CHARACTERISING CHINESE HAMSTER OVARY CELL LINE STABILITY IN BIOPRODUCTION OF THERAPEUTIC PROTEINS

THESIS SUBMITTED FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

BY

ALEXANDER HENDERSON

Supervised by:

Dr. Francesco Saverio Tedesco (UCL), Academic Supervisor

Prof. Patrizia Ferretti (UCL), Secondary Academic Supervisor

Ms. Shahla Salehi, Mr. Remben Talaban and Dr. Daniele Perna (GSK), Industrial Supervisors

Department of Cell and Developmental Biology, University College London, Rockefeller Building,
21 University Street, London, UK

Department of Biopharm Process Research, GlaxoSmithKline Medicines Research Centre,
Gunnels Wood Road, Stevenage, UK

Declaration

I, Alexander Henderson, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Signed……

Abstract

Chinese Hamster Ovary (CHO) cell factories are the platform of choice for manufacturing therapeutic proteins. Cell factories must generate stable levels of high-quality recombinant therapeutic proteins over prolonged periods of time for manufacture. Selecting stable cell lines currently requires between four and six months (70-150 generations) of continuous cell culture and ~60% of cell lines fail to meet stability acceptance criteria (<30% drop in titre). Although production instability is an industry wide concern, a causative pathway of the production instability phenotype has not been elucidated.

Here, I investigate the role of karyotypic heterogeneity in CHOK1a-GS-KO host and therapeutic protein producing cell lines, utilising multi-colour FISH (MFISH) to assess the cell lines mutational landscape. I present data that highlights a robust link between increased gross chromosomal mutations, leading to increased karyotypic heterogeneity within the culture flask, and production instability. Through designating distinct chromosomal populations, into clonal (CCA, genetically stable) and non-clonal (NCCA, genetically unstable/rare) chromosomal aberrations, an overall characterisation of a cell lines mutational landscape can be obtained. Increased % NCCA was shown to robustly correlate with production instability, in a therapeutic protein agnostic manor, in maintenance and production run environments. Utilising this metric to predict productionally unstable cell lines resulted in an 80-100% best case correct prediction rate, within one month of cell culture (~20 generations).

To elucidate a pathway that may be attributed to increased mutations within CHO cells, DNA damage, relative interstitial telomere sequence length and telomere protection was characterised within CHOK1a-GS-KO host and CHOK1a-GS-KO producer cell lines. I hypothesised that elevated DNA damage and improper repair, localised at large blocks of interstitial telomeric sequences present in the CHO genome, leads to gross chromosomal mutations compounding karyotypic heterogeneity and the production instability phenotype. This pathway was investigated across productionally stable and unstable cell lines; however, DNA damage and telomere length changes were not correlated to production instability. The work presented here provides additional evidence that CHO cell lines obtain vast karyotypic heterogeneity and establishes a novel link between increased chromosomal mutation rate and production instability.

Impact Statement

Currently, no published or commercially available CHO cell line stability predictor is available, whilst CHO production stability remains an industry-wide concern. Through manual image analysis of multi-colour FISH (MFISH) images, which allows visualisation of gross chromosomal rearrangements, I have created a methodology that can characterise genomic instability for therapeutic protein producing cell lines. Comparing genetically stable and unstable populations, across productionally stable and unstable cell lines, I identified a causal link between genomic instability and production instability (P=<0.0001).

Using this method, predictions were performed on a blinded panel of cell lines representing a normal stability cycle. Overall correct prediction rates ranged from 70-82.5% with a bias towards unstable cell lines (80-100%). Through a successful collaboration with GSKs artificial intelligence and machine learning group, we industrialised the manual prediction method providing GSK with a truly scalable tool that is ready to be embedded into the cell line development critical path.

Using this method, we can predict cell line stability within one-month (~20 generations) of a cell's life span (compared to waiting four to six months in a traditional stability assessment), leading to biopharmaceutical chemistry, manufacturing and controls (CMC) timeline savings of up to 9 months. This is achieved through the early triaging of unstable cell lines (from 48 to 12), increasing cell line development (CLD) capacity four-fold, by freeing space in bioreactors (e.g. Ambr 15's) for additional cell lines that produce a different therapeutic protein.

Time savings can be achieved by assessing four therapeutic proteins concurrently in a single stability cycle (seven months total, due to capacity restraints), rather than performing four sequential stability studies (16-month timeline). Alleviating the stability assessment bottleneck allows new medicines to reach patients faster and increasing CLD capacity provides cost savings per medicine.

To my knowledge, this is the first study to identify a causal link between karyotypic heterogeneity and production instability. Furthermore, this is the first study to test such novel findings on an industry relevant panel of cell lines (60 cell lines across 5 therapeutic proteins). The work presented here has culminated in a patent application submission, funded by GSK.

Acknowledgements

To Saverio, thank you for providing me the opportunity to complete my masters within your lab, which subsequently lead to my enrolment onto this PhD program. Your passion for science has inspired me throughout these last five years and I am grateful that you provided me a platform for me to grow as a scientist.

To Shahla, Remben, Daniele and Shawal, you have all been my mentors throughout my time at GSK. Without the support network of you all, I would not have been able to achieve what I have over the last four years. Thank you for having patience with me and allowing me to grow both professionally and personally.

To Nicola and Laura, working with you both over the last year on this project was my highlight of this PhD. I am very pleased to have met two inquisitive people, who share the same desire, passion and work ethic, as I. I wish you all the best for your future endeavours.

To Simon and Sonia, your guidance across all statistical aspects of this thesis has been invaluable. In each of our meetings, I learnt something new and I have been able to apply this knowledge to robustly interrogate results generated here. Thank you!

To my closest friends, thank you for supporting me by giving me an avenue to relieve stress and take my mind off anything science related, through days and nights out, stag do's and beautiful weddings. I am looking forward to more of the same, as we all grow together.

To my beloved parents, without your perseverance and support over the last 10+ years of education, I would not have been in a position to complete this PhD. I am proud of both of you for working so hard throughout your lives, to create a better life for us and providing me a head start in life.

To my amazing wife, from the moment we first met, I knew we had something special together. I am grateful to have had you by my side whilst we have gone through peaks and troughs during our PhD's. Once this is over, we can look back on this time with pride. I cannot wait to build a new life with you in New York! Ti amo!

Table of Contents

List of Figures

List of Tables

Chapter 1: Introduction

Over the years of Biopharmaceutical research, Chinese Hamster Ovary (CHO) cells have been adopted as the primary biological factory for producing recombinant therapeutic proteins for a variety of diseases (Wurm, 2004). Despite their wide use in industry, deep cellular characterisation of CHO cells within a bioproduction environment has only recently begun. There have been a variety of optimisation strategies that have been employed (e.g. media and vessel culture conditions) that have pushed CHO cells to increase their therapeutic protein titres (Wurm, 2004). However, despite great advances in titre increasing efforts, CHO production stability (maintenance of titre within a +/- 30% threshold, across a 70-150 generation manufacturing window) is still an industry concern. 8-63% of all CHO cell lines are considered productionally unstable (Tharmalingam et al., 2018), with GSK in-house data placing this value closer to 60% on average. Moreover, these values can also fluctuate stochastically, dependent on the therapeutic protein being produced.

The high percentage of unstable cell lines leads to a greater volume of cell lines being developed and analysed, in a bid to identify a lead line and a suitable back-up for manufacture. The volume of cell lines produced for each project (a new therapeutic protein to be manufactured) extends project timelines significantly, due to the lack of excess capacity within the laboratory to enable cell line development for concurrent projects. Therefore, this represents a significant bottle neck in manufacturing timelines, which delays therapeutics from reaching the market faster.

Understanding CHO biology to an unprecedented level should develop a greater depth of knowledge surround pathways that lead to production instability, which may be exploited by researchers. Industry and research laboratories are now starting to delve into CHO biology, to

find potential bioengineering targets that ameliorate some of CHO unwanted characteristics in bioproduction (Fan et al., 2012; Hefzi et al., 2016; Kim et al., 2004; Lee et al., 2015a; MacDonald et al., 2018; Matasci et al., 2011; Ritter et al., 2017; Singh et al., 2018; Xu et al., 2011; Yamane-Ohnuki et al., 2004).

Here, I characterise CHO host and therapeutic protein producing cell lines at the genomic and transcriptomic level, to better understand CHO cells response to the production environment stress; specifically focussing on gross chromosomal changes. There are few publications outlining chromosomal instability in CHO cells (Vcelar et al., 2018a; Vcelar et al., 2018b) but to my knowledge, it has not been explored specifically in an industry relevant panel of cell lines, producing multiple therapeutics. I also explore how large blocks of interstitial telomere repeats, observed in CHO cell lines by ourselves and others (Krutilina et al., 2001; Smilenov et al., 1998), may increase DNA damage at these sites during therapeutic protein production. DNA damage response (DDR) key players will also be assessed within the producer environment, to investigate differences in their expression, between productionally stable and unstable cell lines. If telomere specific damage is observed, their role in CHO cell line karyotypic heterogeneity will be assessed. Gross chromosomal changes have been shown, in CHO and other cell models, to occur when a cell is under stress (Armstrong et al., 2000; Yusufi et al., 2017; Zhu et al., 2018). Identification of a link between telomere specific DNA damage, gross chromosomal mutations and production instability, would help to elucidate a potential cause for this under-characterised phenotype. Additionally, the expression of these key players within this pathway will be assessed by real time polymerase chain reaction (qRT-PCR), to identify any differential expression profiles between productionally stable and unstable cell lines. In the future, such targets may be investigated for biotherapeutic engineering strategies to create a more robust host.

Recombinant Proteins as Therapeutics (1.1)

Monoclonal antibody development for novel therapeutic targets (1.1.1)

The majority of marketed therapeutics that have been produced by CHO cell lines are recombinant monoclonal antibodies that are of the IgG1 class (Walsh, 2018). Monoclonal antibodies (mAbs) are monovalent and bind to the same epitope. They were first produced in mice hybridomas (Kohler and Milstein, 1975), through the immunisation of mice with a specific druggable antigen. The activated B-lymphocytes were harvested from the spleen of the mouse. B-lymphocytes are then fused (either by chemical or virus induced methods) with an immortalised myeloma cell line with a hypoxanthine-guanine-phosphoribosyl transferase (HGPRT) knock-out. HGPRT enables synthesis of de novo nucleotides and is used as a selection system, as successfully fused B-lymphocyte and myeloma cells should only survive in a selective media containing aminopterin (Kohler and Milstein, 1975). Aminopterin blocks de novo DNA synthesis by inhibiting dihydrofolate reductase, whilst HGPRT provided by the B-lymphocytes provides a salvage pathway to promote cellular division (Kohler and Milstein, 1975).

The resulting cell culture contains cells producing polyclonal antibodies. Single cell cloning is performed, to isolate and assess individual antibodies that were generated to the immunized antigen (Li et al., 2010). The hybridoma technique is a long process, which requires 6-9 months before the identification of a specific mAb that has a high binding affinity to the antigen epitope is identified. Antibodies generated in this way also require a development step of humanisation if the antibody generated elicits an immune response when administered. This has been attributed to mouse antibody sequences, which can affect

overall therapeutic response, as these antibodies are subsequently neutralised (Harding et al., 2010).

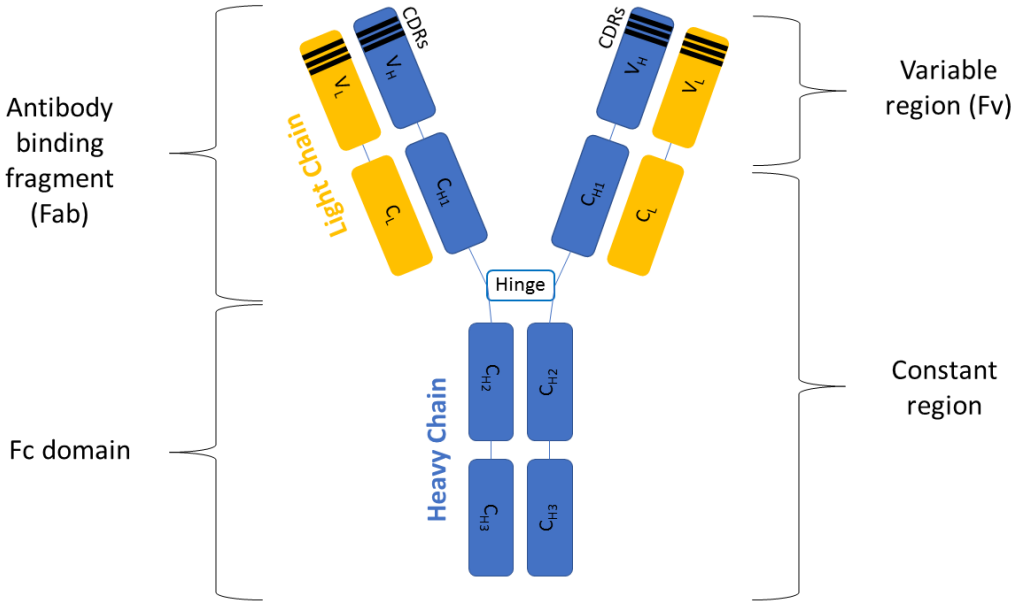


Figure 1.1. Schematic of a monoclonal antibody. Antibody domain regions of heavy and light chains, showing variable and constant regions and the antibody binding fragment (Fab) (adapted from (Schroeder and Cavacini, 2010)).

As the hybridoma technique has long developability timelines and minimal points of action, where antibody developability can be influenced by a researcher, bacteriophage display was repurposed for antibody discovery (Smith, 1985). Originally this method was produced to investigate the enrichment of fusion proteins encoded by cDNA sequences in phage particles (Smith, 1985). Phages display antibodies can be generated from antibody libraries that are either naïve, such as the human combinatorial antibody (HuCal) library (Rothe et al., 2008), or generated from DNA extracted from B lymphocytes of the mouse spleen from an immunised mouse. Non-human derived antibodies require "chimerization" that involves substituting mouse constant regions of the antibody with homologous human sequences. Additionally, "humanization" can also be performed where mouse sequences are reduced only to the complementary determining regions (CDRs, figure 1.1), minimising non-human derived mAb regions (Clementi et al., 2012).

Synthetic libraries utilise random nucleotide sequences within selected CDRs and multiple framework regions in order to replicate antibody diversity of a natural repertoire (Conrad and Scheller, 2005). Phage display of such libraries enables researchers to biopan antibodies that are either fully human or humanised during the construction of the library, reducing immune reactions to the antibodies. Libraries have successfully been constructed from human bone marrow, peripheral lymphoid tissue and blood lymphocytes (Burton et al., 1991; Chang and Siegel, 2001; Clark et al., 1997; Mao et al., 1999; Williamson et al., 1993; Zhang et al., 2004a). Sequences from these libraries are then fused to a sequence encoding the gene III phage coat protein, allowing display of antibody fragments (Chan et al., 2011). Molecules generated from such libraries are usually monovalent Fab fragments or single chain Fv fragments (scFvs). Fabs contain the heavy chain VH and CH1 domains and the whole light chain, whereas scFvs consist of both variable regions of the heavy and light chains fused to a flexible aminoacidic linker (Clementi et al., 2012).

Phages that now express antibody fragments on their coat surface undergo a series of binding and washing protocols to assess binding affinity of the antibody to a desired antigen or epitope that is usually coated onto microtiter plates. Protocols are designed to identify the antibodies that have the highest binding affinity by, for example, using changes in temperature and/or pH to identify antibodies that are presumed to be able to elicit the strongest therapeutic response in vitro/in vivo (Clementi et al., 2012).

DNA of the selected phage is then amplified and transformed into bacteria to obtain a large enough quantity of DNA for further antibody development. Optimisation of a desired antibody sequence can be achieved through PCR site-specific mutation of complementary determining regions (CDRs), error-prone DNA amplification in E.coli or antibody gene amplification using error-prone PCR, which allows specific mutation of genes of interest (Irving et al., 1996; Low et al., 1996; Thie et al., 2009). Once a final sequence has been determined, which shows desired specificity and efficacy, the DNA is transfected into a mammalian cell line factory (e.g. Chinese Hamster Ovary cells (CHO)) for the production of the antibody with the relevant human-like glycosylation profiles (figure 1.2). Cell line development subsequently occurs to identify a cell line that is deemed manufacturable (figure 1.3).

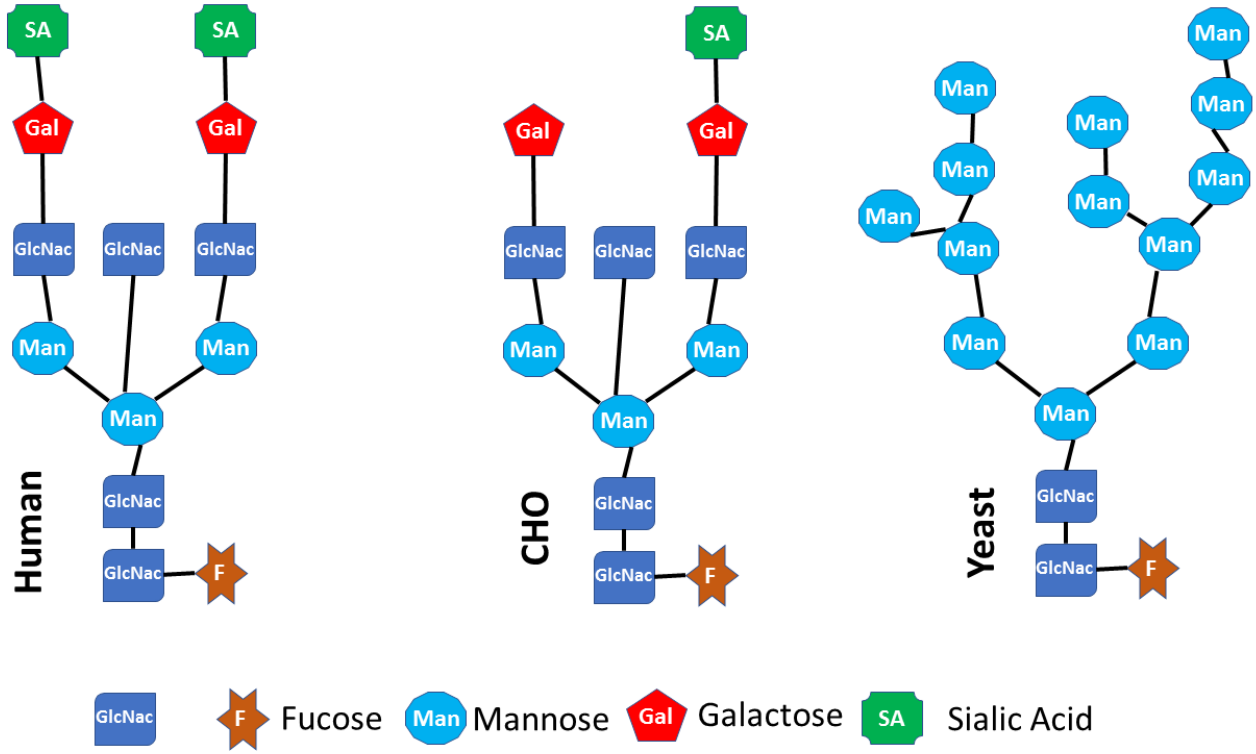


Figure 1.2. Comparison of glycosylation patterns between human, CHO and yeast cells. Human glycosylation patterns are closely mimicked by CHO, except for a missing sialic acid residue. Yeast cells produce a completely diverse glycosylation pattern, dominated by mannose residues, which can lead to immunogenicity of the therapeutic (adapted from (Butler, 2005)).

Clinical impact of monoclonal antibodies (1.1.2)

The discovery of monoclonal antibodies (mAbs) was awarded a Nobel prize in 1985 (Milstein, 1985) having been discovered 10 years prior (Kohler and Milstein, 1975). Although there was sentiment for mAbs to be utilised in therapeutic areas such as oncology, which would allow for highly specific targeting of heterogenous molecular markers on cancer cells, early clinical results exploring mAb-based therapeutics were disappointing (Vaickus and Foon, 1991) and even as early as 25 years ago, some experts questioned the validity of antibody-based therapeutics (Weiner, 2015). Disappointing results were later found to be explained by the immunogenic reaction to the host species from which the mAb was generated. The first cancer mAbs that were evaluated in the clinics were murine that provided intrigue in terms of a potential therapeutic. Ultimately murine based mAbs could not provide sufficient clinical utility due to the immune response generated by the patient, rapid clearance and suboptimal interaction with the cancer cell target (Meeker et al., 1985).

The next iteration of therapeutic mAbs were born out of the development of genetic modification techniques that paved the way for humanised mAbs (chimeric mouse-human). Through editing of the mAb sequence to include human sequences, researchers were able to generate chimeric mAbs, which in most ways mimicked the pharmacodynamics of naturally occurring human IgG, whilst reducing mouse regions within the antibody (LoBuglio et al., 1989; Maloney et al., 1997). Additionally, enhancements in antibody screening technologies, such as human immunoglobulin producing transgenic mice, directed affinity maturation methods and human antibody phage libraries, culminated in improvements to efficiency of development, specificity and reactivity of mAbs to their selected antigens (de Haard et al., 1999; Hoogenboom and

Chames, 2000; Knappik et al., 2000; Ludwig et al., 2003; Mendez et al., 1997; Schier et al., 1996).

These novel mAbs were shown to have similar half-lives of natural IgG1, whilst eliciting a better adaptive immune response (Horowitz et al., 1996) and essentially industrialised the target discovery process, resulting in consistent isolation of high-affinity fully human or humanised monoclonal antibodies.

Due to these improvements, it was clinically feasible to administer patients (weekly or monthly) on a schedule that was more practical for patients, as mAbs were shown to be still circulating at therapeutic levels for weeks/months at a time. Distribution of mAbs were observed in the intravascular and extravascular compartments, present in solid tumour mass for prolonged periods and shown to interact with malignant cells, stromal cells, extracellular matrix, vasculature and benign lymphocytes (Weiner, 2015). Although mAbs have thus far been utilised to target a range of therapeutic areas, to better illustrate how mAbs are used to target a therapeutic area, this section will provide examples for mAbs that target cellular pathways in the oncology therapeutic space.

Antigens that are either expressed or presented on a cancer cells surface have been targeted by a variety of mAb therapeutics (Scott et al., 2012). Certain characteristics of the antigen are desired when investigating novel targets. First and foremost, the consistency, density and specificity of antigen expression is paramount in order to elicit an efficacious response, whilst protecting physiologically vital benign cells (Weiner, 2015). The solubility of the antigen is also an important consideration as this may represent a stumbling block to the specific delivery of the mAb upon administration.

Once a specific antigen target has been identified, dependent on the structure and function of the mAb, numerous pathways of direct/indirect apoptosis of the malignant cell can be activated. Specificity to the antigen is enabled through the variable Fab arms of the molecule, whilst the hinge region of the Fc domain allows for engagement of the humoral immune response to direct complement mediated cytotoxicity (CMC) or antibody dependent cellular cytotoxicity (ADCC) (Clynes et al., 1998; Taylor, 2006; Wang et al., 2018b).

Strength of the elicited response in an in vitro environment should be interpreted with caution as the effect can vary greatly in vivo (Weiner, 2010). This has been suggested to be linked with inadequate in vitro models that try to replicate human pathology, but equally the diverse complexity of mAb interactions on downstream pathways is difficult to fully understand from in vitro assays, making confident extrapolations of results to the clinic, difficult (Weiner, 2015).

CMC targeting antibodies utilise serum as a source of complement and therefore are usually designed to target hematologic malignancies where they are exposed to complement in circulation (Pawluczkowycz et al., 2009). The ability of a mAb to induce CMC is dependent on antigen concentration, orientation of the membrane and antigen structure. It has been shown to be difficult to identify mAbs that can activate CMC in vivo as it has been elucidated that some but not all mAbs can activate CMC, even a mAb with the same isotype and antigen target can have different efficacies in CMC activation (Teeling et al., 2006; Wang and Weiner, 2008).

Further difficulties in CMC activation can also occur through a process termed trogocytosis ("shaving") (Beum et al., 2011). Trogocytosis occurs when a malignant cell has an antibody-complement complex bound to the membrane and circulates through the liver or spleen. The shearing process removes both the antibody-complement complex along with the antigen itself, resulting in cells becoming therapy resistant, albeit temporarily (Beum et al., 2011; Wang et al., 2009). Through Fcγ-Receptor (Fcγ-R) binding, ADCC can be induced through NK cells, granulocytes, monocytes and macrophages via immunoreceptor tyrosine-based activation motif (ITAM) signalling pathways (Beers and Glennie, 2013; Dall'Ozzo et al., 2004; Hernandez-Ilizaliturri et al., 2003; Lefebvre et al., 2006).

mAb coated target cells can induce production and release of cytokines by immune effector cells, leading to the activation of tumour specific adaptive immune response in the surrounding tumour microenvironment (Bowles and Weiner, 2005; Hoffmeyer et al., 1997). ADCC has the potential to elicit a longer lasting response as mAb induced cancer cell lysis can lead to enhanced uptake of targeted antigen and promote cross-presentation by Dendritic cells (Harbers et al., 2007). However, in some cases it has been noted that activating CMC can interfere with the immune system's ability to elicit an ADCC response, by blocking the interaction between mAb and FcγRs (Rogers et al., 2014; Wang et al., 2008).

Although specific targeting of cancer specific antigens holds promise, there has been relatively better success in developing therapeutics that induces an altered host response in the tumour microenvironment. A prime example is Avastin (VEGFA target, Bevacizumab), which reduces intratumoural blood vessel growth by inhibiting angiogenesis by reducing VEGF in the microtumour environment (Ferrara et al., 2004).

Avastin is administered in combination with chemotherapy and starves the tumour of oxygen and nutrients by limiting its blood supply. Upon cell death, VEGF blockade inhibits a tumours ability to re-establish its blood supply after cytotoxic treatment. However, tumours are often observed to escape angiogenesis inhibitors, leading to disease progression and therefore a lot of work has been performed to target pathways of escape mechanisms, with varying success (Sennino and McDonald, 2012; Weiner, 2015).

Another successful target is cytotoxic T-lymphocyte protein-4 (CTLA-4), a T cell checkpoint blockade inhibitor (Peggs et al., 2006). CTLA-4 is a receptor expressed by activated T cells and results in a downregulated T-cell response. In normal homeostasis this is required for avoiding autoimmunity but unwanted when trying to generate a sustained specific T cell response to cancerous cells (Page et al., 2014; Pardoll, 2012; Peggs et al., 2006). Combinatorial therapies that include a specific mAb that elicits an ADCC response to a tumour and CTLA-4 / programmed death ligand 1 (PD-L1) (Brahmer et al., 2012) may represent a viable option when looking to enhance ADCC action in vivo. James Allison, a pioneer for the work on CTLA-4 as a target for cancer immunotherapy therapeutics, was recently awarded a Nobel prize award for his contribution to medicine (Nobel Media, 2019).

Bispecific antibodies, containing two antigen binding regions, is being explored as a method to streamline manufacturing production of mAbs whilst providing a combinatorial therapy option. Most bispecifics involve constructs that bind to an antigen of a cancer cell and CD3 on T cells (Weiner, 2015). Initial studies with intact Fc regions were observed to activate T-cells non-specifically resulting in high toxicity due to auto-immunity. With the improvement of technology and better understanding of bispecific modalities, efficacy has increased with a concomitant reduction in toxicity, leading to the approval of blinatumomab by the FDA (Nagorsen et al., 2009; Segal et al., 1999; Topp et al., 2011).

With the increase in immunological understanding for different therapeutic areas, there has been a big shift into utilising mAbs as a therapeutic agent, compared to the historically common small molecule (Wurm, 2004). With the increasing identification of novel mAb based therapeutics, which are primarily manufactured by Chinese hamster ovary cells (CHO), CHO production platforms have been extensively optimised at the process level (e.g. media optimisation) to drive a greater overall therapeutic mAb titre that can be obtained from a single manufacturing run. The following section introduces how CHO was adopted as the primary biological production factory for mAb therapeutics.

Chinese Hamster Ovary (CHO) Cell Line Utilisation as a Therapeutic Protein Platform (1.2)

CHO cell lines as a basic research model (1.2.1)

Before their use in biopharmaceuticals, Chinese Hamsters (Cricetulus Griseus) were initially introduced into the laboratory for typing pneumococci in the early 1920's. Their popularity grew amongst researchers as they represented a valuable tool in epidemiological research into Leishmania, as they were known carriers of the parasite. Further interest ensued, as a model was gained through spontaneous hereditary diseases obtained by inbreeding, leading to research interest in hamster genetics, which was favourable due to their low chromosome number (2n=22), making them useful models for radiation cytogenetics and tissue culture studies (Yerganian, 1972).

In 1956, CHO cells were first identified within a set of immortalised fibroblasts, which were isolated from Chinese Hamster ovary tissue derived from a partly inbred Chinese Hamster by Theodore Puck and colleagues. These cells were noted as resilient with fast generation times and karyotype heterogeneity amongst their chromosomal populations garnered interest for studying chromosomal abnormalities (Puck et al., 1958). As CHO are highly adaptive and easy to maintain they have been utilised in many basic biological fields, from G-protein coupled receptors and their associated signalling pathways to DNA damage repair through radiation research (Batista et al., 2006; Figler et al., 2003; Fiore et al., 2002; Hornigold et al., 2003; Schulte and Fredholm, 2003).

CHO initial utilisation as a therapeutic protein producing platform (1.2.2)

Before the efforts of scientists at Genentech in the mid-1980's, bacterial expression systems were widely used to produce recombinant therapeutic proteins (RTPs). At the time, CHO cell culture was predominantly adherent resulting in low cell growth capacity, small titres of RTP and also reliant on increased man power for suitable cell culture maintenance. Moreover, cost benefit analysis comparisons between bacterial and CHO expression systems would have favoured bacterial systems, but the potential of post-translation modified RTPs would have been recognised. The consequent adaptation of adherent cell culture to suspension by Genentech culminated in the production of kilogram quantities, for the first time, of recombinant tissue plasminogen activator (Grossbard, 1987).

Using 10,000 litre bioreactors Genentech purified clinically relevant RTP quantities for the first time. This advancement resulted in an increase of therapeutic product, a decrease of man power required to maintain CHO cultures and also a reduction in consumables, culminating in a commercially viable platform (Wurm, 2004). Such a process would have been highly sort after by competitors, leading to the wide adoption of CHO cell suspension technique. With the relatively recent advancements in titre augmentation, CHO expression systems are now widely used to produce mAbs to target disease pathways. For example, an early typical process output of CHO manufacture would result in a maximal density of 2x10^6 cells/ml, high viability of cells for 7 days, max titre of 50mg/L and specific productivity of ~10pg/cell/day. At the time of publication Wurm (Wurm, 2004) noted a dramatic improvement to reach cell densities of 10x10^6/ml, higher viabilities maintained for up to 3 weeks, max titre of ~4.7g/L and a specific productivity of up to ~90pg/cell/day. 15 years later, with a major focus on process optimisation, most CHO suspension culture systems have

surpassed microbial expression systems, delivering 3-10g/L (in-house data). With reports of cell lines reaching titre levels of 13g/L in fed-batch production (Kelley, 2009).

A major advancement in these platform improvements was achieved through the discovery of selection systems. Initial mutational experiments, giving rise to a series of auxotroph's, were initially performed for fundamental research. However, early discoveries could be utilised as selection pressure systems and were readily transferable to industrial bioreactors (Goldfarb et al., 1977; Kao and Puck, 1968; Taylor et al., 1977). In addition to auxotroph's, mutants defective in transcriptional and translational machineries for certain amino acids were also developed, providing further selection mechanisms for exogenously delivered DNA (Adair and Carver, 1979; Chan et al., 1972; Waye and Stanners, 1979).

Of these selection systems, dihydrofolate reductase (DHFR) was one of the first systems to be widely adopted by research laboratories and industry. CHOK1 cell lines were selectively mutated using deoxuridine ([^{3}H]dUrd), which incorporates into DNA and is toxic by radioactive decay. For [^{3}H]dUrd to be incorporated, it must be converted into thymidylate, a product of a series of enzymatic reactions, of which, DHFR conversion of folic acid (in the media) to tetrahydrofolic acid (a cofactor that is required for one-carbon transfers in various biosynthetic reactions) is a rate limiting step (Urlaub and Chasin, 1980). Fully DHFR deficient mutants require glycine, purine and thymidine in the culture media, for growth. However, the initial study produced a mutant that had half of the DHFR enzyme activity that could be further reduced with the addition of methotrexate (MTX, amethopterin), an analogue of tetrahydrofolate (Alt et al., 1992).

MTX resistant clones have been isolated and their resistance has been identified to be due to an increase in DHFR gene amplification (Alt et al., 1992; Urlaub and Chasin, 1980). When selecting a CHO expression

system, such traits were desirable as in theory the inclusion of a DHFR gene in an integrating plasmid, combined with MTX in the media inhibiting native DHFR protein, should lead to a dramatic increase in the expression of the gene of interest (GOI). As the cell naturally drives expression of exogenous DHFR expression (and therefore the desired GOI) to compensate for the lack of endogenous DHFR (due to MTX mediated inhibition), only cells that increase DHFR expression (along with GOI expression) can prosper in a media environment that contains a strong selective pressure. The DHFR expression system is one of the first systems to be widely adopted by molecular biologists and the biopharmaceutical industry for vector-mediated gene integration and recombinant therapeutic protein production in CHO cells (Kingston et al., 2002). Other selection systems were sought after by companies, in order to provide an alternative to the DHFR system, potentially providing an opportunity to protect a company's CHO platform.

In a typical cell line selection protocol, used in research laboratories (Dahodwala and Lee, 2019), limiting dilution seeding is performed to obtain single cell clones. Clones are screened in 96 well plates for their titre and cell lines are triaged. Remaining cell lines are maintained and clone selection is performed to identify a cell line that has a desired titre production and product quality attributes, such as glycosylation profiles (Lingg et al., 2012). Product quality attributes are monitored in industry by high-throughput assays to ensure aspects of the therapeutic, such as molecular weight and fragmentation, does not fluctuate over the manufacturing period.

Another system, used by GSK and others is the glutamine synthetase (GS) knock out CHOK1a host cell line (Fan et al., 2012; Noh et al., 2018), based on a popular GS-system using Methionine sulphoximine (MSX) as a selection agent (Bebbington et al., 1992; Brown et al., 1992; Cockett et al., 1990). ι-Glutamate; ammonia ligase (Glutamine synthetase) is a universal housekeeping enzyme that was

identified in CHO by MSX concentration experiments in a clone that was resistant to 5mM MSX (Sanders and Wilson, 1984). GS is the only enzyme that can synthesise glutamine *de novo* and has been identified in human cells to have tissue specific functions. For example, GS has been shown to regulate toxic levels on ammonia within the brain through the coupling of ammonia to glutamate to form harmless glutamine (Cooper, 2012). There have been various publications implicating aberrant GS activity to Parkinson's disease (Carlsson and Carlsson, 1990), Huntington's chorea (Young et al., 1988) and Alzheimer's disease (Butterfield et al., 2006).

CHO GS is comprised of 6 exons with exon 5 obtaining the critical sequence for GS activity (Hayward et al., 1986; Krajewski et al., 2008). Glutamine has been identified as one of the essential amino acids for CHO cell survival when cultured in vitro (Neermann and Wagner, 1996) and multiple companies have targeted the GS gene as a candidate to knock out in a bid to create a selection system based on depletion of glutamine in the media. However, the GS selection system still retains a basal level of endogenous GS expression, reducing the stringency of this selection method. This has been noted through the maintenance of low producing cells in bulk transfected cultures, maintained in media lacking glutamine (Fan et al., 2012).

Thus, MSX is added to the media as it acts as an inhibitor of GS by binding to the glutamate site of the enzyme (Eisenberg et al., 2000; Shapiro, 1970) and irreversible inhibition of GS can occur through phosphorylation of MSX in the glutamate binding site (Liaw and Eisenberg, 1994; Shapiro, 1970).

CHO transient transfection systems have been shown to produce g/L quantities within a two week period of transient therapeutic protein production (Codamo et al., 2011), making this method of cell line development popular in research laboratories. Transient systems have been utilised as a rapid method to investigate some aspects of protein production within CHO, such a product quality assessments and

generating product for pre-clinical evaluation (Hunter et al., 2019; Rajendra et al., 2015). Although transient systems represent a quick and relatively low-cost platform to generate mAbs, it could be argued that using such models to interrogate CHO biology, in the context of production stability in a manufacturing environment is insufficient. This is primarily due to industry using constitutive expression systems, where the plasmid integrates into the genome, allowing CHO platforms to maintain high therapeutic protein production during manufacture (70-100 generations) (Wurm, 2004). Transient systems do not integrate into the genome and only produce therapeutic proteins for up to two weeks. This may not replicate the biological stresses that a constitutive system may have to endure. Therefore, any discoveries made in a transient system would have to be tested in a constitutively expressing CHO cell line, to understand whether such discoveries are relevant in an industry setting.

With the advancements in CHO expression systems outlined in this section, there is a wealth of methods that research, and industry laboratories can utilise to produce therapeutic proteins in large quantities. However, a caveat to not having a universal system that every laboratory uses, makes comparisons of discoveries between different platforms and cell lines difficult. This difficulty is compounded further when considering CHO cell heterogeneity, which is explained in depth in section 1.3. Despite these caveats, CHO production systems are widely used and have produced ~80% of the so-called “blockbuster” therapeutics (Walsh, 2018). The following section will address how CHO cell lines are utilised as a mAb expression system, primarily within the pharmaceutical industry, to provide a better context to the work in this thesis as it was performed within a pharmaceutical laboratory setting.

CHO cell lines; the therapeutic protein production factory of choice for the pharmaceutical industry (1.2.3)

The vast array of potential mAb based therapeutics stems from their ability to influence the complexity of our immune system. In order for mAbs to be a viable drug for market, they must be produced at gram per litre scales to meet the demand of patients within the therapeutic area. Human-like glycosylation profiles are required, as they heavily influence mAb binding to Fcγ receptor (FcγR) and therefore their ability to elicit a therapeutic action through innate and adaptive response (Bibeau et al., 2009; Moore et al., 2010; Seidel et al., 2013; Wang et al., 2008; Wang et al., 2018b).

Since their identification, CHO cells have been the production system of choice for recombinant therapeutic proteins (RTP). They were adopted due to their relative safety profile, as they do not propagate most human pathogens (Bandaranayake and Almo, 2014), and the relative ease to which exogenous DNA can be delivered and expressed. Through the delivery of a transgene, a multitude of therapeutic proteins can be designed and expressed with the concomitant mammalian post translational modifications (a caveat of bacterial expression systems), to target and ameliorate disease pathways (Butler, 2005; Fliedl et al., 2015).

Although therapeutics produced in other cell lines have gained regulatory approval, such as mouse myeloma (NS0) and human embryo kidney (HEK-293), 80% of mammalian cell culture processes for biopharmaceutical production use CHO suspension cells (Walsh, 2018; Wurm, 2004). CHO cells are preferred when expressing RTPs due to the conservation of mammalian post-translational modifications, which are crucial for mAb-FcγR interactions. Improper post-translational modifications can result in unwanted effects such as altered protein stability, lowered affinity towards a targeted antigen, aberrant clearance rate and immunogenicity profiles. Additionally, the strong track record of

CHO as a biologic factory with regulators allows for a smoother approval process (Walsh, 2018). This section will outline an industry standard cell line development used by biopharmaceutical companies for the production of biologics (e.g. mAbs).

An industrial biopharmaceutical pipeline (figure 1.3) first starts at the target discovery stage where a therapeutic target of interest is identified through various strategies utilising pre-defined cellular and/or animal models that reflect the disease phenotype. Ideally, target discovery would be performed on clinical diseased and healthy tissue samples, but this is largely impractical and, in some cases, unethical. Animal disease models are derived from information from clinical trials and try to replicate the human disease phenotype. Unfortunately, the majority often suffer from a multitude of caveats that lead to the models being poor predictors of human disease (Horrobin, 2003; Lindsay, 2003; Venkitaraman, 2003).

Cellular models share similar caveats to animal models, but they also lack in vivo complexity, as they are usually cultured on tissue culture plastic in a 2D environment. More recently there has been an accelerated effort to recapitulate in vivo complexity utilising human induced pluripotent stem cells (hiPSCs) and organoid technologies. Such technology holds the promise of replicating the 3D complex environment of native tissue, including multiple cell types, which obtain the same disease mutation (Giacomelli et al., 2017; Maffioletti et al., 2018).

Although these models hold promise, they require further maturation before they truly can replicate human in vivo complexity. An additional factor to consider is the contribution of genetic and environmental factor interplay and their impact on disease phenotypes. In spite of the aforementioned caveats, it is widely accepted that most diseases obtain susceptibility genes that could ameliorate the disease phenotype when targeted with a therapeutic (Venkitaraman, 2003),

therefore the majority of therapeutic targets are developed against a single antigen.

In-silico modelling is quickly becoming a high-throughput assay to perform when identifying novel potential molecules that can interact with a desired target epitope (Pinero et al., 2018). Molecular approaches, such as genomics, proteomics and gene association studies generate large amounts of informative data surrounding target interaction with identified compounds, assessing their ability to ameliorate the disease phenotype. For biopharmaceuticals, once a target has been identified, a pre-identified panel of antibodies, selected on the basis of their in-silico modelling scores, are tested against the target to identify the molecule that has the best specificity, efficacy, pharmacokinetic and clearance profiles. Once a desirable variant has been identified, the molecular sequence is progressed for cell line development (Lindsay, 2003; Wurm, 2004).

Following selection, surviving cells are single cell sorted and image confirmed. Clones are then expanded, and their peak expression titre is assessed in media containing many components to drive production, usually across a 4-6-month (70-150 generations) period to assess production stability for manufacture (BioPhorum Survey, 2018). To date, productivity augmentation has largely been achieved through media characterisation and optimisation (Wurm, 2004). The majority of pharmaceutical companies have invested in their own media formulations, whilst there are commercially available media (e.g. CD Opti-CHO, Gibco), which are heavily optimised for each stage of the cell line development process (Mather, 1998) and are used by both research laboratories and biopharmaceutical companies. Usually, researchers use a formulation for sub cultivation periods and use a richer media when required for production runs, which last for two weeks without changing media (Wurm, 2004). Peak titre measurements are obtained at set

timepoints during this two-week period and are periodically assessed across the whole stability assessment (70-100 generations).

Usually, in a robust well-monitored process, a stable cell line can be defined by a retention in 70% or more of volumetric productivity titre over a stability period (typically between 70-100 generations) (Dahodwala and Lee, 2019). Therefore, cell lines are considered stable if they can maintain their RTP production within a threshold of +/- 30% of their original peak titre across the stability period, but these parameters vary between laboratories and companies (BioPhorum Survey, 2018). Additionally, the FDA state that 'no clinically meaningful differences' from the reference product (e.g. structure, function and purity) should occur over the manufacturing process (FDA Orange Book https://www.fda.gov/drugs/ informationondrugs/ucm129662.htm last accessed 28 September 2018). Such parameters identified in antibody product quality, such as glycan analysis, aggregates and fragmentation are utilised to select the best cell line for Good Manufacturing Practice (GMP) manufacture (Wurm, 2004).

Within the field, cell lines are divided into different categories, dependent on the characteristics of their productivity maintenance. Ideal cell lines should maintain therapeutic production within 15% of its initial titre measurement at the end of 100 doubling cycles. Stable cell lines can maintain their productivity within 30% of the initial titre read out after 60 doublings, often exhibiting a decline in productivity after this time point. Gradual loss indicates a cell line that shows measurable and gradual reduction in titre, over a 60-100 doubling window. Immediate loss describes cell lines that show an abrupt decline in productivity within a few populational doublings (Dahodwala and Lee, 2019). However, different laboratories and Biopharmaceutical companies tend to implement their own variation of the above to categorise the stability of their cell lines. This is one of the caveats within the field that makes replication of findings difficult.

Roughly 8-63% of clones generated are considered unstable (Tharmalingam et al., 2018), with GSK in-house data placing the average closer to 60%. Such inefficient stable cell line generation requires a greater number of cell lines to be assessed to identify a single cell line for manufacture. The increase cell line numbers lead to longer project timelines and ultimately represents a bottleneck that inhibits new therapeutics reaching the market faster; an industry-wide concern. To date, no mechanism has been elucidated that explains why cell lines cannot maintain production over prolonged periods. Therefore, the inefficient generation of stable clones demands an increase in cell lines that are subjected to a stability assessment, in order to increase the probability of finding a cell line that meets regulatory and manufacturing needs.

Currently, ~50 cell lines per RTP molecule are assessed to ensure there are multiple stable clones that can be analysed downstream for quality control panels, although this varies from research laboratory to biopharmaceutical companies. Analysing up to 50 cell lines per molecule puts a strain on resources and manpower, seriously limiting the laboratories project/molecule capacity per year. Elucidating a pathway for the production instability phenotype would allow researchers to exploit findings to develop either a novel CHO production platform or allow new assays to be developed to aid the prediction of unstable cell lines at an earlier time point, using identified production instability markers. Both scenarios would allow research and industrial laboratories to increase their analysis capacity and manufacturing capacity, without investing in equipment or increased employee power. Although there is no consensus as to why CHO cell lines struggle to maintain therapeutic protein production, the following section will outline recent publications that have tried to identify a root cause to this phenotype.

