **65**: 1017–1026.

Lee, N., Shin, J., Park, J.H., Lee, G.M., Cho, S., and Cho, B.-K. (2016) Targeted Gene Deletion Using DNA-Free RNA-Guided Cas9 Nuclease Accelerates Adaptation of CHO Cells to Suspension Culture. *ACS Synth Biol* **5**: 1211–1219.

Lei, M. (2005) The MCM Complex: Its Role in DNA Replication and Implications for Cancer Therapy. *Curr Cancer Drug Targets* **5**: 365–380.

Lewis, A.M., Abu-Absi, N.R., Borys, M.C., and Li, Z.J. (2016) The use of 'Omics technology to rationally improve industrial mammalian cell line performance. *Biotechnol Bioeng* **113**: 26–38.

Lewis, N.E., Liu, X., Li, Y., Nagarajan, H., Yerganian, G., O'Brien, E., *et al.* (2013) Genomic landscapes of Chinese hamster ovary cell lines as revealed by the Cricetulus griseus draft genome. *Nat Biotechnol* **31**: 759–765.

Ley, D., Pereira, S., Pedersen, L.E., Arnsdorf, J., Hefzi, H., Davy, A.M., *et al.* (2019) Reprogramming AA catabolism in CHO cells with CRISPR/Cas9 genome editing improves cell growth and reduces byproduct secretion. *Metab Eng* **56**: 120–129.

Li, J., Deng, M., Wei, Q., Liu, T., Tong, X., and Ye, X. (2011) Phosphorylation of MCM3 protein by cyclin E/cyclin-dependent kinase 2 (Cdk2) regulates its function in cell cycle. *J Biol Chem* **286**: 39776–39785.

Lim, U.M., Yap, M.G.S., Lim, Y.P., Goh, L.-T., and Ng, S.K. (2013) Identification of Autocrine Growth Factors Secreted by CHO Cells for Applications in Single-Cell Cloning Media. *J Proteome Res* **12**: 3496–3510.

Loh, W.P., Loo, B., Zhou, L., Zhang, P., Lee, D.Y., Yang, Y., and Lam, K.P. (2014) Overexpression of microRNAs enhances recombinant protein production in Chinese hamster ovary cells. *Biotechnol J* **9**: 1140–1151.

Lonza (2019) The next level of therapeutic protein expression. .

Lynn, D.E. (2001) Novel techniques to establish new insect cell lines. *Vitr Cell Dev Biol - Anim* **37**: 319–321.

Mamo, A., Jules, F., Dumaresq-Doiron, K., Costantino, S., and Lefrancois, S. (2012) The Role of Ceroid Lipofuscinosis Neuronal Protein 5 (CLN5) in Endosomal Sorting. *Mol Cell Biol* **32**: 1855–1866.

Mangalampalli, V.R., Wycuff, D., Chen, M., Berlinger, D., Scheideman, E.H., Menon, A., *et al.* (2015) CHO-DHFR cell line development platform: Application of Clonepix and Automated Mini Bioreactor (AMBR) technologies to meet accelerated timelines. *BMC Proc* **9**: P52.

Marx, N., Grünwald-Gruber, C., Bydlinski, N., Dhiman, H., Ngoc Nguyen, L., Klanert, G., and Borth, N. (2018) CRISPR-Based Targeted Epigenetic Editing Enables Gene Expression Modulation of the Silenced Beta-Galactoside Alpha-2,6-Sialyltransferase 1 in CHO Cells. *Biotechnol J* **13**: 1700217.

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.

McCamish, M., and Woollett, G. (2011) Worldwide experience with biosimilar development. *MAbs* **3**: 209–217.

Mccoy, R.E., Costa, N.A., and Morris, A.E. (2015) Factors that determine stability of highly concentrated chemically defined production media. *Biotechnol Prog* **31**: 493–502.

Mcdermott, M.F. (2016) Anti-TNF Treatment in Rheumatoid Arthritis Anti-TNF treatment in Rheumatoid Arthritis. *NRAS - Natl Rheum Arthritis Soc* .

McMahon, H.T., and Boucrot, E. (2011) Molecular mechanism and physiological functions of clathrin-mediated endocytosis. *Nat Rev Mol Cell Biol* **12**: 517–533.