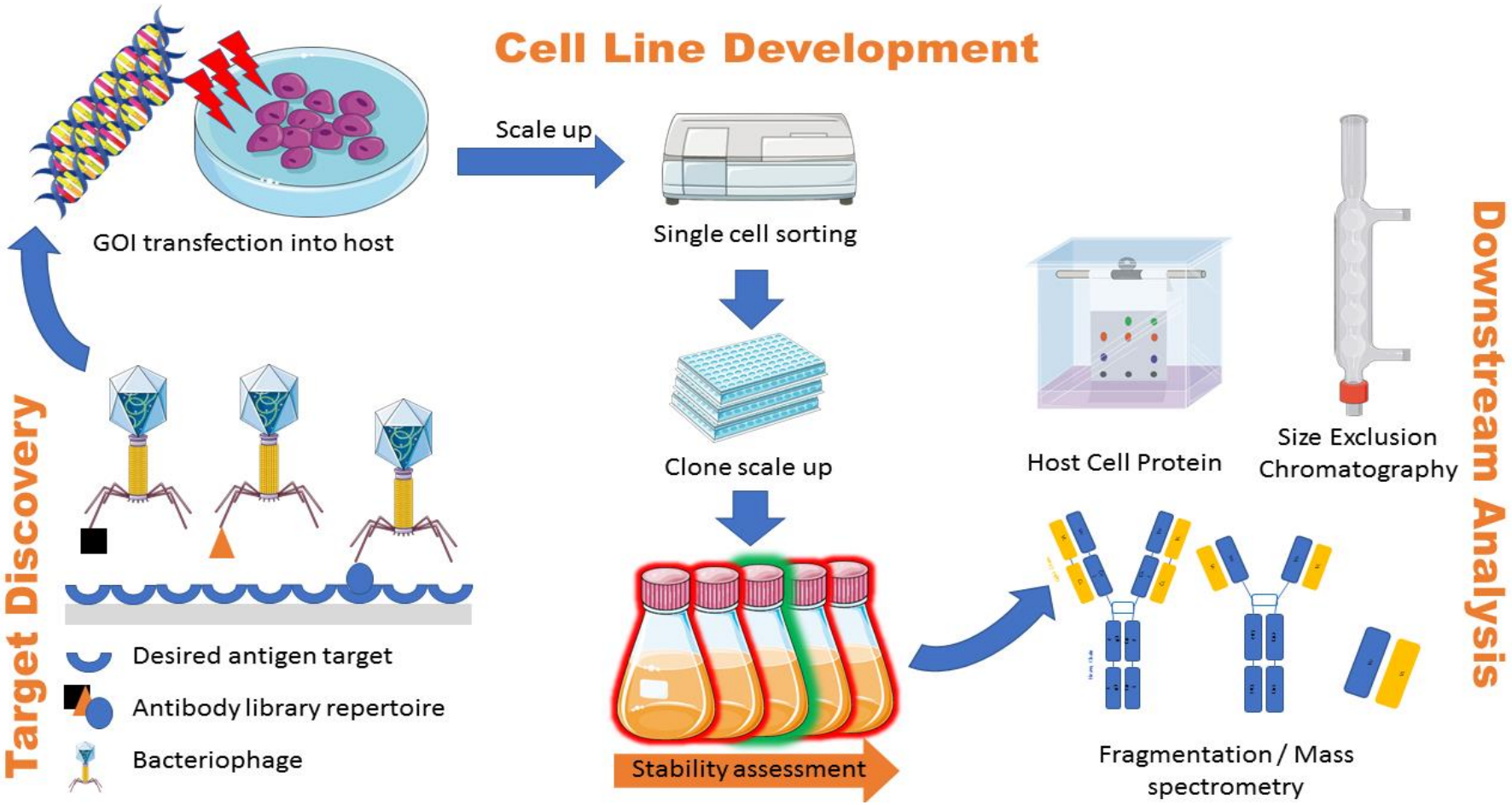


Figure 1.3. Industry biopharmaceutical workflow. Target/antibody discovery performed by phage display / hybridoma techniques. Selected antibodies are progressed to cell line development where a manufacturable cell line is identified by meeting selection criteria based on the stability assessment and downstream analysis of the antibody.

CHO cell lines struggle to maintain consistent therapeutic protein production (1.2.4)

Over the past 30 years there have been many incremental advances (Davy et al., 2017) in biopharmaceutical manufacturing development using CHO as a primary platform, with some cell lines reaching titre levels of 13g/L in fed-batch production (Kelley, 2009). It is estimated that CHO cell line therapeutic protein production instability occurs in 8-63% of all recombinant CHO lines (Heller-Harrison, 2009 (Tharmalingam et al., 2018)). GSK has a wealth of in-house data that also shows CHO cell lines struggle to maintain therapeutic protein production within a +/- 30% threshold, placing the average of unstable cell lines closer to 60% (in-house, unpublished data). Production instability has also been observed to occur across CHO platforms (DHFR and GS systems), making this a CHO cell specific phenomenon (Dahodwala and Lee, 2019), rather than a specific platform issue.

There have been various reports of different causes of observed therapeutic protein production instability. Potential causes of CHO production instability have been observed due to loss of transgene, transcriptional activity, phenotypic/genotypic drift of cells and chromosomal rearrangement (which is covered in depth in section 1.4).

When creating a cell line for therapeutic protein production, most laboratories employ a random integration approach when transfecting a new plasmid containing a therapeutic, when establishing new cell lines for manufacture. This method has been proposed to contribute to the instability of production observed by the field. Random integration strategies can lead to different number of copies of the plasmid being integrated into different loci in alternating chromosomes (Dahodwala and Lee, 2019). Yang et al., report that high producer cells have a higher frequency of plasmid integration within large chromosomes, compared to low producing and/or unstable cell lines having integration sites in small

chromosomes (Yang et al., 2010). This may be due to the type of chromatin the plasmids have integrated into, as there have been studies showing successful integration into sites with high expression lead to increased titre with stable expression (Koduri et al., 2001; Lee et al., 2015b; Zhao et al., 2018). Conversely, DNA methylation has been shown to lead to a loss of productivity within some CHO cell lines (Yang et al., 2010), indicating that plasmid integration into chromatin that is more susceptible to methylation (Illingworth et al., 2008; Zeng et al., 2014) may lead to CHO cell line instability, as titre may dramatically decline upon chromatin methylation surrounding the plasmid integration site. Taking this into account, the process of random integration of a new therapeutic into a CHO host for each cell line development project may result in random expression and stability profiles dependent on where the plasmid has integrated into the cell line. This does not provide strict control of parameters that may be beneficial for productivity.

Plasmid copy numbers has also been shown to be stochastic between cell lines, leading to different numbers of copies being integrated per transfection, potentially affecting overall titre levels within a cell line, as observed by others (Osterlehner et al., 2011). MTX, a commonly used agent for the CHO-GS expression platform, has also been shown to increase plasmid copy numbers, dependent on concentration (Fann et al., 2000). MTX concentration was shown to influence cDNA copy number, with increasing MTX levels correlating to increased copy numbers. However, stability had an inverse correlation, where an increase in MTX resulted in a decrease in the number of unstable clones.

Cultures that had MTX removed from the media showed on average a 60% decline in cDNA copy number but with an increase in the overall number of stable clones (Fann et al., 2000). This study suggests that plasmid copy number may influence overall titre but the increases in copy number may have a detrimental effect on the cell lines stability, through unidentified pathways. Conversely, others have reported cell stress and

apoptosis markers to lead to cell line instability, with plasmid copy numbers having a negligible affect (Bailey et al., 2012). In theory, when cell lines are single cell sorted and selected based on their therapeutic protein expression, the resulting cell line derived from the single cell sort should obtain a flask of homogenous cells that produce similar titre levels. However, this has been reported not to be the case. Phenotypic drift has been observed during cell culturing, where a sub population of low producing cells often becomes a dominant population within the vessel, due to their acquired growth advantage (Dorai et al., 2012; Hammill et al., 2000).

Employing a targeted integration approach into a hotspot of expression within the CHO genome may represent a valid platform to help to reduce unstable cell line generation. As this represents a more controlled process than random integration, there is the potential to insert a defined amount of copies within the same loci (Lee et al., 2015b; Zhao et al., 2018). However, such an approach does not account for the phenotypic drift observed, which leads to titre production decline over a prolonged period of time.

The following section will outline current strategies that have been employed by researchers to reduce production stability and enhance titre production of therapeutic proteins in CHO, through biological and vector engineering efforts.

Biological engineering efforts to ameliorate production stability and enhance titre production of monoclonal antibodies (1.2.5)

There have been various attempts to increase and maintain titre production. One way to achieve this has been explored through recombinant therapeutic protein (RTP) vector development. Various groups have tried to enhance expression through a range of modifications, from investigating different promoter species that could drive enhanced expression, to utilising regulatory elements that maintain euchromatin structure surrounding the gene of interest (GOI) that potentially inhibits silencing effects (Foecking and Hofstetter, 1986; Wang et al., 2016).

To enhance titre expression there have been numerous viral heterologous promoters tested in CHO. In the few available studies that have been performed to test viral promoter strength, human cytomegalovirus promoter (hCMVp) obtained the highest production titres (Foecking and Hofstetter, 1986; Ho et al., 2015; Wang et al., 2016), compared to Simian Virus 40 early promoter (SV40E) and Rous sarcoma virus (RSV) (Foecking and Hofstetter, 1986). Before the annotated CHO reference genome, endogenous promoters were identified using constructed genomic libraries which identified relatively low strength promoters that obtained expression levels 40% compared to SV40E control (Pontiller et al., 2008; Pontiller et al., 2010). Another strategy was to identify endogenous promoters that flanked highly expressed housekeeping genes, such as elongation factor-1α (EF-1α) (Running Deer and Allison, 2004), which was shown to have higher activity in CHO cells than hCMVp and human EF-1α (Running Deer and Allison, 2004). However, a more recent publication highlighting this observation has recently been retracted (Wang et al., 2018a).

Another system which has been explored is the use of the Tet-on system (Gossen et al., 1995) for inducible production of the GOI. Theoretically, the inducible GOI allows for cells to reach their peak viable cell count (VCC) at a faster rate and potentially decreases overall stress during cell maintenance. After GOI induction, high volumetric production should ensue. However, it is unclear whether this expression system can yield the same final titres as the current constitutive process that lasts for up to 70 (+/-10) cellular generations.

Ubiquitous chromatin opening elements (UCOEs) (Antoniou et al., 2003) and scaffold/matrix attachment regions (S/MARs) (Cuvier et al., 1998) have been highlighted as potential regulatory regions of interest as they act to maintain an expressive environment around the GOI in CHO cells (Betts and Dickson, 2016; Kim et al., 2004; Williams et al., 2005). Specifically, the HNRPA2B1-CX3 (A2UCOE) was shown to increase titres and reduce silencing of hCMVp-driven, scFV-Fc fused with enhanced green fluorescent protein (EGFP) production, in CHOK1 cells (Boscolo et al., 2012; Williams et al., 2005). The use of a single chain Fv fragment in these studies, which doesn't faithfully reflect the size of a full IgG antibody (~25kDa vs ~150kDa), suggests that such findings may not be replicated using therapeutic antibodies.

Chromosomal engineering has also been explored to improve CHO platform stability. Ritter et al (Ritter et al., 2016b) through gene expression experiments comparing high and low producer clones in different mAb expressing CHOK1a cell lines, identified loss of chromosome 8 as a recurrent theme that stratified the high and low producers across 3 different molecules. It was noted that genes identified in the microarray as being up-regulated were not shared across the three molecules, highlighting the complexity of pathways that can be utilised by the cell to increase and maintain productivity (or conversely impact production stability). Using an induced evolution approach, based on cyclical treatment of the CHOK1a host with MTX, allowing the viability of

the cell lines to drop and recover, they screened clones for the ipo8 gene which is specific to an ortholog of mouse chromosome 8 (Brinkrolf et al., 2013). Clones that were negative for ipo8 were karyotyped to assess whether chromosome 8 had been lost. Cell lines that obtained a loss of chromosome 8, which was interestingly shown to harbour sub-telomeric genes, were later shown to exhibit increases in productivity and production stability (Ritter et al., 2017; Ritter et al., 2016a; Ritter et al., 2016b).

Direct genome engineering has also been performed to produce a host cell line with preferable phenotypic traits. For example, the utilisation of targeting methods such as zinc finger nucleases (Kim et al., 1996) to disrupt a1, 6-fucosyl transferase (FUT8) through homologous directed repair (HDR) in CHO-DG44 cell lines. FUT8 knock outs have been shown to produce completely defucosylated recombinant antibodies, which has been implicated in antibody-dependent cellular cytotoxicity (Yamane-Ohnuki et al., 2004). Although these have been shown to work in CHO based systems, the efficiency of ZFNs have been reported to be extremely low (~1%). Thus, identifying a successfully targeted cell line requires arduous screening of single cell clones that have successfully formed colonies after cell sorting.

Clustered regularly interspaced short palindromic repeat (CRISPR) (Jinek et al., 2012) is the newest direct targeting technology to be utilised in CHO bioengineering that has been shown to offer increased insertion/deletion (indel) frequencies of 7.6-47.3% in two CHO host cell lines, targeting multiple genetic regions (Lee et al., 2015b; Ronda et al., 2014). CRISPR utilises a guide (g)RNA to specifically target gene sequences and unlike other conventional gene target strategies, based on homologous recombination, in its native form the Cas9 nuclease induces double strand breaks (DSB) and activates DNA repair machinery.

DSBs are highly erroneous, leading to frame mutations that disrupt the gene, therefore providing a quick, inexpensive and targeted

system that can knock out genes (Gaj et al., 2013). As CRISPR suffers from off target events, due to the 20nt gRNA sequence that guides the Cas9 protein to a specific sequence through Watson-crick pairing. A Cas9 mutant was developed (D10A, Cas9n) that obtains a mutation in the RuvC domain, which restricts the cut to the targeted single strand template. Combined with off-set gRNAs (Cong et al., 2013; Ran et al., 2013a), which requires the + and - template strands to be cut simultaneously to induce a successful double strand break, has been shown to reduce off target frequency by 50-1000 fold (Dianov and Hubscher, 2013; Ran et al., 2013a). Currently there has been no CHO based CRISPR screenings published but this methodology has been widely used to identify potential drug targets in oncology.

Utilising CRISPR gRNA screens, genes can be both downregulated, like more traditional RNAi screens (Gargiulo et al., 2014; Ngo et al., 2006; Westbrook et al., 2005), or conversely have their expression upregulated with transcription start site specific guide RNAs (Cheng et al., 2013). This allows the identification of phenotypic changing pathways that are initiated through both activation or suppression of different genes. There are numerous publicly available CRISPR screening libraries, which are focused around both positive and negative genetic modifying screening methods, albeit largely for human cell types (Marceau et al., 2016; Sanjana et al., 2014; Shalem et al., 2014; Wang et al., 2015), but CHO specific CRISPR screens are now being offered by some companies.

Such a CHO based library could have wide reaching utility, allowing identification of key genes that produce a desired manufacture phenotype or characteristic. To account for redundant pathways that could potentially be targeted, co-essential screens have been performed that knock out two or more genes at a time to ensure silencing/activation of complete pathways (Han et al., 2017). However, identification of such robust genes requires well defined stratification of CHO phenotypes, which currently are largely poorly defined at a biological level.

Historically, CHO cell lines have been documented to obtain genomic instability (Puck, 1979; Smilenov et al., 1998) and there have been numerous strategies that have been investigated to either augment titre production or ameliorate production instability. In this thesis, I look into how genomic instability may affect therapeutic protein stability and try to characterise a potential causative pathway that causes an increased mutational rate, as observed in productionally unstable cell lines. To achieve cytogenetic characterisation, fluorescence in-situ hybridisation (FISH) techniques were utilised, using semi-automated and fully automated hardware and image analysis workflows, to allow for relatively fast characterisation on a large number of cell lines. The following section will outline FISH as a technique and how it has been applied both historically and currently in research laboratories.

Chromosomal Heterogeneity in CHO Cell Lines and its Potential Impact on Production Instability (1.3)

Fluorescence in-situ hybridisation (FISH) – A method utilised to interrogate cell line genomic changes at the single cell level (1.3.1)

Fluorescence in-situ hybridisation (FISH) is a molecular technique that allows the visualisation of nucleic acids within a cell. Specific identification of DNA and RNA sequences can be achieved through complementary Watson-crick base pairing using probes conjugated to fluorophores that emit fluorescence when excited by the relevant wavelength. FISH can be used in numerous ways to characterise and observe cytogenetic mutations within a cell. This section describes the development and wide adoption of FISH as a clinical and molecular technique and how its use has waned over time due to the advancements of other molecular technologies, such as the polymerase chain reaction (PCR).

Initial FISH experiments were performed in the late 1960s and utilised radioisotopes. Joseph Gall and Mary Pardue were the first scientists to describe a technique that allowed visualisation of cytological DNA using tritium-labelled ribosomal RNA (Gall and Pardue, 1969). They followed logical principals to ensure "life-like" detection of DNA molecules, such as timely fixation of cells, removal of basic proteins (which interfere with hybridisation (Bolton and Mc, 1962)), whilst retaining cytological integrity during DNA denature. Utilising these principals, they developed a technique that can be used to identify complementary sequences within a cell, allowing the visualisation of gene DNA and RNA transcripts that allow further characterisation of a cell and interrogate genes that may be involved in disease pathology.

Although radioisotope-based FISH was widely utilised and represented an advancement on current cytological techniques, there are obvious draw backs to using radioactive isotopes within the laboratory. Namely, radiation hazard of reagents, regulatory requirements for proper disposal of radioactive materials and the short shelf-life of the labelled reagents, due to radioactive decay (Shan et al., 2000). Such negative aspects of use within the laboratory did not deter researchers from utilising this technique, as it represented a significant advance in resolution and speed of nucleotide identification. However, these negative aspects also prompted researchers to seek out alternative methods of detection.

The first application of FISH, as we use this method today, was performed by Bauman et al., in 1980 (Bauman et al., 1980). This was the first publication to outline flurophore conjugation to the 3' end of a specific RNA probe. Such a technique had advantages over the radiolabelling, such as an increased spatial resolving power and the speed of which samples could be imaged; within one day (Bauman et al., 1980).

Even with this advancement, FISH techniques were limited due to the resolution of DNA that could be obtained from the technique, therefore low copy number genes were difficult to detect due to the relatively high background-noise observed upon image acquisition. A technological advancement, in the form of using amino-allyl modified bases (Langer et al., 1981), which aids attachment to fluorescent dyes to enhance a fluorescent signal, represented a significant development in the technology as it allowed the resolution of low copy genes that was previously left wanting. Increased signal was achieved through secondary labelling using florescent dyes, achieving a specific secondary amplified response, reminiscent of immunofluorescence staining techniques.

Further techniques, utilising the same rational of secondary fluorescent signal amplification, were then developed. Such as nick translation, biotinylated probes and streptavidin conjugates. Such methods were used for the detection of DNA (Manuelidis et al., 1982) and mRNA (Singer and Ward, 1982) in the 1980's. A decade later, improved labelling of synthetic single-stranded DNA probes allowed the chemical preparation of probes carrying significantly more fluorescent molecules per probe, allowing for direct detection of low copy genes (Kislauskis et al., 1993). Two significant representatives of these novel probes (oligomers) are locked nucleic acids (LNAs (Kumar et al., 1998)) and peptide nucleic acids (PNAs (Nielsen et al., 1991)). LNAs consist of monomer nucleotides associated by a 2'-4' methylene bridge. Entropic restriction of this linker leads to increased binding affinity of LNAs to complimentary nucleic acids (Kumar et al., 1998). LNA phosphodiester backbone allows for good aqueous solubility and obtain enhanced resistance to nuclease degradation (Kumar et al., 1998).

PNAs, probes that are primarily used within this thesis, are synthetic DNA analogues of which the phosphodiester backbone is replaced with repetitive units of N-(2-aminoethyl) glycine, to which purine and pyrimidine bases are attached via a methyl carbonyl linker (Nielsen et al., 1991). Although they were originally designed as a potential antisense and antigene oligomer therapeutics (Boffa et al., 1996) their resulting beneficial hybridization properties represented a technological advancement on current probe chemistries. Their neutral charge results in no electrostatic repulsion during hybridisation of PNAs to their nucleic acid sequence target, providing increased stability to PNA-DNA / PNA-RNA duplexes, compared to homo- or heteroduplex interactions of native DNA or RNA oligomers (Jensen et al., 1997). This greater stability is reflected by the increased thermal melting temperature of PNA heteroduplexes compared to their DNA and RNA counterparts (Jensen et al., 1997).

Due to the PNA polyamide backbone, PNA oligomers are not recognised by nucleases within a cell. Combining this property to their increased melting temperature and their ability to be readily conjugated to biotin or fluorophores, provides an extremely stable and robust probe to be utilised in FISH experiments (Demidov et al., 1993). Overall, PNA probes have the added advantage of tighter binding and higher specificity with a faster rate of hybridization using immobilized targets. Additional benefits of PNA chemistry also include lower background signals, enabling the use of mild washing procedures and unlimited stability of the probe mixture in storage conditions (Williams et al., 2002). The PNA-FISH technique was first developed for telomere analysis, as performed within this thesis. Telomere PNA-FISH probes was first used to label human telomeric sequences and data obtained were shown to be accurate estimates of telomere lengths (Williams et al., 2002). PNA probes have also been used for in-situ specific identification of human chromosomes in metaphase and interphase nuclei (Chen et al., 2000), indicating a wide spread use for the PNA chemistry for cytogenetic characterisations of different cell types.

With the combination of probe development and technological advancements in fluorescent microscopy imaging and its associated image-processing algorithms, researchers can now obtain super-resolution images at the sub microscopic level (Carrington et al., 1995). Popularity of the assay increased dramatically in the 1990s (figure 1.4), due to these combined advancements of probe chemistries and computational algorithms. Uses of FISH ranged from single (Bauman et al., 1980), dual (Hopman et al., 1986) and triple (Nederlof et al., 1989) colour identification of targeted chromosomes, genes, exogenous DNA or RNA sequences. Additionally, mRNA localisation mapping can be performed by assaying single mRNA transcripts and parts of RNA (Femino et al., 1998), providing, at the time, a rapid and robust method to interrogate cells cytogenetically.

Earlier studies evaluated technical variances regarding signal-to-noise ratios, spatial resolution of fluorescent signals and hybridisation and detection efficiencies of FISH tests in lymphocytes and amniocytes (Klinger et al., 1992; Ried et al., 1992). These studies led to the commercialisation of FISH probes for standardised labelling. Clinical utility was first shown by Ward et al (Ward et al., 1993), this was the first clinical study to utilise FISH for the rapid detection of chromosome aneuploidy in 4,500 patients, showing robust utility of FISH probes in large experiments. In the clinic, prenatal screens are performed using amniotic fluid using a multiplex FISH probe panel that identifies gains or losses of chromosomes X, Y, 13, 18 and 21 (Ried et al., 1992; Ward et al., 1993)). Pregnant women with single or multiple indications of advanced maternal age, abnormal ultrasound or serum screening have an increased risk (4-30%) for carrying numerical and structural chromosomal abnormalities. Utilising the multiplex panel, 84% of numerical abnormalities were detected with 16% requiring further microarray analysis (Li et al., 2011). Such publications provide evidence that FISH probe chemistries have developed robustly to characterise cytogenetic mutations within clinical trials, hospitals and in research laboratories.

The developments in FISH probe chemistries provides researchers numerous methods to locate certain targets within specific chromosomes and assess whether any cytogenic mutation has occurred during a disease pathology or due to a treatment applied to the investigated cell model. Further colouring schemes were then developed utilising colour ratios of probes within similar nucleotide sequence targets. This allows dozens of different colours to be identified, based on the colour ratio signature of the probes that have hybridised to each chromosomal region and is commonly known as multi-colour FISH (mFISH) (Nederlof et al., 1992; Nederlof et al., 1990). Whole chromosome colouring using MFISH was later performed using DNA fragments extracted from bacterial artificial clones (BACs) that contained cloned human genomic DNA

sequences with sizes ranging between 100-200 kilobases (Kb). The fragments are directly labelled by nick translation to incorporate nucleotides that are coupled with different flurophore (Morrison et al., 2002).

Dependent on the designed probes, this method can be used to interrogate specific regions or genes, creating a painting band pattern of specific chromosomes or the whole genome. Thus, providing rapid identification of chromosomes or chromosome regions that can be qualitatively and quantitatively assessed allowing chromosomal mutations, such as deletions, amplifications and translocations to be observed and tracked in a single cell manor. There is increasing evidence that using bulk cell preparations for gene studies is inadequate for heterogenous cell populations (de Sousa, 2012(Macaulay and Voet, 2014; Saliba et al., 2014; Stahlberg et al., 2013). Using bulk samples has been shown to mask genetic differences in between cell population subsets (Stahlberg et al., 2011) which may result in false positive or false negative differences when comparing across different samples. This observation may be exacerbated when using highly proliferative and highly mutagenic cell lines, such as CHO, due to the increased heterogeneity across millions of cells within a single flask. Therefore, it may be pertinent to apply single cell measures to such heterogeneity, in order to capture populational subsets within culturing flasks.

After these technological advancements had been implemented, publications containing the terms “fluorescence in-situ hybridisation” increased rapidly in the 1990s from 78 publications in 1989 to 2276 publications in 2000 (figure 1.4, Corlan, 2004), indicating a wide spread adoption in FISH as a routine cytological technique.

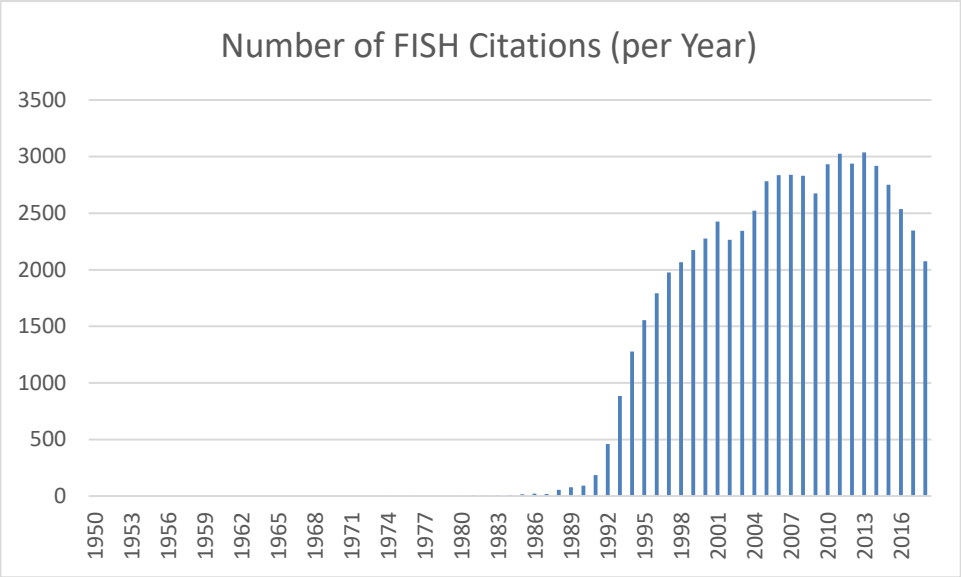


Figure 1.4. Number of citations containing the terms "fluorescence in situ hybridisation" from 1950 to 2018. Citations rapidly increase in volume upon technological advances in the 1990s indicating a wide adoption of FISH techniques for cytological experiments.

With the increase of use in FISH and fluorescence imaging techniques, quantitation and analysis of such images also evolved and advanced in their complexity. The first quantitative analysis of fluorescence images was first used for rudimentary cytogenetic tests, using charge-coupled digital cameras for fluorescent signal detection (Pinkel et al., 1986). Automated algorithms were also tested in the early stages to detect large probes in interphase and metaphase preparations, assisting pathologists for diagnostic probe detection, providing an early platform for making simple diagnostic conclusions (Piper et al., 1994). However, manual cytopathology remains the gold standard for reliable tissue analysis, using multiple scientists to confirm pathological results. There still remains an element of operator bias and/or operator-to-operator variation when qualitatively assessing image samples in this way. Reducing such bias can be achieved through more complex computerised methods, such as artificial intelligence and machine

learning. Automated ways of detecting DNA loci (Lawrence et al., 1988), sites of transcription (Lawrence et al., 1989) and multi-gene transcriptional profiling (Levsky et al., 2002) have all been performed.

Through the technological advancements outlined, FISH assays now provide the ability to robustly characterise cytological nucleic acid sequences. Coupled to the advancements of software algorithms, FISH can now be feasibly deployed as a high-throughput single cell assay, both clinically and in research laboratories. Despite the initial wide adoption of FISH as a technique, upon the development of Polymerase Chain Reaction (Mullis et al., 1986), which offered a rapid method to interrogate nucleic acid sequences from a relatively small DNA/RNA quantity, FISH use within the laboratory declined, as PCR represented a less cumbersome technique for bulk genetic and transcriptomic analyses. With the wide adoption of PCR, FISH has largely remained a technique specifically used for diagnostic tests, rather than used as an investigational tool within research laboratories. However, with the increased understanding of cell heterogeneity in CHO cell lines and other therapeutic areas (e.g. Oncology), one could argue that using PCR from a pooled cell source for DNA/RNA extraction may lead to false positive or negative data. Single cell techniques, such as FISH, provides a robust and relatively quick method to interrogate different cytogenetic populations within a CHO cell line flask, aiding the understanding of CHO biology during manufacturing pressures. The following section will explore the potential impact CHO chromosomal instability may have on therapeutic protein production instability.

CHO inherent chromosomal heterogeneity and its potential impact on production stability (1.3.2)

To date, a multitude of bioengineering techniques have been utilised in a bid to better characterise CHO biology and ameliorate some of the pitfalls associated with the expression platform. There have been varying degrees of success, with results mostly being replicated in a few cell lines. Thus, CHO cell line production instability remains an industry wide concern, as these efforts have not been replicated at the platform scale.

Although the root cause of production instability has not been elucidated. There are several studies that highlight the karyotypic heterogeneity of the CHOK1 line, indicating a highly mutational background. Work by Deavan and Peterson (Deaven and Petersen, 1973) highlighted that 24% of their cells contained a chromosome number that differed from the expected 22 (chromosome numbers ranged from 19-23) and the phenomenon still persists to this day (Auer et al., 2018; Vcelar et al., 2018a; Vcelar et al., 2018b; Yusufi et al., 2017).

In house data indicates CHO cell lines analysed within this thesis often have a modal chromosome number of 19 or 20 that is dominant within the population of cells, suggesting a loss of chromosome number at some point in the cell lines development. During a pharmaceutical CHO cell lifecycle, CHO cells undergo constant genomic modifications (both intrinsic and extrinsic), which has been shown to attribute to phenotypic differences in clonal cell lines (Derouazi et al., 2006). In addition to the natural mutational tendency of CHOK1 cell lines, the use of the MTX/MSX directed expression systems have been shown by many groups to compound mutagenesis. A high frequency of chromosomal disturbances have been observed when culturing cells in MTX/MSX; such as breakages, dicentric chromosomes and disruption to telomeric structures have been well documented in human, mouse and hamster

cell lines (Barbi et al., 1984; Goulian et al., 1980; Kaufman et al., 1983; Markkanen et al., 1982; Schimke, 1988). In CHO, MSX has been shown to amplify cDNA of exogenous plasmid in a concentration dependent manor (Fann et al., 2000). Within this study, the overall effect of MSX on the CHO cell line genome was not assessed, so there may have been other mutations that had occurred within the CHO cell line, as observed by ourselves and others (Vcelar et al., 2018a; Vcelar et al., 2018b).

Although the complex interplay between karyotype and CHO'mic cell systems has not been fully elucidated, it is reasonable to assume that gross changes in karyotype within a cell population will have a downstream effect on its overall cell system profile. As highlighted by Yusufi and colleagues (Yusufi et al., 2017), the 'omic profile of a CHO cell line rapidly changes upon transfection; 10,959 short indels, 3,313 SNPs, 21.91 Mb homozygous deletions, 228.83Mb heterozygous deletions, 2.40 Mb 2x amplifications, 15.39 Mb 4x amplifications and 11,896 genes affected due to the transfection process. To put these numbers into context, there is 2.45Gb of genomic sequence in CHOK1 ancestral cell line, with a total of 24,383 predicted genes (Xu et al., 2011), suggesting around 49% of CHOK1 genes are affected after a single transfection event.

Chromosomal instability observed in CHO production platforms may contribute to such a wide range of mutations, however it has largely been overlooked by researchers when applying more advanced technologies such as next generation and RNA sequencing. These pooled cell assays have the potential to shed light on genomic changes utilising 'omics approaches on a single cell line (Yusufi et al., 2017). Taking into account the genomic heterogeneity when analysing multiple cells in one sample, the average target expression is identified and often there can be false positive or false negative results that may not reflect the overall heterogeneity within the population of cells (Macaulay and Voet, 2014). This is especially pertinent in cell lines that are known to have heterogeneity at the chromosomal level (Vcelar et al., 2018a; Vcelar et al.,

2018b) and also exhibit phenotypic drift over prolonged cell culture (Dorai et al., 2012; Hammill et al., 2000).

Yusufi findings (Yusufi et al., 2017) are yet to be replicated in an industry setting, where ~50 cell lines are produced (per therapeutic protein) to identify a single cell line that is deemed manufacturable. Identifying causative genes between two clones derived from the same parent may provide informative data with regards to that individual cell line lineage. However, there is no evidence to suggest that repeating the transfection and reanalysing the cell lines in this way would provide the same result. It is common to find publications within the CHO field that concentrate on cell lines that have been subcloned from a single parent due to difficulties replicating results across multiple therapeutic producing cell lines, however this does not replicate industry process. There are multiple therapeutic transfections occurring in parallel and any data insights must be therapeutic/transfection agnostic, to have any meaningful impact on CHO platform therapeutic protein manufacture.

A key factor to consider for the potential irreproducibility is the gross chromosomal changes that occur during the transfection of exogenous DNA and inherent chromosomal mutations during prolonged periods of culture. Upon each transfection event there is the possibility for multiple chromosomal rearrangements that occur before a cell can establish itself as a colony, including after single cell sorting (Yusufi et al., 2017). The added heterogeneity at the transcriptomic, proteomic and metabolomic level, perhaps due to chromosomal rearrangements, potentially leads to a situation where the resulting cell line expressing a therapeutic protein, can no longer be reliably compared to its original host and across separate transfection and single cell sorting events.

Thus, poor reproducibility of results when comparing multiple cell lines, may be explained by karyotypic heterogeneity observed within the field. Such heterogeneity may mask true signals that are comparable between cell lines and such a caveat is compounded by inadequate

stratification of different therapeutic producing cell lines. For example, a cell lines stability is based on an arbitrary threshold that defines production titre stability (stability defined as <30% drop in titre across 70-150 +/-10 generations), which realistically bears no indication of the underlying biology. Having a biological marker that is causative of the production instability phenotype will allow for better stratification of cell lines for future experiments looking at potential causative pathways of production instability.

Despite these aforementioned advances in industrial relevant titre and viability increases, a major bottleneck for CHO platforms is the time taken to identify the production stability for a clone. A stability assessment can take up to 4-6 months (70-150 generations) with a substantial number of clones being assessed due to ~60% of clones being considered unstable (in-house data). Without maintenance of titre throughout the manufacturing period, process yield can have a significant impact on timelines, as manufacturing schedules are typically booked up to at least a year in advance. Therefore, unexpectedly low titres can lead to repeat manufacture runs, having an enormous impact on scheduling and a knock-on effect on product distribution (Bailey et al., 2012; Barnes et al., 2003; Betts et al., 2015).

I hypothesise that the highly mutagenic background of the CHO host is transferred to producing cell lines after transfection of the GOI and the rate of mutation has an impact on the cell lines ability to maintain therapeutic protein production over the stability assessment. To assess the mutational background, I used multicolour fluorescent in-situ hybridisation (MFISH), a technique that utilises species and chromosomal specific sequences, conjugated to different fluorophores, which allows combinations of multiple colours to produce karyotype images of 'painted' chromosomes, after hybridisation.

Painting of chromosomes reduces the subjectivity of analysing karyotypes using banding patterns when assessing karyotypic mutations. Compared to comparative genomic hybridisation (CGH), a method for analysing copy number variations relative to ploidy, MFISH has the ability to visualise large structural variants and balanced translocations. Although it suffers from caveats such as the need to culture cells before extracting metaphase spreads, laborious imaging and analysis and it is relatively low through-put in nature. However, it provides a robust method to understand the mutational landscape at a single cell populational level. MFISH has largely been applied in the clinic to better characterise human chromosome biology, such as numerical and structural variations within cancer patient samples. Other specific uses include understanding spontaneous micronucleation compared to irradiation induced mutations (Balajee et al., 2014) and identification of mutually exclusive gene amplifications in gastric cancer patients (Das et al., 2014).

To my knowledge, the use of cytogenetic techniques such as MFISH has not been widely applied to CHO cell line characterisation. Of the few publications, Vcelar et al (Vcelar et al., 2018a) utilise MFISH probes to characterise the number of structural variations within four CHO host cell lines, which are predominantly used across industry. Their data suggests the modal chromosome number distribution across the CHO host cell lines does not massively fluctuate between the hosts. CHO-S cells were the only host cell line to retain the same ploidy as the native Chinese hamster (2n=22). Their observation of the CHOK1 cell line having a modal chromosome number of 19 is consistent with our own observations. It was noted that although the cell lines shared a similar modal chromosome number, there are a range of numerical fluctuations within each CHO host karyotype that resides between 10-45 chromosomes. Although the modal karyotype appears similar between the host cell lines, when MFISH is applied to assess the structural variations, a very different picture appears.

To assess the chromosomal changes that host cell lines have undergone, a Chinese hamster lung fibroblast primary cell line was analysed over a similar time period to provide a reference point. The primary cell line completely lost the original karyotype and 40% of cells contained diverse numerical and structurally distinct karyotypes by 18 population doublings. This heterogeneity only increased with further growth of the cell line. Such propensity for genetic mutations whilst maintaining viability indicates the highly adaptable and robust nature of the Chinese hamster cell genome (Vcelar et al., 2018a).

Whilst this genetically chaotic background allowed the CHO cell line to be readily utilised as a biopharmaceutical expression platform, the nature of mass heterogeneity provides its own issues when trying to characterise and bioengineer these cell lines. A case in point; CHOK1 analysed replicates had two completely different phenotypic shifts during the time of analysis. Replicate A lost its early predominate phenotype to be replaced by a karyotype with an additional chromosome 10. Replicate B predominate karyotype mutated to consist of a combination of chromosome 10 or 8 that has gained a portion of chromosome 1. The only CHO host cell line that showed reasonable genetic stability across the culturing period was CHOK1a adapted to grow in glutamine free conditions (CHOK1a 0mM Gln) (Vcelar et al., 2018a). Overall across the host karyotypes there is a noticeable difference between the number of specific populations within each replicate and there isn't a consensus dominating population. Suggesting if a cell gains a spurious and fortuitous mutation that provides a growth advantage, it can readily establish itself within the flask. One can speculate on the overall impact this may have at the 'omic level, which could potentially be compounded by large structural variations witnessed both in this publication and in our own observations.

Vcelar et al (Vcelar et al., 2018b) extended this analysis further to assess the effect of genomic instability during the generation and subcloning of recombinant protein expressing cell lines. Vcelar et al (Vcelar et al., 2018b) generated multiple clones that derived from the same working cell bank, to assess and track karyotypic changes across a mock cell line development workflow that includes freeze thaw cycles and subcloning steps. In terms of chromosome number, they observed a similar distribution compared to the host cell analysis. The modal chromosome number was 19, flanked by large fluctuations ranging from 8-55 chromosomes within a cell line. The crux of the data indicates that using a selection pressure to retain cells that obtain the GOI promotes homogeneity within the culture.

CHOK1a host line main karyotype was lost over time, with no distinct karyotype present across 6 months of culturing. A subclone expressing cluster of differentiation 4 (CD4) and green fluorescent protein (GFP), which was sorted four times to obtain a 'stable' expressing subclone, maintained its main population from time point 0 (T0) during 6 months of culture. However, there were two other populations that greatly established themselves during the culture period, from 3% at T0 to 53% and 33%, respectively. Again, the use of smaller proteins that do not replicate a full sized mAbs size and conformation has to be taken into consideration, as the size differences may affect protein production kinetics and stress that is exerted on the cell itself. However, their findings suggest that cell populations within the flask are dynamic. Therefore, cell populations that have obtained a mutation that inhibit RTP production, may become more dominant with increasing population doublings, having an overall effect on titre and production stability.

With their findings, they question the requirement of clonality for FDA approval, as there is the assumption that CHO cells retain genetic homogeneity upon a single cell cloning event. This is compounded by another study by Pilbrough et al (Pilbrough et al., 2009) who also show

that recloning a confirmed subclone leads to heterogeneity in production, when logically one would expect homogeneity.

In line with our own observations, data presented here could indicate a potential mechanism explaining therapeutic protein production fluctuation over the stability assessment process. Cells which obtain the GOI but obtain decreased titre production, have the ability to become a significant proportion of the cell culture, effecting the overall titre obtained from a single manufacturing run. Based on this data and ours, there appears to be growing evidence that CHO cell lines cannot retain genomic homogeneity, however, divergent populations can be shown to be genetically related and this was deemed an acceptable criterion for clonality by the FDA (in-house data, unpublished).

Understanding and characterising how CHO cells obtain the ability to freely mutate whilst retaining high viability and production of exogenous proteins will be key to future platform enhancements, in a bid to build a stronger CHO platform that negates current pitfalls in production stability. There is a multitude of pathways that could be cause such instability. Here, I will investigate how telomeres and DNA damage have a downstream effect on genomic heterogeneity and therapeutic protein production, assessing how these profiles may differ between productionally stable and unstable cell lines. In the following section I will introduce how telomeres can elicit DNA damage within cell models and how such damage can lead to increased genomic mutations.

Interstitial telomeric repeats represent potential hot spots for DNA damage response activation and may lead to increased chromosomal mutations in CHO cell lines (1.3.3)

A potential hot spot for DNA damage and resulting structural damage at the genome level is the lack of telomeres at the extreme ends of chromosomes and the presence of interstitial telomeric repeats witnessed in CHO cell lines (Fernandez et al., 1995; Slijepcevic and Hande, 1999; Smilenov et al., 1998). In the majority of other species, telomeres usually exist at the extreme ends of chromosomes and act to cap and stabilise the chromosome structure. Critically short telomeres initiate a DNA damage response (DDR), directed through primarily ATM and ATR kinases, leading to events such as non-homologous recombination, end-joining and translocations.

Such events contribute to the overall chaotic nature of the CHO genome which may potentially have a direct impact on its stability during production of RTPs. Although there are a multitude of pathways in which large structural variations can occur within the genome, this thesis will investigate the potential causal link between aberrant telomere homeostasis and DNA damage response (DDR) activation, in relation to inherent CHO cell genomic instability and its potential effect on production instability.

In human, mouse and yeast (where most of work has been performed), telomeres are known to be situated at the extreme end of chromosomes, formed of inert G rich repeats ($TTAGGG^{n}$) encased by a specialised 6-protein complex termed Shelterin. Telomeres were first identified by Muller and McClintock in the 1930s in fly and corn chromosomes, respectively. Telomere structures were elucidated through exposing chromosomes to high doses of x-rays, which caused DNA breakage and/or deletions, and witnessing fusions of chromosomes.

Interestingly, Muller coined the term 'telomere' deriving the word from Greek; telos (terminus) and meros (part) (McClintock, 1941; Montpetit et al., 2014). Telomeres are composed of duplex g-rich tamdem repeats which possess three main configurations; g-overhang, T-loop (and D-loop) or G-quadruplex DNA. The 3' overhang can be formed through dissociation of the extreme end RNA primer which is used for initiating transcription for the lagging strand (Lu et al., 2013).

The Shelterin complex is a specialised 6 membered protein complex that encompasses double stranded telomeric DNA, which confers the appropriate stability and homeostasis of the telomeres (de Lange, 2005). Shelterin enables cells to distinguish between double stranded breaks (DSBs) and natural chromosome termini through the repression of DDR response whilst regulating telomerase-based telomere maintenance and elongation. Shelterin protein has been shown to be stable in the absence of telomeric DNA, as demonstrated by its isolation from nuclear cell extracts (de Lange, 2005).

Shelterin is abundant at telomeres through specific localisation, however it has been shown not to function anywhere else within the nucleus. The formation and localisation to the telomeres is dependent on the interaction of three key structure binding proteins; telomeric repeat binding factor 1 and 2 (TRF1/2), TRF2 interacting nuclear protein 2 (TIN2) and tripeptidyl peptidase 1 (TPP1). TRF1 and TRF2 bind to double stranded telomere DNA and initiate the recruitment of the other four complexes of shelterin (TIN2, RAP1, TPP1 and POT1) (Ye and de Lange, 2004; Ye et al., 2004a; Ye et al., 2004b).

TRF1 and TRF2 share a common domain structure termed the TRF homology domain (TRFH) connected by a flexible hinge domain which allows them to bind to dsDNA as a homodimer or oligomer through homotypic interaction of TRFH domain (Bianchi et al., 1997; Broccoli et al., 1997). POT1 is the final piece of the shelterin complex and arguably the most important. It binds to single stranded G-rich DNA and has a

binding preference for TAGGGTTAG sites at the 3' end, through two OB-folds situated at the N-terminus (de Lange, 2005; Wang et al., 2007). POT1 confers DDR damage pathway inhibition through the sequestering of the 3' single strand DNA (ssDNA) at the extreme ends of chromosomes.

Shelterin function not only depends on the ability of its individual constituents to bind and function, it is also dependent on the interaction with shelterin accessory factors and mediators of the DNA damage response (DDR), which have a direct effect on telomere homeostasis. It is important to note that these accessory factors have other cellular functions and their effect on telomere length is minimal in normal conditions and usually only occurs transiently and associated at different stages of the cell cycle (shelterin is present at all stages) (de Lange, 2005). DDR is formed of a highly intricate web of positively and negatively interacting pathways that contain hundreds of upstream and downstream proteins. Homeostasis of these pathways is imperative to the progression normal cellular life. The genome encounters multiple insults that require repair before progression to S phase in highly proliferative cells. Insults can arise from endogenous and exogenous factors and can include DNA mismatches, introduced through DNA replication or DNA stand breaks caused by abortive topoisomerase I and II activity (Jackson and Bartek, 2009).

Cells defective in DDR machinery generally display heightened sensitivity towards DNA-damaging agents and in extreme cases can culminate in increased non-homologous end joining (NHEJ) and homologous recombination (HR) events. Some lesions may be repaired through the action of a single enzyme, however, the majority of lesions are repaired through a series of concomitant actions by multiple proteins (Jackson and Bartek, 2009). For example, in mismatch repair, detections of mismatches and insertion/deletion loops trigger a single-strand incision that it subsequently acted upon by nuclease, polymerase and ligase enzymes to facilitate its repair (Jiricny, 2006).

Although there are many mechanisms by which the telomeres are repaired by DNA damage pathways, this section will highlight the actions of DSB repair mechanisms, as this is the predominant mechanism to which telomeres are subjected to. For DSB repair, there are two principle mechanisms, which are used – non homologous end-joining (NHEJ) (Lieber, 2008) and homologous recombination (HR) (San Filippo et al., 2008).

In NHEJ, DSBs are recognised by Ku80 and Ku70 protein accumulation (Koike and Koike, 2008) which in turn binds and activates the protein kinase DNA-PKcs. This leads to the disassociation of the Ku heterodimer with the concomitant recruitment and activation of end-processing enzymes polymerases and DNA ligase IV which facilitates NHEJ and HR repair (Lee et al., 2016). An alternative pathway termed microhomology-mediated end-joining (MMEJ) or alternate end-joining also exists. Although it is less characterised, it is known that this pathway always results in sequence deletions (McVey and Lee, 2008). Both NHEJ and MMEJ pathways are error-prone but can operate at any stage of the cell cycle. Conversely, HR is generally restricted to S and G2 stages as it uses sister-chromatid sequences as templates to mediate faithful repair. HR has many sub-pathways and is always initiated by ssDNA generation which can be promoted by various proteins including the MRE11-RAD50-NBS1 (MRN) complex (San Filippo et al., 2008).

If NHEJ and HR pathways are working correctly, they can resolve DSB within the genome without a substantial risk of carrying forward phenotype-changing mutations. Cell cycle progression must be slowed to allow time for repair to occur. Ataxia-telangiectasia mutated (ATM) and ataxia-telangiectasia and Rad3 related (ATR) protein are key protein kinases that modulate DDR signalling pathways. They are recruited to and activated by DSBs and replication protein A (RPA) coated ssDNA, respectively (Bartek and Lukas, 2007; Cimprich and Cortez, 2008; Shiloh, 2003). Chek1 (Chk1) and Chek2 (Chk2) protein kinases are the

most characterised downstream targets of ATM and ATR. They act in tandem with ATM and ATR to reduce cyclin-dependent kinase (CDK) activity by various mechanisms, some of which are mediated by p53 (Bartek and Lukas, 2007; Kastan and Bartek, 2004; Riley et al., 2008).

Inhibition of CDK inhibits or slows cell cycle progression at each cell cycle checkpoint. ATM and ATR also act in parallel to enhance repair by inducing DNA-repair proteins transcriptionally or post-transcriptionally; by recruiting repair factors to the damage and activating them through phosphorylation, acetylation, ubiquitylation or SUMOylation (Huen and Chen, 2008). Additionally, chromatin structure has an impact on DDR and is modulated in response to DNA damage (Bartek and Lukas, 2007).

ATM/ATR/DNA-PK mediated phosphorylation of serine-139 of histone H2.A variant H2.AX on chromatin sites flanking DSBs prevents DNA replication machinery from interacting with the insult before it is repaired (figure 1.5) (Bartek and Lukas, 2007). Ubiquitin adduct formation subsequently follows and leads to the recruitment of DDR factors and chromatin-modifying components which are thought to promote DSB repair and amplify DSB signalling (Huen and Chen, 2008).

Specifically at the telomeres, when telomere sequences reach the Hayflick limit (Hayflick, 1965; Hayflick and Moorhead, 1961) (the critical length of telomere where apoptosis is triggered) or significant loss of shelterin occurs, the 3' G-rich overhang of the telomere is then recognised by the DDR pathway dependent on MRN and ATM action (O'Sullivan and Karlseder, 2010). The overhang is excised and the chromatin structure changes through the dimethylation of Lys20 of histone H4 and of Lys79 of H3, amongst others.

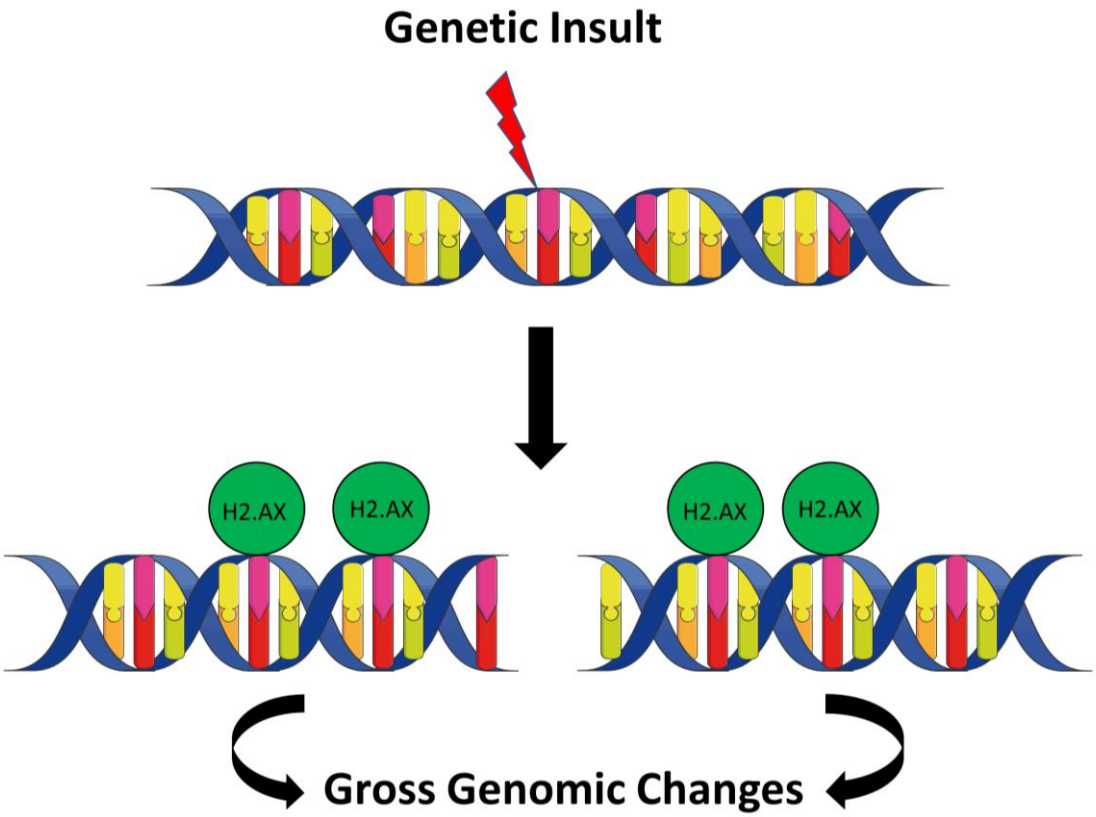


Figure 1.5. Schematic depicting DNA double stranded break (DSB) and the subsequent phosphorylation of histone 2.AX. Environmental or genetic insults can lead to DSB within the genome. Improper resolution of the double strand break leads to gross genomic changes, potential mutations and genomic instability. Increased DNA damage may result in increased chromosomal changes that in turn may affect production stability.

Tumour suppressor p53-binding protein (TP53BP1) is recruited and facilitates NHEJ dependent covalent fusion of chromosome ends, in the absence of p53 and retinoblastoma (RB)-dependent tumour suppressor pathways (O'Sullivan and Karlseder, 2010). When the fused chromosomes pass through mitosis a series of break-fusion-bridge (BFB) cycles (Marotta et al., 2013) occur, where the chromosomes break randomly during mitosis leading to non-reciprocal translocations and genome instability (O'Sullivan and Karlseder, 2010). BFB cycles have also been shown to lead to whole chromosome loss (Thomas et al., 2018), DNA amplification (Lo et al., 2002) and has been attributed to a potential cause of intratumor heterogeneity (Gisselsson et al., 2000).

Generally speaking, if these events allow for effective DNA damage repair, DDR inactivation ensues leading to the resumption of normal cell functionality. However, if the damage cannot be repaired in a prompt fashion, a chronic DDR signal triggers cell death through apoptosis or cellular senescence pathways (Jackson and Bartek, 2009). As telomeres obtain a 3' ssDNA overhang at the extremities of the telomere caps, shelterin complexes located at the telomeres sequester the 3' ssDNA G-rich overhang, preventing them from engaging in NHEJ-mediated fusions or activating ATM/ATR signalling pathways (de Lange, 2005).

Interestingly, the same DDR components actually play a role in the normal homeostasis of telomere length regulation (table 1.1). It has been shown that mammalian telomeres are recognised by ATM and MRN complex during G2 phase possibly triggering a localised DDR that promotes telomere end processing and shelterin complex formation (Jackson and Bartek, 2009). Establishing whether a telomere specific DDR response is transient or chronic, may be able to offer some insight into chromosomally chaotic nature within CHO cell lines.

Table 1.1 DDR pathway proteins. (Taken from (Jackson and Bartek, 2009))

DDR Protein	**Role at Normal Telomeres**
MRN Complex	Telomere length regulator and a role in end processing
ATM, ATR and CHK2	Telomere length maintenance phosphorylate shelterin components, possible roles in telomerase activation and recruitment
Ku and DNAPKcs	Telomere length and component regulators, possible capping function
RAD9-RAD1-HUS1 (9-1-1)	Telomere component and telomere-length regulator; aids telomerase recruitment/activation
Nucleases EXO1, FEN1, XPF/ERCC1 and Apollo	Telomere termini processing for telomerase action promotion, regulates telomere integrity
PARP1	Potential telomere-length regulator
BRCA1	Telomere maintenance
RPA	Telomere component with role in telomerase recruitment
WRN	Maintains telomere structure and functions in telomere replication
RAD51D and other HR Proteins	Regulate telomeric integrity

Smilenov et al (Smilenov et al., 1998) are one of the few groups who have tried to characterise CHO telomeres and their role in DDR activation. As observed within this project and by others (Bouffler et al., 1993; Bouffler et al., 1996), CHO cell lines obtain large tracts of interstitial telomeric repeats (ITS), whose fluorescence intensity often masks the telomeres at the extreme ends of the chromosomes when applying the appropriate thresholding. These interstitial blocks have been shown in other cellular models to be hot spots for illegitimate recombination (Katinka and Bourgain, 1992) and undergo spontaneous amplification (Pandita and DeRubeis, 1995).

Smilenov et al (Smilenov et al., 1998) were the first to investigate whether Chinese hamster (ch)TRF1 is co-localised to telomere repeats, which can serve as an indication that the TTAGGGn sequences are protected from DDR pathways via the shelterin complex. Unlike mouse TRF1, chTRF1 shows 97.5% identity with human TRF1 (Smilenov et al., 1998) suggesting that they could function in a similar way to their human counterparts. It was observed that chTRF1 was not seen at all chromosome ends of CHO cell lines and also was found to be completely absent from internal tracts of telomere sequences (Smilenov et al., 1998). This implies that the interstitial telomere repeats are not protected by the shelterin complex and the extreme telomeres only have partial protection.

Vulnerable telomere sequences may lead to an increase in DDR pathways at these sites, causing the genomic instability observed in CHO. This was observed in four CHO cell lines by Slijepcevic et al (Slijepcevic et al., 1996) where 25/27 (93%) of breakpoints occurred from spontaneous terminal deletions and were localised in regions containing interstitial telomeric repeats, such an observation has been corroborated by other groups (Alvarez et al., 1993; Bertoni et al., 1996; Fernandez et al., 1995). Fragments of CHO ITS have been shown to range from 1kb to >100kb, the majority of sequences shown to be extensive and continuous arrays of telomeric-like sequences. Short ITS, comprised of 29-126bp

TTAGGGn repeats, were shown to be AT rich indicating fragile sites across the genome (Faravelli et al., 2002). These sites pose another mechanism by which DDR pathways can be activated at ITS sites in CHO cell lines. Interestingly, Rivero et al., noted that the re-joining of DSBs at ITS is slower compared to the whole genome in wild type CHO cells, demonstrating an intragenomic heterogeneity in DSB repair (Rivero et al., 2004). In a highly proliferative and mutagenic context (such as CHO cells in therapeutic protein production), escape from key cell cycle check point pathways could lead to a situation where ITS DSBs are left un- or mis-repaired. Thus, upon mitosis the unrepaired ITS could lead to translocations and/or large deletions of chromosomes, contributing to the overall genetic instability of the cell line, as witnessed by ourselves and others (Vcelar et al., 2018a).

As a pathway of CHO therapeutic protein production instability has not been elucidated, this thesis will explore any potential differences in telomere length, overall and telomere specific DNA damage levels and how such DDR pathway activation may influence the overall genomic mutational profile of CHO producing cell lines. If a link between telomere damage, genome mutations and therapeutic protein production instability can be established, this will be the first identified causative pathway of the CHO production instability phenotype. The following section will highlight the commercial impact CHO derived therapeutics currently obtains in the drug market and why having a robust CHO production platform is highly sought after by the Biopharmaceutical industry.

CHO Global Impact on Drug Markets (1.4)

Despite the issues of cell line stability, mAb expression by CHO cell lines has become a market dominating pharmaceutical powerhouse. In 2013, adalimumab (trade name; Humira) entered Forbes list of best selling drugs of all time (http://www.forbes.com/sites/simonking/2013/01/28/the-best-selling-drugs-of-all-time-humira-joins-the-elite/#a4bd2e6193dc), with annual sales of $11.02 billion. Furthermore, global revenue of products from CHO cell lines have increased to more than $100 billion and biopharmaceutical sales grew accumulatively to $651 billion between 2014-2017, whereas 2017 alone reached $188 billion (Walsh, 2018).

45 biologics have reached "blockbuster" status (sales reaching >$1 billion) with monoclonal antibodies representing the most lucrative single product class, thus highlighting the widescale adoption of this cell line across big pharma (Jadhav et al., 2013; Walsh, 2018). An important consideration for companies is the ability to maintain a dominant market share within the drugs therapeutic market. Due to mammalian post-translational modifications, even after a drug patent expires, it is extremely difficult for a rival company to produce a cell line that mimics the safety profile and productivity of the original drug. Whereas drugs based on chemistry can be easily synthesised once the full annotation of the molecule is known and with a relative certainty it will possess the same immunogenic and pharmacokinetic profiles as its predecessor, leading to generics being produced. This allows companies to hold their market share for periods beyond the original patent without the opposition of generics.

The time and cost of R&D into new drugs or repurposing drugs for new indications is ever increasing (Measuring the return from pharmaceutical innovation, Deloitte, 2018). Thus, there is a big effort into stream lining manufacturing practices in order to increase R&D output and many companies are looking for internal and external solutions to better understand the biology of the CHO expression platform with a view to bioengineer more stable and highly producing cell lines or create methodologies that decrease cell line development timelines. Developing a novel CHO platform that does not obtain the aforementioned caveats would allow for faster drug development timelines with R&D savings potentially passed on to the patient. Therefore, many groups are looking to characterise CHO biology to a greater extent than has been performed previously in a bid to identify targetable pathways for bioengineering efforts.

Aims and hypotheses (1.5)

Hypothesis (1.5.1)

As outlined in this introduction, there are numerous reports of CHO cell line instability at both the genomic and therapeutic production level. At the time of starting this thesis, there were no publications that utilised MFISH to identify karyotypically distinct populations within CHO cell lines. Vcelar et al (Vcelar et al., 2018a; Vcelar et al., 2018b) have recently published data identifying heterogeneity within CHO host cell lines, but have not investigated fully chromosomal heterogeneity and its role in therapeutic protein production. Moreover, a pathway that explains CHO's natural tendency to mutate at the chromosome level is still wanting.

Within this thesis, I attempt to bridge the knowledge gap by investigating CHO chromosomal heterogeneity and how it may impact therapeutic protein production instability. Although there are many causes for genomic mutations within a cell, I will specifically explore the overall and telomere specific DNA damage levels in productionally stable and unstable cell lines, to understand if there is any correlation between DNA damage and chromosomal mutations.

I hypothesise that interstitial telomeric sequences (ITS) within CHO chromosomes are hot spots for DNA damage and their resulting pathways (DNA damage response, DDR). Improper resolution of DNA damage, in a highly proliferative context, leads to gross chromosomal rearrangements, resulting in a heterogenous culture in the bioreactor, creating fluctuations in the maximum titre a cell line can produce over the manufacturing period (i.e. production instability, figure 1.6).

De-novo heterogenous karyotypes within a culture flask, during prolonged cell culture, may lead to heterogenous production of therapeutic proteins potentially due to low producing populations gaining proliferative advantage within the bioreactors (Dorai et al., 2012; Hammill et al., 2000). As shown in previous studies (Deaven and Petersen, 1973; Derouazi et al., 2006; Vcelar et al., 2018a), CHO host cell lines have a natural propensity for genetic mutations and these are only exacerbated during prolonged culture (Vcelar et al., 2018a). Prolonged culture is a requirement to assess the production stability of a CHO producing cell line, as a manufacturable cell line must maintain its titre over the manufacturing process. If the hypothesis holds true, this will represent a causal link between genomic instability and production instability, witnessed by ourselves and others.

If a correlation between ITS specific DNA damage, genomic instability and production instability can be identified, future work could be performed to characterise this pathway further. Building upon this pathway will be imperative to gain a better understanding of CHO as a production platform and may identify potentially robust targets for bioengineering an inherently stable host. Additionally, markers of this pathway may be used to define cell line stability, rather than using arbitrary titre thresholds, to better stratify CHO producing cell lines based on the cell's internal biology.

Aims and objectives (1.5.2)

Within this thesis I will investigate a potential causative pathway of the CHO production instability phenotype through the following aims and objectives:

1. Characterise early (~10 generations) and late (~150 generations) timepoints of CHO-GS host cell line to identify basal levels of;
    a. Chromosome number distribution
    b. Chromosome rearrangements and cell population identification (karyotype)
    c. Telomere length
    d. Shelterin component expression
    e. Telomere co-localisation with Shelterin
    f. DNA damage levels and telomere specific damage
2. Characterise the same parameters between productionally stable and unstable cell lines, across early and late time points, to identify potential differentials that may be utilised in platform bioengineering or utilised in unstable cell line prediction methods
3. Perform a transcriptomic screen of genes involved in DNA damage response and telomere homeostasis pathways, to identify genes that may be differentially expressed between productionally stable and unstable cell lines, across ~140 generations
4. Investigate how a production environment affects cell lines at the chromosomal level, comparing the mutation profile to cell lines in normal passaging culture. Additionally, levels of DNA damage within these samples will be characterised to elucidate whether increases in DNA damage can lead to greater heterogeneity through increased chromosomal mutations
5. Develop and automate a cell line instability prediction method, utilising CHO cell karyotype heterogeneity as a marker, to aid cell line triage during cell line development and decrease Chemical, Manufacture and Controls (CMC) timelines.

Summary (1.6)

CHO has become the predominant work horse of the pharmaceutical industry, but it is poorly characterised at any 'omic level, compared to other human, mouse and yeast cell lines. Inherent genomic instability and phenotypic drift/diversity in CHO cell lines represent a complex and poorly understood problem for the bioproduction of therapeutic proteins. Although CHO has been known to be genomically unstable since its derivation (Cox and Puck, 1969; Puck, 1971; Puck et al., 1964), their ability to produce therapeutic protein at g/L scales has led to their wide adoption across academia and industry. Although much has been achieved with media and process optimisation (Wurm, 2004), increasing viable cell densities and overall titre, it appears that laboratories are seeking further large-scale beneficial changes to the CHO production platform, focussing specifically on the underlying biology of the CHO cell.

Much progress has been made with regards to sequencing the CHO genome (Maccani et al., 2013; Xu et al., 2011), however, with the knowledge that ~49% of the CHO genome acquires mutations in a single transfection event (Yusufi et al., 2017), it is yet to be determined whether sequencing efforts can be robustly utilised in a multitude of CHO cell lines in laboratories across the field. Additionally, mass heterogeneity of CHO host cell lines and phenotypic drift has been observed over long term culture of these lines (Vcelar et al., 2018a; Vcelar et al., 2018b), making reproducibility of findings and data comparisons between different cell lines and time points difficult. Despite the increased genomic and transcriptomic characterisation made by the field, a production instability pathway has not yet been identified.

Here, I attempt to elucidate a pathway that connects CHO chromosomal instability and therapeutic protein instability. Although there are many avenues to which a cell can acquire chromosomal mutations, within this thesis I focus specifically on DNA damage and telomere homeostasis pathways and how their interplay may impact chromosomal instability in CHO during maintenance and production environments. Furthermore, I will investigate how karyotypic heterogeneity may affect therapeutic protein instability by analysing the frequency of different populations within cell lines to assess any differential characteristics between productionally stable and unstable cell lines. To alleviate aforementioned caveats, this thesis has tried to focus on single cell characterisation strategies, in a bid to characterise CHO cell lines at a cell populational level. This will allow for comparisons between karyotypically distinct cell populations within and across cell lines, to identify any potential markers that may lead to the production instability phenotype. If a pathway can be identified, it will provide the basis for further characterisation in the future that may identify robust bioengineering targets to create an inherently stable CHO host cell line. Moreover, any biological markers identified within this thesis that robustly identify production unstable cell lines, could be utilised to create predictive methods that may reduce stability assessment time lines.

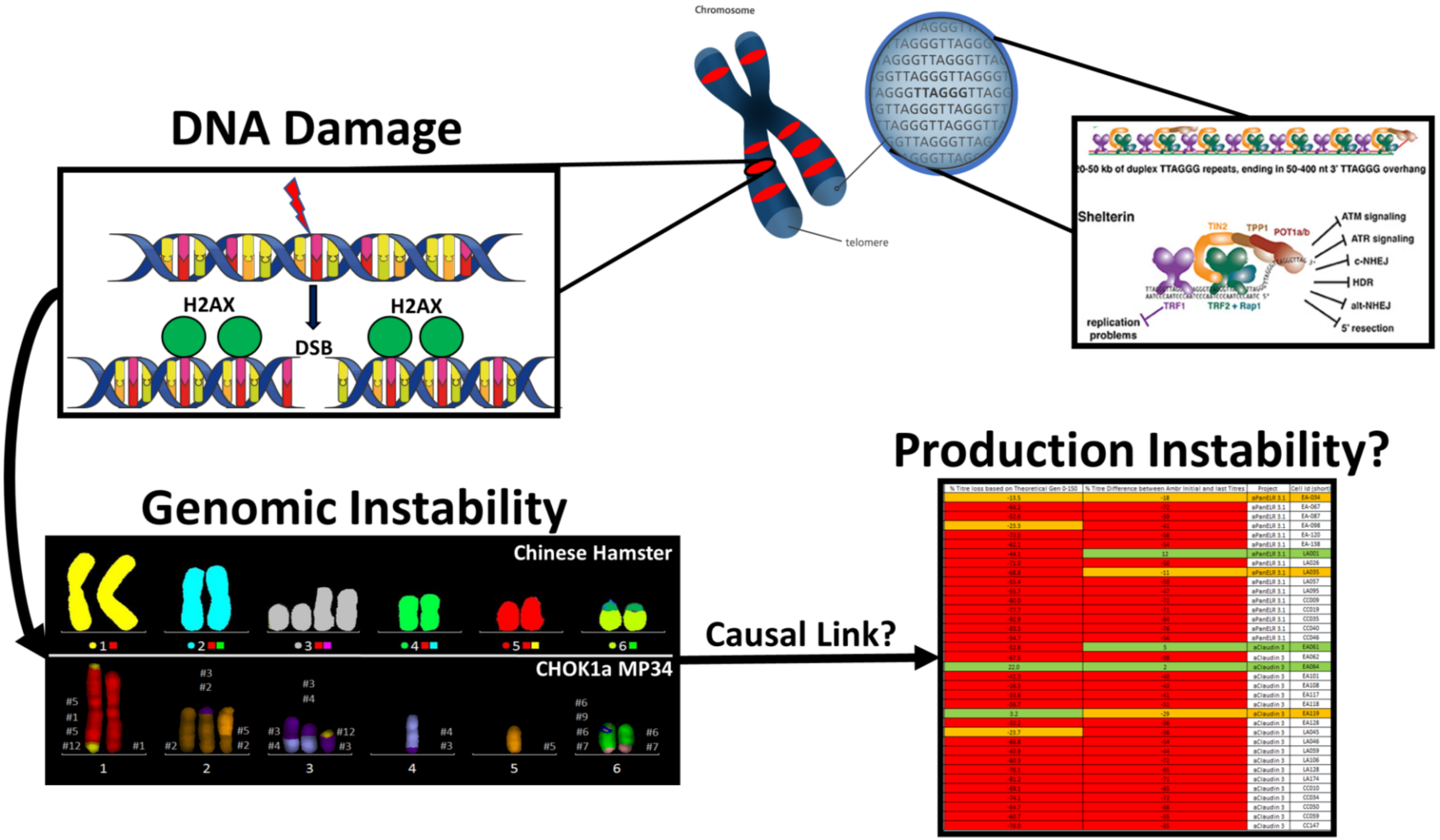

DNA Damage
H2AX
DSB
H2AX
Chromosome
telomere
10-50 kb of duplex TTAGGG repeats, ending in 50-400 nt 3' TTAGGG overhang
Shelterin
TIN2
TPP1
POT1a/b
TRF1
TRF2 + Rap1
replication problems
ATM signaling
ATR signaling
c-NHEJ
HDR
alt-NHEJ
5' resection
Genomic Instability
Chinese Hamster
CHOK1a MP34
Causal Link?
Production Instability?

Figure 1.6. Schematic depicting the overview of hypothesis. Interstitial telomere sequences observed within CHO cell lines may be amenable to increased DNA damage response pathways (top right). The potential increase in DNA damage at these sites may lead to genomic instability in CHO cell lines, as observed using multi-fluorescence in situ hybridisation (MFISH), in the form of gross chromosomal mutations (e.g. translocations, top and bottom left schematic, top MFISH image provided by Metafer, bottom MFISH image acquired during this project). Such mutations may lead to therapeutic production instability observed within GSK and the CHO biotherapeutic platform field.

Chapter 2: Methods

Cell Culture (2.1)

Cell lines (2.1.1)

Cell lines whose production stability had already been determined were obtained from liquid nitrogen stocks from Biopharm Process Research (BPR) department, GlaxoSmithKline (GSK), Stevenage R&D laboratories. Cell lines used are listed in table 2.1. Cell lines for therapeutic proteins 2, 3 and 5 were used for stable and unstable comparison experiments and proteins 1 and 4 were used for blinded validation of MFISH production stability prediction method.

Table 2.1. List of cell lines used throughout this thesis.

Host Cell line	Therapeutic Protein	IgG Class	Cell Line ID	Media
CHOK1a-GS-KO	N/A - Host	N/A	CHOK1a-GS-KO	CD-CHO + 8mM Glutamine
CHOK1a	N/A – Host	N/A	CHOK1a	CD-CHO + 8mM Glutamine
CHOK1-SV	N/A – Host	N/A	CHOK1-SV	CD-CHO + 8mM Glutamax + 5µg/ml Puromycin
HEK293T	N/A – Host	N/A	HEK293T	Freestyle 293 + 8mM Glutamax
CHOK1a-GS-KO	Protein 1	IgG1	Cell Line 19	CD-50 + 25µM MSX
CHOK1a-GS-KO	Protein 1	IgG1	Cell Line 20	CD-50 + 25µM MSX
CHOK1a-GS-KO	Protein 1	IgG1	Cell Line 21	CD-50 + 25µM MSX
CHOK1a-GS-KO	Protein 1	IgG1	Cell Line 22	CD-50 + 25µM MSX
CHOK1a-GS-KO	Protein 1	IgG1	Cell Line 23	CD-50 + 25µM MSX
CHOK1a-GS-KO	Protein 1	IgG1	Cell Line 24	CD-50 + 25µM MSX

CHOK1a-GS-KO	Protein 1	IgG1	Cell Line 25	CD-50 + 25µM MSX
CHOK1a-GS-KO	Protein 1	IgG1	Cell Line 26	CD-50 + 25µM MSX
CHOK1a-GS-KO	Protein 1	IgG1	Cell Line 27	CD-50 + 25µM MSX
CHOK1a-GS-KO	Protein 1	IgG1	Cell Line 28	CD-50 + 25µM MSX
CHOK1a-GS-KO	Protein 1	IgG1	Cell Line 29	CD-50 + 25µM MSX
CHOK1a-GS-KO	Protein 1	IgG1	Cell Line 30	CD-50 + 25µM MSX
CHOK1a-GS-KO	Protein 1	IgG1	Cell Line 31	CD-50 + 25µM MSX
CHOK1a-GS-KO	Protein 1	IgG1	Cell Line 32	CD-50 + 25µM MSX
CHOK1a-GS-KO	Protein 1	IgG1	Cell Line 33	CD-50 + 25µM MSX
CHOK1a-GS-KO	Protein 1	IgG1	Cell Line 34	CD-50 + 25µM MSX
CHOK1a-GS-KO	Protein 1	IgG1	Cell Line 35	CD-50 + 25µM MSX
CHOK1a-GS-KO	Protein 1	IgG1	Cell Line 36	CD-50 + 25µM MSX
CHOK1a-GS-KO	Protein 1	IgG1	Cell Line 37	CD-50 + 25µM MSX
CHOK1a-GS-KO	Protein 1	IgG1	Cell Line 38	CD-50 + 25µM MSX
CHOK1a-GS-KO	Protein 1	IgG1	Cell Line 39	CD-50 + 25µM MSX
CHOK1a-GS-KO	Protein 1	IgG1	Cell Line 40	CD-50 + 25µM MSX
CHOK1a-GS-KO	Protein 1	IgG1	Cell Line 41	CD-50 + 25µM MSX
CHOK1a-GS-KO	Protein 2	IgG1	Cell Line 1	CD-50 + 25µM MSX
CHOK1a-GS-KO	Protein 2	IgG1	Cell Line 2	CD-50 + 25µM MSX
CHOK1a-GS-KO	Protein 2	IgG1	Cell Line 3	CD-50 + 25µM MSX
CHOK1a-GS-KO	Protein 2	IgG1	Cell Line 4	CD-50 + 25µM MSX
CHOK1a-GS-KO	Protein 2	IgG1	Cell Line 5	CD-50 + 25µM MSX
CHOK1a-GS-KO	Protein 2	IgG1	Cell Line 6	CD-50 + 25µM MSX
CHOK1a-GS-KO	Protein 3	IgG1	Cell Line 7	CD-50 + 25µM MSX

CHOK1a-GS-KO	Protein 3	IgG1	Cell Line 8	CD-50 + 25µM MSX
CHOK1a-GS-KO	Protein 3	IgG1	Cell Line 9	CD-50 + 25µM MSX
CHOK1a-GS-KO	Protein 3	IgG1	Cell Line 10	CD-50 + 25µM MSX
CHOK1a-GS-KO	Protein 3	IgG1	Cell Line 11	CD-50 + 25µM MSX
CHOK1a-GS-KO	Protein 3	IgG1	Cell Line 12	CD-50 + 25µM MSX
CHOK1a-GS-KO	Protein 4	IgG1	Cell Line 42	CD-50 + 25µM MSX
CHOK1a-GS-KO	Protein 4	IgG1	Cell Line 43	CD-50 + 25µM MSX
CHOK1a-GS-KO	Protein 4	IgG1	Cell Line 44	CD-50 + 25µM MSX
CHOK1a-GS-KO	Protein 4	IgG1	Cell Line 45	CD-50 + 25µM MSX
CHOK1a-GS-KO	Protein 4	IgG1	Cell Line 46	CD-50 + 25µM MSX
CHOK1a-GS-KO	Protein 4	IgG1	Cell Line 47	CD-50 + 25µM MSX
CHOK1a-GS-KO	Protein 4	IgG1	Cell Line 48	CD-50 + 25µM MSX
CHOK1a-GS-KO	Protein 4	IgG1	Cell Line 49	CD-50 + 25µM MSX
CHOK1a-GS-KO	Protein 4	IgG1	Cell Line 50	CD-50 + 25µM MSX
CHOK1a-GS-KO	Protein 4	IgG1	Cell Line 51	CD-50 + 25µM MSX
CHOK1a-GS-KO	Protein 4	IgG1	Cell Line 52	CD-50 + 25µM MSX
CHOK1a-GS-KO	Protein 4	IgG1	Cell Line 53	CD-50 + 25µM MSX
CHOK1a-GS-KO	Protein 4	IgG1	Cell Line 54	CD-50 + 25µM MSX
CHOK1a-GS-KO	Protein 4	IgG1	Cell Line 55	CD-50 + 25µM MSX
CHOK1a-GS-KO	Protein 4	IgG1	Cell Line 56	CD-50 + 25µM MSX
CHOK1a-GS-KO	Protein 4	IgG1	Cell Line 57	CD-50 + 25µM MSX
CHOK1a-GS-KO	Protein 4	IgG1	Cell Line 58	CD-50 + 25µM MSX
CHOK1a-GS-KO	Protein 4	IgG1	Cell Line 59	CD-50 + 25µM MSX
CHOK1a-GS-KO	Protein 4	IgG1	Cell Line 60	CD-50 + 25µM MSX

CHOK1a-GS-KO	Protein 4	IgG1	Cell Line 61	CD-50 + 25µM MSX
CHOK1a-GS-KO	Protein 4	IgG1	Cell Line 62	CD-50 + 25µM MSX
CHOK1a-GS-KO	Protein 5	IgG1	Cell Line 13	CD-50 + 25µM MSX
CHOK1a-GS-KO	Protein 5	IgG1	Cell Line 14	CD-50 + 25µM MSX
CHOK1a-GS-KO	Protein 5	IgG1	Cell Line 15	CD-50 + 25µM MSX
CHOK1a-GS-KO	Protein 5	IgG1	Cell Line 16	CD-50 + 25µM MSX
CHOK1a-GS-KO	Protein 5	IgG1	Cell Line 17	CD-50 + 25µM MSX
CHOK1a-GS-KO	Protein 5	IgG1	Cell Line 18	CD-50 + 25µM MSX

Viable cell counting (2.1.2)

500µl of cell suspension in standard media were decanted into a 4ml sampling tube. 500µl (1:1 volume) of TrypLE (Gibco, #12605010) is added to the cell suspension and sample processed by a Vi-Cell XR (Beckman Coulter), providing metrics of total and viable cell counts, percent viability and cell diameter.

Cell line thawing (2.1.3)

Cell vials were thawed in 37°C PBS and resuspended in 10ml of the appropriate media (table 2.1). Cell lines were counted on a ViCell (Beckman Coulter) by adding 500µL TrypLE (Gibco, #12605036) to 500µL cell suspension (1:1 volume). Culture flasks were seeded with 0.5x10^6 cells in 20ml media and incubated in a humidified shaking incubator set at 37°C, 5% CO_2 (v/v) and 5.5 x g.

Cell culture maintenance (2.1.4)

Upon recovery of cells to >95% viability, cell lines were maintained and passaged in the appropriate media (table 2.1) at 0.3x10^6 cells in 30ml every 3 or 4 days. Seeding density was calculated using a ViCell (Beckman Coulter), following previously outlined method (2.1.2).

Cell line freezing (2.1.5)

10x10^6 cells per vial were cryopreserved in 1.8ml freezing media, comprised of the appropriate media (table 2.1) for each cell line, supplemented with 7.5% DMSO (v/v, JT Baker, 9033-04). Vials were placed in Mr Frosties (ThermoFisher, #5100-0036) and then placed in a -80^{o}C freezer. After 48 hours, cells were transferred to liquid nitrogen for long term storage.

Cell line transfection (2.1.6)

A day prior to transfection, a therapeutic protein producing cell line was seeded at a density of 0.8x16^6 cells/ml in the relevant media. Cells were counted and an appropriate volume of culture for 5x10^6 cells/ml decanted into a 15ml falcon tube. Supernatant was removed and cell pellet washed with 10ml Dulbecco's phosphate buffer solution (D-PBS) (D8537). The cell pellet was resuspended with 100µl of Amaxa 4D cell line nucleofection buffer with the addition of supplement 1 (SF Cell Line 4D-Nucleofector X kit, Lonza, #V4XC-2012).

10µg of modified PX458-telogRNA (Telo-Cas9) and PX458-scrambled plasmid (Scrambled-Cas9) was pipetted into each nucleocuvette. 10µl water and 10µl GFP-positive control plasmid (SF Cell Line 4D-Nucleofector X kit, Lonza, #V4XC-2012) were used as positive and negative transfection controls. X unit, for suspension transfections, was selected on the Amaxa 4D, followed by buffer (SE) and CHOK1 nucleofection programme options.

Upon pulsing the cuvettes, sterile Amaxa pastettes were used to transfer the nucleofected cells into 6-well suspension plates containing 2ml conditioned media (media that contained cells for 72-hours of culture growth). Plates were then transferred into a static incubator at 37^{o}C, 5% CO_2 (v/v) in humidified conditions (80% humidity).

Cell line scale-up (2.1.7)

After 48 hours, cells were topped up with 2ml fresh selective media (CD50 + 25µM MSX and 400µg/ml G418, table 2.1, Gibco #10131027). Around 1 week after selection, 2ml fresh selective media is added. Once cell lines were confluent (~2 weeks post transfection, time can vary) cells were transferred to a T25 flask and selective media was added to reach a final volume of 10ml.

Viability and cell counts were quantified at this stage. After 3-4 days, 2ml fresh media is added. After ~3 weeks post transfection, cells were scaled up to 125ml shake flasks if viability was >30-40% and cell counts >0.2x10^6 cells/ml and topped up to 20ml final volume with selective media. Once stabilised in shaking flasks, cells were maintained as previously outlined in 2.1.4.

Shake Flask Production Runs (2.1.8)

Shake flask production runs are performed in 125ml shake flasks (Corning, CLS431143) to assess production stability of newly transfected therapeutic protein producing cell lines. Production media and supplements (listed in table 2.2) are prewarmed to 35^{o}C. Cell counts were performed as previously described (2.1.2). Cells were seeded at 1x10^6/ml in 50ml production media (table 2.2). 8.8ml/L of 3XC25+Asn supplement was added before placing flasks into a shaking incubator set at 35^{o}C and 5.5 x g. Feeding and sampling schedule for the duration of the production run was performed as outlined in table 2.3. On sample days, cells were counted on the ViCell as previously described (2.1.2) and 0.5ml of sample is collected and placed into a centrifuge at 13226 x g for 3 minutes. The resulting supernatant is then analysed on Cedex Bio HT (Roche) to obtain IgG titre, glucose and lactate measurements within the media. Glucose is fed back to 7g/L or 10g/L depending on the sampling day (table 2.4) and supplementations were performed (table 2.3). Final titre measurement is performed at day 15, however, the final sampling day may be prolonged if viability has not dropped below 50% and titre hasn't plateaued.

Table 2.2. List of supplements for production media during the production run process. Volumes of components to be added to the media are listed in table 2.3.

Component	Catalogue
CD51 Media	N/A – Proprietary formulation
100uM L-Methionine Sulphoxamine (MSX)	EMD Millipore, GSS-1015-F
3XC25-L-asparagine monohydrate	Hyclone #RR15010.01, Sigma #A7094
Cystine-Tyrosine	Sigma #RES1520C-A7, Sigma #RES3156T-A7
25mM Ferrous Sulphate solution	Sigma #F8633
1000mM Magnesium sulphate heptahydrate solution	Sigma #RES0089M-AZ02X
300g/L glucose solution	Sigma #G7021
28.75mM Zinc sulphate heptahydrate	Sigma Z0635

Table 2.3. Supplement volumes to be added to media. Additional components are added to the media upon the stated day of the production run (ml/L). Concentrations of supplements are scaled accordingly to the number of flasks in production (total volume required).

Supplement	**Day (ml/L)**					
	0	**3**	**6**	**8**	**10**	**14**
3xC25+Asn	8.8	16	28	28	48	20
Cystine-Tyrosine	-	3.2	3.2	3.2	6.4	3.2
Ferrous sulphate solution	-	0.2	-	-	-	-
Magnesium sulphate solution	-	1.2	-	-	-	-
Zinc sulphate solution		0.8				

Table 2.4. Glucose feed concentrations (g/L) required on specified days of sampling. Volume of glucose is dependent on required media volume.

	Day (g/L)					
	0	**3**	**6**	**8**	**10**	**12**
300g/L Glucose	-	7	7	7	10	7

24 Deep Well Production Runs (2.1.9)

24 Deep-well plates (Invitrogen, CS15124) can also be used to assess cell line production stability, in a more high-throughput manor. Cell counts were performed as previously described (2.1.2). Production media and supplements (table 2.2 and 2,3) are prewarmed to 35^{o}C. Cells were seeded in duplicate wells at 1x10^6/ml in 4ml production media. Production run plates should be set up twice for following experimental assays. One plate is used for sampling on day 6 and the other for day 15 sampling. Supplements listed in table 2.3 were added to each well following the supplement schedule, plates are then placed into a shaking incubator set at 35^{o}C and 5.5 x g for culture incubation. Feeding and sampling schedule can be found in table 2.3 and 2.4. On sample days, 0.5ml of culture media was mixed with 0.5ml TrypLE and cells were counted on the ViCell. The supernatant of each well is collected using 0.5ml of sample and placing samples into a centrifuge set at 13226 x g for 3 minutes. Resulting supernatant analysed on Cedex Bio HT (Roche) to obtain IgG titre, glucose and lactate measurements within the media. Day 6 plate is sacrificed for sampling and the feeding occurs in the second plate. Final titre measurement is performed at day 15.

Table 2.5. 24-Deep well plate production feed components and volumes (ml/L). Component volume is dependent on media requirements for experiment.

Supplement	Day (ml/L)	
	0	6
3xC25+Asn	-	28
Cystine-Tyrosine	-	3.2
Ferrous sulphate solution	0.2	-
Magnesium sulphate solution	1.2	-
Zinc sulphate solution	0.8	
300g/L Glucose solution	27	33

Molecular Cloning (2.2)

Overall cloning strategy for telomere and scrambled gRNA PX458-Neomycin plasmid (2.2.1).

Zhang's laboratory PX458 CRISPR-Cas9 plasmid has been modified to include a neomycin resistance gene cassette to enable selective growth of cells that obtain the PX458 plasmid. A telomere specific single gRNA ($TTAGGG_4$) and scrambeled gRNA was inserted into the PX458-Neomycin plasmid. Telomere specific Cas9 was used to assess how telomere specific DNA damage may impact production instability in CHO therapeutic producing cell lines, with PX458-Scrambeled being used as a control.

PX458 plasmid was linearised using NarI (NEB, R0191S) directly after the CRISPR-GFP-bGHPoly A sequence to insert neomycin resistance gene directly after the CRISPR protein construct coding sequence, using a restriction digest protocol (section 2.2.2). Neomycin resistance cassette was generated with homologous arms that correspond to the NarI restriction digest at 3' and 5' ends, facilitating homologous recombination when using In-fusion cloning kit protocol outlined in section 2.2.8 and 2.2.9 (Takarabio, #638920). Telomere and scrambled gRNAs were cloned into the PX458 gRNA site using Bbsi-HF (NEB #R3539) to linearise the plasmid (figure 2.1, section 2.2.2). Resulting clones were pre-screened for successful inserts using PvuII restriction enzyme (section 2.2.2, 2.2.3 and figure 2.2), amplified using a Maxi preparation kit (section 2.2.7) and sequence confirmed, before transfection into a producing cell line. All procedures to fulfil this cloning strategy are outlined below.

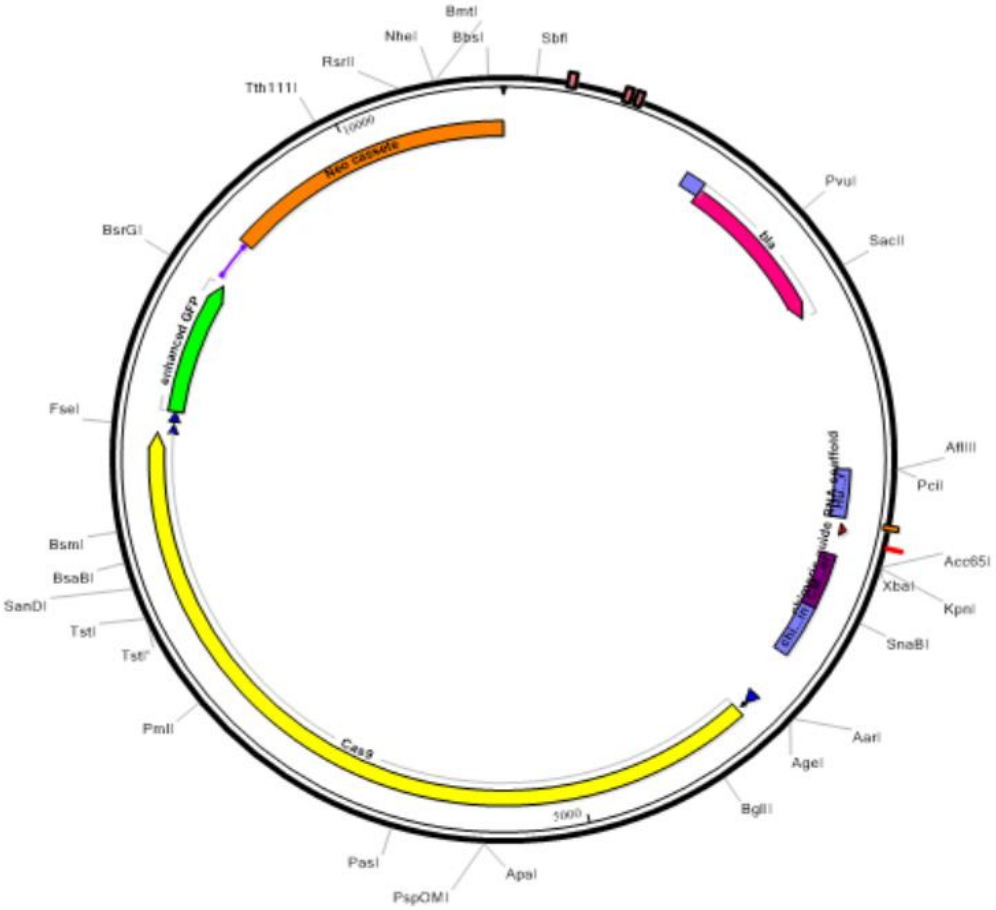


Figure 2.1. Representative plasmid map of PX458-Telo and PX458-Scrambled. Plasmids contain a Cas9 cassette (yellow), enhanced GFP (green), Neomycin resistance gene (orange) and telomere and scrambled gRNAs (red triangle). Unique enzymatic restriction sites highlighted on the perimeter of the plasmid.