Medendorp, K., Vreede, L., Groningen, J.J.M. van, Hetterschijt, L., Brugmans, L., Jansen, P.A.M., *et al.* (2010) The mitotic arrest deficient protein mad2b interacts with the clathrin light chain a during mitosis. *PLoS One* **5**: e15128.

Milkereit, R., and Rotin, D. (2011) A role for the ubiquitin ligase Nedd4 in membrane sorting of LAPTM4 proteins. *PLoS One* **6**: e27478.

Mocciaro, A., Roth, T.L., Bennett, H.M., Soumillon, M., Shah, A., Hiatt, J., *et al.* (2018) Light-activated cell identification and sorting (LACIS) for selection of edited clones on a nanofluidic device. *Commun Biol* **1**: 41.

Mohan, C., Park, S.H., Chung, J.Y., and 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.

Mora, A., Zhang, S.S., Carson, G., Nabiswa, B., Hossler, P., and Yoon, S. (2018) Sustaining an efficient and effective CHO cell line development platform by incorporation of 24-deep well plate screening and multivariate analysis. *Biotechnol Prog* **34**: 175–186.

Mordor Intelligence (2018) Biopharmaceuticals market - growth, trends and forecast (2019 - 2024). *Mordor Intell* .

Na, W., Park, N., Yeom, M., and Song, D. (2015) Ebola outbreak in Western Africa 2014: what is going on with Ebola virus? *Clin Exp Vaccine Res* **4**: 17.

Nienow, A.W. (2006) Reactor engineering in large scale animal cell culture. *Cytotechnology* **50**: 9–33.

Nissom, P.M., Sanny, A., Kok, Y.J., Hiang, Y.T., Chuah, S.H., Shing, T.K., *et al.* (2006) Transcriptome and proteome profiling to understanding the biology of high productivity CHO cells. *Mol Biotechnol* **34**: 125–40.

Oskouian, B., Soonyakumaran, P., Borowsky, A.D., Crans, A., Dillard-Telm, L., Tam, Y.Y., *et al.* (2006) Sphingosine-1-phosphate lyase potentiates apoptosis via p53- and p38-dependent pathways and is down-regulated in colon cancer. *Proc Natl Acad Sci U S A* **103**: 17384–17389.

Park, H.S., 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.

Park, J.H., Jin, J.H., Lim, M.S., An, H.J., Kim, J.W., and Lee, G.M. (2017) Proteomic analysis of host cell protein dynamics in the culture supernatants of antibody-producing CHO cells. *Sci Rep* **7**: 44246.

Patel, T., Josse, L., Smales, M., and Young, R. (2016) Manipulation and exploitation of MicroRNAs for enhanced recombinant protein production in CHO cells. .

Peden, A.A., Oorschot, V., Hesser, B.A., Austin, C.D., Scheller, R.H., and Klumperman, J. (2004) Localization of the AP-3 adaptor complex defines a novel endosomal exit site for lysosomal membrane proteins. *J Cell Biol* **164**: 1065–1076.

Pekle, E., Smith, A., Rosignoli, G., Sellick, C., Smales, C.M., and Pearce, C. (2019) Application of Imaging Flow Cytometry for the Characterization of Intracellular Attributes in Chinese Hamster Ovary Cell Lines at the Single-Cell Level. *Biotechnol J* **14**: 1800675.

Pereira, J., Rajendra, Y., Baldi, L., Hacker, D.L., and Wurm, F.M. (2011) Transient gene expression with CHO cells in conditioned medium: a study using TubeSpin® bioreactors. *BMC Proc* **5**: P38.

Petříček, R., Moucha, T., Rejl, F.J., Valenz, L., Haidl, J., and Čmelíková, T. (2018) Volumetric mass transfer coefficient, power input and gas hold-up in viscous liquid in mechanically agitated fermenters. Measurements and scale-up. *Int J Heat Mass Transf* **124**: 1117–1135.

Pilbrough, W., Munro, T.P., and Gray, P. (2009) Intraclonal protein expression heterogeneity in

recombinant CHO cells. *PLoS One* **4**: e8432.

Pipalia, N.H., Hao, M., Mukherjee, S., and Maxfield, F.R. (2007) Sterol, protein and lipid trafficking in Chinese Hamster ovary cells with Niemann-Pick type C1 defect. *Traffic* **8**: 130–141.

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.

Prinz, S., Hwang, E.S., Visintin, R., and Amon, A. (1998) The regulation of Cdc20 proteolysis reveals a role for the APC components Cdc23 and Cdc27 during S phase and early mitosis. *Curr Biol* **8**: 750–760.

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**: 1700216.

Qian, Y., Khattak, S.F., Xing, Z., He, A., Kayne, P.S., Qian, N.X., *et al.* (2011) Cell culture and gene transcription effects of copper sulfate on Chinese hamster ovary cells. *Biotechnol Prog* **27**: 1190–1194.

Recker, R.R., Benson, C.T., Matsumoto, T., Bolognese, M.A., Robins, D.A., Alam, J., *et al.* (2015) A randomized, double-blind phase 2 clinical trial of blosozumab, a sclerostin antibody, in postmenopausal women with low bone mineral density. *J Bone Miner Res* **30**: 216–224.

Renner, W.A., Lee, K.H., Hatzimanikatis, V., Bailey, J.E., and Eppenberger, H.M. (1995) Recombinant cyclin E expression activates proliferation and obviates surface attachment of chinese hamster ovary (CHO) cells in protein-free medium. *Biotechnol Bioeng* **47**: 476–482.

Rosano, G.L., Morales, E.S., and Ceccarelli, E.A. (2019) New tools for recombinant protein production in Escherichia coli: A 5-year update. *Protein Sci* **28**: 1412–1422.

Roth, W., Deussing, J., Botchkarev, V.A., Pauly-Evers, M., Saftig, P., Hafner, A., *et al.* (2000) Cathepsin L deficiency as molecular defect of furless: Hyperproliferation of keratinocytes and pertubation of hair follicle cycling. *FASEB J* **14**: 2075–2086.

Rupp, O., MacDonald, M.L., Li, S., Dhiman, H., Polson, S., Griep, S., *et al.* (2018) A reference genome of the Chinese hamster based on a hybrid assembly strategy. *Biotechnol Bioeng* **115**: 2087–2100.

Saarenpää, T., Jaakola, V.P., and Goldman, A. (2015) Baculovirus-mediated expression of GPCRs in insect cells. In *Methods in Enzymology*. Academic Press, pp. 185–218.

Sanchez-Garcia, L., Martín, L., Mangues, R., Ferrer-Miralles, N., Vázquez, E., and Villaverde, A. (2016) Recombinant pharmaceuticals from microbial cells: a 2015 update. *Microb Cell Fact* **15**: 33.

Sauna, Z.E., Lagassé, H.A.D., Alexaki, A., Simhadri, V.L., Katagiri, N.H., Jankowski, W., and Kimchi-Sarfaty, C. (2017) Recent advances in (therapeutic protein) drug development. *F1000Research* **6**: 113.

Scarcelli, J.J., Hone, M., Beal, K., Ortega, A., Figueroa, B., Starkey, J.A., and Anderson, K. (2018) Analytical subcloning of a clonal cell line demonstrates cellular heterogeneity that does not impact process consistency or robustness. *Biotechnol Prog* **34**: 602–612.

Schaub, J., Clemens, C., Schorn, P., Hildebrandt, T., Rust, W., Mennerich, D., *et al.* (2010) CHO gene expression profiling in biopharmaceutical process analysis and design. *Biotechnol Bioeng* **105**: 431–8.

Schillberg, S., Raven, N., Spiegel, H., Rasche, S., and Buntru, M. (2019) Critical analysis of the commercial potential of plants for the production of recombinant proteins. *Front Plant Sci* **10**: 720.

Schoberer, J., and Strasser, R. (2018) Plant glyco-biotechnology. *Semin Cell Dev Biol* **80**: 133–141.

Serpieri, F., Inocencio, A., Oliveira, J.M. de, Pimenta, A.A., Garbuio, A., Kalil, J., *et al.* (2010) Comparison of Humanized IgG and FvFc Anti-CD3 Monoclonal Antibodies Expressed in CHO Cells. *Mol Biotechnol* **45**: 218–225.

Shen, C.C., Sung, L.Y., Lin, S.Y., Lin, M.W., and Hu, Y.C. (2017) Enhancing Protein Production Yield from Chinese Hamster Ovary Cells by CRISPR Interference. *ACS Synth Biol* **6**: 1509–1519.