Restriction enzyme digestion (2.2.2)

Restriction digests were performed for plasmid linearization and isolation of neomycin resistance gene fragments to be utilised in infusion cloning strategies to prepare the final plasmids for transfection (section 2.2.8). Restriction digests were carried out using New England Biolabs (NEB) restriction enzymes as listed in table 2.6 (representative digests outlined in figure 2.2 c and d).

Table 2.6. Restriction enzymes used in cloning.

Restriction Enzyme	Catalogue Number
Bbsi-HF	R3539S
NarI	R0191S

Agarose gel electrophoresis (2.2.3)

Agarose gel electrophoresis was used to separate restriction digests to isolate DNA fragments for cloning and sequencing. 100g (1%, w/v) Agarose powder (Sigma, A9539) was dissolved in 100ml TAE (1X) buffer (Roche, #11666690001) using a microwave as a source of heat. 20µl of SYBR safe (ThermoFisher, #S33102) is added to the dissolved agarose and the gel is then cast with the appropriate well sizes. Once set, gels were removed from the cast and placed into a horizontal gel box (Axygen) and the tank filled with 1X TAE buffer until the gel is completely immersed. DNA loading buffer was added to samples to a final concentration of 1X (NEB, #B7021S) and 20µl of sample added to wells in addition to an appropriate ladder (100bp – Invitrogen, #15628019, 1kb – NEB, #N3232). Voltage was set at 10V/cm for 1-2hours and stopped when the DNA ladder reached ¾ of the gel length. Gel extraction of desired DNA fragments was performed using Qiagen's QIAquick Gel Extraction Kit (#28704) (representative images in figure 2.2).

Plasmid amplification (2.2.4)

PX458-CRISPR vector was transformed in One Shot Stbl3 E. Coli (Invitrogen, #C737303) chemically competent E. coli as per manufacturer's instructions. In brief; 1 vial of One Shot Stbl3 was thawed on ice. 500ng PX458-CRISPR was added to One Shot Stbl3 vial and incubated for 30 minutes on ice. Bacterial cells were heat shocked for 45 seconds at 42^{o}C in a water bath. 250µl vLB broth was added to the vial and placed in a shaking incubator at 37^{o}C for 1 hour at 14 x g to amplify the bacterial cells containing the plasmid. 50µl and 100µl were spread on separate premade agar plates containing 100ug/ml ampicilin to ensure outgrowth of single cell colonies. Plates were inverted overnight in a 37^{o}C incubator. Colonies were then picked the following day, 7ml or 250ml warm vLB was added, dependent on stage in the cloning process, and placed in a shaking incubator at 37^{o}C overnight. Samples were then processed using Qiagens Mini Prep (#27106) or Maxi prep kits (#12362).

Polymerase chain reaction (PCR) (2.2.5)

PCR was performed using AmpliTaq gold 360 master mix (Applied Biosciences, #4398876). Each sample reaction was set up as in table 2.7 and PCR cycling performed in a BioRad thermocycler (table 2.8 and figure 2.2b).

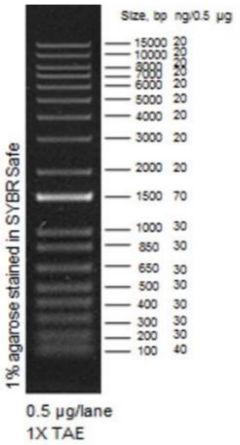
Table 2.7. Reaction master mix components.

Component	25µl Reaction
Distilled water	Up to 25µl
AmpliTaq Gold 360 master mix	12.5µl
10µM forward primer	0.5µl
10µM reverse primer	0.5µl
Template DNA	varies

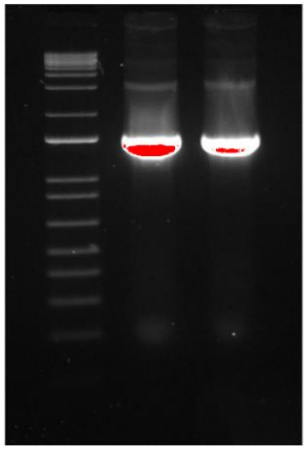
Table 2.8. Cycling conditions for PCR amplifications.

Step		Temperature (°C)	Time
Initial Denaturation		95	10 min
25-40 PCR Cycles	Denature	95 95	Amplicons >2kb: 15s Amplicons <2kb: 30s
	Anneal	~55 (dependent on Tm)	30s
	Extend	72	1min/Kb
Final extension		72	7min
Hold		4	indefinitely

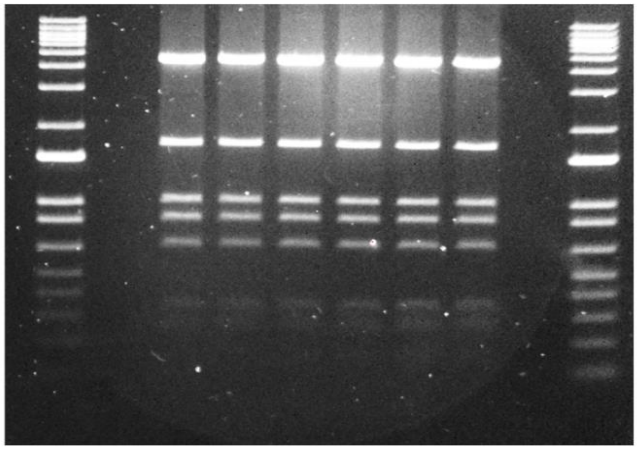
A)



B)



C)



D) Telomere gRNA Scrambled gRNA

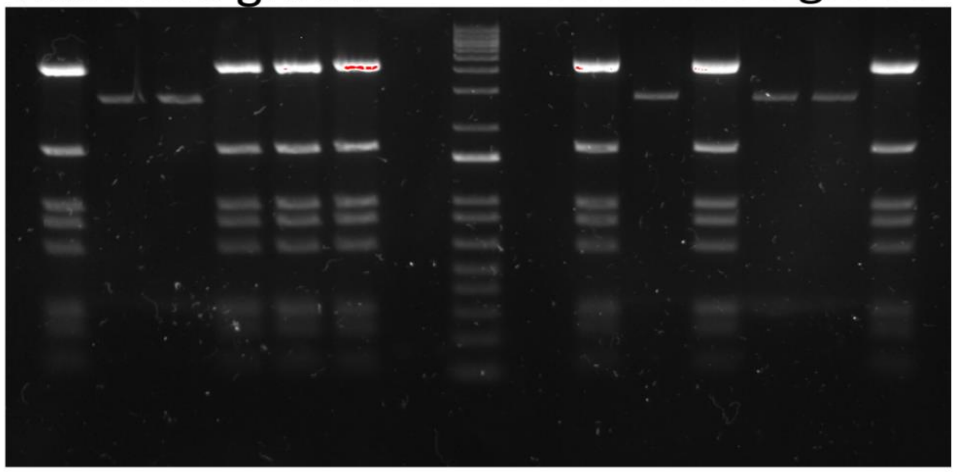


Figure 2.2. PX458-Telo and PX458-Scrambeled restriction digests. A) 1kb+ ladder used in gels. B) In-fusion cloning PCR of Neomycin cassette with homology arms corresponding to NarI linearization of PX458-Cas9 plasmid. Expected band; 1480bp C) Restriction digest with PvuII of in-fusion cloning between Neomycin gel extracted cassette (B) and linearised PX458-Cas9. Expected bands; 4542bp, 1700bp, 1008bp, 859bp, 663bp, 330bp, 240bp, 147bp. D) PX458-Cas9-Neomycin restriction digest with PvuII after inserting Telomere and Scrambled gRNAs to linearised plasmid using Bbsi. Expected bands; 4542bp, 1700bp, 1008bp, 859bp, 663bp, 330bp, 240bp, 147bp. Full Sequencing of final PX458-Cas9-Neomycin-Telo and Scrambled gRNA can be found in the supplementary data.

Bacterial Glycerol Stock Preparation (2.2.6)

Bacterial glycerol stocks were prepared for each plasmid for long term storage. 1ml of overnight culture of transformed bacteria was added to 1ml 50% glycerol (v/v), aliquoted and placed in the -80^{o}C. For reamplification of plasmids, glycerol stocks were thawed and spread on selection plates containing 100µg/ml ampicilin.

Maxi preparations of plasmid DNA (2.2.7)

Supernatant from 249ml overnight growth culture collected after centrifugation at 4000 x g for 1 hour at 4°C. Qiagen's EndoFree Plasmid Maxi Kit (Qiagen, #12362) was used to extract plasmid DNA from the growth culture as per manufacturer's instructions. In brief; Cell pellet was resuspended in 10ml buffer P1. 10ml buffer P2 was added and mixed by vigorous inverting 4-6 times and then incubated at room temperature. 10ml chilled buffer P3 was added to the lysate and mixed by vigorous inverting. Lysate was poured into the barrel of a QIAfilter cartridge and incubated at room temperature for 10minutes. Filter the lysate into a 50ml tube. Add 2.5ml buffer ER to the lysate, mix by inverting and incubate on ice for 30minutes. Equilibrate QIAGEN-tip 500 by applying 10ml buffer QBT and allow the column to empty by gravity flow. Apply filtered lysate to the tip and allow it to enter by gravity flow. Wash tip with 2 x 30ml buffer QC. Elute DNA with 15ml buffer QN. Precipitate DNA by adding 10.5ml of isopropanol, mix and centrifuge at 4000 x g for 1 hour at 4°C. Carefully decant the supernatant and wash DNA pellet with 5ml 70% ethanol (v/v). Spin at 4000 x g at 4°C for 30minutes. Air dry pellet and re-dissolve with 500ul nuclease free water. Nanodrop solution and dilute plasmid to ~1ug/ul.

In-Fusion Primer Design (2.2.8)

Primer design and quality were critical for the success of the In-Fusion (Takarabio, #638920) reaction. In-Fusion allows the joining of any combination of fragments, e.g. vector and insert (or multiple fragments), as long as there are 15 bases of homology at each end. Therefore, In-Fusion PCR primers must be designed in such a way that they generate PCR products containing ends that were homologous to those of the vector, depending on the restriction enzyme used for linearisation of the plasmid (table 2.9). Although 3 primers were designed, NarI associated primer was used for Neomycin resistance gene cloning into PX458 plasmid.

Table 2.9. Primer list for Neomycin cassette amplification with homology arms corresponding to the overhangs created by each restriction enzyme in question.

Cut with PluTI insert beta globin promotor + neo + poly A	
P1	TGCCTGCAGGGGCGCAGCTTTGCTTCTCAATTTCTTATTT
P2	TACCGCATCAGGCGCTCGAGCCCCAGCTGGTTC
Cut with SfoI insert beta globin promotor + neo + poly A	
P1	TGCCTGCAGGGGCGCAGCTTTGCTTCTCAATTTCTTATTT
P2	TACCGCATCAGGCGCTCGAGCCCCAGCTGGTTC
Cut with NarI insert beta globin promotor + neo + poly A	
P1	TGCCTGCAGGGGCGCAGCTTTGCTTCTCAATTTCTTATTT
P2	TACCGCATCAGGCGCTCGAGCCCCAGCTGGTTC

When designing In-Fusion PCR primers, they must have two characteristics; the 5' end of the primer must contain 15 bases that are homologous to 15 bases at one end of the DNA fragment to which it will be joined (i.e., the vector or another insert). The 3' end of the primer must contain sequence that is specific to the target gene.

The 3' portion of each primer should be gene-specific, between 18-25 bases in length, and have a GC-content between 40–60% and have a melting temperature (Tm) between 58–65°C. The Tm difference between the forward and reverse primers should be ≤ 4°C, or amplification will be affected. Tm should be calculated based upon the 3' (gene-specific) end of the primer, and not the entire primer. If the calculated Tm is too low, increase the length of the gene-specific portion of the primer until you reach a Tm of between 58–65°C.

Avoid primers that contain identical runs of nucleotides. The last five nucleotides at the 3' end of each primer should contain no more than two guanines (G) or cytosines (C). Fragments of desired insert DNA with homology arms were generated using the following PCR conditions (table 2.10 and 2.11):

Table 2.10. In-Fusion PCR reagent components.

Reagent	Volume (µl)
In-Fusion PCR Mix	12.5
Primer 1	0.75 (10µM)
Primer 2	0.75 (10µM)
Plasmid DNA	50ng
Water	Up to 25µl
Total	25µl

Table 2.11. PCR cycling conditions.

Temperature (°C)	Time (s)	Cycles
98	10	30-35 cycles
55	15	
72	5	

In-Fusion Cloning (2.2.9)

Plasmid DNA, for which the fragment(s) will be inserted into, is linearised by a restriction digest using an appropriate enzyme (table 2.9). The overhangs generated from this digest must be identified as this determines the primer design, as explained above. Fragments generated with appropriate overhangs were gel extracted as previously described (2.2.3), along with plasmid DNA, for which the fragment will be inserted. A ligation reaction is set up with the following conditions (table 2.12).

Table 2.12. Infusion cloning reaction components.

Reaction Component	Reaction volume/weight	Negative control reaction	Positive control reaction
Purified PCR fragment	10-200ng	-	2µl of 2kb control insert
Linearised Vector	50-200ng	1µl	1µl of pUC19 control vector
5X In-Fusion HD Enzyme Premix	2µl	2µl	2µl
Deionized Water	To 10µl	To 10µl	To 10µl

Reaction mixture was then incubated at 50°C for 30 minutes and placed on ice. 1µl of mixture was then transformed into E. Coli as previously described (2.2.4). Multiple clones were picked and amplified in culture. Positive clones were assessed by restriction digesting extracted DNA and culture is grown for a maxi preparation.

Cytogenetics (2.3)

Chromosome Harvest (2.3.1)

Chromosome harvests are performed on cell lines to capture the cytogenetic state of the cell for analysis using MFISH or telomere FISH. To harvest the chromosomes, 3x10^6 cells per cell line was added to T25 flasks containing 5ml of fresh media (table 2.1). Cells were then placed into a static incubator (37°C, 5% CO_2, v/v) and cultured for three days. 2ml of media is then replaced by 2ml of fresh media in each T25 and 100µl of KaryoMAX colcemid (Gibco, #15212012) is added to the T25s. T25s were placed into the shaking incubator (37°C, 5% CO_2, v/v) overnight.

Cell pellets were collected by centrifuging 15ml falcons containing the cell suspension at 279 x g for 5 minutes at room temperature (RT). The supernatant was discarded, and the pellet resuspended with 5ml of warm (37°C) 0.075M KCL (Sigma, #P5405) and placed into a static incubator (37°C) for 5 minutes. 2ml of pre-chilled fixing solution (-20°C, 3:1 solution of methanol (Sigma, #34860): acetic acid (Sigma, #A6283) was added and supernatant discarded after centrifugation at 279 x g for 5 minutes.

The supernatant was again discarded, and the pellet resuspended in 5ml fixing solution, then incubated at -20°C for 30 minutes. Cells were resuspended in an appropriate volume/density to apply the metaphase spreads to slides to ensure single cell chromosome deposition. Slides were then stored at -20°C until probes were applied.

Telomere Fluorescent in-situ hybridisation (TeloFISH) (2.3.2).

Telomere FISH was utilised to identify telomere sequences within CHO cell lines. After chromosome harvesting and slide preparation (section 2.3.1), slides containing sample metaphases were placed into a coplin jar containing 40ml TBS solution (Agilent Dako, K532711-8) and incubated at room temperature for 2 minutes. Slides were placed into another coplin jar containing 40ml TBS solution and incubated for a further 2 minutes. Slides were treated with an ethanol series of 70%, 90% and 100% (v/v) for two minutes each to dehydrate the slides. Slides were removed from the chambers and left to dry before applying the telomere probe.

5µl of Telomere probe (Agilent Dako, K532711-8) was added to the slides and covered with 18x18mm cover glass and sealed with fixogum (VWR, ICNA11FIXO0125). Slides were placed upright into a humidified chamber (ThermoBrite) and incubated for two hours at 37^{o}C. Slides were removed from the humidified chamber and fixogum and cover slip removed. Slides were placed into a coplin jar containing 40ml rinse solution (Agilent Dako, K532711-8) and incubated for 2 minutes.

Slides were then incubated in 40ml wash solution (Agilent Dako, K532711-8) for 5 minutes at 65^{o}C. Slides were treated by an ethanol series of 70%, 90% and 100% (v/v) for 2 minutes each. Slides were allowed to dry and prewarmed (37^{o}C) 20µl DAPI II counterstain (Abbott Molecular, 06J50-001) applied. Slides were covered with a 22 x50mm cover slide and sealed with fixogum. Images were captured using an Axio Z2 imager using metasystems software (V5.7.4).

Telomere FISH performed on Thermo Brite Elite (TBE) (2.3.3)

The Thermo Brite Elite is an automated FISH staining machine that provides robust solution temperature and incubation time control throughout the staining process, providing consistent hybridisation of the probe. The following protocol is programmed as a protocol within the TBE software.

Slides containing sample metaphases were placed face down into the incubation chamber of the TBE. 30ml TBS solution is added per chamber and incubated at room temperature for 2 minutes under rocking conditions (12 cycles/min). Chambers were drained and TBS is re-added to the chambers and incubated for a further 2 minutes under rocking conditions.

Slides were treated with an ethanol series of 70%, 90% and 100% (v/v) for two minutes each. Slides were removed from the chambers and left to dry. 5µl of Telomere probe (Agilent Dako, K532711-8) was added to the slides and covered with 18x18mm cover glass and sealed with fixogum (VWR, ICNA11FIXO0125). Slides were placed upright into the chambers and chambers were filled with water and incubated for two hours at 37^{o}C. Slides were removed from the chambers and fixogum and cover slip removed.

Slides were placed face down into the chamber and chambers were filled with 30ml rinse solution and incubated for 2 minutes. Chambers were drained and filled with 30ml per chamber of wash solution and slides were incubated for 5 minutes at 65^{o}C. Chambers were drained and slides were treated by an ethanol series of 70%, 90% and 100% (v/v) for 2 minutes each. Slides were left to dry and prewarmed (37^{o}C) 20µl DAPI II counterstain (Abbott Molecular, 06J50-001) applied. Slides were covered with a 22 x50mm cover slide and sealed with rubber cement. Images were captured using an Axio Z2 imager using metasystems software (V5.7.4).

Multicolour-FISH (MFISH) (2.3.4)

MFISH was performed using Metasystems 12XCHamster (D-1526-060-DI) probe set. In brief, coplin jars with 0.1X SSC (Invitrogen, #15557044) and 2X SSC were placed at 4°C, with an additional 2X SSC prewarmed at 70°C. Prepared slides (section 2.3.1) were placed into 70°C 2X SSC for 30 minutes, then removed from the water bath and left to cool for 20 minutes. During this step, 5µl per slide of 12XCHamster probes was prepared in a PCR machine using a program of 75°C for 5 minutes, 10°C for 30 seconds, 37°C for 30 minutes.

Slides were then transferred to 0.1X SSC at room temperature (RT) for 1 minutes and denatured in 0.07N NaOH (Sigma, #S2770) at RT for 1 minute subsequently. Slides were then placed sequentially into 0.1X SSC and 2X SSC at 4°C for 1 minute each and dehydrated in an ethanol (Sigma, #51976) series of 70%, 80%, 90% and 100% (v/v) for 1 minute each. After air drying, 5µl of denatured and prehybridized probe was placed onto metaphase spreads, overlaid with a coverslip and sealed with rubber cement. Slides were incubated in a humidified chamber (ThermoBrite, Leica Biosystems) at 37°C for 1-2days.

After incubation the rubber cement and coverslips were removed, and slides were placed into prewarmed (72°C) 0.4X SSC for 2 minutes. Slides were then placed in 2X SSCT (2XSSC, pH 7-7.5 containing 0.05% Tween20, v/v) at RT for 1-2 minutes. Slides were washed briefly in double distilled water to avoid crystal formation and air dried. 20µl of DAPI/antifade (D-0902-500-DA) was applied to metaphases and a coverslip overlaid. Slides were captured using metasystems automated acquisition platform and images analysed as outlined in the population determining section.

Quantitative Reverse Transcription Polymerase Chain Reaction (qRT-PCR) (2.4)

Cell harvesting for RNA extraction and cDNA conversion (2.4.1)

Cell line samples for productionally stable and unstable gene expression comparison were cultured as outlined in section 2.1. Cell pelleting was performed by decanting the relevant volume of cell culture that equated to 1x10^6 cells in a 15ml falcon tube. Media was discarded by centrifuging cell suspension at 279 x g at room temperature for 5 minutes and resuspended in 1ml PBS (4^{o}C) and transferred to a 1.5ml Eppendorf tube. The 1.5ml Eppendorf tube was centrifuged at 13226 x g for 5 minutes at 4^{o}C to reduce potential transcriptomic changes within the cell before RNA extraction. PBS supernatant was then removed, and cell pellets stored at -80^{o}C for RNA extraction (section 2.4.2).

RNA Extraction (2.4.2)

RNA extraction was performed using RNeasy plus kits (Qiagen, #74134). In brief; cells were resuspended with 350ul buffer RLT plus and decanted into a QIAshredder to homogenise the cells. Homogenised lysate was transferred to a gDNA eliminator spin column and centrifuged at 13226 x g at room temperature for 30 seconds. The column was discarded and 350µl (1 volume) of 70% ethanol (v/v) was added to the flow-through and mixed by pipetting. 700µl of the sample was transferred to a RNeasy spin column and spun down at 13226 x g for 15 seconds. Flow through was discarded and 700µl buffer RW1 was added to the column and spun down at 13226 x g for 15 seconds. Flow through was discarded and 500µl buffer RPE was added to the column and spun down at 13226 x g for 15 seconds. Flow through was discarded and an

additional 500µl buffer RPE was added to the column and spun down at 13226 x g for 15 seconds. Flow through was discarded and columns spun down again to remove any excess buffer. Column was placed in a new 1.5ml Eppendorf tube and 30ul of RNase-free water was added directly to the membrane. Column was spun at 13226 x g for 1 minute to elute the RNA. RNA concentration was elucidated using nanodrop.

RNA to cDNA conversion (2.4.3)

RNA to cDNA conversion was performed using applied biosystems high capacity RNA-to-cDNA kit (Applied Biosystems, #10400745). In brief; PCR reaction tubes were set up to contain reverse transcriptase positive and reverse transcriptase negative reactions per sample conversion. Master mixes were set up as follows (per reaction);

Table 2.13. RNA to cDNA conversion master mix components

Component	Volume/Reaction (uL)	
	+RT	-RT
2X RT buffer	10	10
20X RT Enzyme	1	-
Nuclease Free Water	Up to 20ul	Up to 20ul
Sample (1ug RNA in 5ul water)	5	5
Total per reaction	20	20

On ice, 15µl of RT+ and RT- master mixes were decanted into the relevant tubes. 5µl of the relevant RNA was subsequently added to the reaction tubes. Tubes were sealed and transferred to a thermo cycler. Thermo cycler settings were as follows; 37^{o}C for 60 min, 95^{o}C for 5min, 4^{o}C on infinite hold. Resulting cDNA was diluted to 10ng/6µl using nuclease free water and stored at -80^{o}C until use.

Primer pair specificity assessment and optimisation (2.4.4)

Primers were designed against a panel of genes that encompass DNA damage pathways and telomere homeostasis (table 2.15). Where feasible, primers were designed to span exon-exon boundaries to ensure mRNA template specific amplification. Light cycler 480 (Roche Life Science) 96 well plates were used to conduct qRT-PCR experiments. An example of a plate plan can be found in table 2.14. Master mixes were comprised of the following (per sample); SYBR green (Roche, #04707516001) (2X) 10µl, primer pair solution (2.5uµM) 4µl. Volumes of components were multiplied by the number of samples. 14µl of the master mix was decanted into the relevant wells according to the plate plan. Triplicates of each sample were performed to account for pipetting variation and human error. 6µl cDNA of the relevant cell line, passage and dilution was then decanted into the plates according to the plate plan. Plates were then sealed using Light Cycler 480 sealing foils and spun down at 279 x g for 5 seconds. qRT-PCR was then performed using a Roche Light Cycler 480.

A consecutive dilutional linearity series was performed by diluting cDNA (10ng/6µl) by 1in10, then 1in100, using the 1in10 as a stock solution. This series tests the primer pair's amplification efficiency across a decreasing level of template. A cycle threshold (CT) is the point at which fluorescence of the amplified transcript exceeds background fluorescence (threshold is automatically determined by the software). A small CT signifies greater amount of targeted template within the analysed sample. Conversely, a large CT signifies a smaller amount of targeted template, therefore it requires extra amplification cycles to exceed background fluorescence.

Table 2.14. Roche Light Cycler 480 cycling conditions for SYBR green based qRT-PCR.

Name	Cycles	Temperature (Celsius)	Acquisition mode	Hold (mm:ss)	Ramp Rate (C/s)	Acquisitions (per C)
Pre-incubation	1	95	none	05:00	4.4	-
Amplification	40	95	none	00:10	4.4	-
		55	none	00:30	2.2	-
		72	Single	00:10	4.4	-
Melting Curve	1	95	none	00:05	4.4	-
		65	none	01:00	2.2	-
		97	Continuous	-	0.11	5
Cooling	1	37	none	00:30	2.2	-

Selection criteria for a primer pair was an initial manual assessment of differences between the CTs between each dilution. CTs should differ around 3.3 cycles between each dilution sample. Once a primer pair met this criterion, the amplification efficiency was assessed by plotting a linear curve and calculating the average slope of the dilution series, which is subsequently used to calculate an amplification efficiency percentage;

*(=(10^(-1/slope)-1)*100)*

Primers that obtained an amplification efficiency between 90-110% spanning a range of 10ng to 0.1ng of template, with a specific melting temperature curve (Tm), were accepted into the gene panel. Specificity (Tm) of a primer pair can be defined as a unanimous single peak across all wells used for primer efficiency assessment, during a melting curve protocol.

Table 2.15. Genetic screen primer pairs that have met selection criteria, grouped by their action with DNA damage pathways and telomere homeostasis complexes.

Gene	Primer Sequence		Primer Efficiency (%)	Gene Group
GAPDH 1	Forward:	CAGTGCCAGCCTCGCT	98.7	House Keeping Genes
	Reverse:	TGAAGGGGTCATTGATGGCA		
GAPDH VC	Forward:	GCCAAGAGGGTCATCATCTC	101.2	
	Reverse:	CCTTCCACAATGCCAAAGTT		
Tubulin	Forward:	GGCAACCAGATCGGTGCTAA	98.3	
	Reverse:	ACTTGCCACCTGTGCGTTC		
B2M	Forward:	ACGGAGTTTACACCCACTGC	99.6	
	Reverse:	CAGACCTCCATGATGCTTGA		
Terf2IP	Forward:	GCGCCTCACCTACACGAAC	94.7	Slx4 complex
	Reverse:	CCTTGAGAGACTGCCAGGAG		
Slx4	Forward:	GTGCCTGCCCAGATACTTGT	99.7	
	Reverse:	CACACATGAGAGGAGCCCAG		
Xpc	Forward:	GAGAACAACAAAGCAGCCCG	99.4	DNA Damage Binding
	Reverse:	TCAAAATCATCCGCCTCGCT		
Xrcc1	Forward:	CCTTCTGAGAGCCGAAGTGG	99.7	
	Reverse:	AGGCCATAGGGTGAGTCCTT		

Ercc1	Forward:	TGGACCTTGGGAAAGACGAG	113.4	
	Reverse:	CAAACTTCTTCCTGGTGGGC		
Rad51b	Forward:	CTTGCTGAAGTACCTGGCTGA	100.2	
	Reverse:	GCAGTTCGTTGCCTTCAATCC		
Brca1	Forward:	CAAACCACCTCTCTGGGAGC	102.1	
	Reverse:	CCTCAGAAAACTCACAAGCAGC		
Lig4	Forward:	TGTGTGATTGCAGGCTGTGA	113.2	
	Reverse:	ATCATAAAGCGGGGTTGCCA		
H2ax	Forward:	GCGCGCCGGTCTACAATTC	98.9	
	Reverse:	TCGGCATAGTGGCCTTTCCT		
Xpa	Forward:	TGCCAACATGTGATAGCTGC	94.9	
	Reverse:	CGATAAACCTAAGCGCAGGC		
Rad51	Forward:	AAAGATGCGGAGGTCTGCTG	97.6	
	Reverse:	CCTGGCGGACCTGTAATCTC		
Smc3	Forward:	ACATTGAAGAGCGGTTGCAC	98.6	Cell Cycle
	Reverse:	GCACGCCTCATCTTATCCCA		
Rad17	Forward:	CCCTGACAGTGGGGATGAAC	107.9	
	Reverse:	CAGGTTTCTGGGTCAGCAGT		
Tlk1	Forward:	AGAGTCGGAGACACCAGAGA	102.0	
	Reverse:	GCCACGTCCCCCAATACTTT		
Parp1	Forward:	CAGTAAGCTGGAAAAGGCCC	110.7	
	Reverse:	TCCAAGATCGCAGACTCTCCAG		
Mbd4	Forward:	CCAGTGCTCTGAGCTGTGTT	99.2	
	Reverse:	TTCTCTAGGGGTGACGGAGG		
Rbbp4	Forward:	GTCCAGCTTCCGAACGATGA	100.9	
	Reverse:	GACAGAGCCAAAACCTCCGA		
Rad21	Forward:	AGTACCCCCAAGAACCCAGA	101.0	Shelterin Complex Regulation
	Reverse:	TCTTCCCGTCAAGGGACTCA		
Rassf1	Forward:	TGCATCCCTCTGCCCCTTAT	97.0	
	Reverse:	CCCTGCAAACAGAACCTCGT		
Krit1	Forward:	TGGGAAAGTTGAGGCCACTC	106.0	
	Reverse:	AGCAAAGTGAAGAGGGGAGC		
Dclre1b	Forward:	TGCAATCCAGCCTTGGTTCT	106.9	
	Reverse:	GCTAGTTGCTCCAGCAGTGA		
Rapgef1	Forward:	CCTACTGGGTACGCTCCATC	95.6	
	Reverse:	CCGAATCCAGAGCTGAGAGG		
Dkc1	Forward:	TGGTGGTCAGATGCAGGAAC	97.2	Telomerase Associated Complex
	Reverse:	AACAACACGCCGCAAGTAAC		
Gar1	Forward:	CTCCGGAACGTGTCGTCTTA	102.4	
	Reverse:	AGGAGCGTTGAAGTAAGGCA		
Nhp2	Forward:	AGCTTCATGACCCGGAAGTG	95.6	
	Reverse:	GAGACTGCGGCAATCCAGTA		
Nop10	Forward:	GGCGATCGCGTCTATACGTT	99.3	
	Reverse:	CTGGGTCATGAGCACCTTGA		

Ercc4	Forward:	CCCCGACGTGTGACAAATGA	99.5	Telomerase Maintenance
	Reverse:	ACACCAGGATGCCGGTTATT		
Acd	Forward:	AGGTTCAAACTGCCAGGCTT	100.2	
	Reverse:	CAAGTTCAGGGCCCAAGCTA		
Blm	Forward:	TACCAAGAATCTGGCCGAGC	94.2	
	Reverse:	CTCTTCAGTCTGGTCACATCGT		
Ptges3	Forward:	TCACAAGACAGCGATGACGA	89.7	
	Reverse:	CCTGGCGATGACAACATTCC		
Xrcc6	Forward:	GGGCCAGGACCAAAGCTAAT	96.8	
	Reverse:	AGTGAACCCCGAGATCCTCA		
Tnks2	Forward:	GTTACGGCTCTGACCCTTCC	99.6	
	Reverse:	CATTGCCCATTTGTGCTGCT		
Rtel1	Forward:	GACAGACCCAAACAGCTTGC	98.6	
	Reverse:	GGGATCCCAGCACACATACC		
Tep1	Forward:	ACATCTGTCCACCCAGACATT	96.1	
	Reverse:	CAGCGGAGAGGGTAGCTAGA		
Hnrnpd	Forward:	AGCCAAGGTTACGGTGGTTA	102.5	Telomere associated genes
	Reverse:	CCTCGCCTGGATACTTTCCC		
Cdk2	Forward:	CGGATCTTCCGAACTCTGGG	95.3	
	Reverse:	TCATCCAGTGGAGGCACAAC		
Hsp90aa1	Forward:	CTGCGTATTTGGTTGCCGAG	97.5	
	Reverse:	CGACCCATTGGTTCACCTGT		
Hat1	Forward:	TACGCTCTTTGCGACCGTAG	95.4	
	Reverse:	ATCTGACTTACACGTGGCCG		
Hnrnpa2b1	Forward:	AGCTGAGGCAATTTTTGGTGT	122.3	
	Reverse:	AGTTTGCGGAACTGCTCCTT		
Rif1	Forward:	CTACTTTGCTCGTACACCCCT	99.7	
	Reverse:	CCAGTGCCTGAGTTGTTGGT		
Sun1	Forward:	TTCCTAGTTCAACTTTTTCAAGTGT	93.4	
	Reverse:	AGTGACCAAGATGGGCTTCG		
Pinx1	Forward:	TCGGCAGATTTGCGACAATG	100.4	Telomerase Regulation
	Reverse:	CCACCCCATCTTCTCAAGCA		
Esf1	Forward:	GGAGCACCAGAATCTGAGCA	103.5	
	Reverse:	TGCCTGAAATCGTGAGTCGT		
Men1	Forward:	GCAGTGACTGGGAACCTTCA	102.2	
	Reverse:	CAAGTGGGAGGAATGCCGAA		
Ppp2r1b	Forward:	TTTAATCGACGAGCTCCGCA	100.4	
	Reverse:	TCGAGTCCTTTCCACTCCGA		
Ssb	Forward:	TGTGGTGACTGTGGGATCAG	98.3	
	Reverse:	GCACCCAGCCTTCATCCAAT		
Atp5a1	Forward:	AAGAGAACGGGTGCCATTGT	98.0	
	Reverse:	ACTCGTCTACGGGTCTTGGA		
Pparg	Forward:	TTATTCTCAGTGGAGACCGCC	101.0	

	Reverse:	CCTCGATGGGCTTCACGTTC		
Sart1	Forward:	CAGCGGTGAGAAGGTGGTAG	100.9	
	Reverse:	CGGCAAAACTCAGAAGTGGC		
Smad3	Forward:	GATGCAGGCTCTCCAAACCT	100.3	
	Reverse:	AATGTCTCCCCAACTCGCTG		
Rb1	Forward:	CCTGCACTACTCCGAGAACC	89.6	
	Reverse:	GGGTGTTCGAGGTGAACCAT		
Ppp2r1a	Forward:	CGTGCGTGAATATTGCCCAG	105.1	
	Reverse:	GGTGGGCATCACTAAAGCCT		
Mtch1	Forward:	GGCCCATCCTTTACACGTCA	98.0	
	Reverse:	AAGCCGACGAAGAATCCCAG		
Eif5b	Forward:	CCCGGACATGAGTCTTTCAGT	97.3	
	Reverse:	TCTGGGGCTCTAAACCATGC		
Akt1	Forward:	GTGAGCGTGTGTTTTCCGAG	95.7	
	Reverse:	GATAGTCCAGGGCTGACACG		
Atrx1	Forward:	ACTCACCCCTGTTTCGCTTT	104.1	Genes related to DNA repair
	Reverse:	ACGGCCATCCTTCTCTTGTG		
Rev1	Forward:	ACAAGTAACAGAGGCACGGG	97.2	
	Reverse:	CTCCCATACTGATGAATCGGGT		
Tdg	Forward:	TCCTGGTGGCGTTCCAAATG	90.8	
	Reverse:	CTTCTTCCTGCTCCTGTGTCC		
Smc1a	Forward:	AGAACGCTTGACAGAGGAGC	98.1	
	Reverse:	ACTGCACCTGACGTAGTTCG		
Trip12	Forward:	AAGACCAGAACAAGGCCAGG	99.3	
	Reverse:	TGAAACGAGCAGCCCAATCA		
Trip11	Forward:	CGGCATCGAGAGGAGCTAAG	104.7	
	Reverse:	CCTGAGCCACCTTGCTGTAA		
Rbm41	Forward:	ACCACAAAAAGACAAGTGCAGA	101.3	
	Reverse:	GGGGGTTCATCTTTTGTGCC		
Rbm43	Forward:	GGACCTTCTTCCGACACGTT	104.7	
	Reverse:	CCCCGCCTTTAGTGACCTTT		
Polh	Forward:	TCGTGAATCCCGTGGGAAAG	101.2	
	Reverse:	TTGCACAGCACTGGTCAGAT		
Apex1	Forward:	GGATATAGTGGCGTGGGCTT	99.9	
	Reverse:	CATGCTCTTCCTCACCAATGC		
Gtf2h1	Forward:	ACCAGATGGTGCCAAACGAT	97.4	
	Reverse:	ATGCCGCAGAAGTTCTCCAA		
Mif	Forward:	GCCACCATGCCGATGTTCAC	98.4	
	Reverse:	CGTGCACTGCGATGTACTGG		
Rad9a	Forward:	GCTGATGTCCTACCTTGCCC	98.0	
	Reverse:	TGACCCCATACTTGCAGTGG		
Pttg1	Forward:	ACCAGATTGCACATCTCCCC	97.0	
	Reverse:	AAGGGATTGGATTCCCACGG		

Rad18	Forward:	GGCTGGCGATGATGAAGACA	103.8	
	Reverse:	CAAAGCAAATCCCACAGCGA		
Pttg1	Forward:	TCTTGCCTAAAGCCAGCAGA	103.6	
	Reverse:	TTGCTTCTGTTTGAGGGGTCT		
Trip12	Forward:	AGCTGGAGGCAGCTTTTCTC	96.6	
	Reverse:	TCAATCTTGACAGGACCGCC		
Trip10	Forward:	AACTGCAGGTGGTTCCCATT	102.3	
	Reverse:	GGCGAACCCTGACTTGTGTA		
Smc1a	Forward:	GTCGCTCTGGGGAATTAGCA	100.5	
	Reverse:	TATCGGTCCGCCTCCTCTTT		
Trip11	Forward:	GAGTCAGAAGTGCCAGACCC	95.4	
	Reverse:	CTCCAACCTCGCTACCTGTG		
Ube2a	Forward:	GCTGGAGCCCAACCTATGAT	98.4	
	Reverse:	GCTGTTCGCTGGACTATTGG		
Tdg	Forward:	GTCAGGACTGAGTGAGGTTCA	103.3	
	Reverse:	GCCAATACCATACTTTCCGGG		
Cry2	Forward:	TTCTACTACCGCCTGTGGGA	102.3	
	Reverse:	TCCATTCGGTCAAACCTGGG		
Apex1	Forward:	ACTTACTGACAGCTTCCGGC	102.2	
	Reverse:	CGACATTCTTAGCCCGAGCA		
Wrn	Forward:	GAAGGACAAGTCCATCCGCT	103.4	
	Reverse:	ACACTGGCACAGTATCACCC		
Ung	Forward:	CCTGCTCTGGGGCTCTTATG	102.7	
	Reverse:	CTCTGTACACCGACAACGGG		
Rbm45	Forward:	AGAGCACTCTTGGCTGAACC	102.2	
	Reverse:	CCCCAACGAAGGGAAACAGA		
Fen1	Forward:	ATGTTCTACCGCACCATCCG	98.6	
	Reverse:	TTTTCAGCTTCAGCTCGCCT		
Polh	Forward:	GTCCTGGCAAAACTCGCTTG	92.9	
	Reverse:	ACGAATTGGCATCTGGCTGA		
Rbm42	Forward:	GCCGATGACTTCCGGATCTT	97.4	
	Reverse:	GCCATAGCCCTTGGTTTTGC		
Pot1	Forward:	ACCTTTCAGCACCATTCCTCA	98.2	Shelterin Complex
	Reverse:	TCCTAGAACTTCTGCCACAGC		
Tpp1	Forward:	ATACCTGATGAGTGCTGGTGC	102.6	
	Reverse:	GTGTCACCTGAGGCAAAAAGTA		
Trf1	Forward:	TCTGATGAAGGCAGCAACAAA	100.1	
	Reverse:	TCAGTTTCTGACTGTTGGTCAT		
Trf2	Forward:	TCATGCAGGCGTTGCTTGTC	94.6	
	Reverse:	CTGATTCCAGGGGTGTGAGC		
Tin2	Forward:	GCCAAGAAACCAATGCAGTCC	109.3	
	Reverse:	TGCAGGGGTAGATGGAGAGAT		
Rap1a	Forward:	AATGGCCAAGGGTTTGCACTA	98.9	
	Reverse:	TCTTTGCCAACTACCCGCTC		

RAp1b	Forward:	GCTAGTCGTTCTTGGCTCCG	112.2	
	Reverse:	TTGCTGTAAATTGTTCCGTTCCT		
Tert	Forward:	CATAGCACCTGCCACCGAT	116.3	
	Reverse:	CAGCAGCCTGTTTGACTTCTTC		
Mre11a	Forward:	TCTCAGAGAGGCCGAGACAC	96.7	MRN Complex
	Reverse:	GGAAGGCTGCTGTCGAGTAG		
Nbn	Forward:	CCGAAACCATGCTGTGCTAAC	99.9	
	Reverse:	TTCAAAGTGTAGGGAAGGCCA		
Rad50	Forward:	CGGGGAAGACGACCATCATT	105.7	
	Reverse:	TCGAAACTGCAGGCGAATCT		

Flow Cytometry Analysis (2.5)

gH2AX staining (2.5.1)

10x10^6 cells were inoculated into T25s in the appropriate media (table 2.1). CHOK1a-GS-KO host cells was used as a gating control to identify single cell gates and used as a negative control for gH2AX staining. CHOK1a-GS-KO supplemented with 1ng/ml Neocarzinostatin (Sigma, N9162), a DNA damage inducer that is incubated for 1 hour at 37^{o}C in a shaking incubator, was used as a positive control. 10x10^6 cells of each sample were transferred into separate 15ml falcon tubes, placed on ice to avoid increases in DNA damage due to cell handling procedure, and washed with PBS at 4^{o}C, topping up to the falcon tube to 15ml. Supernatant was discarded after centrifugation at 279 x g at 4^{o}C for 5 minutes. The pellet was resuspended using 2ml cold PBS and transferred to a 2ml Eppendorf tube. Samples were centrifuged again at 176 x g at 4^{o}C for 5 minutes.

The supernatant was discarded again, and the pellet resuspend using a pulse vortex. 100µl of intracellular fixation buffer (Invitrogen, #00-5523-00) was added to the samples and incubated for 20 minutes. 2ml of 1X permeabilisation buffer (Invitrogen, #00-5523-00) was added to the samples and centrifuged 176 x g for 5 minutes at room temperature. The supernatant was discarded, and the wash step repeated. The pellet was resuspended with 100µl of permeabilisation buffer and 5µl of gH2AX-PE (Ebioscience, #12-9865-42) was added per sample and incubated for 1 hour at room temperature in the dark. Samples were then washed twice as previously described and resuspended in 300µl flow cytometry staining buffer (Invitrogen, #00-4222-26) and passed through a 40µM cell strainer (Falcon, 352340) to

produce a single cell suspension. Samples were analysed on a BD Aria II, using BD FACSDiva software.

Intracellular IgG staining (2.5.2)

Samples were prepared, permealised and prepared for staining as outlined in section 2.5.1. Cell samples were stained an anti-human IgG-647 (Invitrogen, Cat# A-21445) antibody, to assess intracellular IgG expression levels.

BD Aria Gating Strategy (2.5.3)

Voltages of the corresponding lasers were set against the appropriate negative controls and positive controls. Voltages were manipulated to ensure single cell populations were observable on the scatter graph using forward and side scatter axes. Negative and positive samples for the appropriate staining were used to designate positive and negative gates, ensuring there was a large shift between the two corresponding peaks. Once voltages had been set, they remained constant for all sample analysis, allowing data to be directly compareable. A full list of controls can be found in table 2.16.

Fluorescence-activated cell sorting (FACS) for CRISPR-GFP enrichment (2.5.4)

After transfection and scale up of Telomere and scrambeled CRISPR-Cas9 transfected cell lines (outlined in sections 2.2.1, 2.1.6, 2.1.7 and 2.1.4), 3x10^6 live cells of each sample were passed through a 40μM cell strainer (Falcon, 352340) and analysed on the BD Aria for GFP expression. As GFP is conjugated to the Cas9 protein, GFP provides a method to assess the number of cells that express the Cas9 protein within a cell line. Single cell and negative gates were assessed as previously described (section 2.5.1 and 2.5.3) using CHOK1a-GS-KO host cell line and non-transfected producing cell line control (protein 3). Positive GFP gates were used to single cell sort clones of PX458-Telo and PX458-Scrambeled into 24-well plates containing conditioned media (section 2.1.6), to obtain clones of the transfected pool that was used to assess telomere specific damages impact on production instability. The method outlined in this section was also used to enrich cell lines for GFP positive cells, as GFP expression drift was observed during maintenance of cell culture (explained in section 5.2.2, figure 5.3 and 5.4). Cell lines were maintained as outlined in section 2.1.4.

Table 2.16. List of controls used for appropriate gating strategy for Flow cytometry analysis.

Control Name	Stained?	Positive or Negative?	Control for?
Live CHOK1A-GS-KO	Unstained	Negative	All staining's
Fixed and permeabilised CHOK1A-GS-KO	Unstained	Negative	All staining's
Fixed and permeabilised Protein producing cell line	Unstained	Negative	Intracellular IgG-647
CHOK1A-GS-KO Treated with Neocarzinostatin	Stained	Positive	anti gH2AX-PE
CHOK1A-GS-KO Un-treated	Stained	Negative	anti gH2AX-PE
CHOK1A-GS-KO IgG	Stained	Negative	Intracellular IgG-647
Lead cell line	Stained	Positive	Intracellular IgG-647
CHOK1A-GS-KO Treated with Neocarzinostatin	Stained	Negative	Intracellular IgG-647 anti gH2AX-PE
Live Protein 3	Unstained	Negative	CRISPR-GFP

Immuno-Fluorescence (2.6)

TRF2 and Telomere Co-Staining (2.6.1)

Samples were fixed and permeabilised as per the flow cytometry analysis (section 2.5.1). 20,000 cells were cytospun onto a super frost slide (Thermo Scientific, #10149870) and blocked for 30 minutes using 0.5% donkey serum (v/v, Sigma, #D9663) in permeabilisation buffer (Invitrogen, #00-5523-00). Blocking and permeabilisation allows for increased staining efficiency (permeabilisation) and reduction of non-specific signals (blocking). Slides were washed 3 x 5 minutes with DPBS (Ca^-/Mg^-) (D8537) and 1/200 TRF2-Alexfluor488 diluted in permeabilisation buffer was added and incubated over night at 4°C. Slides were washed 3 x 5 minutes with DPBS and 4% paraformaldehyde (v/v, PFA) was added to the slides for 10 minutes at room temperature. The addition of PFA is used to stabilise the TRF2 bound antibody to prevent signal decrease upon hybridisation of the telomere probe.

Slides were washed 1 x 5 minutes with DPBS and DAPI (50µg/ml stock) was added to the slides at final concentration of 1µg/ml and incubated for 10 minutes. Slides were washed a further two times to remove excess DAPI. Slides were then dehydrated with an ethanol series of 70%, 80%, 90% and 100% (v/v) for 1 minute in each condition. Slides were allowed to dry and 5µl of Telomere (PNA-TTAGGG(n)-Cy3) probe (Agilent, #K532611-8) was added and covered with a coverslip. Slides were then placed into a ThermoBrite (Leica Biosystems) and incubated at 85°C for 5 minutes, followed by 37°C for 2 hours, to allow probe hybridisation under humidified conditions. Slides were imaged using a Zeiss Imager Z.2 and analysed as outlined in the image analysis section (2.7).

gH2AX and Telomere Co-Staining (2.6.2)

gH2AX and Telomere co-staining was used to assess overall and telomere specific DNA damage within stable and unstable cell lines. Co-localisation of telomere and DNA damage signals were quantified to represent telomere specific DNA damage. To create a DNA damage positive control, CHOK1-GS-KO host was treated with Neocarzinostatin as previously described (2.5.1). CHOK1-GS-KO untreated host was utilised as a baseline negative control for comparison of DNA damage against stable and unstable producing cell lines. 3x10^6 cells of each cell line sample were aliquoted into 15ml falcon tubes and placed onto ice to reduce increases of DNA damage due to cell handling procedure. Ice cold DPBS (Ca^-/Mg^-) (D8537) was added up to 15ml Tubes to was cell samples.

The supernatant was discarded after centrifuging samples at 279 x g for 5 minutes at 4^{o}C and this step was repeated for a second time. Samples were fixed using -20^{o}C 70% ethanol (v/v) solution in a dropwise manor under slow vortexing conditions. 5ml of -20^{o}C 70% ethanol (v/v) solution was then added to the samples and then incubated at -20^{o}C for 1 hour to fix the cells. 20,000 cells were cytospun onto a super frost slide (Thermo Scientific, #10149870) for staining. Slides were blocked for 30 minutes using 1% donkey serum (v/v, Sigma, #D9663) in permeabilisation buffer DPBS (Ca^-/Mg^-) (D8537).

After blocking and permeabilisation, slides were washed 3 x 5 minutes with DPBS (Ca^-/Mg^-) (D8537) and then incubated with 1/5,000 gH2AX primary antibody diluted in permeabilisation buffer over night at 4°C. Slides were washed 3 x 5 minutes with DPBS and then stained with 1in500 anti-rabbit IgG, conjugated with Alexafluor488 and diluted in permeabilisation buffer, for 1 hour at room temperature under humidified conditions. Slides were washed 3x5minutes and 4% paraformaldehyde (v/v, PFA) was added to the slides for 10 minutes at room temperature, to stabilise

Slides were washed 1 x 5 minutes, DAPI (50µg/ml stock) was added to the slides at final concentration of 1µg/ml and the slides were washed a further two times. Slides were then dehydrated with an ethanol series of 70%, 80%, 90% and 100% (v/v) for 1 minute in each condition. Slides were allowed to dry and 5µl of PNA-TTAGGG(n)-Cy3 probe (Agilent, #K532611-8) was added and covered with a coverslip. Slides were then placed into a ThermoBrite (Leica Biosystems) and incubated at 85°C for 5 minutes, followed by 37°C for 2 hours, under humidified conditions. Slides were imaged using a Zeiss Imager Z.2 and analysed as outlined in the image analysis section.

Image Analysis (2.7)

Manual analysis (2.7.1)

Population determining (manual) (2.7.1.1)

Single cell karyotypes of cell lines identified and captured using MFISH staining (section 2.3.4) were used to determine karyotypically distinct populations within the cell line culture flask. Populations were elucidated through analysing each individual image that represents a single cell. A new population is defined by witnessing a mutagenic event (such as a translocation). Cells that obtain identical karyotypes are considered a single population and provided a population identification number (e.g. population 1). When observing a mutation event that does has not been witnessed before, a new population identifier is designated (e.g. population 2).

A translocation is confirmed by reviewing the DAPI channel image to ensure that the chromosomes were attached to each other, rather than in close proximity. A quantitative measure of fluorescent intensity of each image channel can be obtained using cell profiler (https://cellprofiler.org/, section 2.7.2.1). Fluorescent intensity colours shift between mutated chromosome and the modal chromosome (chromosome configuration that is obtained with the highest frequency within the sample analysed, regardless of population), is then assessed. The colour percentage change between the chromosome colour profiles, in addition to the qualitative assessment of DAPI, to ensure the chromosomes represent a single entity, allow the confirmation of a mutation within a new population. Note; multiple mutations can occur within a new population.

Missing or increased numbers of chromosomes must be confirmed by witnessing 3 metaphase spreads that contain the same aberration. This is to ensure that the aberration is not an artefact of the metaphase spread preparation – as outlined in the European Cytogenetics association guidelines (https://www.e-c-a.eu/en/GUIDELINES.html). The frequency of each metaphase belonging to each population forms the basis of clonal or non-clonal chromosomal aberration (CCA/NCCA) designation. A clonal chromosomal aberration (CCA) is defined as a population that comprises >5% of the total population and considered as a chromosomally stable population, as it has established itself as a dominant population. Non-clonal chromosomal aberrations (NCCA) were defined as <=5% of the total population (Henry Heng et al, Molecular Cytogenetics, 2016) and represents a chromosomally unstable population. Increased numbers of NCCAs in the total population indicate an increased mutagenic background leading to greater overall chromosomal instability (CIN). 50 images per cell line are analysed.

Automated (2.7.2)

Confirmation of mutations using Cell Profiler (2.7.2.1)

The following workflow is performed on Cell Profiler (https://cellprofiler.org/) (V3.0.0). Single channel images were exported from the Metafer software (Metasystems, V5.7.4) in .tif format. Images were selected based on their spread. Chromosomes that were in close proximity or crossing each other were not selected for analysis as they do not segment well with this workflow. 6 single channel images were thresholded using the thresholding module with the Global-Otsu algorithm selected using a 1.1 correction factor threshold value to remove background fluorescence. Images were smoothened after thresholding, using a gaussian filter.

Edges of chromosomes within the image were enhanced, to improve identification of chromosomes, by using a sobel algorithm module. Identify primary objects is used to identify chromosomes within an image using the modules automatic threshold strategy. Resulting image masks were then manually edited using the edit image module with DAPI channel image as a guide, to allow for faithful masking of the original image. Each chromosome is arbitrarily assigned a number and this chromosome identifier stays consistent throughout all populations, even if a mutation event has occurred. The numbering system of chromosomes is kept consistent across cell lines to allow comparisons between different populations from different cell lines, as outlined in section 2.7.1.1.

Fluorescent intensity values of each chromosome and single channel were extracted using the measure object intensity module and exported to an excel sheet. Fluorescent intensities were converted into percentages with the sum of all channels equalling 100%. Mutations within chromosomes that have been visually identified were then

confirmed using the fluorescent intensities colour combinations extracted by this method, as outlined in section 2.7.1.1.

Telomere length quantification using Cell Profiler (2.7.2.2)

Telomere length was quantified in a semi-automated fashion using a similar strategy as outlined in section 2.7.2.1, however telomere and DAPI fluorescent intensity is only quantified. The following workflow is performed on Cell Profiler (https://cellprofiler.org/)(V3.0.0). Single channel images (figure 2.3) were thresholded using the thresholding module with the global Otsu algorithm selected using a 1.1 threshold correction value to remove background fluorescence. Identify primary objects is used to identify chromosomes within an image using the automated strategy within the module to create chromosome masks. Chromosome masks were manually edited using the edit mask module to ensure faithful masking of the image. Telomere signals were then identified within the chromosomes with the identify secondary objects using the global Otsu algorithm. The two thresholded images were related using the relate images module (figure 2.4), to ensure fluorescent intensity quantification of telomeres that reside within the chromosome mask regions. Values of the number of chromosomes and fluorescent intensity of telomeres within said chromosomes were extracted by the measure object intensity module and exported in an excel. Telomeres were quantified by representing their fluorescent intensity as a proportion of DAPI staining.

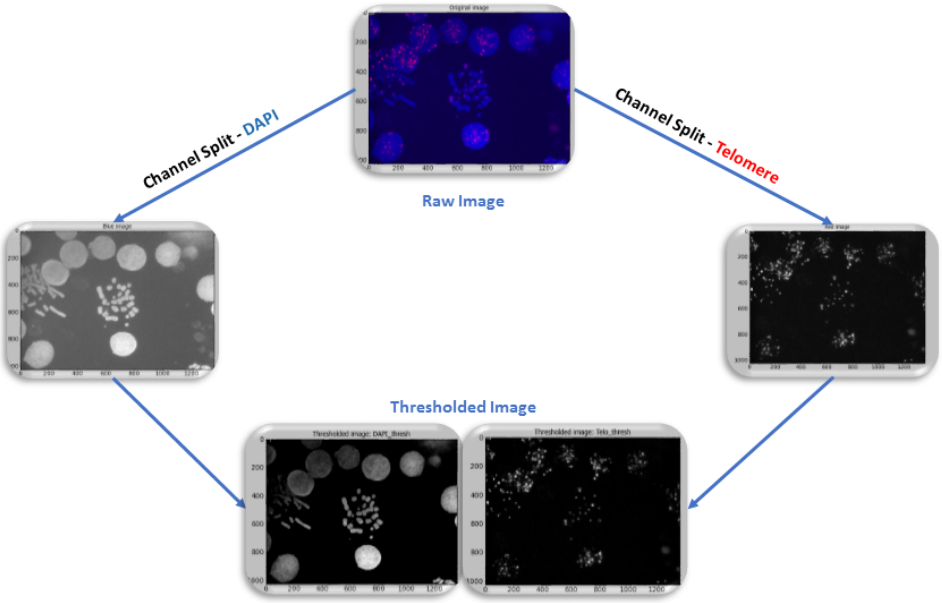


Figure 2.3. Channel Splitting. Image splitting into the separate channels and automated thresholding of image to remove fluorescent background of image

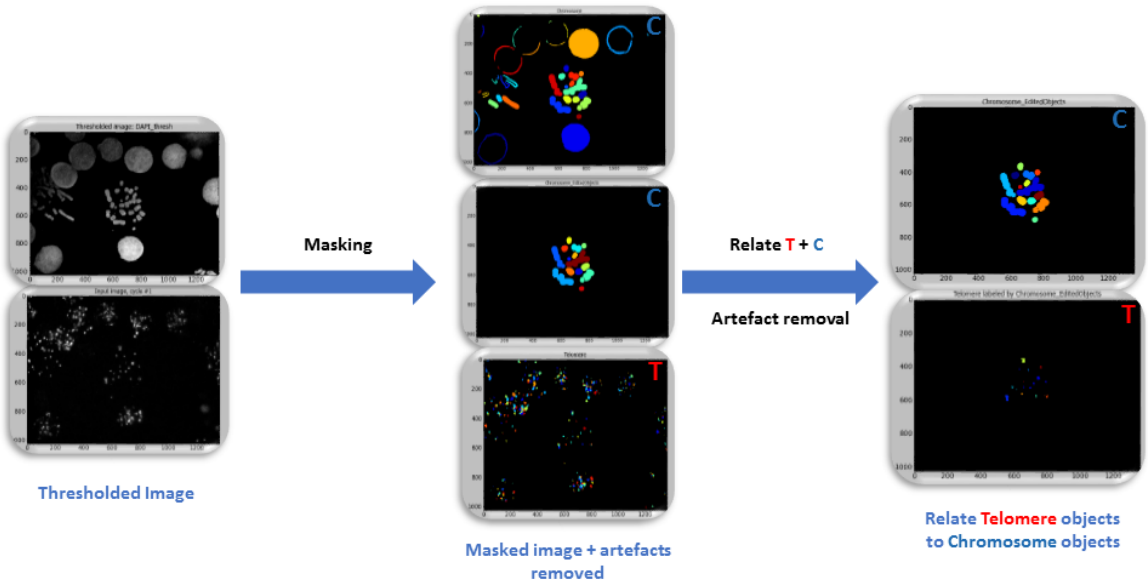


Figure 2.4. Image editing and relating. Thresholded images are masked and semi-automated edited to solely retain chromosome masks. Telomere and chromosome images are related so only the telomere signals within chromosomes are calculated.

Population determining (automated) (2.7.2.3)

Population determination, as outlined in section 2.7.1.1, is a laborious process. Therefore, this method was recreated using machine learning algorithms (figure 6.9) to reduce analysis time, increase analysis output and reduce analyst bias. The automated workflow relies on accurate segmentation of chromosomes in images (chromosome masks). Image segmentation is achieved through a uNet convolutional network (https://arxiv.org/abs/1505.04597) with a contracting and expansive path that results in a feature vector mapped to a desired number of classes. A weight map, identified through a manually segmented training set (as outlined in 2.7.2.1), forces the network to learn small separation boarders between closely located chromosomes. The deep learning network can automatically segment chromosomes that were in close proximity and naturally ignores cells that may be present in the input image, providing a faithfully segmented image.

Once chromosomes have been faithfully masked, a pseudo colour is applied to pixels within the chromosome mask region to reduce heterogeneity of colours that may be incurred due to different hybridising efficiencies across different staining experiments. Chromosomes were pseudo coloured by clustering their pixels in 5D greyscale space using a 12-component Gaussian Mixture model. They were described by an 11D vector where the *i*th element is the proportion of the chromosome with pseudo colour *i*. Each individual chromosome of two images were compared using a linear assignment algorithm, generating a matching cost matrix where similar chromosomes return a low matching cost (genetic/colour similarity) and mismatched chromosomes return a high matching cost (genetic/colour dissimilarity, figure 6.11).

The smallest cost for a single chromosome (when compared to all other chromosomes in another image) is returned to the user. A low cost represents chromosomes that were similar based on their fluorescent colour make up within the chromosome mask. A high cost indicates a mutation event which has occurred as the fluorescent colour of the new chromosome deviates heavily from the original chromosome. As the colour codes are chromosome specific, we related the shift in colour to a genetic change within that chromosome. Therefore, a high matching cost identifies genetic mutations between two populations. Each subsequent image can be assigned to either a new population or is assigned one that has already been identified, providing a frequency score of each population that can be designated CCA or NCCA (as outlined in section 2.7.1.1).

Chromosome number counting (2.7.2.4)

Chromosome number counting was performed using Fiji (Image J, version 1.5.1) by utilising the cell counter module of the software. 50 images of each time point were loaded into Fiji and the cell counter initialised. Images containing appropriately spread metaphase chromosomes were used to ensure all chromosomes were derived from a single cell source. After counting, analysed images were saved to include the counter markers and data saved into an excel sheet.

TRF2 and gH2AX co-localisation quantification with telomere foci (2.7.2.5)

TRF2 and gH2AX colocalization with telomeres were quantified from staining protocols outlines in sections 2.6.1 and 2.6.2. Single foci of TRF1 or gH2AX (green), Telomere signals (red) and co-localised signals (orange) was quantified using Fiji (Image J, V.1.5.1, utilising the cell counter module of the software. Overlapping signals were quantified when green and red signals occupied the same location within the cell nucleus. Adjacent signals were counted as individual foci. Overlapping signals were confirmed through visual inspection of split channel Gray scale images. Quantifications were performed on 50 images per cell line.

Data Analysis and Graph Generation (2.8)

All graphs presented here were produced with JMP software (version 14), unless otherwise stated.

Statistical Analyses (2.9)

All statistical analysis was performed with either JMP or InVivoStat software (version 3.7, http://invivostat.co.uk/). Explanations of all statistical tests performed are outlined in the relevant result sections.

Chapter 3: Baseline characterisation of CHOK1a-GS-KO Host cell line to assess telomeric, mutational and DNA damage baseline profiles.

Introduction (3.1)

In normal homeostasis, telomeres are situated at the extreme ends of chromosomes. Telomeres are formed by G-rich repeats (TTAGGGn) and protected by shelterin, a 6-membered protein complex, which binds specifically to telomeres and inhibits DNA damage pathways at the single stranded telomeric DNA through its sequestering by POT1 (de Lange, 2005). In Chinese Hamster Ovary (CHO) cell lines, interstitial telomere sequences (ITS) are in abundance compared to the extremities. Shelterin complex is known to bind to ITS, however, its inhibitory action on localised DNA damage is less well defined (Schmutz and de Lange, 2016).

At the extreme ends, upon cellular division, telomere length shortens until they reach the Hayflick limit, the critical length of telomeres where apoptosis is triggered (Hayflick, 1965; Hayflick and Moorhead, 1961). Shortening of the telomeres to this critical length results in a significant loss of shelterin complex and de-protection of the ssDNA leading to the activation of DNA damage response (DDR) pathways. DDR pathways, through the action of Ataxia-telangiectasia mutated (ATM) and ataxia telangiectasia and Rad3-related protein (ATR), usually lead to genetic insult repair before progression through mitosis by inhibition of CDK proteins that slow down cell cycle progression (Huen and Chen, 2008). Upon repair, cell cycle progresses without the activation of apoptotic pathways (Roos and Kaina, 2006).

CHO cells represent a highly proliferative and immortalised cell line, reminiscent of cancer cell lines such as HeLa and indirectly HEK293T. Although not derived from a cancer tissue, HEK293T cells express Ad5 E1A/E1B proteins which deregulate pRetinoblastoma (RB) and p53 pathways, disrupting the cell cycle (Berk, 2005; Sha et al., 2010). If genetic insults are not corrected appropriately and the immortalised cell line has acquired mutations that allow cell cycle progression, genetic instability can occur. If a genetic insult occurs specifically at the telomeres, tumour suppressor p53 binding protein (TP53BP1) is recruited and facilitates non-homologous end joining (NHEJ) at chromosomal ends. TP53BP1 action is only made possible in absence of p53 and RB pathways (O'Sullivan and Karlseder, 2010).

Cells that obtain fused chromosomes and acquired the ability to pass through mitosis lead to break-fusion-bridge cycles whereby chromosomes break non-reciprocally to create two genetically distinct daughters (Marotta et al., 2013). BFB cycles have been implicated to intratumor heterogeneity and been shown to promote DNA amplification and chromosome loss (Gisselsson et al., 2000; Lo et al., 2002; Thomas et al., 2018). This may represent a pathway that leads to the genomic instability of CHO cell lines (Vcelar et al., 2018a; Vcelar et al., 2018b).

To elucidate a potential pathway, based on telomeric promoted genetic instability that may drive the CHO therapeutic protein production instability phenotype, a characterisation of the host cell line must first be performed. Here, CHOK1 host variants are assessed by their telomeric profiles and compared to a cancerous-like cell line – HEK293T. The baseline of these results will be used to compare telomeric profiles of the host to therapeutic protein expressing cell lines, to assess any changes that may occur during the cell line development process of a therapeutic protein producing cell line. As CHOK1a-GS-KO host is used for subsequent analyses of stable vs unstable producing cell lines, CHOK1a-

GS-KO host is further analysed for baseline chromosomal mutations, telomere protection and DNA damage profiles.

Telomere FISH profiles of CHOK1a, CHOK1a-GS-KO, CHOK1-SV and HEK293T cell lines across a 6-month stability assessment (3.2).

Aims and hypothesis (3.2.1)

Telomere sequence profiles will be qualitatively assessed and compared against HEK293T cell line. HEK293 represents a normal telomeric signal profile expected in mammalian cell lines and is therefore used as a reference point when analysing CHO host lines. Changes in telomeric profile, across the 6-month culture period, may represent an indication of genomic instability.

Results (3.2.2)

CHOK1a, CHOK1a-GS-KO, CHOK1-SV and HEK293T cell lines were thawed and revived using the appropriate media as listed in the methods section (table 1, 2.1.1). Once cells had reached >98% viability, chromosomes were harvested from each cell line at passage 6. Chromosome harvesting was performed in 10 passage increments to mimic six months of cell culture, as performed in therapeutic protein productivity stability assessments. This mock stability assessment using commonly used CHOK1 hosts was performed to elucidate whether there are significant changes in telomere profile across the culturing period (figure 3.1). Cell viability and diameter of each cell line assessed remained

consistent for the duration of the experiment, indicating the cell lines phenotype did not fluctuate.

Compared to HEK293T, all CHOK1 host variants have most telomere sequences interstitially, with varying degrees of distinct patterns between each CHOK1 host, as highlighted by the signature chromosome column. CHOK1-SV obtains a large block of TTAGGGn repeats on one chromosome, compared to CHOK1-GS-KO which obtains a telomere pattern that indicates BFB cycles may have occurred leading to non-reciprocal translocations or amplifications. Noticeably upon thresholding, there are no visible telomere signals at the extreme ends of the chromosome, whist interstitial telomeric repeats exist in large blocks of repeats. Lack of telomere sequence at the extreme ends of chromosomes may lead to increased telomere specific DNA damage response pathway activation promoting CHO chromosomal instability. CHOK1-GS-KO is the host platform of choice and further analyses based on therapeutic producing proteins are derived from this host, the cell line was characterised further to establish a baseline comparison against the therapeutic protein producing cell lines.

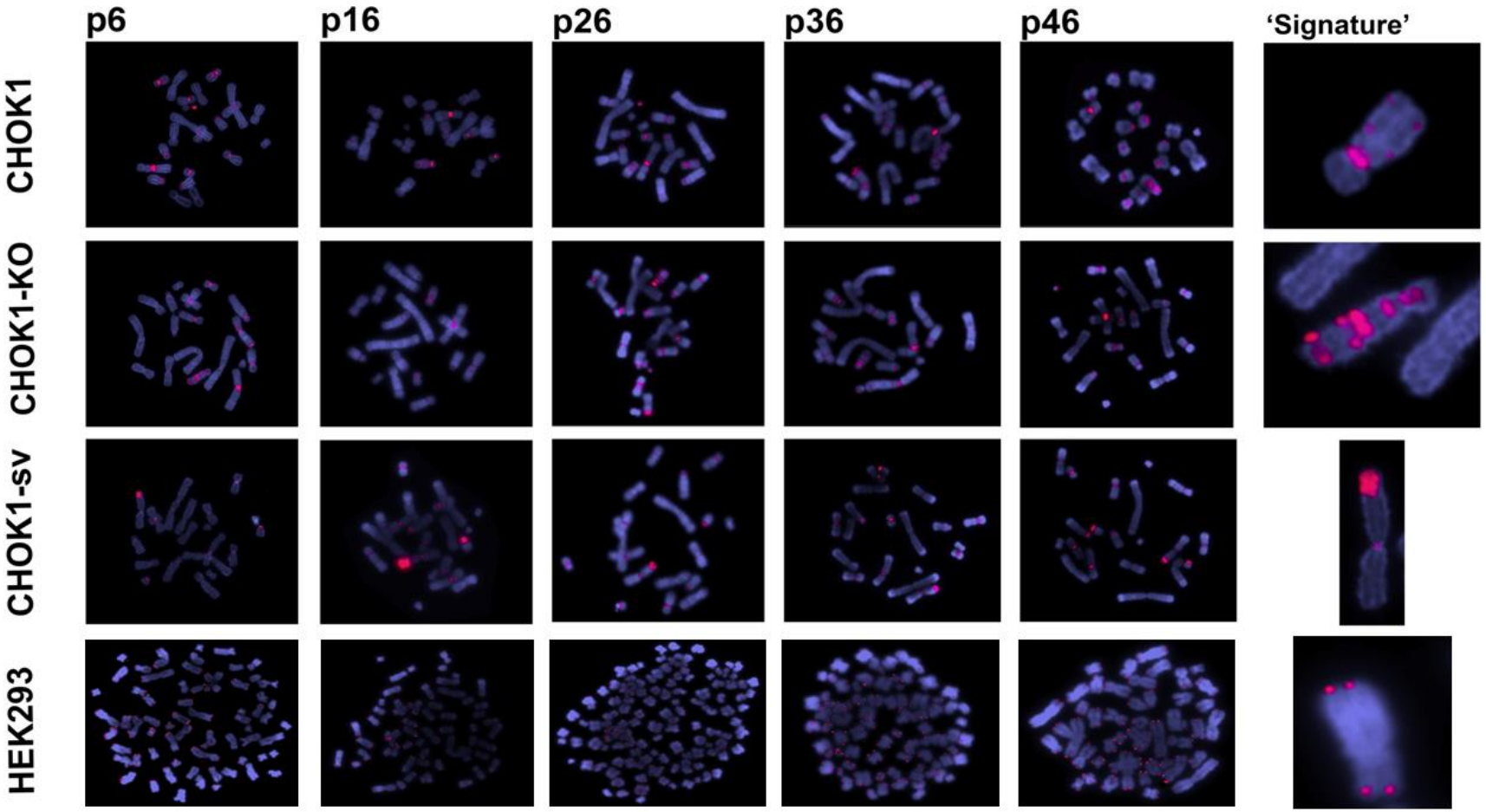

p6
p16
p26
p36
p46
'Signature'
CHOK1
CHOK1-KO
CHOK1-sv
HEK293

Figure 3.1. Representative images characterising telomere profiles of CHOK1, CHOK1a-KO, CHOK1-sv and HEK293T lines across a 6-month culturing period, mimicking a cell line stability assessment. Chromosomes were harvested every 10 passages and their qualitative chromosome and telomeric signal was assessed. Each cell line obtained a signature telomere profile that was maintained throughout the 6-month stability period. Telomere signatures are predominantly interstitial telomere repeats (ITS) within CHO based cell lines, whereas telomeres exist only at the extremities of HEK293 cells.

Chromosome number distribution and Telomere FISH quantification of CHOK1a-GS-KO host cell line across a 6-month stability period (3.3).

Aims and hypotheses (3.2.1)

Chromosome number distribution and telomere sequence fluorescent signals will be quantified across a 6-month stability period to generate a baseline characterisation of CHOK1a-GS-KO host cell line to be utilised as a comparator against CHOK1a-GS-KO therapeutic protein producing cell lines. Host cell lines should be telomerically stable to promote genomic stability during the manufacturing process. Fluctuations in chromosome number and telomere length may suggest an increase in genetic instability within the host over the 6-month stability period.

Results (3.2.2)

Chromosome number distribution of the CHOK1a-GS-KO host was assessed using 50 randomly selected images per timepoint and quantified using Fiji (image J, methods 2.7.6). Metaphase images were selected only when a distinct chromosome population could be identified, to ensure the chromosomes being counted belonged to a single cell source. Images with overspread chromosomes were excluded from analyses. Chromosome number distribution was assessed at both early and late timepoints to elucidate whether there is a change in distribution over time, which may highlight genomic instability.

After thawing from a development cell bank, CHOK1a-GS-KO host cell line was passaged until viability reached >98% and chromosomes were harvested and counted as previously described. At the early timepoint the median chromosome number was 19 (figure 3.2), which is reflective of previous reports (Vcelar et al., 2018a; Vcelar et al., 2018b). Modal chromosome range is between 18-21 chromosomes with an overall chromosome number range of 15 to 37. Outlier frequency (figure 2, black dots) of the modal chromosome range occurred in 7 cells with most of the data distributed between the model range (43). Conversely, at the late timepoint the median chromosome number increased to 20 (2-sample T-test, P=0.0384), indicating a gain of an additional chromosome across the 6-month culturing period.

Modal chromosome range remained the same, however there was an increase in the overall chromosome number range (7-39 chromosomes), and outlier frequency (12, black dots, figure 3.2). This may suggest increased chromosome instability across the 6-month stability period, as there is an increase in cells that obtain an abnormal number of chromosomes. If this can be attributed to chromosomal instability, this data suggests that it is innate to the host cell line.

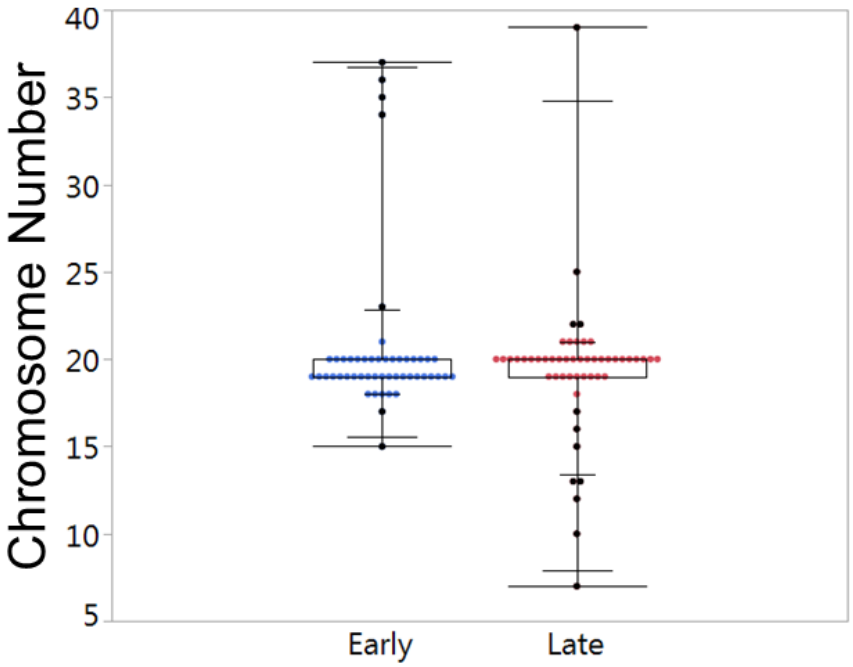


Figure 3.2. Box plot of chromosome number distribution of CHOK1a-GS-KO host cell line at early and late time points. Median chromosome number increased from 19 to 20 (blue and red dots) across the 6-month culture period (2-sample T-test, P=0.038). Number of outliers (black dots), representing cells with abnormal chromosome numbers, increase from 7 to 12 when comparing early to late time points.

Using cell profiler, a semi-automated telomere quantification workflow was created to analyse telomere fluorescent intensities as outlined in methods section 2.7.3 (figure 10 and 11). Telomere probes are formed of PNA-TTAGGG(n) repeats with a conjugated Cy3 fluorophore. Through hybridisation, probes bind to complementary telomere sequences through Watson-crick pairing. Fluorescent intensity is therefore proportional to telomere signal and changes in fluorescent intensity is in relation to telomere sequence changes present within the chromosomes. Telomere signals that reside within the chromosome masks generated on DAPI images are measured (figure 3.3a), providing quantifications of telomere length that is chromosome specific.

50 images were analysed per timepoint and ratios of telomere fluorescent intensity to DAPI intensity (telomere proportion %) were compared between time points. Although the ratio does not provide an absolute value, the relative signals can be used to assess whether there has been a change in telomere length across the stability period. Mean telomere proportion obtained at the early time point was 2.9%. Over 6 months of continuous culture, mean telomere proportion increased to 8.9% (T-test, P=<0.0001, figure 3.3b). Telomere signals quantified solely resided within the chromosomes as interstitial telomere repeats, suggesting that the amplification of these sequences remained within the chromosome (ITS amplification) itself rather than at the extremities. This was visually confirmed by inspecting the late images for telomere signals at the extreme end of chromosomes.

Data presented here suggests there is an inherent genetic instability within the CHOK1a-GS-KO host. This is corroborated by the increase in median chromosome number across 6 months of culture (figure 1, 2-sample T-test, P=0.0384) that suggests a shift towards a dominant chromosomal population that obtains an additional chromosome. Moreover, an increase in telomere sequence (figure 5, P=<0.0001), suggesting amplifications have taken place at interstitial telomere repeats has occurred. These traits are indicative of genetic instability at the chromosomal level, as highlighted in previous studies (Gisselsson et al., 2000; Lo et al., 2002; Thomas et al., 2018).

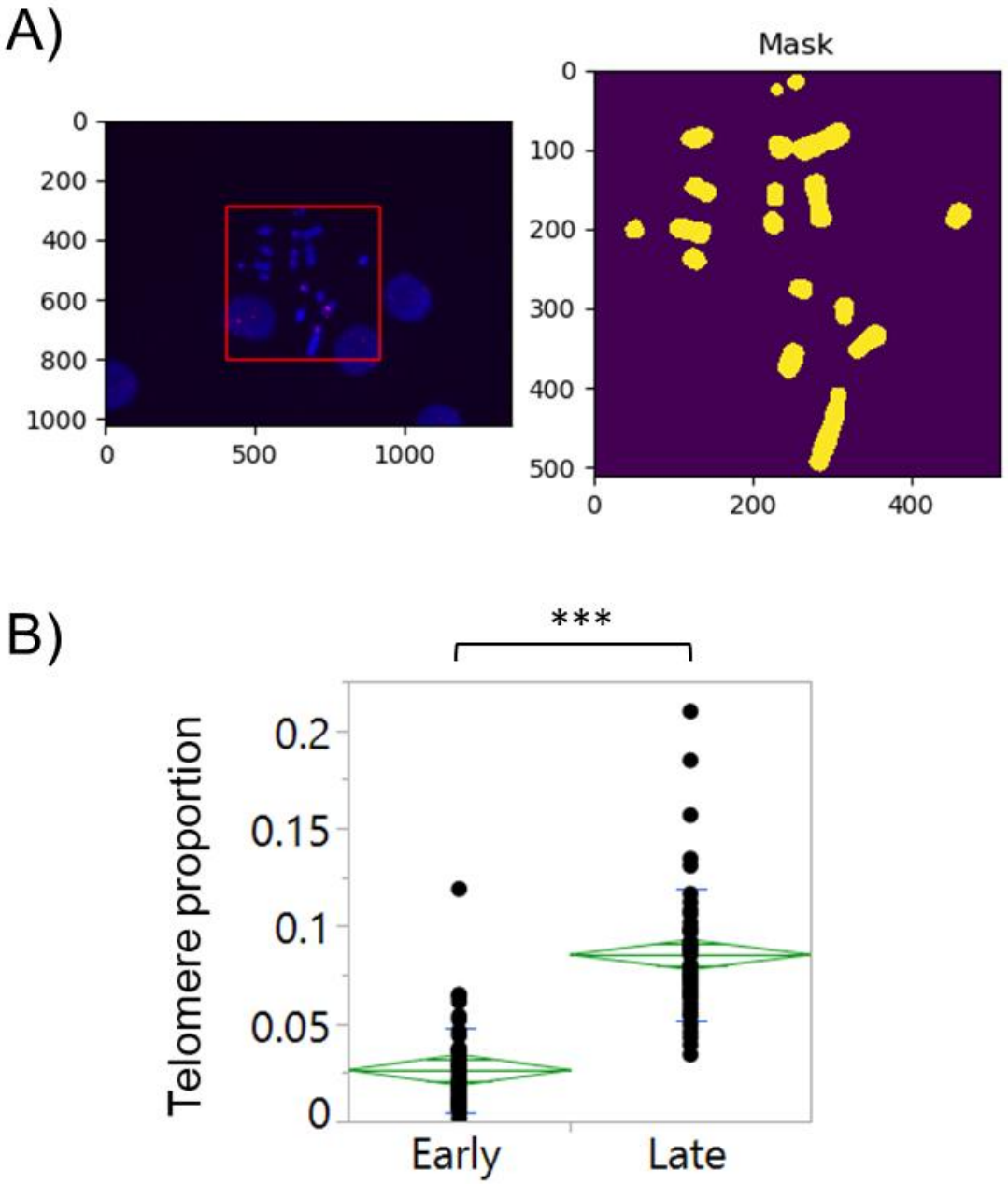


Figure 3.3. Telomere quantification in CHO-GS-KO Host across ~140 generations. A) Representative image of semi-automated masking of metaphase images, telomere sequences are quantified within chromosome masks. **B)** Telomere proportion represents the ratio of telomere signal to DNA signal obtained by PNA-TTAGGGn and DAPI staining's, respectively. An increase in telomere length of 6.1% occurred, from 2.8% at the early time point to 8.9% after 6 months of culturing (P = <0.0001). Green triangles represent mean and 95% confidence intervals, blue lines represent standard deviation

Using Multi-colour Fluorescent in-situ Hybridisation (MFISH) to Assess Karyotypic Changes of CHOK1a-GS-KO Host Across a 6-Month Stability Period (3.3).

Aim and hypotheses (3.3.1)

To assess homogeneity of CHOK1a-GS-KO karyotype and how the karyotype may fluctuate over time. Host cell lines used for therapeutic protein production should retain genetic homogeneity from single cell cloning and maintain genetic stability during routine culture. Heterogeneity found at the host level may be passed onto derived producing cell lines.

Results (3.3.2)

Multi-colour fluorescent in-situ hybridisation (MFISH) was performed on CHOK1a-GS-KO cell lines at early and late timepoints, as previously described (methods, 2.3.4). MFISH 'paints' chromosomes to allow the visualisation of chromosome constituents. Probes specific to the Chinese hamster genome were generated against a primary cell line (Metasystems) and provides a colour code for each individual chromosome (e.g. chromosome 1 = red, 2 = brown, etc). Thus, it provides the means to assess chromosomal mutations within host cell cultures and allows comparison both internally, between cell lines and time points. Mutations can be tracked at a single cell level and specific chromosome mutations may be able to be attributed to phenotypic traits.

Cell culture populations were manually determined using the methodology previously described (method, 2.7.1).

Karyotypically distinct cells obtain a unique population ID and matching karyotypes are grouped together under the same population identifier. 40 randomly selected images were analysed, and the frequency of each population assessed. Based on this frequency, clonal chromosomal aberration (CCA, >5%) or non-clonal chromosomal aberration (NCCA, <=5%) is assigned to each population, reflecting on the population's genetic stability (further described in the methods section 2.7.1).

At an early timepoint, 18 distinct populations were identified with population 1 and 2 representing the majority of the culture obtaining 45% and 13%, respectively (figure 3.4a, 14 and 15). Population 1 and 2 are designated CCA populations (green segments) whereas 15 populations had a frequency of <=5% and were classed as NCCA populations (red segments, figure 3.4a). Analysing karyotype populations after 6 months of continuous culture revealed 16 distinct populations, suggesting 2 populations have been lost during the culture process, although this may be an artefact of the number of images analysed. 3/16 populations are designated CCA populations compared to 13 NCCA. 6 out of the original 18 populations were maintained throughout the 6-month process, with 10 de novo populations arising (figure 3.4b). Although NCCA populations 6, 8, 13 and 14 were maintained over the culture period, their NCCA designation did not change, suggesting that their acquired mutation did not provide a growth advantage.

Of the 10 de novo populations, population 4 became dominant in the host culture. De novo population 4 gained a proliferative advantage and became the second largest population within the culture, surpassing the retained population 2 from the early timepoint. Population 2 frequency decreased from 13% to 8%, whilst population 4 obtained a 15% frequency (figure 3.4b). Comparisons of karyotypes between early and

late timepoints revealed early population 2 may have been the pre-requisite of de novo population 4 as their karyotypes are identical apart from an apparent duplication of chromosome 6 (figure 3.4c, d enlarged images in figure 3.6 and 17).

Duplication of chromosome 6 may have provided a proliferative advantage for the cell to allow it to establish itself as a predominate population within the flask.

Chromosome mutations that led to the creation of a newly distinct population was quantified (figure 3.5a). Chromosomes 2, 4, 5, 7, 10, 11, 14, 15, 18, 19 did not obtain any translocations that created a new population over the 6-month culture period, suggesting the majority of CHOK1a-GS-KO host chromosomes have maintain genomic stability. Chromosome 8 was the most frequently mutated compared to any other chromosome, accounting for 11 distinct populations across both time points, suggesting an inherent instability within this chromosome that contributes to the natural heterogeneity of the CHOK1a-GS-KO host cell line. There was a noticeable mutation increase in chromosome 6 and 13 (in addition to chromosome 8) after 6 months of continuous culture that contributed to 7 newly distinct populations (13 populations in total, when including chromosome 8).

Inherent weaknesses within chromosome 6, 8 and 13 may provide a mechanism by which mutations can be elicited to gain competitive advantage (figure 3.7). This is corroborated by the duplication of chromosome 6 that allowed de novo population 4 to establish itself as the second most prominent population at the end of the culturing period (figure 3.4a, b, c, d, e).

Chromosome mutations were categorised into mutation types and coloured by chromosome to assess the predominant mode of mutation that creates heterogeneity within the host. Translocations in chromosome 1, 8, 9, 12 and 16 at the early timepoint and chromosome 3, 6, 8 and 13

at the late timepoint contributed to 19 newly distinct populations (figure14b). Deletions (chromosome 8 and 13) and chromosomal breaks (chromosome 3 and 6) only occurred during the late time point, suggesting that these mutations may be indicative of long-term culture stress.

When analysing the overall CCA and NCCA frequency of populations within each time point, the ratio of CCA to NCCA remained similar. CCA frequency increased from 57.5% to 67.5% from early to late time points, suggesting a small shift to genetic stability through the creation of de novo population 4 which contributed to the CCA increase (figure 3.4a, b, and g). Conversely, NCCA decreased from 42.5% to 32.5%, due to the increase in population 4 and the loss of two NCCA populations from the early timepoint (figure 3.5c).

Overall, the data presented here highlights a single cell cloned host that has acquired mutations during routine culturing at both early and late time points. Prolonged culturing of the host seems to exacerbate this issue, maintaining the genomic heterogeneity as witnessed at the early stages of culturing. Chromosome 8 seems to play a role in the creation and maintenance of the karyotypic heterogeneity, with translocations being the predominate type of mutation that creates de novo populations. Transfecting therapeutic proteins into a heterogenous host creates a scenario where upon single cell sorting, the clonal outgrowths will be genetically dissimilar as the plasmid may enter any one of the distinct populations. In this manor, the background genomic heterogeneity of the host creates an environment where clones single cell sorted from the same host may have divergent CHO'mic profiles that may impact phenotypes in manufacturing conditions.

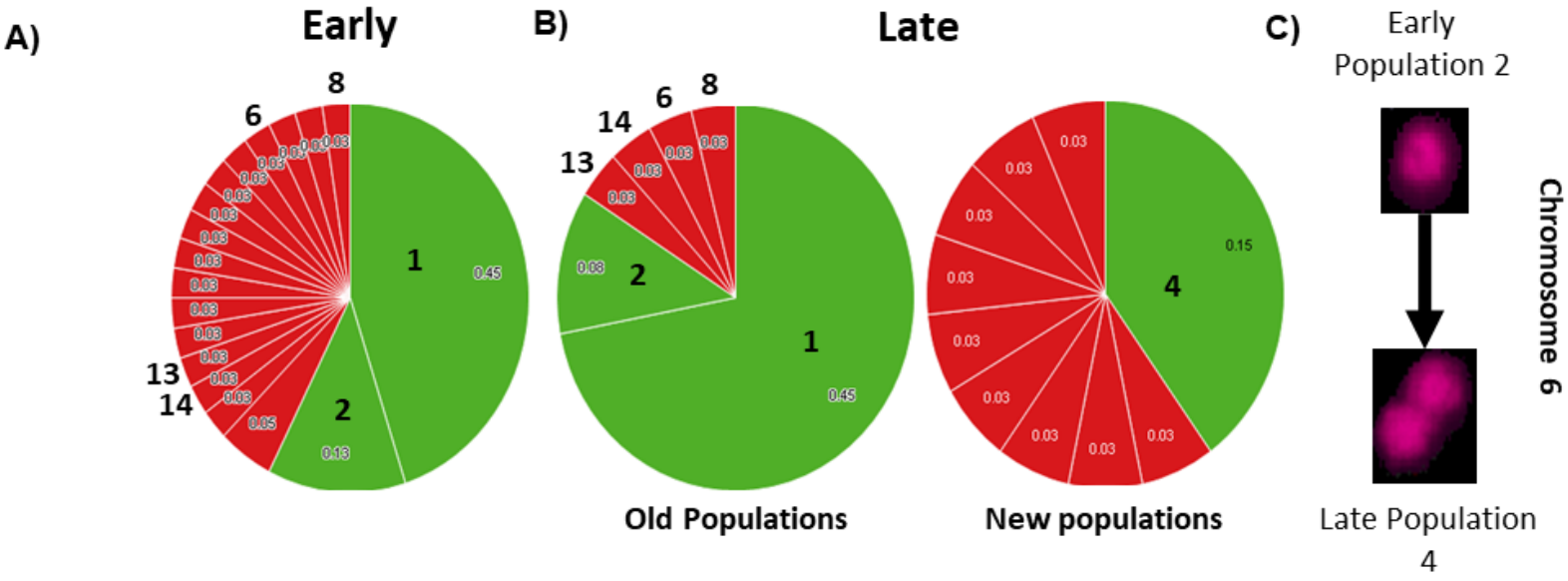

A)
Early
8
6
1
0.45
2
0.13
13
14
B)
Late
Old Populations
New populations
C)
Early
Population 2
Late Population
4
Chromosome 6
D)
Early Population 1
Late Population 4
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18

Figure 3.4. CHOK1a-GS-KO host karyotype characterisation across ~140 generations. A) 18 distinct karyotypes were identified at the early time point with population 1 and 2 representing most of the culture flask with 45% and 13%, respectively. B) 16 karyotypes were identified after 6 months of culture. 6 of the original 18 karyotypes remained at the late timepoint with 10 newly distinct karyotypes identified. Population 1 frequency remained the same at 45%, whereas population 2 frequency decreased to 8%. De novo population 4 represented 15% of the culture surpassing population 2 as the second most predominate karyotype. C) Representative chromosomal difference between early population 2 and late population 4; a duplication of chromosome 6. D) Representative images of main karyotypes at early and late time points. Population 1 remained the most prominent population across the 6-month stability period (enlarged images in figures 8-10).