Shen, D., Kiehl, T.R., Khattak, S.F., Li, Z.J., He, A., Kayne, P.S., *et al.* (2010) Transcriptomic responses to sodium chloride-induced osmotic stress: A study of industrial fed-batch CHO cell cultures. *Biotechnol Prog* **26**: 1104–1115.

Sheridan, C. (2017) Ablynx's nanobody fragments go places antibodies cannot. *Nat Biotechnol* **35**: 1115–1117.

Shridhar, S., Klanert, G., Auer, N., Hernandez-Lopez, I., Kańduła, M.M., Hackl, M., *et al.* (2017) Transcriptomic changes in CHO cells after adaptation to suspension growth in protein-free medium analysed by a species-specific microarray. *J Biotechnol* .

Singh, V. (1999) Disposable bioreactor for cell culture using wave-induced agitation. *Cytotechnology* **30**: 149–58.

So, A.D., and Katz, S.L. (2010) Biologics boondoggle. *New York Times* 7(03).

Steger, K., Brady, J., Wang, W., Duskin, M., Donato, K., and Peshwa, M. (2015) CHO-S Antibody Titers >1 Gram/Liter Using Flow Electroporation-Mediated Transient Gene Expression followed by Rapid Migration to High-Yield Stable Cell Lines. *J Biomol Screen* **20**: 545–551.

Stettler, M., Zhang, X., Hacker, D.L., Jesus, M. De, and Wurm, F.M. (2007) Novel orbital shake bioreactors for transient production of CHO derived IgGs. *Biotechnol Prog* **23**: 1340–1346.

Strotbek, M., Florin, L., Koenitzer, J., Tolstrup, A., Kaufmann, H., Hausser, A., and Olayioye, M.A. (2013) Stable microRNA expression enhances therapeutic antibody productivity of Chinese hamster ovary cells. *Metab Eng* **20**: 157–166.

Strzalka, W., and Ziemienowicz, A. (2011) Proliferating cell nuclear antigen (PCNA): A key factor in DNA replication and cell cycle regulation. *Ann Bot* **107**: 1127–1140.

Sur, S., and Agrawal, D.K. (2016) Phosphatases and kinases regulating CDC25 activity in the cell cycle: clinical implications of CDC25 overexpression and potential treatment strategies. *Mol Cell Biochem* **416**: 33–46.

Takatsu, H., Yoshino, K., and Nakayama, K. (2000) Adaptor γ ear homology domain conserved in γ-adaptin and GGA proteins that interact with γ-synergin. *Biochem Biophys Res Commun* **271**: 719–725.

Tan, H.K., Lee, M.M., Yap, M.G.S., and Wang, D.I.C. (2008) Overexpression of cold-inducible RNA-binding protein increases interferon-γ production in Chinese-hamster ovary cells. *Biotechnol Appl Biochem* **49**: 247.

Taylor, P. (2019) Evenity back on track with FDA panel blessing in osteoporosis. *PMLive* .

Tekoah, Y., Tzaban, S., Kizhner, T., Hainrichson, M., Gantman, A., Golembo, M., *et al.* (2013) Glycosylation and functionality of recombinant β-glucocerebrosidase from various production systems. *Biosci Rep* **33**: 771–781.

Tian, W., Ye, Z., Wang, S., Schulz, M.A., Coillie, J. Van, Sun, L., *et al.* (2019) The glycosylation design space for recombinant lysosomal replacement enzymes produced in CHO cells. *Nat Commun* **10**: 1785.

Toussaint, C., Henry, O., and Durocher, Y. (2016) Metabolic engineering of CHO cells to alter lactate metabolism during fed-batch cultures. *J Biotechnol* **217**: 122–131.

Turan, S., Zehe, C., Kuehle, J., Qiao, J., and Bode, J. (2013) Recombinase-mediated cassette exchange (RMCE) — A rapidly-expanding toolbox for targeted genomic modifications. *Gene* **515**: 1–27.

Veith, N., Ziehr, H., MacLeod, R.A.F., and Reamon-Buettner, S.M. (2016) Mechanisms underlying epigenetic and transcriptional heterogeneity in Chinese hamster ovary (CHO) cell lines. *BMC Biotechnol* **16**: 6.

Vidarsson, G., Dekkers, G., and Rispens, T. (2014a) IgG subclasses and allotypes: from structure to effector functions. *Front Immunol* **5**: 520.

Vidarsson, G., Dekkers, G., and Rispens, T. (2014b) IgG subclasses and allotypes: From structure to effector functions. *Front Immunol* **5**: 520.

Vito, D., Eriksen, J.C., Skjødt, C., Weilguny, D., Rasmussen, S.K., and Smales, C.M. (2019) Defining lncRNAs Correlated with CHO Cell Growth and IgG Productivity by RNA-Seq. *iScience* **0**: 100785.

Vito, D., and Smales, C.M. (2018) The Long Non-Coding RNA Transcriptome Landscape in CHO Cells Under Batch and Fed-Batch Conditions. *Biotechnol J* **13**: 1800122.

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

Walsh, G. (2018) Biopharmaceutical benchmarks 2018. *Nat Biotechnol* **36**: 1136–1145.

Wang, B., Albanetti, T., Miro-Quesada, G., Flack, L., Li, L., Klover, J., *et al.* (2018) High-throughput screening of antibody-expressing CHO clones using an automated shaken deep-well system. *Biotechnol Prog* **34**: 1460–1471.

Wang, C., Gong, B., Bushel, P.R., Thierry-Mieg, J., Thierry-Mieg, D., Xu, J., *et al.* (2014) The concordance between RNA-seq and microarray data depends on chemical treatment and transcript abundance. *Nat Biotechnol* **32**: 926–932.

Wang, Q., Chen, Y., Park, J., Liu, X., Hu, Y., Wang, T., *et al.* (2019) Design and Production of Bispecific Antibodies. *Antibodies* **8**: 43.

Weinstein, J., Jacobsen, F.W., Hsu-Chen, J., Wu, T., and Baum, L.G. (1994) A novel mammalian protein, p55CDC, present in dividing cells is associated with protein kinase activity and has homology to the Saccharomyces cerevisiae cell division cycle proteins Cdc20 and Cdc4. *Mol Cell Biol* **14**: 3350–3363.

Wildt, S., and Gerngross, T.U. (2005) The humanization of N-glycosylation pathways in yeast. *Nat Rev Microbiol* **3**: 119–128.

Wilkens, C.A., Vishwanathan, N., Baltes, N.J., Lucero, A.T., Hu, W.S., and Gerdtzen, Z.P. (2019) An LDHa single allele CHO cell mutant exhibits altered metabolic state and enhanced culture performance. *J Chem Technol Biotechnol* **94**: 1488–1498.

Winters, A., McFadden, K., Bergen, J., Landas, J., Berry, K.A., Gonzalez, A., *et al.* (2019) Rapid single B cell antibody discovery using nanopens and structured light. *MAbs* **11**: 1025–1035.

Wong, D.C.F., Wong, K.T.K., Lee, Y.Y., Morin, P.N., Heng, C.K., and Yap, M.G.S. (2006) Transcriptional profiling of apoptotic pathways in batch and fed-batch CHO cell cultures. *Biotechnol Bioeng* **94**: 373–382.

Wong, M.S., Wu, S., Causey, T.B., Bennett, G.N., and San, K.Y. (2008) Reduction of acetate accumulation in Escherichia coli cultures for increased recombinant protein production. *Metab Eng* **10**: 97–108.

Woo, R.A., McLure, K.G., Lees-Miller, S.P., Rancourt, D.E., and Lee, P.W.K. (1998) DNA-dependent protein kinase acts upstream of p53 in response to DNA damage. *Nature* **394**: 700–704.

Wozniak, E., and Biesecker, K. (2017) Perfusion Solutions for Therapeutic Development. *Genet Eng Biotechnol News* **37**: 20–21.

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

Wurm, F.M., and Hacker, D. (2011) First CHO genome. *Nat Biotechnol* **29**: 718–720.

Xu, M., Motabar, O., Ferrer, M., Marugan, J.J., Zheng, W., and Ottinger, E.A. (2016) Disease models for the development of therapies for lysosomal storage diseases. *Ann N Y Acad Sci* **1371**: 15–29.

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

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

Yee, C.M., Zak, A.J., Hill, B.D., and Wen, F. (2018) The Coming Age of Insect Cells for Manufacturing and Development of Protein Therapeutics. *Ind Eng Chem Res* **57**: 10061–10070.

Yee, J.C., Gerdtzen, Z.P., and Hu, W.-S. (2009) Comparative transcriptome analysis to unveil genes affecting recombinant protein productivity in mammalian cells. *Biotechnol Bioeng* **102**: 246–263.