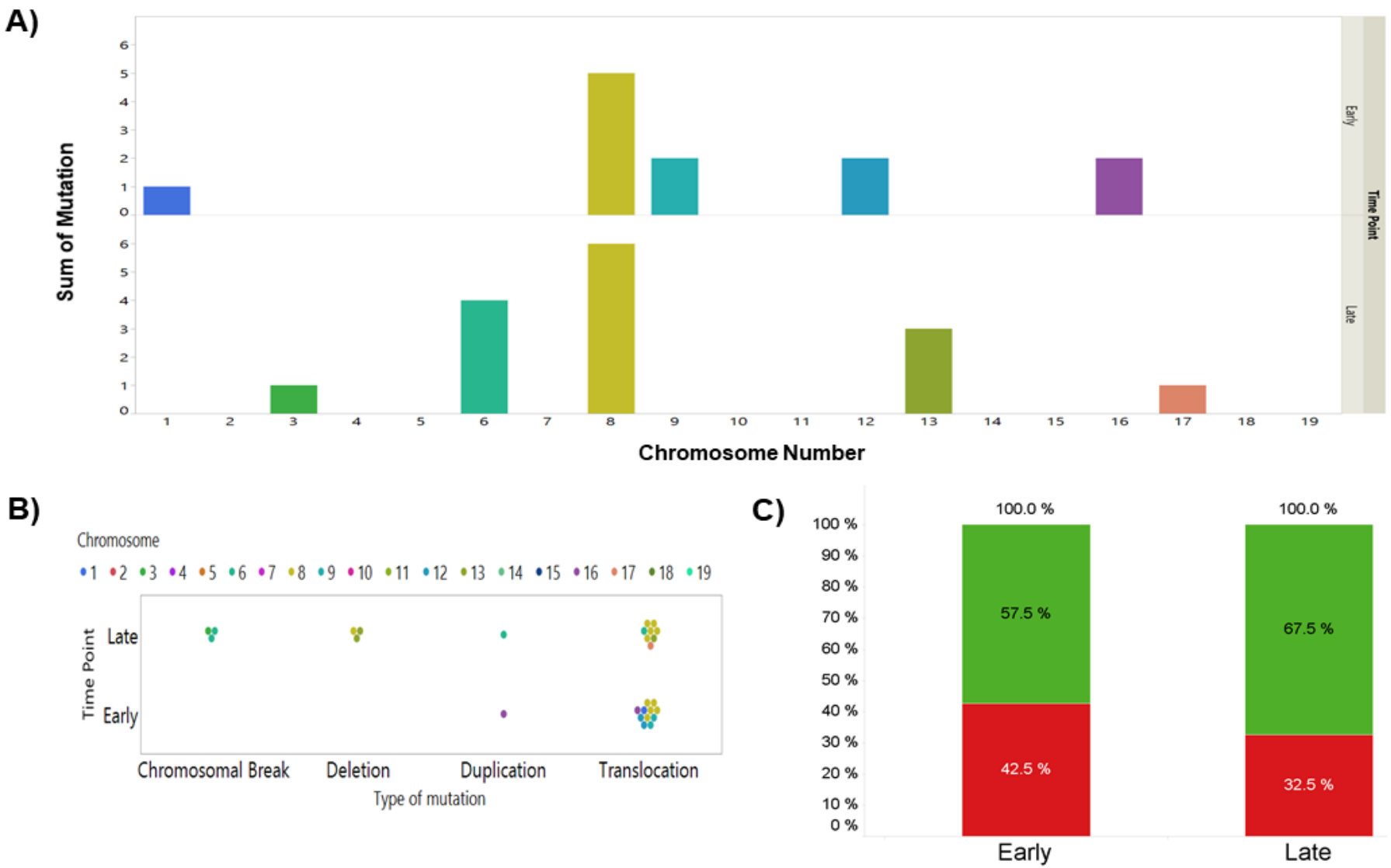

A)
Sum of Mutation
Chromosome Number
Time Point
Early
Late
B)
Chromosome
Time Point
Late
Early
Chromosomal Break
Deletion
Duplication
Translocation
Type of mutation
C)
100.0 %
100.0 %
57.5 %
67.5 %
42.5 %
32.5 %
Early
Late

Figure 3.5. Characterisation of mutations in CHOK1a-GS-KO host across ~140 generations. A) Summary of mutations per chromosome, which resulted in a new distinct population, compared across early and late time points. Chromosome 8 appears to contribute to the majority of heterogeneity in both time points. B) Overall percentage frequency of clonal chromosomal aberrations (CCA) and non-clonal aberrations (NCCA) between early and late time points. CCA frequency increased from 57.5% to 67.5%, whereas NCCA decreased from 42.5% to 32.5%. C) Quantification of mutation types which resulted in a new population, across early and late time points. Translocations contribute to most of the heterogeneity within the host cell line across both time points.

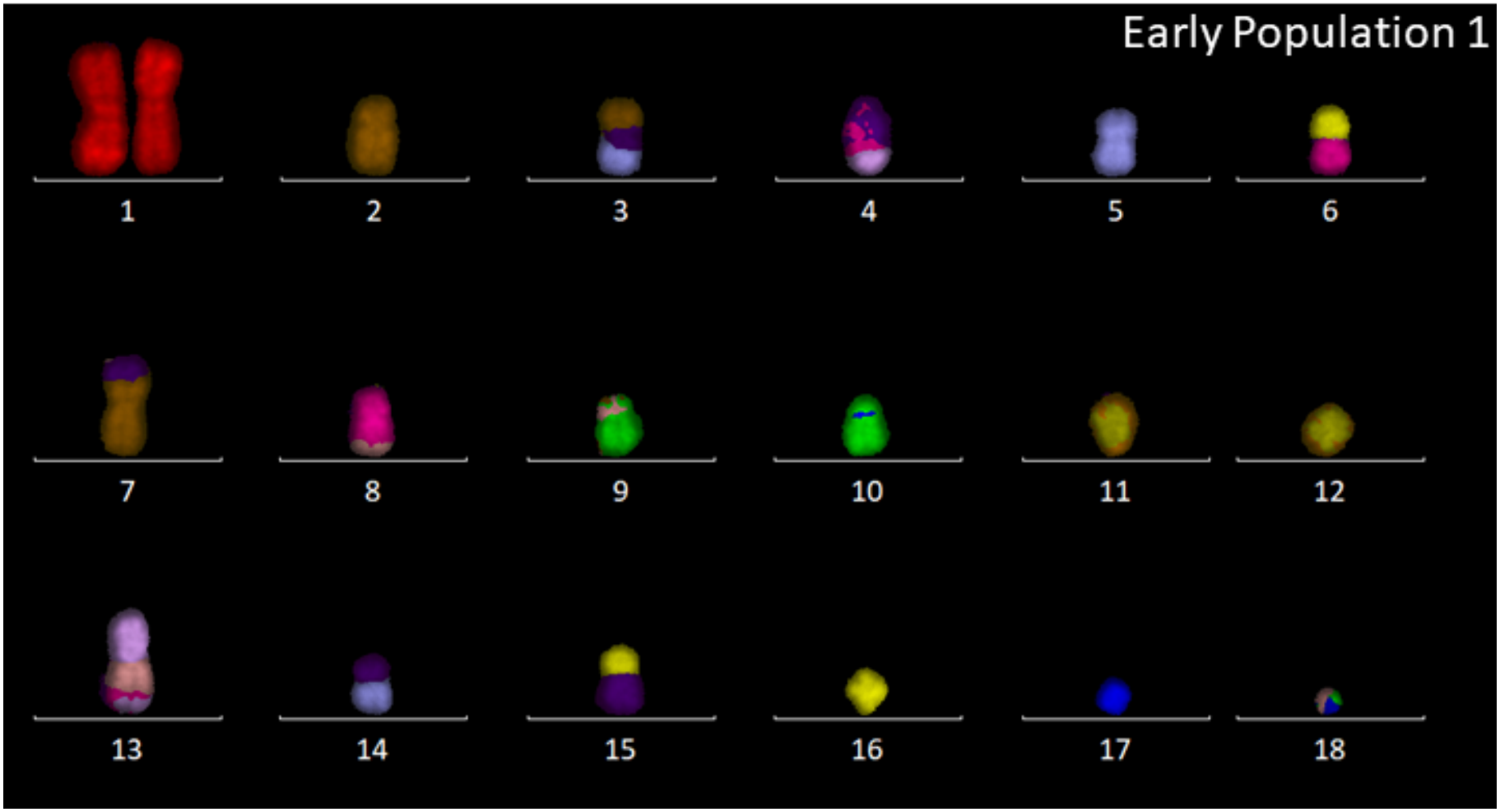


Figure 3.6. Representative full karyotype of CHOK1a-GS-KO host early population

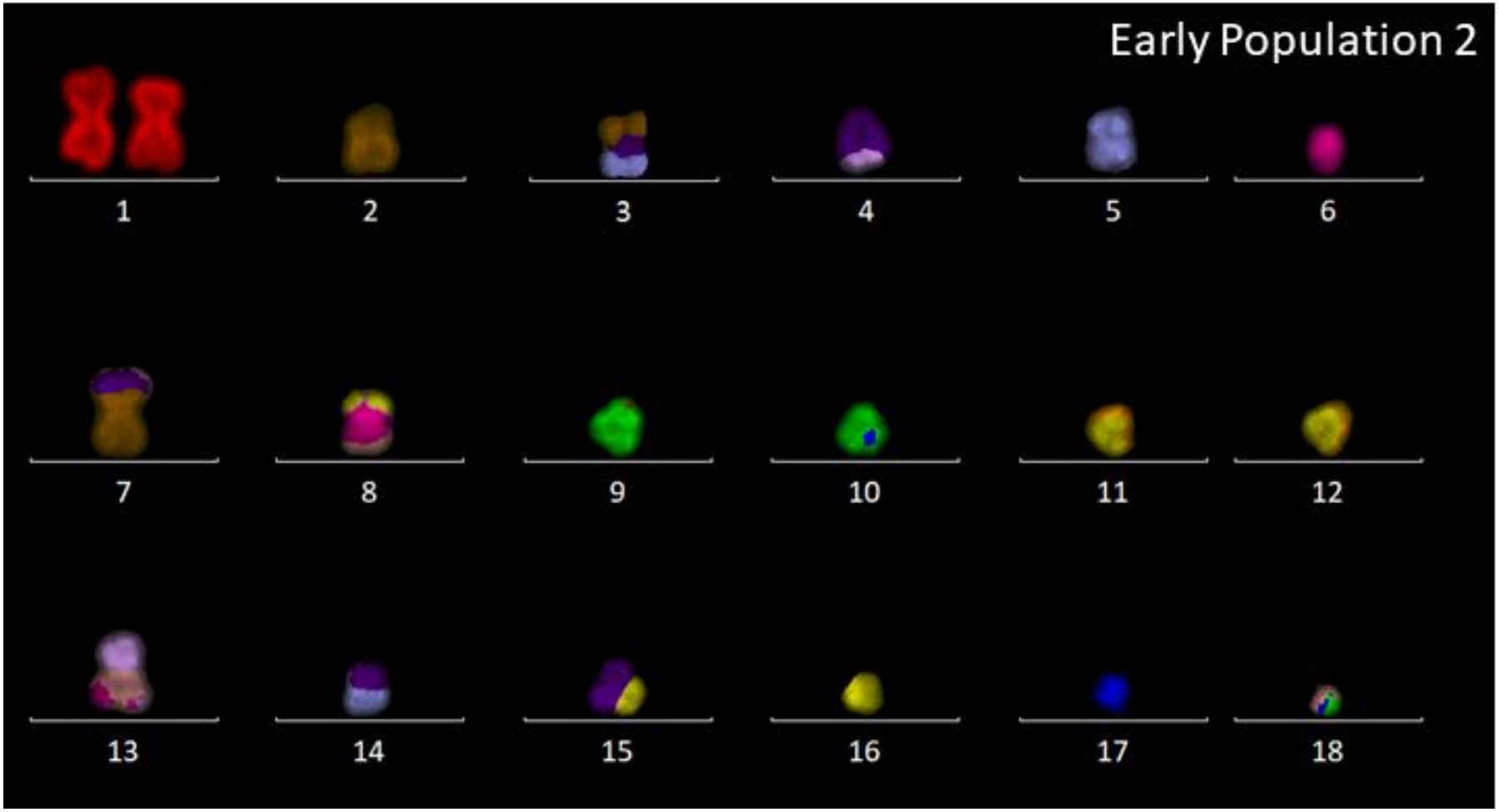


Figure 3.7. Representative full karyotype of CHOK1a-GS-KO host early population 2

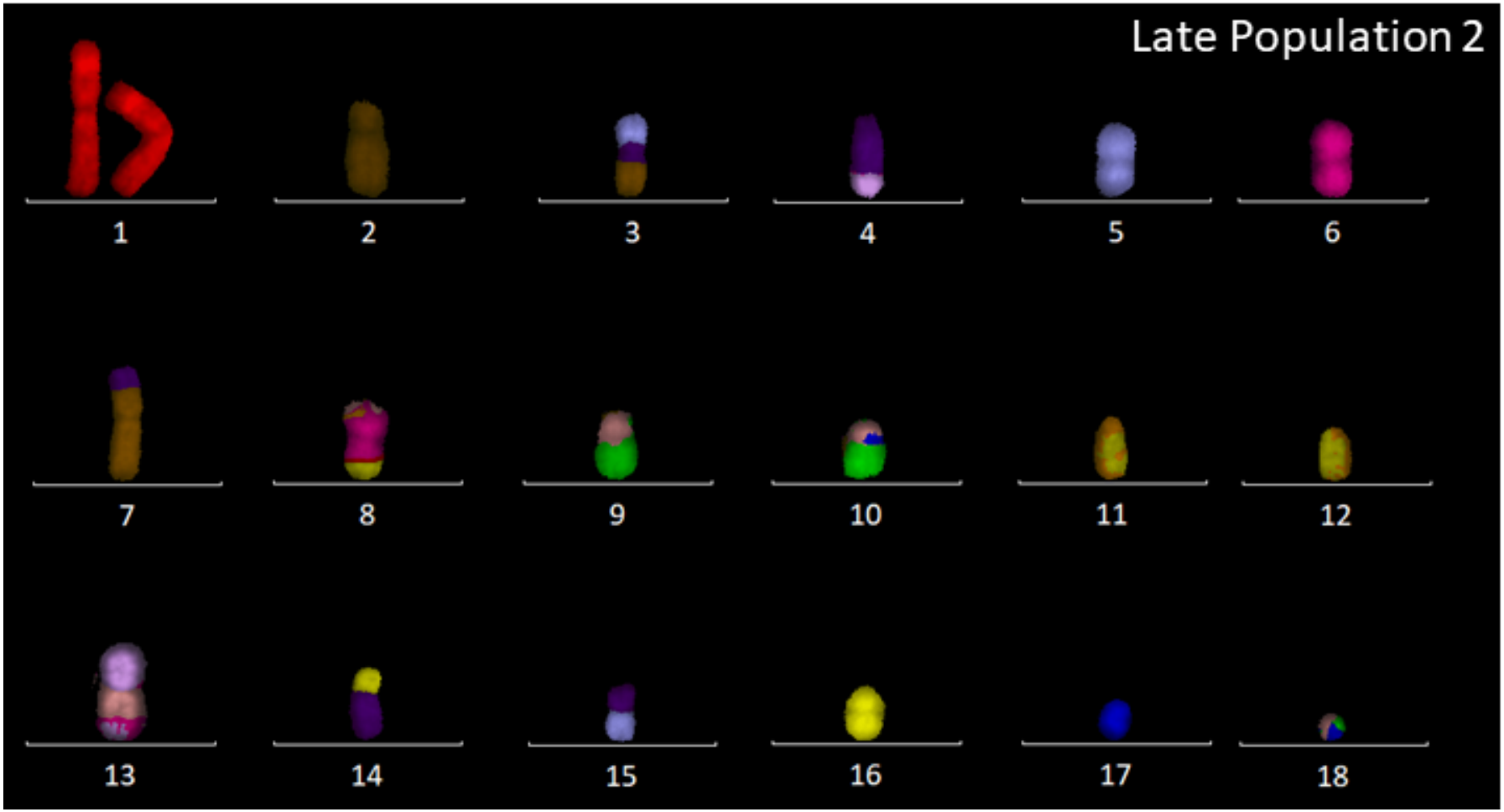


Figure 3.8. Representative full karyotype of CHOK1a-GS-KO host late population 4 depicting chromosome 6 duplication.

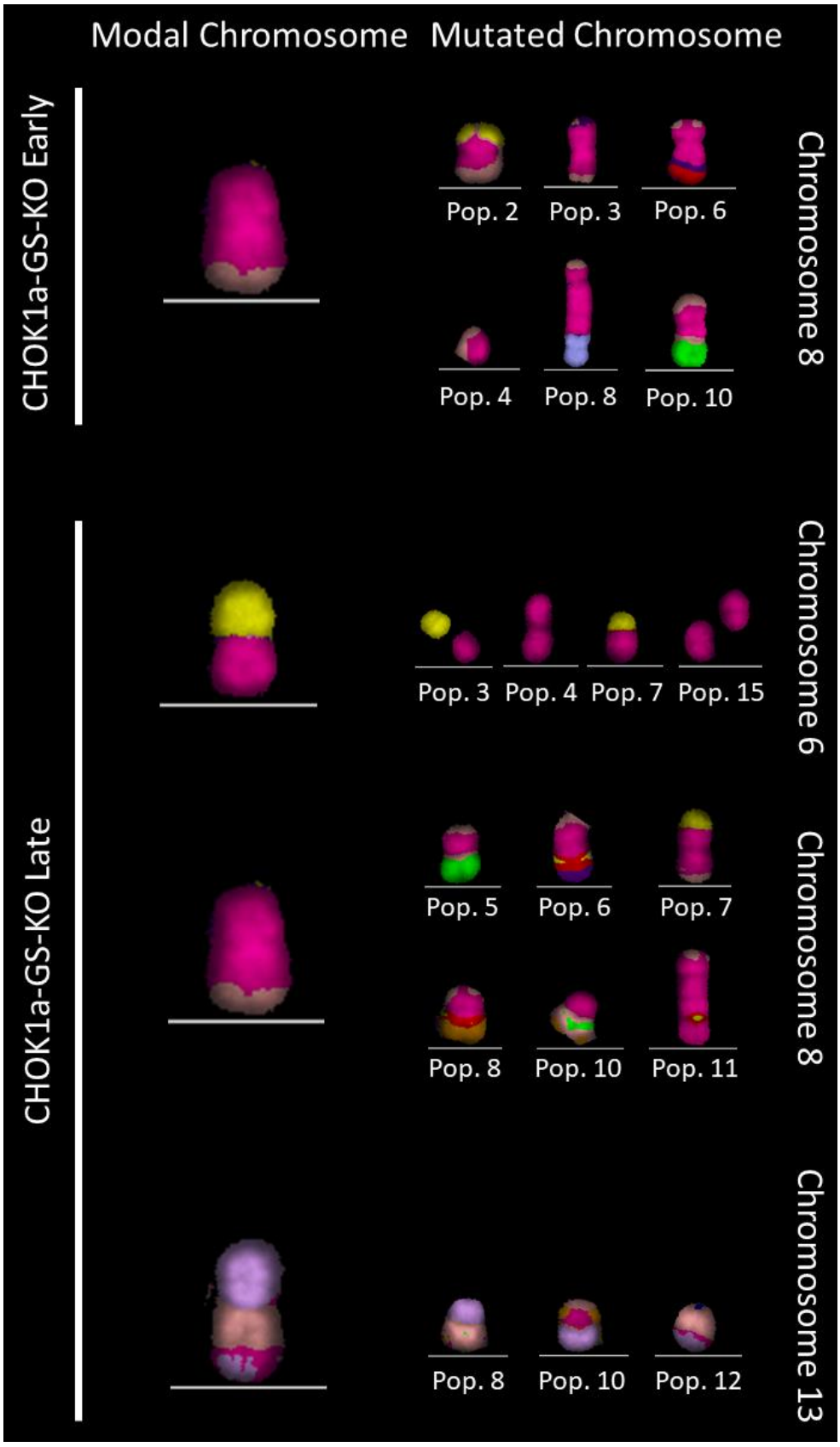


Figure 3.9. Partial karyotypes of frequently mutated chromosomes within CHOK1a-GS-KO host at early and late time points.

Extent of telomere protection by Shelterin complex in CHOK1a-GS-KO host cell line (3.4).

Aims and hypotheses (3.4.1)

To characterise the extent of interstitial telomere, repeat protection by the shelterin complex within the CHOK1a-GS-KO host. A baseline quantification of TRF2 and gH2AX localised with telomere sequences will be performed. TRF2 will provide an indication of the amount of shelterin that is bound to telomeres and telomere specific DNA damage (TIF) will be quantified to assess the baseline TIF of the host. Protection of telomere sequences should confer telomere stability against DNA damage pathways. Inadequate protection can lead to increased DNA damage and induce genomic instability.

Results (3.4.2)

To assess the extent of interstitial telomere sequence (ITS) protection by the shelterin complex, immunofluorescence co-staining of TRF2 and telomere sequences was performed on the CHOK1a-GS-KO host cell line as previously described (methods, 2.6.1). TRF2 represents a key component of the Shelterin complex, which confers telomeric protection specifically at telomere sequences. Shelterin binds to telomere sequences through the action of TRF1 and TRF2 and recruits TIN2, TPP1 and POT1 to complete the telomere bound shelterin complex (Ye and de Lange, 2004; Ye et al., 2004a). Co-localisation of telomere and TRF2 immunofluorescent signals allows the characterisation of baseline telomere bound TRF2 protein that may confer protection of ITS repeats. Although TRF2 staining is not indicative of the full shelterin complex

being bound to ITS repeats, it identifies the possibility for full shelterin complex assembly at these sequences, as TRF1 and TRF2 are essential for telomere specific binding and shelterin assembly (Ye and de Lange, 2004; Ye et al., 2004a; Ye et al., 2004b).

The proportion of Shelterin complex assembly in relation to bound TRF2 protein has not been established, however knock down of TRF2 has been shown to coincide with telomere length shortening, due the deprotection of the telomeres (as Shelterin complex assembly can no longer occur) and a concomitant increase in DNA damage (Kim et al., 2009). TRF2 (green) and telomere (red) co-localisation along with single TRF2 or telomere signals were quantified using image J (methods, 2.7.7).

Overlapping of TRF2 and telomere (figure 3.10a, blue triangles) signals indicate that ITS repeats have TRF2 bound. However, there were instances where single telomere signals were observed (figure 3.3a, white triangles), indicating ITS repeats that do not obtain bound TRF2 and are therefore without Shelterin conferred protection. CHOK1a-GS-KO telomere foci without co-localised TRF2 signal occurred 1.9 times per cell out of ~100 cells analysed at an early time point. After 6 months of cell culture, this increased to 2.2 times per cell, however this difference was non-significant (pooled T-Test, figure 3.10b). Co-localised signals were quantified as 6 and 4.7 times per cell at early and late time points, respectively (figure 3.10c, pooled T-Test, non-significant difference).

These results indicate that the levels of co-localised and non-co-localised signals of TRF2 and telomeres do not fluctuate within the host during a 6-month culturing period. As single telomere signals were witnessed across both time points, it indicates there are potentially unprotected ITS repeats that may be more susceptible to DNA damage events. Thus, DNA damage will be quantified through gH2AX staining of cells and its co-localisation with telomeric signals assessed to understand whether telomeres represent a preferential site for DNA damage within the CHOK1a-GS-KO host.

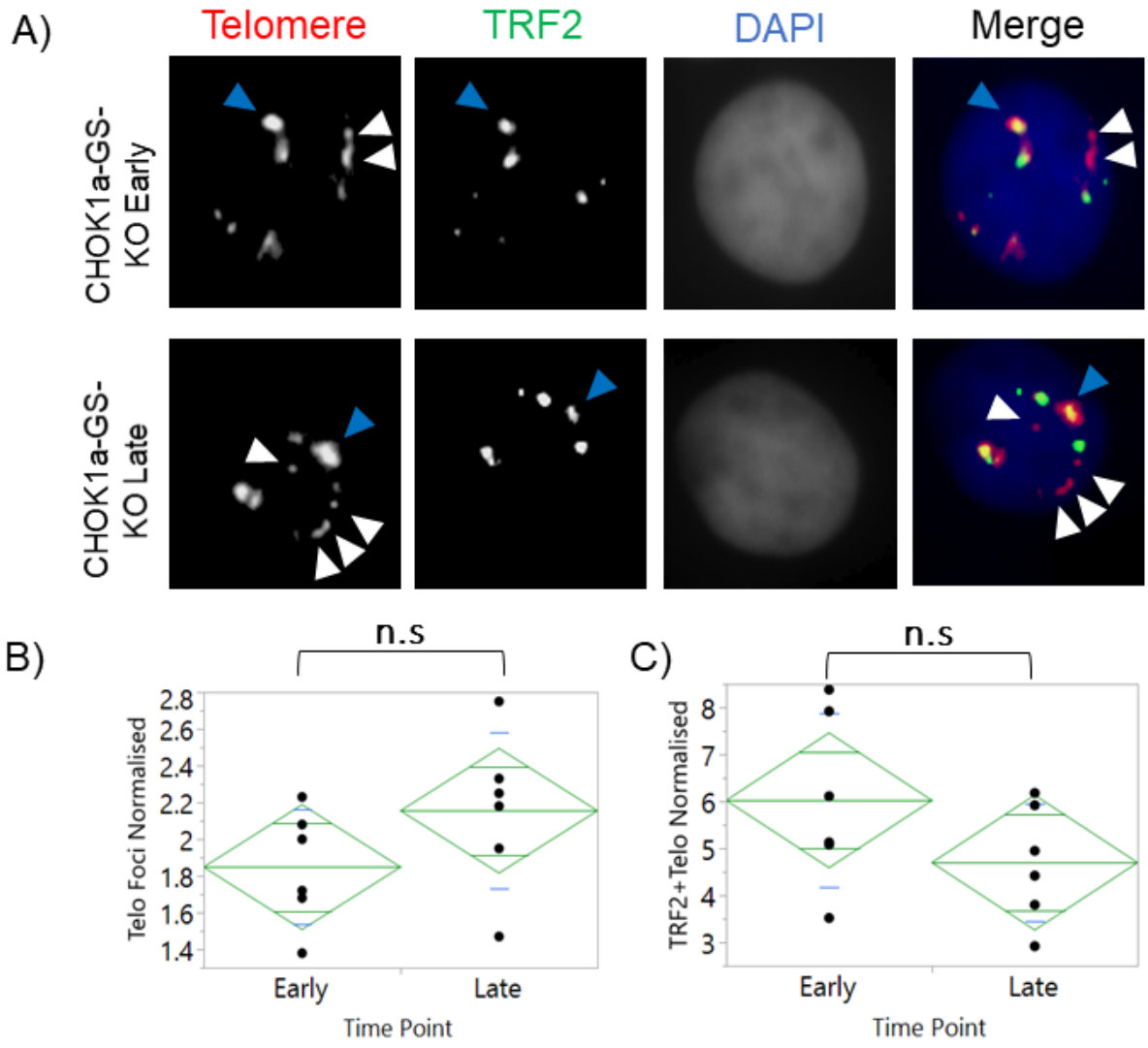


Figure 3.10. Quantification of telomere protection by shelterin in CHOK1a-GS-KO host. A) Representative images of immunofluorescent staining for TRF2 (green) and telomeres (red). Overlapping red and green signals indicate telomeres have bound TRF2 (blue triangles) whereas individual red signals indicate a telomere sequence without bound TRF2 (white triangles). B and C) Co-localisation of signals was quantified using image J. Individual telomere foci and TRF2 and telomere co-localisation levels remained similar across a 6-month culturing period (P=>0.05).

To assess the extent of DNA damage in CHOK1a-GS-KO host and its specific localisation to the telomeres, gH2AX and telomere induced foci (TIF) was quantified using image J (methods, 2.7.7). TIFs are confirmed by observing co-staining gH2AX and telomere sequences within the nucleus and has been established as a common assay to assess telomere dysfunction in relation to DNA damage (de Lange, 2002, 2005; Mender and Shay, 2015; Takai et al., 2003). Increased TIFs within the host may elicit elevated DNA damage response pathways at the telomeres, leading to increases in mutations, representing a mechanism by which CHOK1a-GS-KO host cell lines obtain their heterogenous karyotypes (figure 3.4a).

gH2AX and telomere co-staining was performed on CHOK1a-GS-KO early and late samples. A positive control to assess the sensitivity of the assay was obtained by treating protein 5 cell line with 1ng/ml Neocarzinostatin for 1 hour, as previously described (methods, 2.6.2). Quantification of co-localised signals (TIF, figure 3.11a, blue arrows) and individual gH2AX signals (white arrows) was performed using image J.

CHOK1a-GS-KO obtained 0.68 gH2AX foci per cell, compared to 5.78 for the positive control (figure 3.11b, pooled T-test, P=<0.0001), indicating that the TIF assay is sensitive to increases in DNA damage elicited by Neocarzinostatin. TIF foci was also shown to increase from 0.3 TIFs per cell in the host compared to 1.2 TIFs in the positive control (figure 3.11c, pooled T-Test, P=<0.001), suggesting that the DNA damaging agent is unspecific in its action. TIF foci was shown to slightly decrease over a 6-month culturing period, however this was a non-significant change (figure 3.11d, pooled T-test, P=<0.05). Quantification of the CHOK1a-GS-KO host DNA damage and TIF foci has allowed the establishment of a baseline level that can be utilised when comparing the DNA damage and TIF levels of therapeutic protein producing cell lines. This data will be used to assess whether there are elevated global DNA damage foci or whether the DNA damage is localised to telomeric signals.

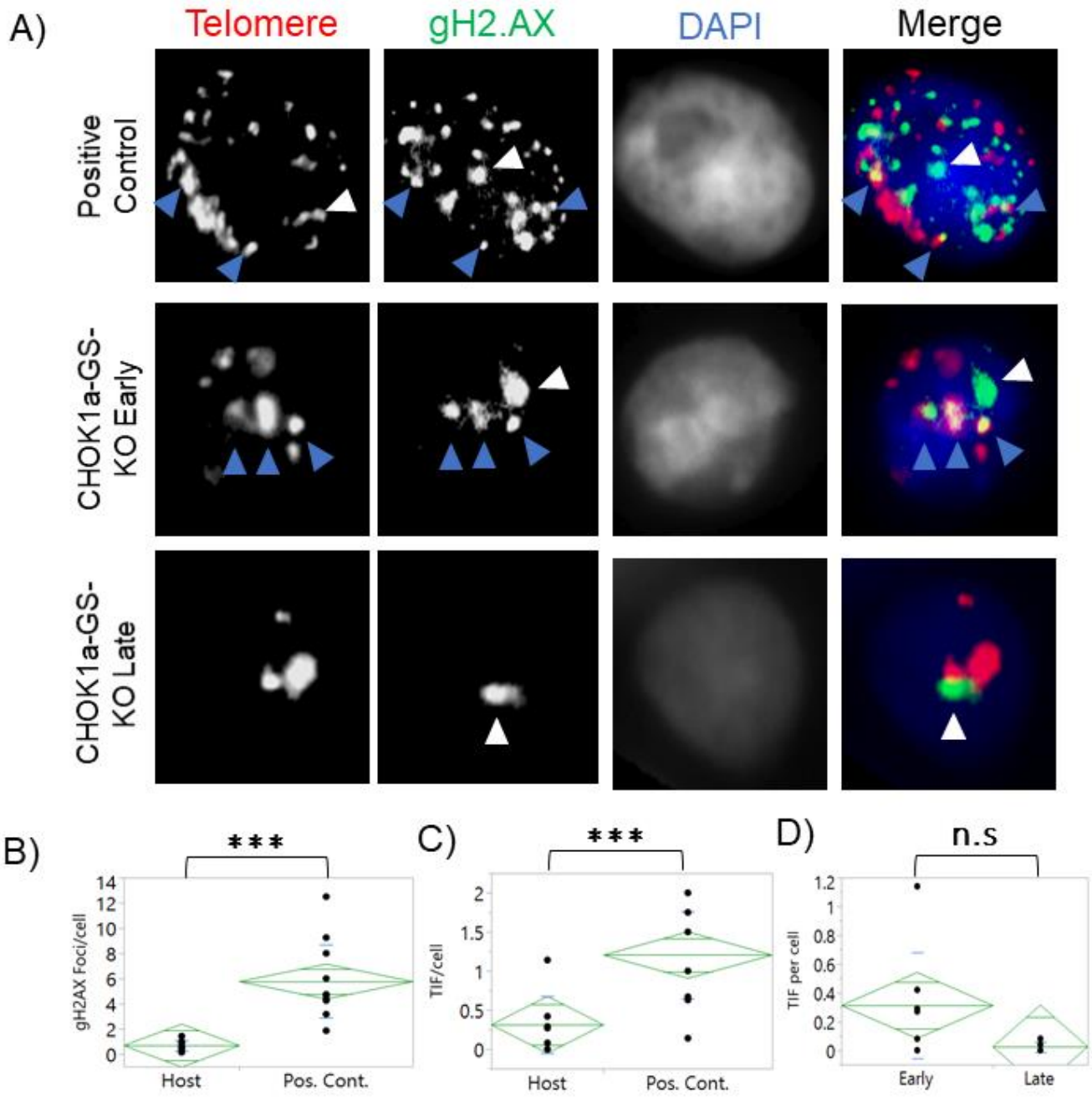

A)
Telomere
gH2.AX
DAPI
Merge
Positive Control
CHOK1a-GS-KO Early
CHOK1a-GS-KO Late
B)
***
gH2AX Foci/cell
Host
Pos. Cont.
C)
***
TIF/cell
Host
Pos. Cont.
D)
n.s
TIF per cell
Early
Late

Figure 3.11. Quantification of DNA damage and telomere specific damage in CHOK1a-GS-KO host. A) Representative telomere induced foci (TIF, blue arrows) with greyscale images of telomere, gH2AX, DAPI and colour merge. B) Quantified gH2AX foci per cell, in CHOK1a-GS-KO host and a positive control cell line that is treated with DNA damaging agent Neocarzinostatin (1ng/ml) for 1 hour. Results show the assay can detect an increase in gH2AX damage through gH2AX foci quantification (pooled T-Test, P=<0.0001). C) TIF foci per cell quantified between the host and positive control. Results indicate an increase in TIFs in the positive control (pooled T-test, P=<0.0001) D) TIF foci comparison between early and late time points. TIF foci per cell was found to be insignificantly different over 6 months of cell culture.

Chapter 4: Characterisation and comparison of productionally stable and unstable cell lines to identify differential patterns that lead to the production instability phenotype.

Introduction (4.1)

Characterisation of the CHOK1a-GS-KO host cell line revealed an inherent genomic instability highlighted by heterogenous karyotypes and interstitial telomeric repeat amplification. Karyotypic heterogeneity is not unique to this specific host cell line but appears to be a fundamental characteristic of CHO cell lines in general, as identified in previous studies (Vcelar et al., 2018a; Vcelar et al., 2018b).

CHO host cell lines are used as mammalian cell factories to create therapeutic protein producing cell lines. Utilising antibody discovery techniques such as phage libraries (Rothe et al., 2008), identified antibody therapeutics are validated against a desired antigen for their therapeutic effect within the relevant disease model. Once validated, the sequence that encodes the antibody is cloned into an expression plasmid and subsequently transfected into the CHO host cell line. Transfected pools are bulked, and single cell sorted, outgrowth of these cell lines is then assessed for their protein production (IgG titre). Clones are ranked based on their titre and undergo a series of triage events until ~50 cell lines are selected to enter a production stability assessment (4-6 months).

Cell line production stability can be assessed by various modes, such as deep well plates, shake flasks and mini-bioreactors. Production stability of a cell line is essential, as biopharmaceutical companies have to be confident that any cell line progressed to the manufacture stage can produce a consistent amount of therapeutic protein across the

manufacturing window. A traditional production stability assessment involves culturing the cell lines across a 4-6-month period, which reflects the manufacturing window. Max titre reads from each timepoint are obtained and the percent titre change is calculated. Based on an arbitrary threshold that varies from company to company (~+/- 30% fluctuation in titre), cell lines are designated a productionally stable or unstable. Stable cell lines are then used for downstream analysis to ensure aspects of the antibody harvesting, such as fragmentation and host cell protein content, satisfy FDA criterion for market approval.

Inherent genetic instability may account for roughly ~50% of producing cell lines (in house data) generated in CHOK1a-GS-KO host being productionally unstable. However, percent frequency of unstable cell lines can fluctuate on a protein by protein basis. This creates an environment where a vast number of cell lines have to be generated to ensure enough stable cell lines are available to be assessed in order to meet FDA requirements for approval. A combination of the amount of cell lines and the length of stability assessment seriously limits cell line development capacity and prolongs chemical, manufacturing and controls (CMC) timelines for antibody manufacture.

Various groups have tried to characterise the production instability phenotype, in a bid to identify causative genes and/or pathways that can be targeted to create a productionally stable host (Betts and Dickson, 2017; Hefzi et al., 2016; Kremkow et al., 2015; Lee et al., 2015a; Lee et al., 2015b; MacDonald et al., 2018; Patel et al., 2018; Ritter et al., 2017; Ritter et al., 2016a; Ritter et al., 2016b; Ronda et al., 2014; Yusufi et al., 2017).

As highlighted by Yusufi and colleagues (Yusufi et al., 2017), the ’omic profile of the cell rapidly changes upon transfection; 10,959 short indels, 3,313 SNPs, 21.91 Mb homozygous deletions, 228.83Mb heterozygous deletions, 2.40 Mb 2x amplifications, 15.39 Mb 4x amplifications and 11,896 genes affected due to the transfection process.

To put these numbers into context, there is 2.45Gb of genomic sequence in CHOK1 ancestral cell line, with a total of 24,383 predicted genes (Xu et al., 2011), suggesting around 49% of CHOK1 genes are affected after a single transfection event. It is important to note that this study profiled a single transfection event, so the aforementioned values are likely to change on a transfection-by-transfection basis.

To our knowledge, most studies exploring genetic pathways involved in unstable therapeutic protein production are performed either in daughter cell lines, derived from a single parent, or performed on cell lines that produce a single therapeutic protein (Betts and Dickson, 2017; Hefzi et al., 2016; Kremkow et al., 2015; Lee et al., 2015a; Lee et al., 2015b; MacDonald et al., 2018; Patel et al., 2018; Ritter et al., 2017; Ritter et al., 2016a; Ritter et al., 2016b; Ronda et al., 2014; Yusufi et al., 2017). There is an important question as to whether such findings can be replicated at an industrial scale, where ~100 cell lines are produced per therapeutic protein and the therapeutic protein being assessed changes constantly, based on strategic portfolio decisions. Studies based on a single transfection event do not replicate the vast heterogeneity at the 'omic level, created by multiple transfection events, which is more akin to the industrial scale. Therefore, identifying a single causative factor, which can be bioengineered into the host to ameliorate production instability, is a difficult task.

Although CHO production instability is well documented (Bailey et al., 2012; Barnes et al., 2003; Betts et al., 2015; Betts and Dickson, 2016; Pilbrough et al., 2009; Ritter et al., 2017), a causative pathway is yet to be identified. There are numerous potential pathways that may impact CHO genomic and/or production stability, this thesis will concentrate on DNA damage levels found at interstitial telomeric sites. In normal homeostasis in human, mouse and yeast cell models, telomeres are protected by the Shelterin complex, which confers inhibitory functions against DNA damage pathways (de Lange, 2005; Schmutz and de Lange,

2016; Wang et al., 2007). To understand how the telomeres may represent hotspots of DNA damage within these highly proliferative cell lines. Telomere repeat-binding factor 2 (TRF2), a main constituent of shelterin, will be investigated for its co-localisation to telomere sequences to assess interstitial telomere sequence (ITS) protection. Additionally, gH2AX co-localisation to telomeres will also be assessed to investigate whether there are any differential changes in telomere specific damage that may lead to increased heterogeneity in unstable cell lines, due to the damage eliciting mutational events.

Within this chapter, I have set out to identify a high-level pathway that could be present in all therapeutic protein producing cell lines, regardless of molecule. I hypothesise that interstitial telomere sequences within CHOK1a-GS-KO chromosomes represent sites that obtain preferential DNA damage, leading to an increase in genomic instability, which in turn promotes production instability.

Chromosome number distribution and relative telomere length changes between stable and unstable therapeutic protein producing cell lines, across early and late time points (4.2).

Aims and Hypotheses (4.2.1)

Chromosome number distribution within CHOK1a-GS-KO host obtained a median chromosome number of 19 and 20 at early and late time points, respectively. At both time points, a large range of chromosome numbers was observed. This heterogeneity will be investigated within CHOK1a-GS-KO producing cell lines to assess any fluctuations of chromosome number between productionally stable and unstable groups. Additionally, telomere length was shown to increase over time within the host, therefore interstitial telomere sequence (ITS) length will be quantified in stable and unstable producing cell lines, across early and late time points, to assess whether there are any ITS length changes characterising the different groups. I hypothesise that telomere length should be more abundant in unstable cell lines and at later time points, leading to greater chromosomal instability due to increased DNA damage.

Results (4.2.2)

Following on from CHOK1a-GS-KO host characterisation, 18 cell lines producing three different therapeutic proteins (protein 2, 3 and 5) were characterised for their chromosome distribution and telomere length. Cell lines were selected based on their production stability that was previously elucidated using Ambr 15's - an industry standard mini-bioreactor used to assess production stability.

To understand whether there are fundamental chromosome number differences that may be indicative of a stable or unstable cell line, chromosome numbers were quantified as previously described. 14 out of 18 cell lines retained a median chromosome number of 19 or 20 that reflects the CHOK1a-GS-KO host cell line. 4 out of 18 cell lines had a median of chromosome number between 35 and 38 chromosomes, suggesting that the single cell sorted clone was derived from a transfected cell in the host that obtained a 'aneuploid' number of chromosomes (table 4.1, figure 4.1). 'Aneuploid' cell lines had the greatest spread of chromosome number with 90% confidence interval (CI) ranges spreading between 17 to 41 chromosomes, indicating that these cell lines have largely heterogenous karyotypes compared to 'haploid' cell lines.

Interestingly, 3 of the 4 cell lines that are considered 'diploid' were productionally stable, suggesting the increased genetic material provides a mechanism that allows the cell lines to cope better with production stress across the stability assessment period. Both the modal and 90% CI range of chromosome number distribution were similar comparing back to the host cell line, suggesting the use of a selection agent does not have a drastic impact on chromosome numbers.

A 2-way ANOVA approach was utilised to compare the median chromosome numbers of cell lines within different therapeutic proteins, to assess any significant difference between stable and unstable groups. Comparing therapeutic protein and stability as factors, there was no significant difference (P=0.108). Pairwise comparisons using a planned comparison approach (table 4.2), where pairwise comparisons are first unadjusted and a post-hoc test is applied for the comparison pairs of choice. Hochberg's procedure was performed to compare chromosome number distribution for each therapeutic protein across stable and unstable groups, with no pairwise comparison being statistically significant. This indicates that chromosome number distribution does not fluctuate between stability and time point groups, suggesting the selection pressure method and different media composition of producing cell lines potentially confers chromosomal stability at the numerical level.

Table 4.1. Median chromosome numbers of productionally stable and unstable cell lines. Modal range of chromosome numbers and 90% confidence interval chromosome numbers are listed. Modal range shows the full range of chromosomal number of the images analysed. 90% CI range shows the range of chromosome number that applies to 90% of the images analysed.

Protein	*Cell Line*	*Median Chromosome Number*	*Modal Range*	*90% CI range*	*Stability*
2	1	20	13 to 22	19 to 21	Unstable
	2	19	11 to 23	16 to 21	Unstable
	3	20	11 to 41	18 to 30	Stable
	4	20	18 to 38	19 to 23	Unstable
	5	34	5 to 70	17 to 38	Stable
	6	38	11 to 43	20 to 41	Stable
3	7	19	13 to 38	18 to 35	Stable
	8	19	12 to 23	18 to 21	Stable
	9	19	16 to 38	18 to 35	Unstable
	10	19	17 to 29	18 to 20	Unstable
	11	19	14 to 36	18 to 30	Unstable
	12	35	14 to 41	19 to 38	Stable
5	13	19	11 to 40	16 to 20	Stable
	14	19	13 to 36	18 to 22	Stable
	15	19	15 to 37	19 to 20	Stable
	16	20	13 to 39	18 to 22	Unstable
	17	20	18 to 23	18 to 21	Unstable
	18	38	12 to 52	23 to 40	Unstable

Table 4.2. Stable vs. Unstable chromosome number distribution comparison. Data was analysed using a 2-way ANOVA approach with therapeutic protein and stability as factors. Unadjusted p-value represents all pairwise comparisons without adjustment for multiplicity (LSD test). Planned comparisons using Hochberg's procedure of therapeutic protein cell lines within stable and unstable groups forms the adjusted p-value. All pairwise comparisons were insignificant.