Yee, J.C., Leon Gatti, M. de, Philp, R.J., Yap, M., and Hu, W.-S. (2008) Genomic and proteomic exploration of CHO and hybridoma cells under sodium butyrate treatment. *Biotechnol Bioeng* **99**: 1186–204.

Yeong, F.M. (2013) Multi-step down-regulation of the secretory pathway in mitosis: A fresh perspective on protein trafficking. *BioEssays* **35**: 462–471.

Yongky, A., Xu, J., Tian, J., Oliveira, C., Zhao, J., McFarland, K., *et al.* (2019) Process intensification in fed-batch production bioreactors using non-perfusion seed cultures. *MAbs* **11**: 1502–1514.

Yuk, I.H., Zhang, J.D., Ebeling, M., Berrera, M., Gomez, N., Werz, S., *et al.* (2014) Effects of copper on CHO cells: Insights from gene expression analyses. *Biotechnol Prog* **30**: 429–442.

Yusufi, F.N.K., Lakshmanan, M., Ho, Y.S., Loo, B.L.W., Ariyaratne, P., Yang, Y., *et al.* (2017) Mammalian Systems Biotechnology Reveals Global Cellular Adaptations in a Recombinant CHO Cell Line. *Cell Syst* **4**: 530-542.e6.

Zhang, F., Sun, X., Yi, X., and Zhang, Y. (2006) Metabolic characteristics of recombinant Chinese hamster ovary cells expressing glutamine synthetase in presence and absence of glutamine. *Cytotechnology* **51**: 21–28.

Zhang, X., Han, L., Zong, H., Ding, K., Yuan, Y., Bai, J., *et al.* (2018) Enhanced production of anti-PD1 antibody in CHO cells through transient co-transfection with anti-apoptotic genes Bcl-x L and Mcl-1. *Bioprocess Biosyst Eng* **41**: 633–640.

Zhang, X., Stettler, M., Reif, O., Kocourek, A., DeJesus, M., Hacker, D.L., and Wurm, F.M. (2008) Shaken helical track bioreactors: Providing oxygen to high-density cultures of mammalian cells at volumes up to 1000L by surface aeration with air. *N Biotechnol* **25**: 68–75.

Zhang, Y., Baycin-Hizal, D., Kumar, A., Priola, J., Bahri, M., Heffner, K.M., *et al.* (2017) High-Throughput Lipidomic and Transcriptomic Analysis To Compare SP2/0, CHO, and HEK-293 Mammalian Cell Lines.

Anal Chem **89**: 1477–1485.

Zhao, L., Fu, H.Y., Raju, R., Vishwanathan, N., and Hu, W.S. (2017) Unveiling gene trait relationship by cross-platform meta-analysis on Chinese hamster ovary cell transcriptome. *Biotechnol Bioeng* **114**: 1583–1592.

Zhao, M., Wang, J., Luo, M., Luo, H., Zhao, M., Han, L., *et al.* (2018) Rapid development of stable transgene CHO cell lines by CRISPR/Cas9-mediated site-specific integration into C12orf35. *Appl Microbiol Biotechnol* **102**: 6105–6117.

Zhao, S., Jiang, E., Chen, S., Gu, Y., Shangguan, A.J., Lv, T., *et al.* (2016) PiggyBac transposon vectors: the tools of the human gene encoding. *Transl lung cancer Res* **5**: 120–5.

Zhou, M., Crawford, Y., Ng, D., Tung, J., Pynn, A.F.J., Meier, A., *et al.* (2011) Decreasing lactate level and increasing antibody production in Chinese Hamster Ovary cells (CHO) by reducing the expression of lactate dehydrogenase and pyruvate dehydrogenase kinases. *J Biotechnol* **153**: 27–34.

Zimran, A., Gonzalez-Rodriguez, D.E., Abrahamov, A., Cooper, P.A., Varughese, S., Giraldo, P., *et al.* (2018) Long-term safety and efficacy of taliglucerase alfa in pediatric Gaucher disease patients who were treatment-naïve or previously treated with imiglucerase. *Blood Cells, Mol Dis* **68**: 163–172.

Zitzmann, J., Sprick, G., Weidner, T., Schreiber, C., and Czermak, P. (2017) Process Optimization for Recombinant Protein Expression in Insect Cells. In *New Insights into Cell Culture Technology*. InTech, .