Comparison	*Unadjusted p-value*	*Adjusted p-value*
'2' and Stable vs. '2' and Unstable	0.0734	0.2937
'5' and Stable vs. '5' and Unstable	0.2357	0.4215
'3' and Stable vs. '3' and Unstable	0.3603	0.4215
'2' and Stable vs. '5' and Unstable	0.4215	0.4215

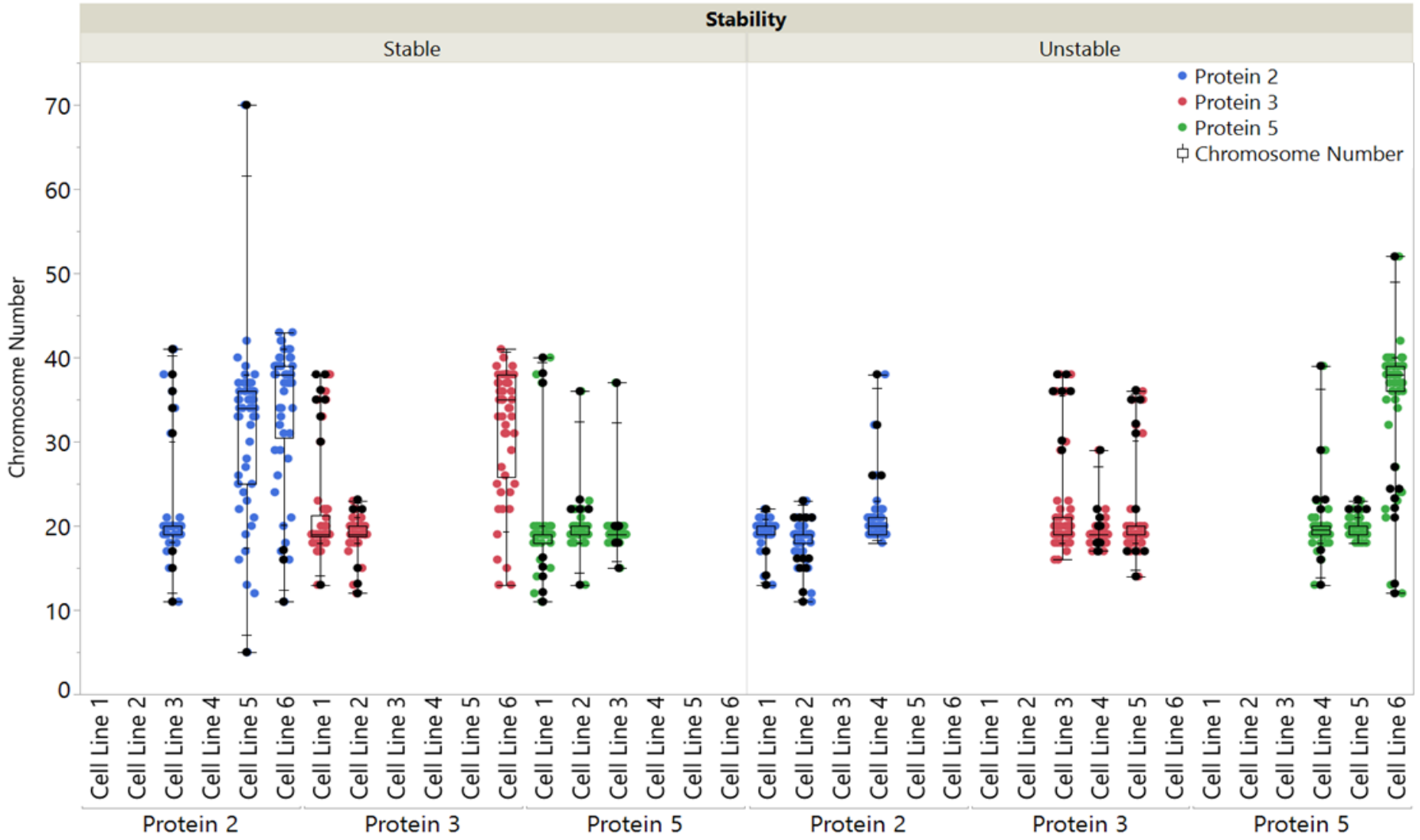


Figure 4.1. Chromosome number distribution across 18 cell lines divided equally into a stable and unstable cell line panel. Boxplots depict the median and percentile values of chromosome numbers for each cell line. Data for this figure is summarised in table 4.1.

To characterise telomere length changes in productionally stable and unstable cell lines, telomere quantification was performed as previously described for CHOK1a-GS-KO host telomere analysis. To identify whether telomere length plays a role in production stability, the same 18 cell lines were stained using a TTAGGGn fluorescent probe. Telomere length was calculated for 200 images of each cell line, across early and late time points, to assess whether telomere length fluctuates between stable and unstable cell lines, in addition to time.

A least square means (LSM) model was applied to the telomere length data set which considers numerous variables of the data. In contrast to the arithmetic mean, LSM is an average based on a linear model that is adjusted for covariates (e.g. time point, chromosome number, protein etc.), providing a better estimate of the true population mean.

LSM calculation of telomere length, considering stability, early and late time points, across modal chromosome numbers is plotted in figure 4.2a. Protein 2 obtained a larger difference in telomere proportion mean when comparing stable and unstable cell lines, however, 95% confidence limit bars indicate that the differences between the means heavily overlap across the data set. Protein 2 difference observed between the stable and unstable telomere proportion LSM was not shared with protein 3 and 5, indicating the increase in telomere proportion for stable cell lines may only be a protein specific difference.

Overall, patterns in telomere length changes does not appear to be consistent across this panel of cell lines. Protein 2 telomere length proportion decreases from early to late timepoints whereas protein 3 and 5 have mixed profiles (increase and decreases in telomere length) dependent on chromosome number category. The varying profiles across early and late time points identified in the LSM plot is reflected in the insignificance of comparing pooled data into early and late categories (pooled T-test, P=0.58, figure 4.2b), corroborating the notion that there is no difference in telomere length proportion over a prolonged culture period.

To assess whether there is an overall difference in telomere length across stable and unstable cell lines, data was pooled into stable and unstable categories (figure 2c). Mean telomere length of unstable cell lines was shown to increase by 0.3% from 2% in the stable cell line category. This difference was found to be highly significant (P=<0.0001, figure 2c), however, the large number of images analysed for each group may have contributed to the increased sensitivity of the statistical test. Additionally, 0.3% may not be a large enough increase to elicit a physiological response.

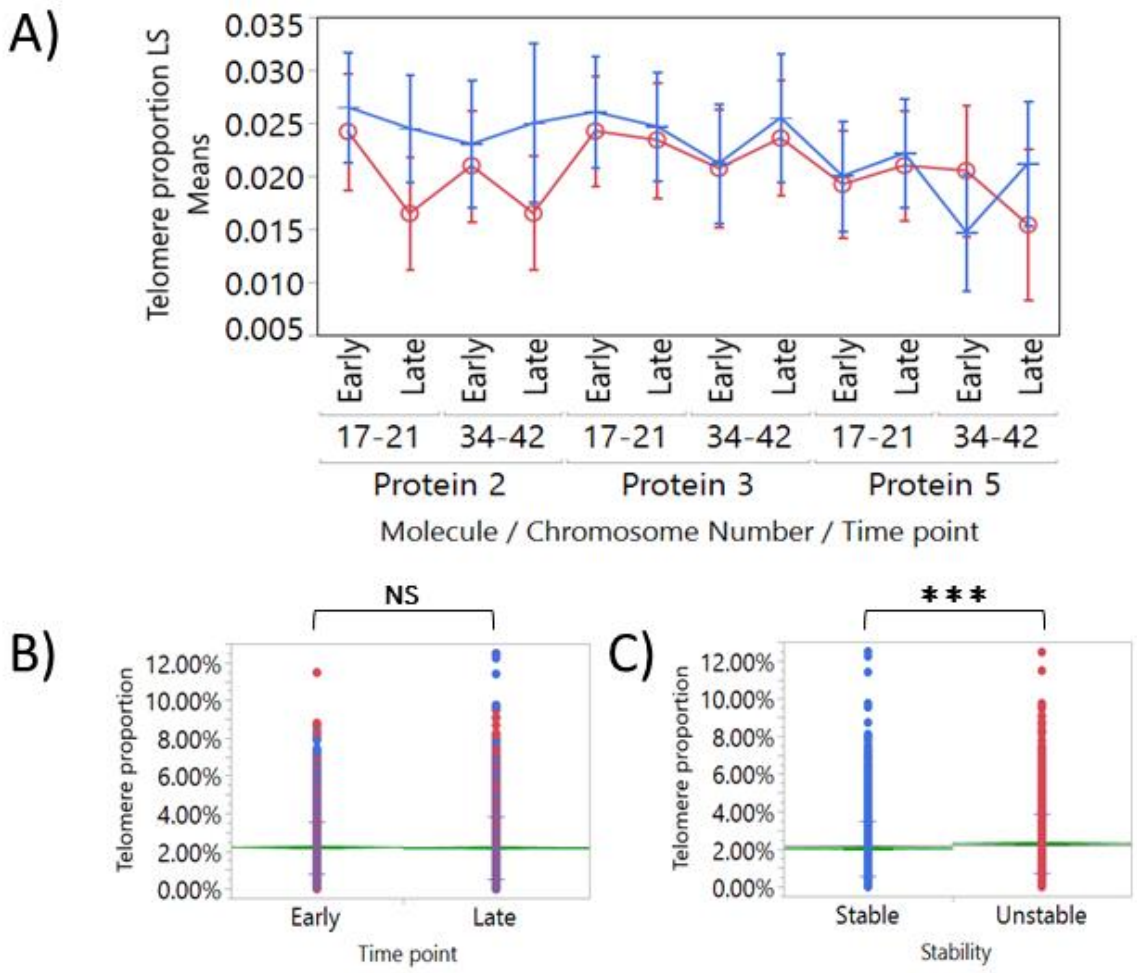


Figure 4.2. Stable vs. Unstable telomere length quantification and comparison across stability groups and timepoints. A) Least square means (LSM) plot of telomere length proportion of cell lines. LSM considers the relationship of telomere proportion in stable (blue) and unstable (red), early and late and chromosome number categories. 95% confidence intervals heavily overlap for each LSM; therefore, no significant difference can be observed. B) Data from each cell line was pooled into early and late categories, to assess overall differences in telomere length proportion over prolonged cell culture. No significant difference is observed (pooled T-test, P=0.58). C) Telomere length data was pooled into stable and unstable categories. Mean telomere length for stable cell lines is 2% and increased 0.3% to 2.3% in the unstable cell line category (P=<0.0001).

Characterising stable and unstable cell line karyotypes to understand the genomic mutation landscape of the productionally unstable phenotype (4.3).

Aims and hypotheses (4.3.1)

As previously identified, CHOK1a-GS-KO host, used for therapeutic protein production, has a heterogeneous karyotype that is maintained over a 6-month culturing period. Here, MFISH has been utilised to characterise productionally stable and unstable cell lines across early and late time points to identify any differences or commonalities in genomic instability profiles across the different groups. I hypothesise that genomic instability is intrinsically linked to production instability. Therefore, identification of potential genomic instability markers at an early stage will be investigated for their robust prediction power of production stability.

Results (4.3.2)

Metaphase chromosomes from a panel of stable and unstable cell lines were harvested and 'painted' using MFISH, as previously described (methods, 2.3.4). Chromosome populations for each cell line was assessed across early and late time points, using the population determining method (2.7.1), as outlined previously.

6 stable and 8 unstable cell lines expressing different therapeutic proteins were selected based on their pre-determined production stability as assessed in automated mini bioreactors (Ambr 15). Cell lines were thawed and then passaged three times to allow for recovery (>98% viability).

Cell lines with a median chromosome number of 19 to 20 were selected for the proceeding analysis, cell lines with chromosome numbers that are considered 'aneuploid' were excluded from analysis as these cell lines are not representative of general cell line population identified here (figure 4.1 and table 4.1) and elsewhere (Deaven and Petersen, 1973; Kao and Puck, 1968; Thomas et al., 2018; Vcelar et al., 2018b).

Across the panel of stable and unstable cell lines, all obtained multiple karyotypically distinct populations, only differing in the proportion of CCA and NCCA population frequency (figure 4.3a). This indicates that the propensity for gross chromosomal mutations is maintained after transfection of the host and single cell sorting events. Comparisons between the populational composition of stable and unstable groups indicate there is a higher proportion of NCCA populations within the unstable group.

Calculating the overall CCA and NCCA percent frequency for stable and unstable categories indicates a high percentage frequency of CCA populations correlates to productionally stable cell lines (figure 4.3b). Conversely, a greater percentage frequency of NCCA populations was retained in the unstable arm of the cell line panel (figure 3b and table 4.3 and 4.4, Two-way ANOVA, P=0.0003). The distinct groupings of %CCA and %NCCA for stable and unstable cell lines indicate that this genomic metric could be utilised as a production stability predictor.

After 6-months of culture, cell lines were re-analysed, and their populations re-determined using the same methodology. Protein 3, cell line 7 (figure 4.3a, P3.C7) late population data was excluded from analysis in figure 4.3c. This is due to the cell line becoming 'aneuploid' over the 6-month culturing process and is therefore not comparable with the rest of the data set (figure 4.4). The distribution of NCCA populations increased drastically regardless of stability except protein 5, cell line 16 (P5.C16) (figure 4.3a).

Comparisons of overall CCA and NCCA percent frequency shows a concomitant decrease in CCA and increase in NCCA frequency across the cell culturing period (figure 4.3c and table 4.3, Two-way ANOVA, P=<0.0001). Increases in NCCA populations indicates cell lines become more heterogenous over time, regardless of their production stability (table 4, P=0.4434). Heterogenous cultures may obtain cells that produce differing amounts of therapeutic protein that may lead to the fluctuations in overall titre over the stability assessment, thus causing the production instability witnessed within CHO cell lines. In a cell line development environment, this data suggests that a cell line identified to have a prominent level of genetic instability at an early timepoint will become increasingly heterogenous and genetically unstable with time, having a major impact on its ability to homogenously express its therapeutic protein, ergo impacting its expression stability.

To understand whether there are common chromosome mutations across cell lines that may be able to identify stability groups, confirmed mutations from the early time point were compiled by chromosome number and coloured by cell line and stability (figure 4.3d and e, respectively). All chromosomes analysed obtained a mutation in one or more cell lines, this indicates that all chromosomes are amenable to deletions, amplifications, rearrangements and/or translocations with no obvious pattern being recognised.

Differentiating the mutations by stability indicates that chromosome 6 and 8 have the highest mutation rates overall and the majority of the mutations belong to unstable cell lines. 11 out of 14 cell lines obtained a mutation in chromosome 6, 3 of the 11 are productionally stable and 8 are productionally unstable. 2 out of 3 stable lines are derived from the same therapeutic protein, which may identify a therapeutic protein specific difference with regards to chromosome 6s potential ability to confer production stability in 57% of total cell lines analysed. 5 out of 8 cell lines that obtained a mutation in chromosome 8 are considered productionally unstable, indicating that a mutation in this chromosome could account for 36% of unstable cell lines analysed. Taken together, these results indicate a potential causal-link between production and genomic instability and highlight the prediction power of this method for determining production stability at early time points.

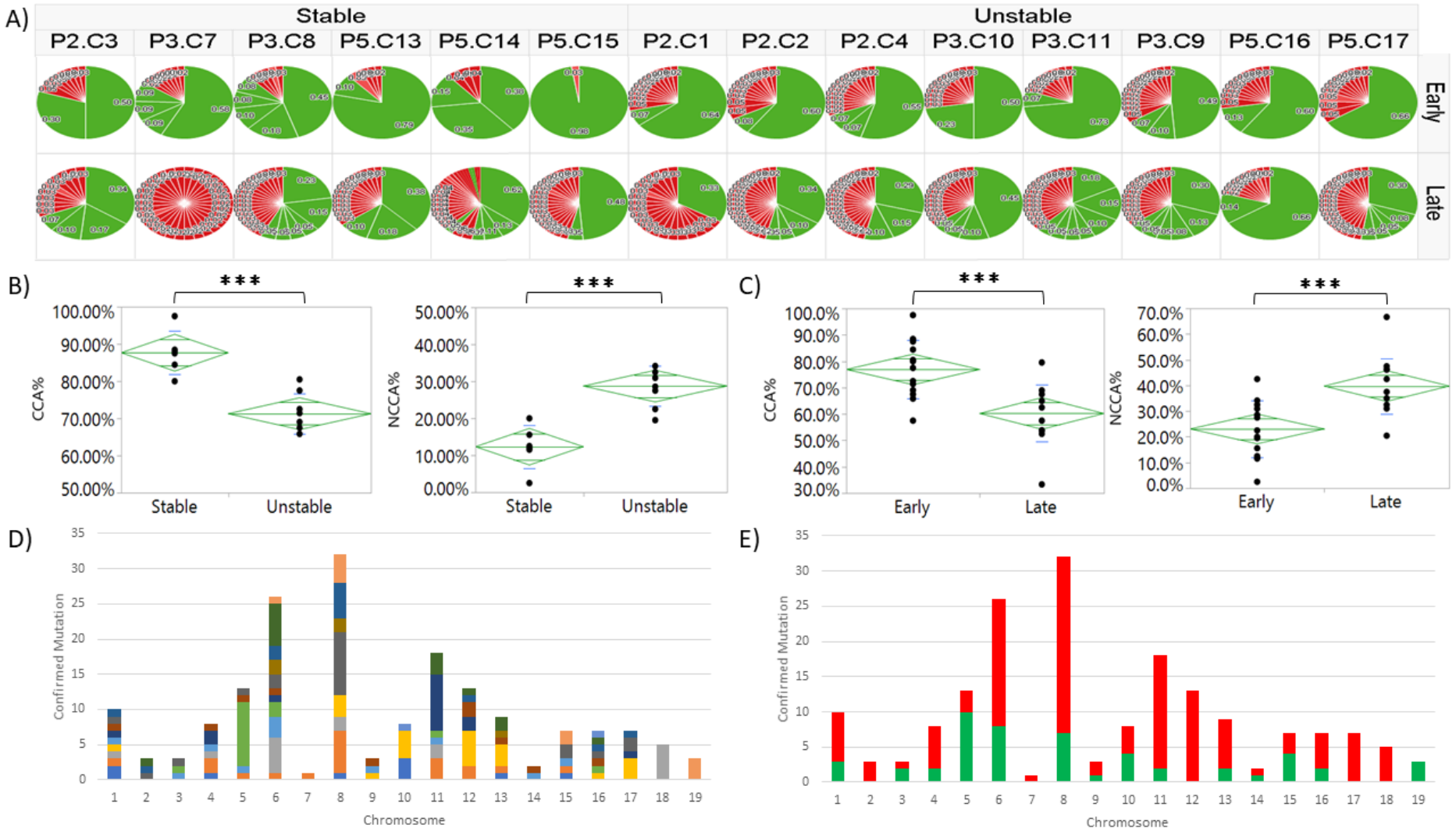

A)
Stable
Unstable
P2.C3
P3.C7
P3.C8
P5.C13
P5.C14
P5.C15
P2.C1
P2.C2
P2.C4
P3.C10
P3.C11
P3.C9
P5.C16
P5.C17
Early
Late
B)
***
CCA%
100.00%
90.00%
80.00%
70.00%
60.00%
50.00%
Stable
Unstable
***
NCCA%
50.00%
40.00%
30.00%
20.00%
10.00%
0.00%
Stable
Unstable
C)
***
CCA%
100.0%
90.0%
80.0%
70.0%
60.0%
50.0%
40.0%
30.0%
Early
Late
***
NCCA%
70.0%
60.0%
50.0%
40.0%
30.0%
20.0%
10.0%
0.0%
Early
Late
D)
Confirmed Mutation
35
30
25
20
15
10
5
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19
Chromosome
E)
Confirmed Mutation
35
30
25
20
15
10
5
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19
Chromosome

Figure 4.3. Stable vs. Unstable comparison of karyotypic heterogeneity. A) Population pie charts of each cell line divided into stability and time point categories. CCA (green) and NCCA (red) pie segments highlight an increase in NCCA populations when comparing stable to unstable and early too late. B) Overall CCA and NCCA frequencies were calculated across each stability group and differences between each group was statistically significant (Two-way ANOVA, P=0.01). The grand mean was calculated at 78% indicating a potential threshold for production stability designation. C) CCA and NCCA population frequency difference between early and late time points are statistically significant (Two-way ANOVA, P=<0.0001), indicating that NCCA populations increase over prolonged periods of cell culture, leading to more heterogeneity. Green triangles represent the population mean and 95% confidence intervals, blue lines indicate standard deviation. D) Mutations categorised by chromosome, cell lines are represented by the different colour segments. Chromosome 6 and 8 retain the most mutations with chromosome 6 being mutated in 11 out of 14 cell lines. E) similar bar chart as D except sorted by stability. All chromosomes except 2, 17, 18, and 19 obtained mutations in both stable and unstable cell lines. No pattern of specific chromosome mutations was observed.

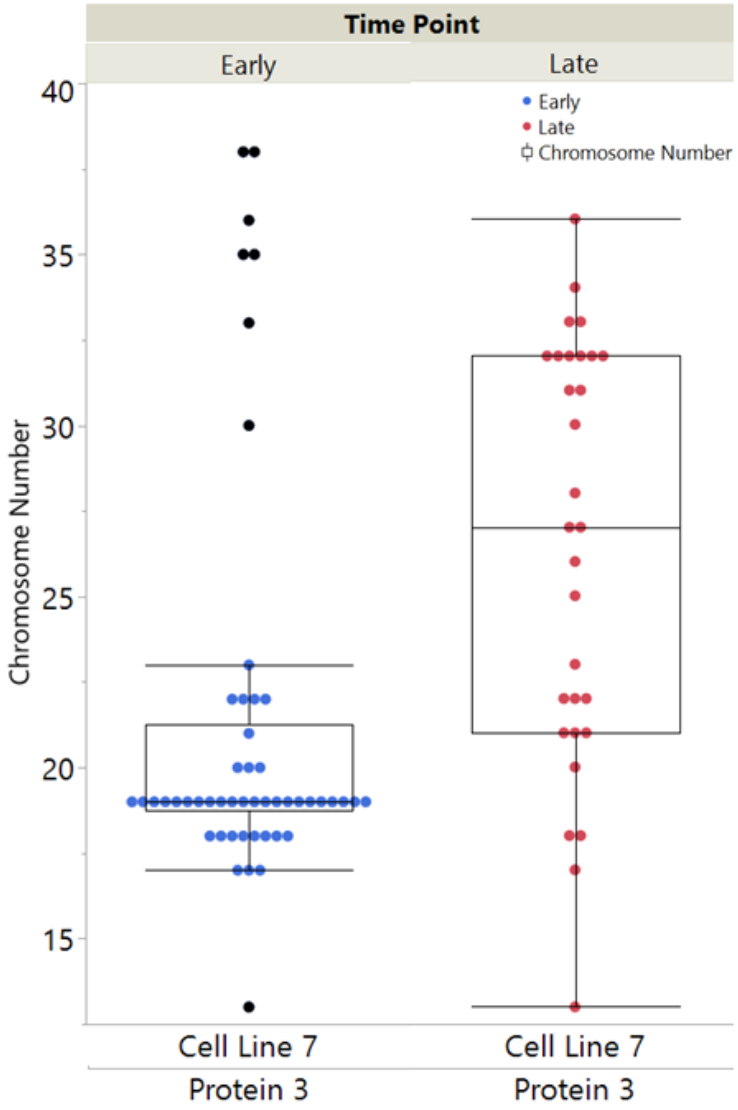


Figure 4.4. Chromosome distribution plot of protein 3, cell line 7 (P3.C7) between early and late time points. Median chromosome number has increased from 19 to 27 over a 6-month culturing period.

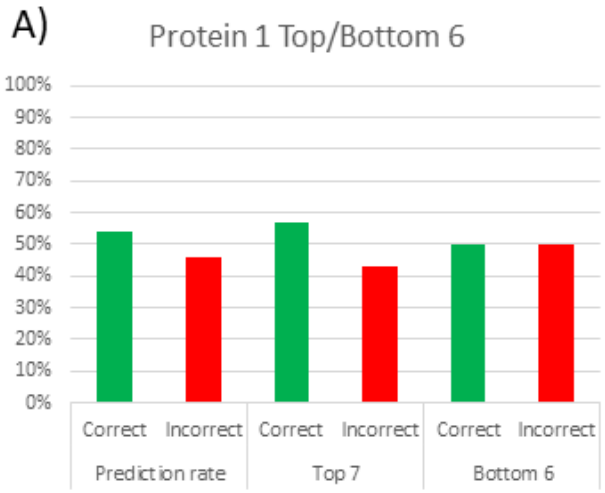
Table 4.3. ANOVA table of CCA% comparisons between stability and time point. Statistically significant differences in CCA% was obtained in stability (P=<0.01) and time point (P=<0.0001).

	Sums of squares	Degrees of freedom	Mean square	F-value	p-value
Stability	0.08	1	0.084	10.78	0.0033
Timepoint	0.22	1	0.222	28.71	< 0.0001
Stability * Timepoint	0.03	1	0.025	3.26	0.0840
Residuals	0.18	23	0.008		

Table 4.4. Hochberg's pair-wise comparison adjusted P-values for stability and time point. Significant differences are observed when comparing %CCA between stable and unstable cell lines across early and late time points. There was no significant difference observed between %CCA of late stable and unstable cell lines.

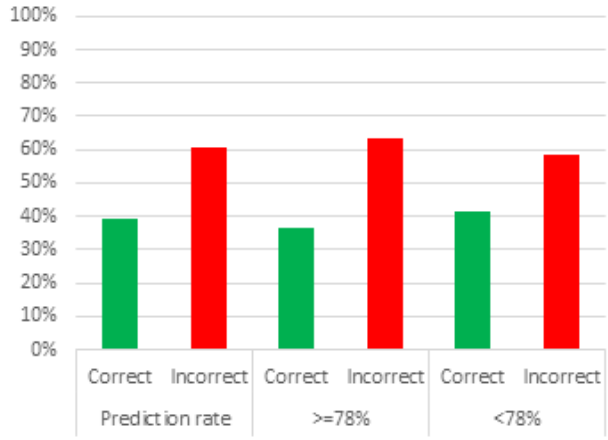
	Comparison	Unadjusted p-value	Adjusted p-value
1	Stable Late vs. Stable Early	7.58516402437071e-05	0.0003
2	Unstable Early vs. Stable Early	0.0021817011643801	0.0065
3	Unstable Late vs. Unstable Early	0.00667007241142281	0.0133
4	Unstable Late vs. Stable Late	0.443449756696863	0.4434

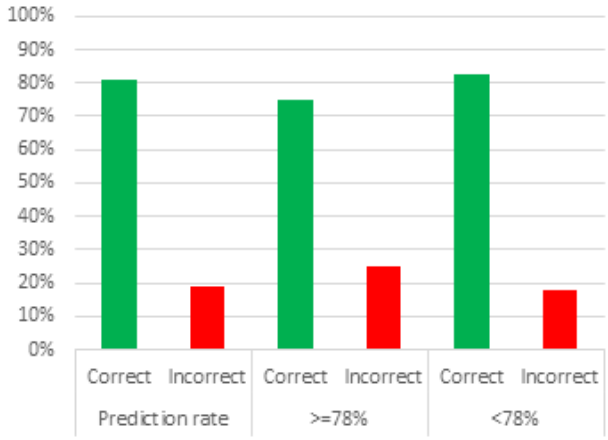
A)


Protein 1 Top/Bottom 6
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
Correct
Incorrect
Correct
Incorrect
Correct
Incorrect
Prediction rate
Top 7
Bottom 6

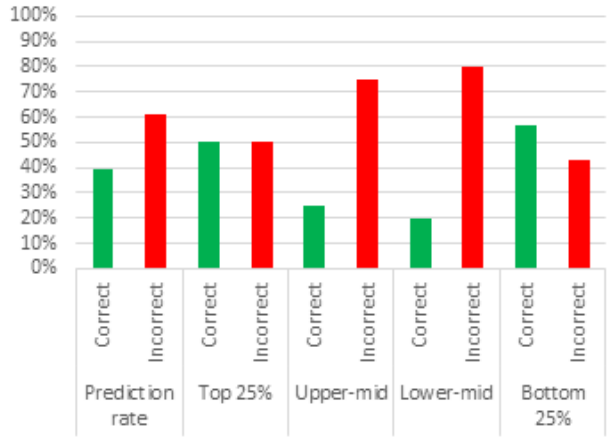

Protein 4 Top/Bottom 6
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
Correct
Incorrect
Correct
Incorrect
Correct
Incorrect
Prediction rate
Top 6
Bottom 6

B)


Protein 1 Threshold Prediction
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
Correct
Incorrect
Correct
Incorrect
Correct
Incorrect
Prediction rate
>=78%
<78%


Protein 4 Threshold Prediction
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
Correct
Incorrect
Correct
Incorrect
Correct
Incorrect
Prediction rate
>=78%
<78%

C)


Protein 1 Quartiles Prediction
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
Correct
Incorrect
Correct
Incorrect
Correct
Incorrect
Correct
Incorrect
Correct
Incorrect
Prediction rate
Top 25%
Upper-mid
Lower-mid
Bottom 25%

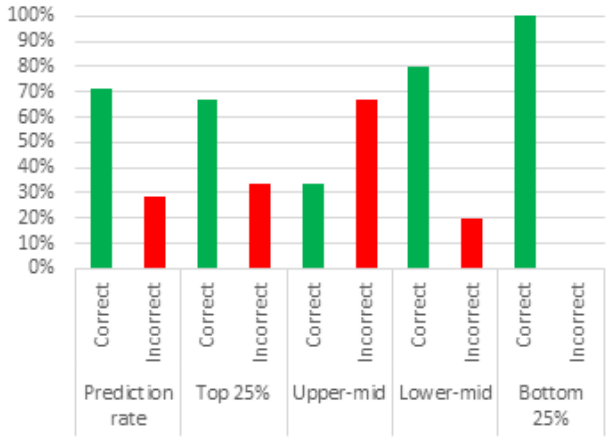

Protein 4 Quartiles Prediction
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
Correct
Incorrect
Correct
Incorrect
Correct
Incorrect
Correct
Incorrect
Correct
Incorrect
Prediction rate
Top 25%
Upper-mid
Lower-mid
Bottom 25%

Figure 4.5. Prediction of Stable and Unstable cell lines using CCA% as a stability marker. Three different prediction methods were devised before the unblinding of cell lines after analysing results. Cell lines were sorted by CCA% from high to low and the different prediction methods were applied and the prediction success rate calculated. A) Top and bottom 25%, utilised to identify the most stable and unstable cell lines. B) Threshold prediction based on initial stable and unstable cell line panel, threshold set at CCA 78%. CCA >=78% is considered a productionally stable cell line, conversely <78% is considered as a productionally unstable cell line. C) Cell lines sorted by CCA % are divided into quartiles to identify top 25% and bottom 50% for cell line triaging.

Thus far, a distinct separation of CCA and NCCA % frequencies between stable and unstable cell line groups (figure 4.3b), has been identified. As the cell lines were analysed at an early time point (~20 generations) the possibility that CCA vs NCCA % frequency could be utilised as a genomic stability metric, predictive of production stability at an early timepoint, was investigated. This may be beneficial to cell line development timelines as it could provide the means to triage cell lines at a much earlier time point (20 generations) compared to completing the whole stability assessment (70-150 +/- 10 generations).

48 cell lines, split between two different therapeutic proteins, were selected to represent a normal distribution of stable and unstable cell lines for any given new live project and their stability remained blinded until CCA and NCCA populations were analysed. The ability to predict cell line production stability, based on their ranking of %CCA, provides real world evidence of the methods prediction power as it mimics it use in the critical path of CLD for triaging cell lines with unknown production

stability. As this method is novel and there is no defined prediction method, three separate methods were tested before unblinding the data (figure 4.5).

Top 6 and bottom 6 (figure 5a) predictions based on the ranking of %CCA for each cell line has the potential to quickly identify stable (for cell line progression) and unstable (for triaging) cell lines. Overall Protein 4 cell lines had a correct prediction rate of 82.5% but this was skewed towards correctly identifying unstable cell lines (100% correct) compared to stable (67.5% correct). Protein 1 cell line prediction obtained an overall prediction rate of 52.5%, which is inadequate to confidently identify and triage cell lines.

A second prediction method based on a %CCA threshold, defined from our previous panel of cell lines (78% threshold, figure 4.3b), showed a similar trend in prediction success when comparing Protein 4 and Protein 1 cell lines (figure 4.5b). Anything equal to or above 78% CCA was considered to be productionally stable, less than 78% CCA was considered as unstable. Protein 4 cell lines obtained an overall correct prediction of ~80% which was more evenly balanced between stable and unstable correct predictions – 75% and 82.5%, respectively. Protein 1 incorrect predictions increased to ~ 60%. A potential benefit of this prediction method is the threshold of CCA% could be better refined as more data is generated, providing a potential increased prediction rate.

Quartile predictions (figure 4.5c) could be utilised to readily identify the top 25% stable cell lines and the bottom 50% productionally unstable lines. Robustly triaging the bottom 50% could drastically increase cell line development capacity by freeing up limited mini bio-reactor space. Protein 1 had an overall 60% incorrect prediction, as with the threshold prediction, with the majority of incorrect predictions residing within the upper and lower middle 50% of ranked cell lines (excluding top and bottom 25%). Protein 4 obtained a 70% correct prediction overall which was largely obtained in the bottom 50% (lower-mid = 80% correct, bottom

25% = 100% correct) whilst obtaining 67.5% correct prediction rate for the top 25% of stable cell lines.

Overall, Protein 4 predictions were successful whereas Protein 1 cell lines were predicted poorly. Initially, we hypothesised the discrepancy between successful predictions could be due to the generation number of the cell lines at the time of sampling (figure 4.6a and b). Protein 4 cell lines were thawed from a development cell bank and already undergone cell culturing for around 10 generations. Whereas, Protein 1 cell lines were thawed from a master cell bank where their generation number starts from 0. Hence, there is a difference of around ~10 generations between these samples at the time of analysis potentially effecting the success rate of prediction as their mutational profiles have not stabilised at an earlier generation number.

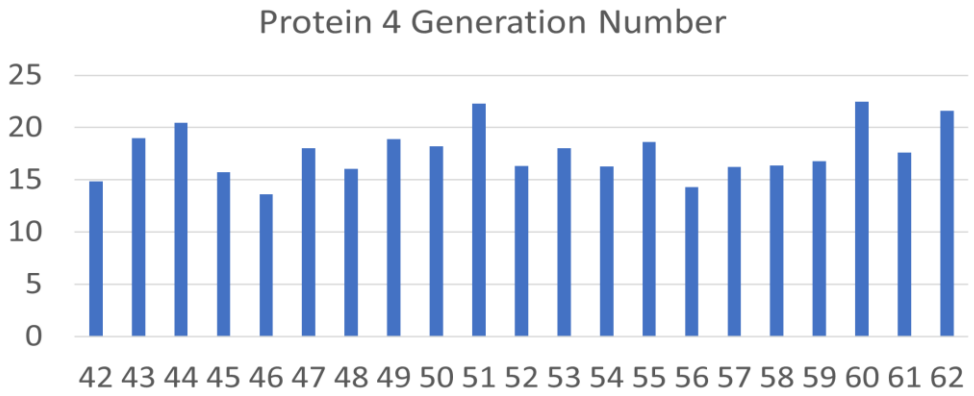
Taking this into account, %CCA and %NCCA differences in cell line production stability was re-assessed (figure 4.6c and d). Taking all data together (figure 6c) there is still a highly significant correlation between %CCA and production stability (pooled T-test, P=<0.005), even when including protein 1 poor prediction results. The statistical significance increases dramatically when protein 1 data is excluded from the analyses (pooled T-test, P=<0.0001).

Data presented here indicates that there is a significant difference between the heterogeneity of productionally stable and unstable cell lines, when populations are grouped by CCA and NCCA designations. Interestingly, all cell lines obtain heterogenous karyotypes that is exacerbated over prolonged culture periods, reflecting observations of cell line titre drastically decreasing after ~100 generations (data not shown). Increases of NCCA populations leads to increased genetic heterogeneity which seems to impact a cell lines ability to maintain production of its therapeutic protein. Conversely, a de novo mutation that is acquired but allows the cell to establish itself within the culture flask (>=5% frequency) appears to be correlated with production stability, as cell lines with

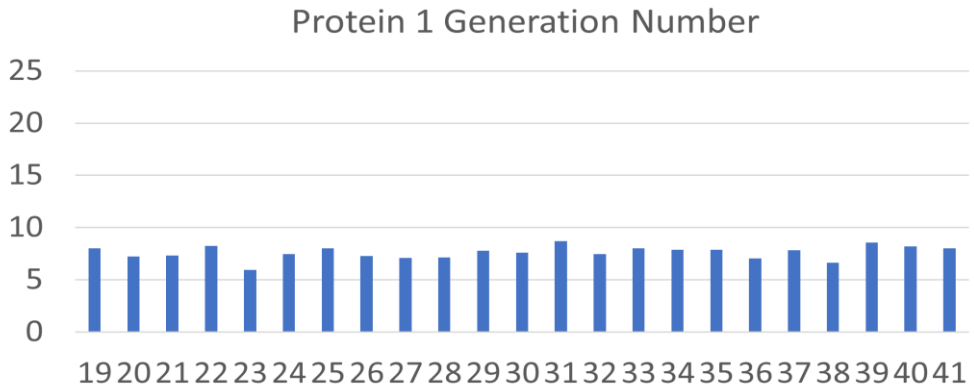
heterogenous populations that are predominantly CCA are on the whole productionally stable (figure 4.3a, b and table 4.3).

To our knowledge, the data presented here represents the first study to investigate novel findings on a potential mechanism for cell line production stability in an industry relevant panel of cell lines (60 cell lines across five different therapeutic proteins). Promising production stability prediction results, across multiple therapeutic protein expressing cell lines, provides evidence that the prediction method could be robust enough to be utilised in an industry setting.

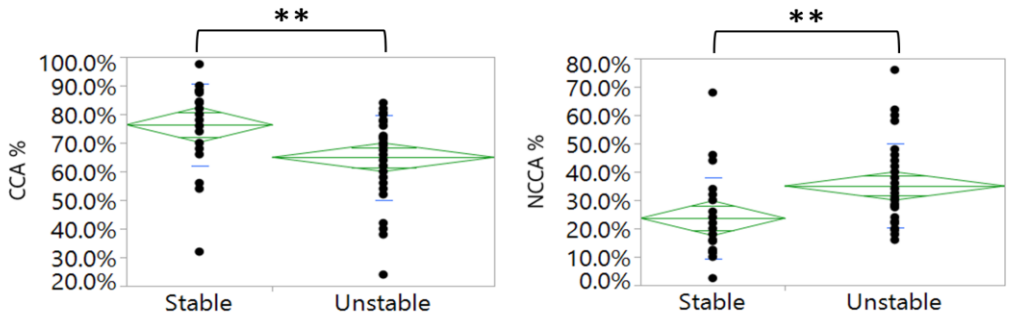
A)


Protein 4 Generation Number
25
20
15
10
5
0
42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62

B)


Protein 1 Generation Number
25
20
15
10
5
0
19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41

C)


**
CCA %
100.0%
90.0%
80.0%
70.0%
60.0%
50.0%
40.0%
30.0%
20.0%
Stable
Unstable
**
NCCA %
80.0%
70.0%
60.0%
50.0%
40.0%
30.0%
20.0%
10.0%
0.0%
Stable
Unstable

D)

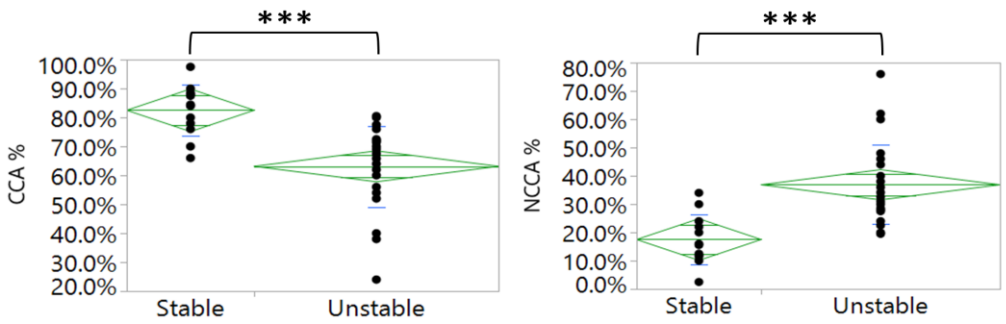

***
CCA %
100.0%
90.0%
80.0%
70.0%
60.0%
50.0%
40.0%
30.0%
20.0%
Stable
Unstable
***
NCCA %
80.0%
70.0%
60.0%
50.0%
40.0%
30.0%
20.0%
10.0%
0.0%
Stable
Unstable

Figure 4.6. Protein 1 and 4 comparison of generation number and comparison between all cell line CCA% data and data with excluded outliers. A and B) Protein 4 and 1 generation number from pre-master cell bank. Protein 4 cell line generation numbers reside within 15-20 generations, whereas protein 1 cell line generation numbers are <10 generations. C) Comparison of %CCA and %NCCA between all therapeutic proteins (initial stability panel) in stable and unstable groups, including protein 1 data (pooled T-test, P=<0.01). D) Same comparison as C but with protein 1 data excluded (pooled T-test, P=<0.0001).

In an attempt to elucidate a link between genetically heterogenous subpopulations observed in stable and unstable cell lines, to fluctuating IgG titre, witnessed in productionally unstable cell lines, a flow cytometry experiment was performed to analyse intracellular IgG content of cells. Cell lines from the same stable and unstable panel were sampled, fixed, permeabilised and stained using an anti-human IgG conjugated to an Alexafluor-647, as previously described (methods, 2.5.2).

Cell gating strategy was performed initially on forward scatter (FSC-A) and side scatter (SSC-A) dot plots to identify cells within the suspension, then a subsequent gate was performed on FSC-A and FSC-H to exclude cell doublets (figure 4.7a). IgG positive and IgG negative gates were set using a CHOK1a-GS-KO stained with anti-human IgG-647 (negative control) and a producer cell line stained with the same antibody (figure 4.7b, red and blue peaks, respectively). Staining of stable and unstable samples was performed as previously described. All samples obtained a peak within the IgG positive gate, as expected (figure 4.7c). When grouping the peaks based on production stability, an interesting pattern emerges, as stable cell lines (figure 4.7d) appear to have more consistent IgG positive fluorescence peaks, compared to unstable cell

lines (figure 4.7e) that obtained a wider spread of IgG peaks across the fluorescent spectrum.

Although a difference can be visually observed, when comparing group IgG positive mean variance across stable and unstable groups, Bartlett's variance test was insignificant (P=>0.05). However, P3.C7 (figure 4.7d, green peak) was observed to be an outlier, its externally studentised residual (Bate and Clark, 2014) was calculated and obtained a value of 17 (similar to 17 SD away from stable group mean), indicating it is an outlier within this sample set. Upon exclusion, Bartlett's variance test was performed again and was found to be significant (P=<0.0001). Taking these results together, it suggests that there may be a difference in the consistency of intracellular IgG production between stable and unstable groups, however, due to the relatively low sample size, there is a discrepancy with the statistical tests. Further samples must be analysed to confirm or disprove this observation.

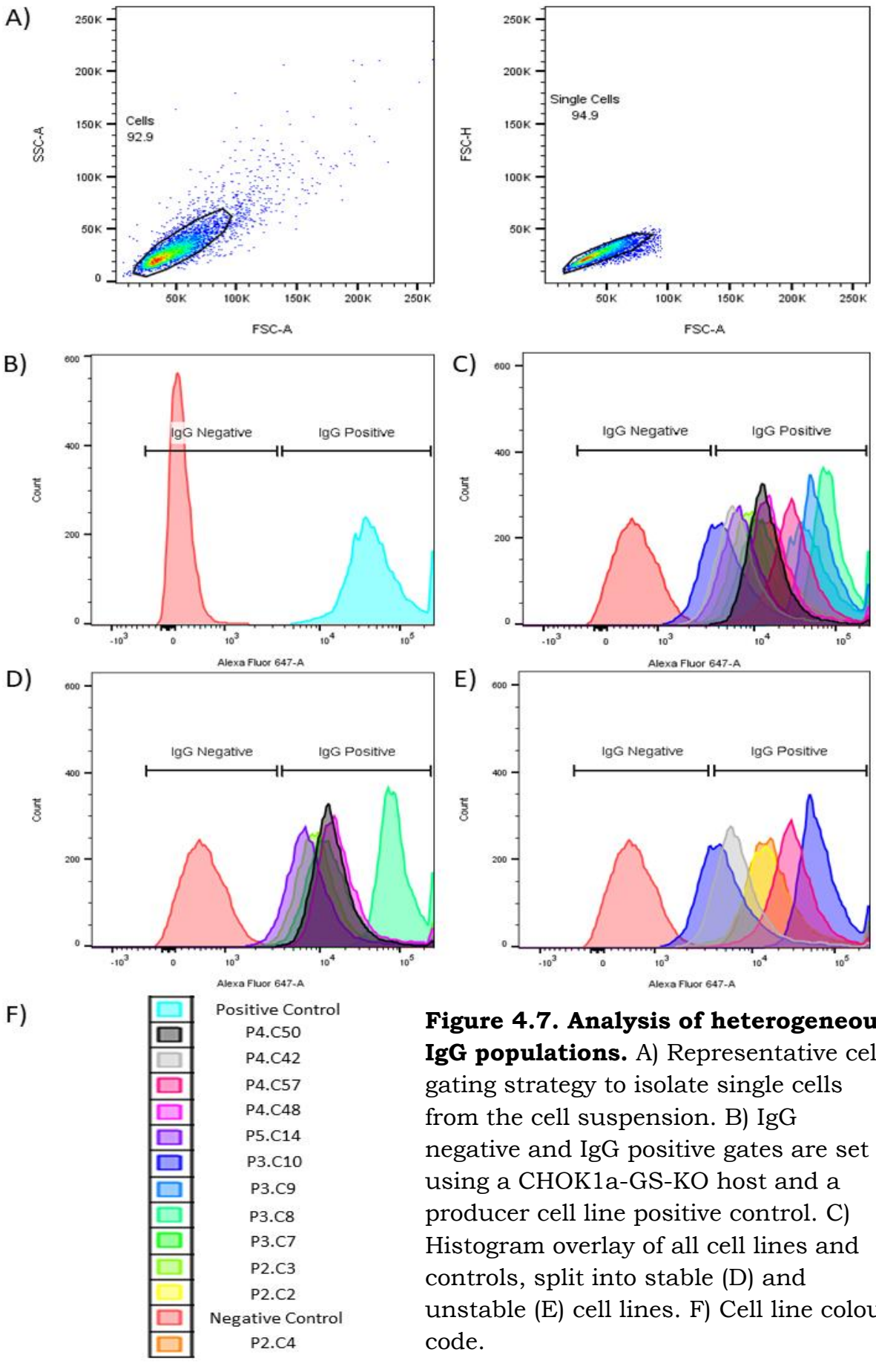


Figure 4.7. Analysis of heterogeneous IgG populations. A) Representative cell gating strategy to isolate single cells from the cell suspension. B) IgG negative and IgG positive gates are set using a CHOK1a-GS-KO host and a producer cell line positive control. C) Histogram overlay of all cell lines and controls, split into stable (D) and unstable (E) cell lines. F) Cell line colour code.

Characterising Shelterin protection of telomeres in stable and unstable cell lines, investigating whether this impacts telomere specific DNA damage (4.4).

Aims and Hypotheses (4.4.1)

CHOK1a-GS-KO analysis of telomere protection (TRF2 + telomere co-localisation) and telomere specific damage (TIF, gH2AX and telomere co-localisation) characterised the hosts baseline level of DNA damage and telomere protection. A panel of stable and unstable cell lines will be investigated for any differential patterns of DNA damage, TIF and telomere protection between productionally stable and unstable cell lines. Increased telomere specific damage and / or deprotection could lead to increased DNA damage signals leading to greater mutational events. Increased mutations will result in a more heterogenous cell line, potentially leading to fluctuations in titre over the stability period, thus impacting cell line production stability across a 4-6-month stability assessment.

Results (4.4.2)

To assess potential differences in gH2AX, telomere dysfunction (TIF, gH2AX and telomere co-localisation) and telomere protection (TRF2 + telomere co-localisation), stable and unstable cell lines were sampled and stained for TRF2 or gH2AX and telomere sequences, as previously described (figure 4.8, representative images). gH2AX, TIF and TRF2 and telomere foci were quantified using Image J and foci per cell was calculated.

Comparing CHOK1a-GS-KO host gH2AX foci per cell to the positive control (treated for 1 hour with 1ng/ml Neocarzinostatin) there was a statistically significant difference showing the assay can detect increases in DNA damage through gH2AX foci quantification (figure 4.8, Two-way ANOVA, Hochberg adjusted P-value, P=<0.0001).

Comparing producer cell lines in stable and unstable groups revealed no significant difference in gH2AX foci in therapeutic protein producing cell lines in a cell culture maintenance environment (figure 4.9, Two-way ANOVA, P=>0.05). Moreover, gH2AX foci levels did not fluctuate compared to CHOK1a-GS-KO host, regardless of production stability (Two-way ANOVA, P=>0.05), suggesting a low basal level of DNA damage is maintained within CHO producing cell lines during routine maintenance. All cell lines were significantly different compared to the positive control (Two-way ANOVA, P=<0.0001).

Quantification of telomere induced foci (TIF) showed a similar trend as gH2AX foci quantification. Overall there was a minimal occurrence of TIF within therapeutic protein producing cell lines, regardless of stability (figure 4.10, Two-way ANOVA, P=>0.05). Cell line 9 and 10 of therapeutic protein 3 showed instances of elevated TIFs in the cells analysed, increasing the mean to 0.2 TIFs per cell. The increase of TIFs within these cell lines were still statistically insignificant when comparing therapeutic protein 3 stable and unstable cell lines against the host, positive control and each stability group (table 4.5, two-way ANOVA, Hochberg adjusted P-values).

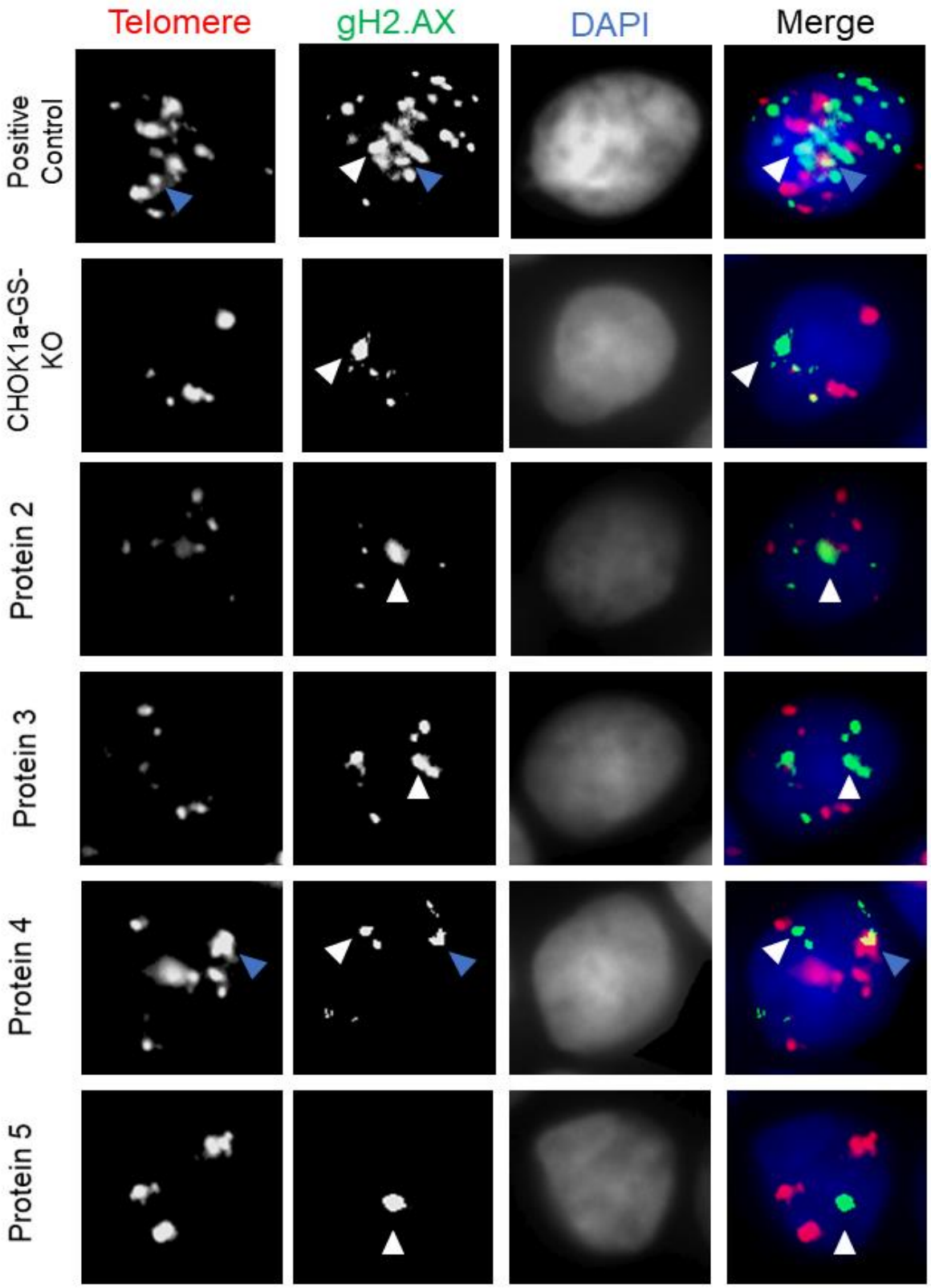


Figure 4.8. Representative immunofluorescent images used to quantify overall and telomere specific DNA damage, from each therapeutic protein expressing cell line. Grey scale images are shown for telomere (red), gH2AX (green, white arrows) and DAPI (blue) for coloured merged images. TIFs are quantified when telomere and gH2AX

signals overlap (blue arrows). Data quantified from these representative images was used to compare differences in DNA damage between Stable and Unstable cell lines.

Table 4.5. Two-way ANOVA with Hochberg's adjusted P-values of therapeutic protein 3 stable and unstable cell lines compared against the host, positive control and stability groups.

	Comparison	Unadjusted p-value	Adjusted p-value
1	Protein 3 Stable vs. Pos. Cont. Control	2.22044604925031e-16	< 0.0001
2	Protein 3 Unstable vs. Pos. Cont. Control	1.33226762955019e-15	< 0.0001
3	Protein 3 Unstable vs. Protein 3 Stable	0.12318087281857	0.2721
4	Protein 3 Stable vs. CHOK1a-GS-KO Control	0.136059992207968	0.2721
5	Protein 3 Unstable vs. CHOK1a-GS-KO Control	0.973836328666101	0.9738

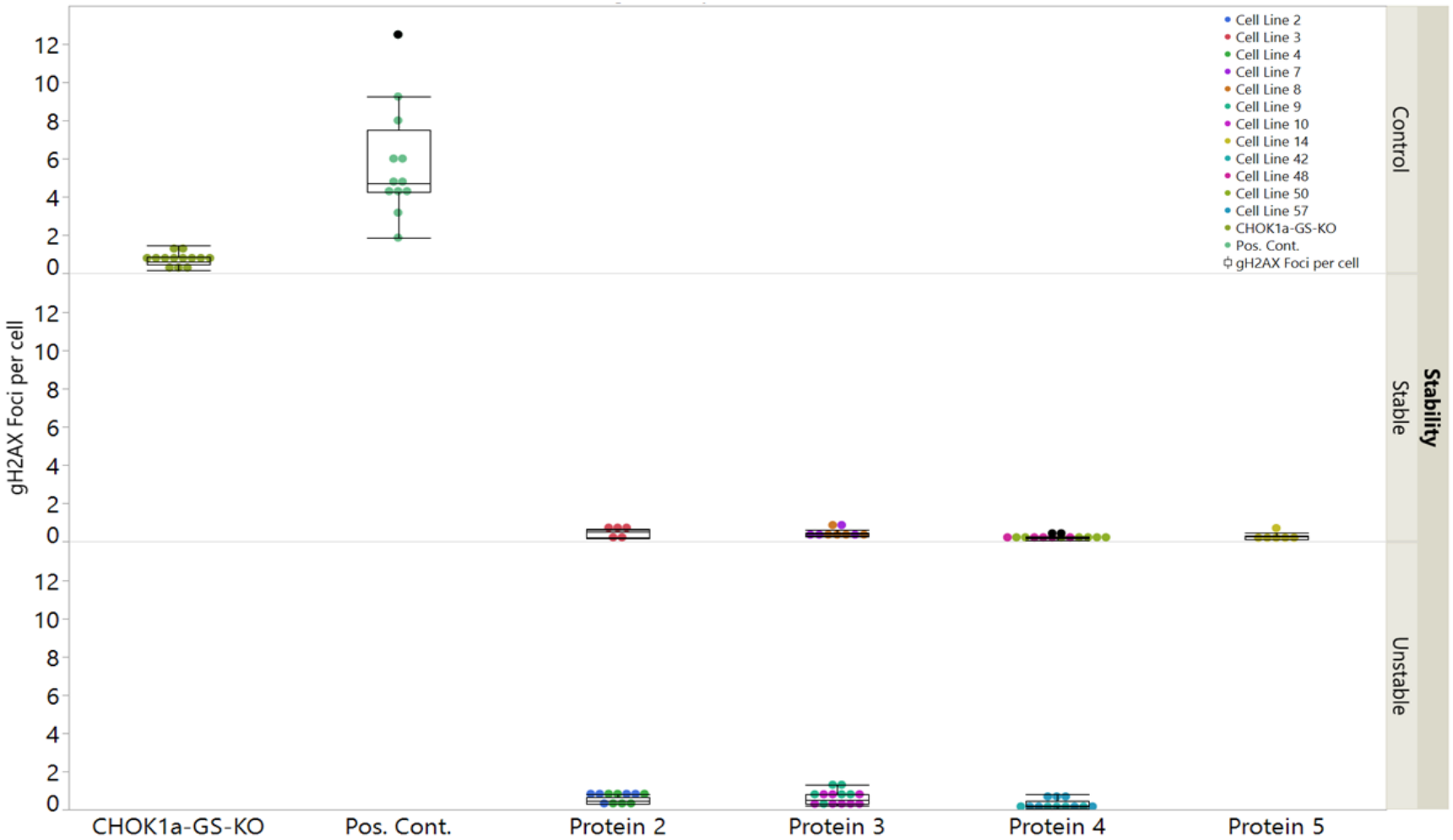


Figure 4.9. gH2AX foci per cell quantification of CHOK1a-GS-KO host, positive control and stable and unstable cell lines, across four different therapeutic proteins. There was no significant difference between producing cell lines when compared against the host (P=>0.05). gH2AX foci did not fluctuate between stable and unstable groups (P=>0.05).

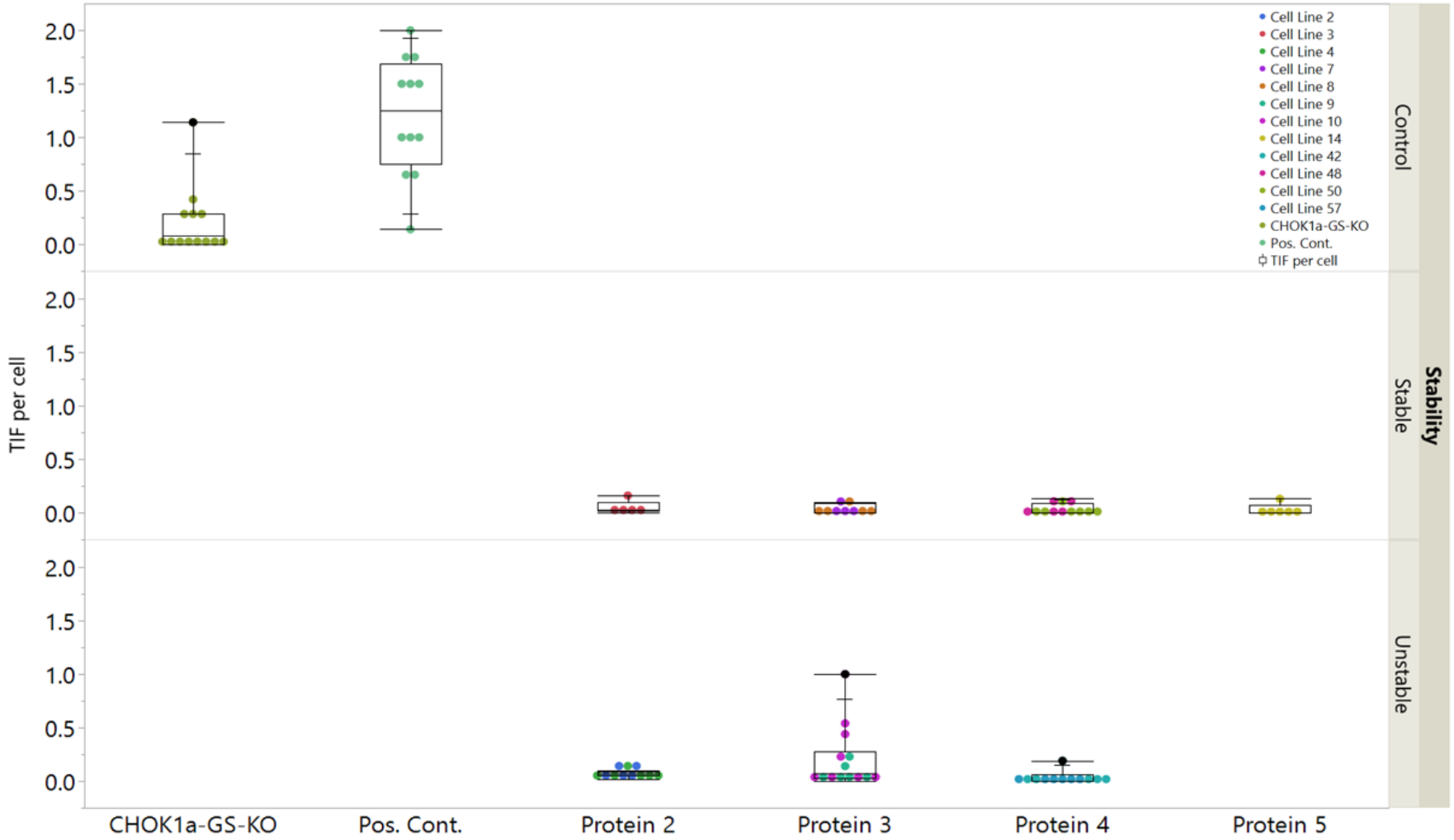


Figure 4.10. TIF quantification of CHOK1a-GS-KO, positive control and stable and unstable cell lines across four different therapeutic proteins. There was no significant difference between all groups analysed (P=>0.05).

Thus far, gH2AX has been quantified in ~100 cells per sample, to assess any preferential DNA damage at telomere sites, in addition to understanding whether there is elevated DNA damage within productionally unstable cell lines. To confirm the observation of low levels of gH2AX, present in producing cell lines, flow cytometry analysis was performed using a different gH2AX conjugated to a PE fluorophore. Flow cytometry allows for the analysis of a greater number of single cells, in this case 10,000 cells were analysed per sample, which may provide a better representation of the culture flask that contains millions of cells. Cells of stable and unstable cell lines, including controls for positive and negative gate setting were harvested, fixed, permeabilised and stained as previously described (methods, 2.5.1).

Cell gating strategy was first performed using side scatter (SSC-A, Y-axis) and forward scatter (FSC-A, X-axis) and applied to all samples (figure 4.11a, representative plots). gH2AX negative and positive gates were set using untreated CHOK1a-GS-KO cells (figure 11b, negative gate, red peak) and CHOK1a-GS-KO cells treated with 1ng/ml Neocarzinostatin for 1 hour (positive gate, blue peak), respectively. As the positive control has shifted to the right of the spectra, due to increases in gH2AX-PE binding, the flow cytometry experimental gates are adequate for gH2AX quantification on stable and unstable cell lines. Upon staining said cell lines, no difference in gH2AX was observed across all cell lines regardless of stability (table 4.6, figure 4.11c-e, one-way ANOVA, P=>0.05).

Flow cytometric analysis of gH2AX has provided an orthogonal confirmation of gH2AX low levels in producer cell lines during maintenance culture, as identified in gH2AX foci quantification using fluorescent images. This provides further evidence that CHOK1a-GS-KO based producer cell lines are either protected from DNA damage due to constituents in the highly optimised media or have acquired favourable mutations that allow for rapid DNA damage resolve.

Table 4.6. Raw value data for gH2AX and IgG positive and negative gate.

Protein	Cell line	Stability	Single Cell Count	gH2Ax Negative %	gH2AX Positive %	IgG Negative %	IgG Positive %	IgG Negative Mean	IgG Positive Mean
Protein 2	Cell Line 4	Unstable	9471	98.8	1.01	0.25	99.8	2360	25765
Protein 2	Cell Line 2	Unstable	9237	96.9	2.65	1.03	99.2	2572	26999
Protein 2	Cell Line 3	Stable	9379	98.8	0.92	0.91	99.4	2657	19169
Protein 3	Cell Line 7	Stable	9039	97.3	2.36	0.37	99.7	2535	19573
Protein 3	Cell Line 8	Stable	9370	98.7	1.1	0.011	100	412	94515
Protein 3	Cell Line 9	Unstable	9157	98.5	1.17	0	100	0	69770
Protein 3	Cell Line 10	Unstable	9473	97.7	1.95	22.8	81.7	2545	9690
Protein 5	Cell Line 14	Stable	9565	98.6	1.19	3.87	97.2	2646	13986
Protein 4	Cell Line 48	Stable	9638	98	1.55	0	100	0	25870
Protein 4	Cell Line 57	Unstable	9415	94.5	4.63	0.074	99.9	2536	37637
Protein 4	Cell Line 42	Unstable	9532	97.8	1.85	5.53	96.2	2684	11659
Protein 4	Cell Line 50	Stable	9529	98	1.66	0.17	99.8	1843	18413
gH2AX Positive Control	gH2AX Positive Control	gH2AX Positive Control	9512	0.78	98.4	100	0	562	10245
Negative Control	Negative Control	Negative Control	9382	97	2.71	100	0.011	203	10506

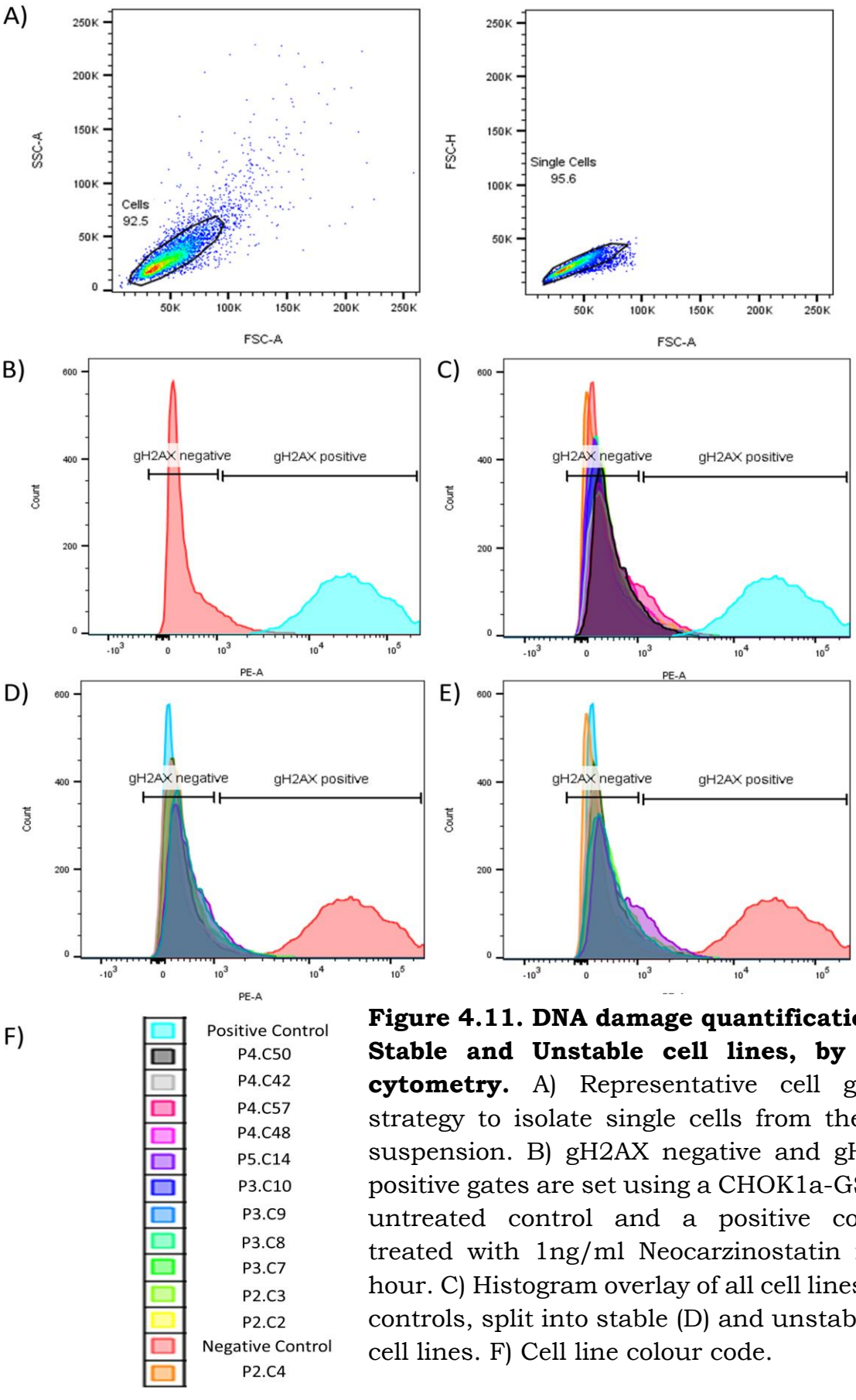


Figure 4.11. DNA damage quantification in Stable and Unstable cell lines, by flow cytometry. A) Representative cell gating strategy to isolate single cells from the cell suspension. B) gH2AX negative and gH2AX positive gates are set using a CHOK1a-GS-KO untreated control and a positive control treated with 1ng/ml Neocarzinostatin for 1 hour. C) Histogram overlay of all cell lines and controls, split into stable (D) and unstable (E) cell lines. F) Cell line colour code.

To understand the extent of telomere protection of ITS repeats within CHOK1a-GS-KO producing cell lines, a co-localisation staining of TRF2 (shelterin component) and telomere sequence probes was performed. Overlapping signals and single telomere signals were quantified in image J to assess the protection of telomeres across the stable and unstable panel of cell lines (figure 4.12, representative images of therapeutic proteins). To assess the significance of co-localisation and unprotected telomeres (single telomere foci) between each group, the data was analysed using a Two-way ANOVA approach using therapeutic protein class and stability groups as factors. Unadjusted P-values were generated using the least significant difference procedure (LSD) and adjusted P-values generated using the Hochberg's multiple comparison procedure (table 4.7).

Table 4.7. Adjusted P-values for TRF2 and telomere co-localisation quantified in the host and therapeutic protein producing cell lines.

	Comparison	Unadjusted p-value	Adjusted p-value
1	Protein 2 Stable vs. CHOK1a-GS-KO Control	5.43e-07	< 0.0001
2	Protein 2 Unstable vs. Protein 2 Stable	1.11e-05	0.0001
3	Protein 5 Stable vs. CHOK1a-GS-KO Control	4.94e-05	0.0004
4	Protein 3 Stable vs. CHOK1a-GS-KO Control	8.50e-05	0.0006
5	Protein 4 Stable vs. CHOK1a-GS-KO Control	0.001	0.0071
6	Protein 4 Unstable vs. CHOK1a-GS-KO Control	0.0157	0.0642
7	Protein 3 Unstable vs. CHOK1a-GS-KO Control	0.0160	0.0642
8	Protein 2 Unstable vs. CHOK1a-GS-KO Control	0.1077	0.2274
9	Protein 3 Unstable vs. Protein 3 Stable	0.1137	0.2274
10	Protein 4 Unstable vs. Protein 4 Stable	0.4635	0.4636

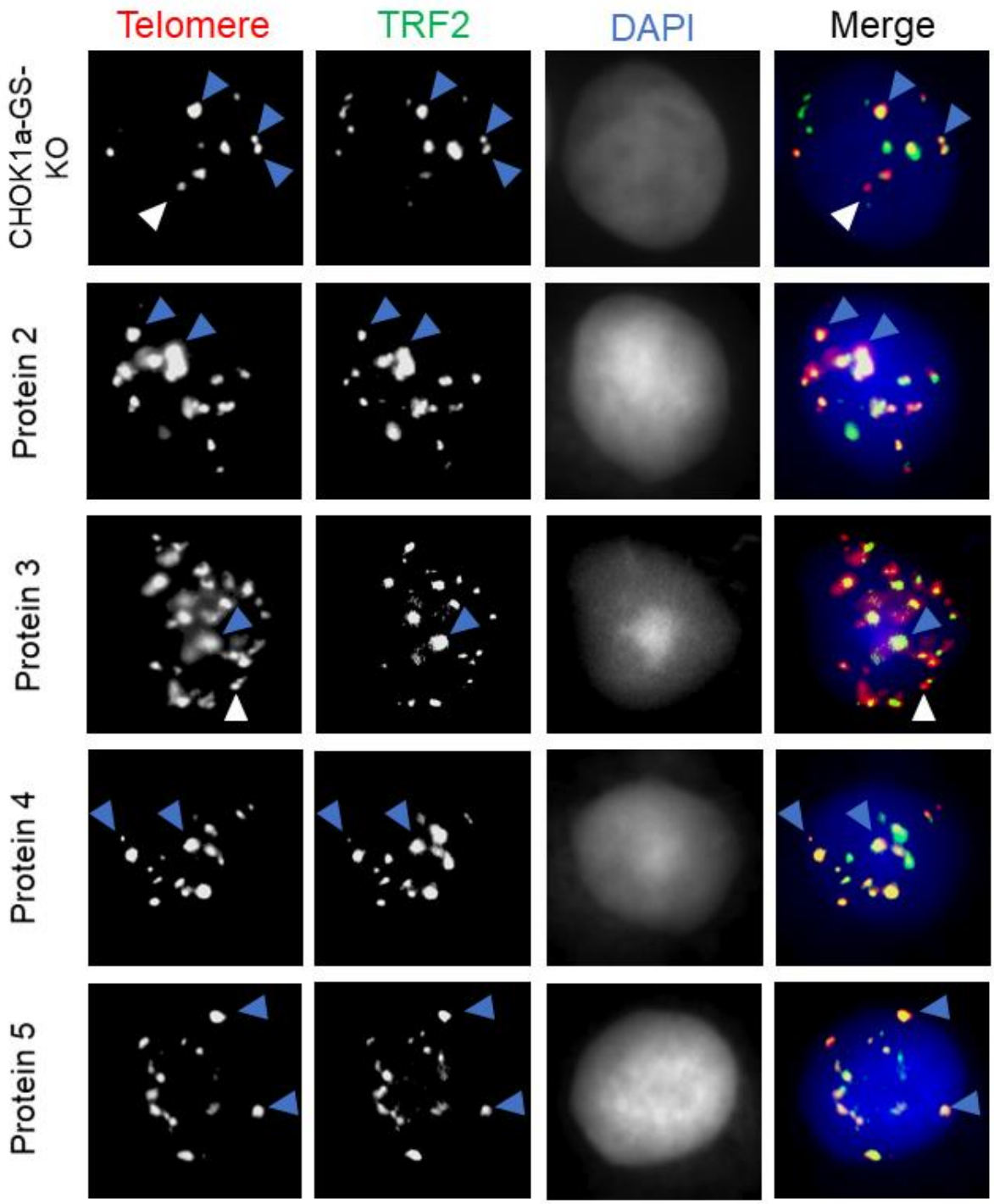


Figure 4.12. Representative images of cell lines used to quantify telomere protection by shelterin component TRF2. TRF2 (green) and telomere probes (red), to assess co-localisation of signals (blue arrows) and single telomere signals (white arrows). Data is used to compare

differences in telomere and TRF2 co-localisation between stable and unstable cell lines.

Quantification of telomere and TRF2 co-localisation across stability groups (figure 4.13) highlighted a statistical significance between therapeutic protein 2 stable and unstable cell lines (table 4.7, P=0.0001). However, this difference was not mirrored when comparing stable and unstable cell lines from therapeutic proteins 3 and 4. Protein 2, 3, 4 and 5 stable cell lines obtained an elevated level of TRF2 and telomere co-localisation, compared to the CHOK1a-GS-KO host (table 4.7, P=<0.01). Suggesting an elevated level of TRF2 and telomere co-localisation may represent a potential mechanism by which stable producing cell lines maintain their production stability. However, to corroborate this notion, one would expect there to be a significant difference when comparing TRF2 and telomere co-localisation between protein stability groups, but this was not observed except for therapeutic protein 2 cell lines.

If telomere protection (as signified by TRF2 and telomere co-localisation) impacted production stability, one should observe an increase in single telomere foci (unprotected telomeres) in unstable cell lines. To test this theory, single telomere foci was quantified in the same panel of cell lines and statistical analysis performed (figure 4.14, table 4.8). All adjusted group comparisons were statistically insignificant (P=>0.05) indicating there is not a concomitant increase in unprotected telomeres compared to CHOK1a-GS-KO host compared to protected telomere sequences that was observed initially (figure 4.13 and table 4.7).

Table 4.8. Adjusted P-values for single telomere foci quantified in CHOK1a-GS-KO and therapeutic protein producing cell lines.

	Comparison	Unadjusted p-value	Adjusted p-value
1	Protein 3 Unstable vs. Protein 3 Stable	0.009	0.0901
2	Protein 5 Stable vs. CHOK1a-GS-KO Control	0.042	0.3867
3	Protein 4 Unstable vs. CHOK1a-GS-KO Control	0.063	0.5084
4	Protein 4 Unstable vs. Protein 4 Stable	0.167	0.6346
5	Protein 3 Stable vs. CHOK1a-GS-KO Control	0.189	0.6346
6	Protein 2 Unstable vs. CHOK1a-GS-KO Control	0.255	0.6346
7	Protein 4 Stable vs. CHOK1a-GS-KO Control	0.392	0.6346
8	Protein 3 Unstable vs. CHOK1a-GS-KO Control	0.394	0.6346
9	Protein 2 Stable vs. CHOK1a-GS-KO Control	0.520	0.6346
10	Protein 2 Unstable vs. Protein 2 Stable	0.634	0.6346

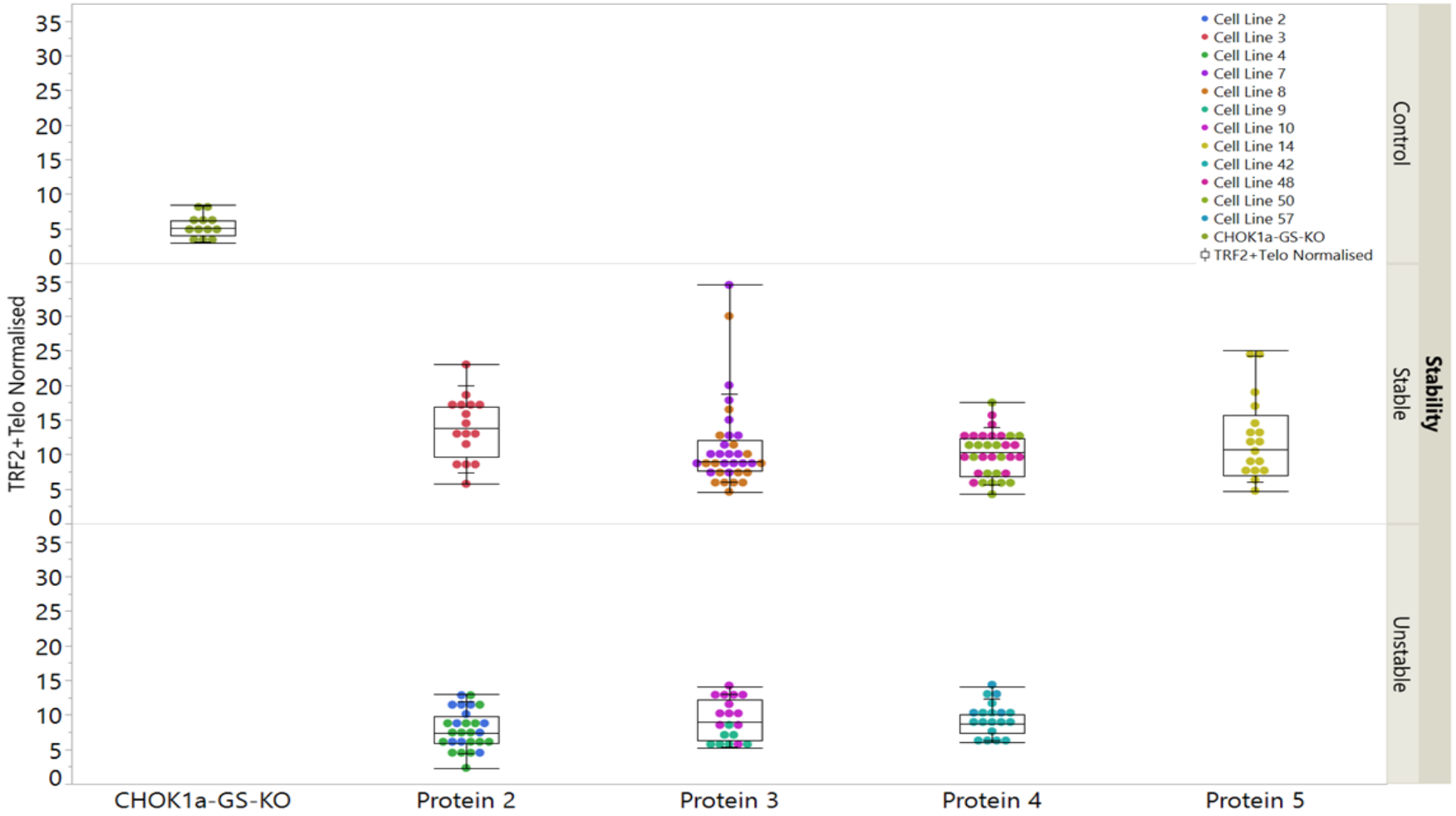


Figure 4.13. TRF2 and telomere co-localisation quantification in CHOK1a-GS-KO host and therapeutic protein producing cell lines.

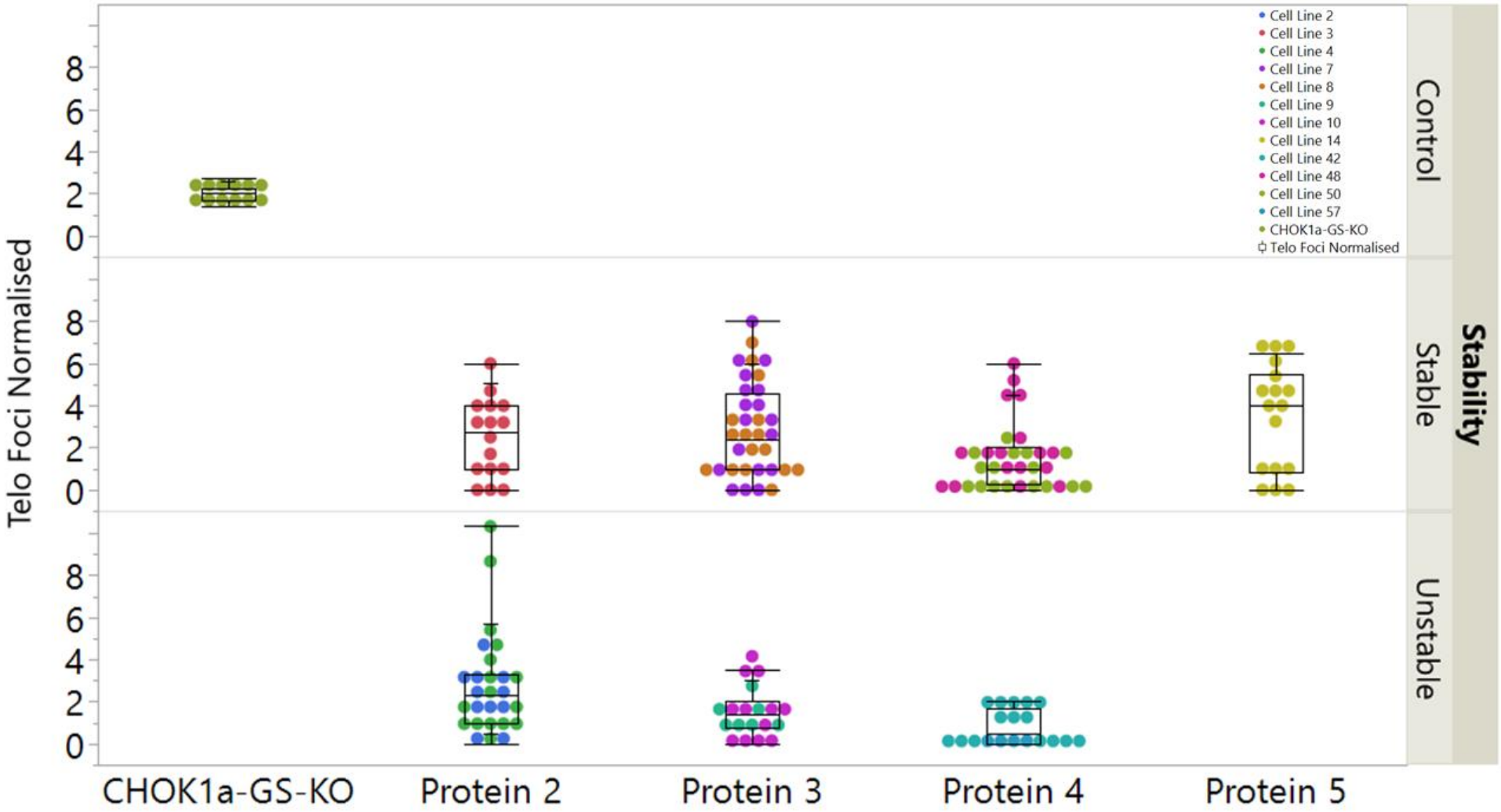


Figure 4.14. Single telomere foci quantified in CHOK1a-GS-KO host and therapeutic protein producing cell lines.

Quantitative reverse transcription polymerase chain reaction (qRT-PCR) screen of mRNA expression of genes involved in telomere homeostasis and DNA damage repair (4.5).

Aims and hypotheses (4.5.1)

To assess transcriptomic landscape surrounding telomere homeostasis and DNA damage pathways, an mRNA expression screen was designed to elucidate any potential differential expression patterns in stable and unstable therapeutic protein producing cell lines, across early and late time points. Any confirmed differentially expressed genes that are shown to be involved in the stability phenotype could reveal potential pathways for further characterisation, which could also be targeted or bioengineered in the host to potentially create an inherently stable host cell line.

Results (4.5.2)

A panel of 100 genes was designed to screen expression patterns in genes related to cell cycle, telomere homeostasis, DNA binding proteins and DNA damage repair. Differential expression patterns in these vital cellular functions may identify potential pathways that could be interrogated further for bioengineering the CHOK1a-GS-KO host to create a more robust platform. Primer pairs, when feasible, were designed to span exon-exon boundaries to ensure mRNA specific binding.

Primer pair efficiency was assessed by performing a serial dilution of sample cDNA (10ng/6µl stock, methods, 2.4.4) to assess the shift in cycle threshold across 10-fold dilutions. This dilution series tests the primer pair's amplification efficiency across a decreasing level of template. Selection criteria for a primer pair was an initial manual assessment of differences between the CTs between each dilution. CTs should differ around 3.3 cycles between each dilution sample. Amplification efficiency was assessed by plotting a linear curve and calculating the average slope of the dilution series, which is subsequently used to calculate an amplification efficiency percentage;

Amplification efficiency (%) = *((10^(-1/slope)-1) *100)*

Primers that obtained an amplification efficiency between 90-110% spanning a range of 10ng to 0.1ng of template, with a specific uniform melting temperature curve (Tm), were accepted into the gene panel (table 24).

As this screen is based on relative quantification against a house keeping gene, several house keeping gene (HKG) candidates were assessed across all samples, stabilities and timepoints to ensure minimal fluctuation of HKG expression in all conditions. This ensures minimal bias of the housekeeping gene when calculating fold changes of genes when normalising against the HKG. Beta-2-microglobulin (B2M), Glyceraldehyde 3-phosphate dehydrogenase (GAPDH) and Tubulin (TUB) were selected for testing HKG CT fluctuations (figure 4.15). B2M, GAPDH and TUB are traditionally used in gene expression analysis for other cellular models (e.g. mouse or human) and these genes have been tested elsewhere in CHOK1a cell lines and are shown to obtain constitutive expression (Bahr et al., 2009). Across all samples B2M, GAPDH and TUB all obtained <3.5% coefficient of variation (CV). B2M obtained the smallest variance of expression (2.01% CV) and was therefore selected as the HKG of choice for the duration of gene screen.

Components of the shelterin complex (RAP1a, TRF1, TRF2, TPP1, TIN2 and POT1) were assessed for their expression levels across a 6-month stability period in the CHOK1a-GS-KO host to identify baseline expression fluctuations of shelterin components. Obtained CT values were normalised against B2M housekeeping gene and time point 0, to allow for comparisons of mRNA expression across all genes and time points. Normalisation was performed using the following equation (Schmittgen and Livak, 2008);

$$2^{-\Delta\ \Delta C_T} = [(C_T \text{ gene of interest} - C_T \text{ internal control}) \text{ sample A} - (C_T \text{ gene of interest} - C_T \text{ internal control}) \text{ sample B})]$$

This form of the equation is used to compare two different samples in relation to the selected house keeping gene and can be applied to assess expression fold changes of a gene of interest directly between a stable and unstable cell line across different time points by keeping a consistent normalising sample. For each therapeutic protein set of cell lines, a consistent stable cell line at time point 0 was selected to normalise all other samples against.

mRNA fold expression differences were only considered as significant if fold changes were + or – 2-fold from the normalising samples (set at 1). Fold changes within this range are considered as assay and/or biological variation. Genes that met the primer efficiency criteria were tested in a panel of stable and unstable cell lines across early and late time points that span the 4-6-month stability assessment. Across all genes analysed there was only one gene that showed a similar trend in mRNA expression across all samples analysed. All shelterin components obtained similar levels of expression across a 6-month maintenance culture period (figure 4.16).

Reflecting TRF2 and telomere co-localisation results, that showed no significant difference in co-localisation across early and late time points within the CHOK1a-GS-KO host.

TRF1-interacting nuclear factor 2 (Tin2), a component of the shelterin complex, was shown to have >2-fold expression in 7/9 unstable cell lines, compared to 2/9 in the productionally stable group (figure 4.17). Interestingly, this trend reverses after the 6-month culture period as late stable cell lines show an elevated Tin2 mRNA expression in 6/9 lines compared to 3/9 late unstable lines.

To investigate whether trends observed at the transcript level was reflected proteomically, Tin2 protein from stable vs unstable samples was quantified using surface plasmon resonance (SPR, Biacore 8k, figure 4.18). SPR is a commonly used method in the pharmaceutical industry to assess levels of harvested therapeutic protein and host cell protein contaminants, from mini-bioreactor runs. Biacore 8k is a novel SPR machine that allows for rapid and automated quantification of multiple samples from a 96 well plate format. SPR utilises a capture surface, in this case bound anti-Tin2 antibody, associated with a gold film to detect light refraction as a means of protein quantification.

A light source is beamed through a prism and dependent on the amount of protein bound, light is refracted from the gold film surface and this change in refraction is captured on a detector. Light path refraction is proportional to the amount of protein bound to the immobilised anti-Tin2 antibodies, providing a rapid means for Tin2 specific protein quantification from sample cell lysates.

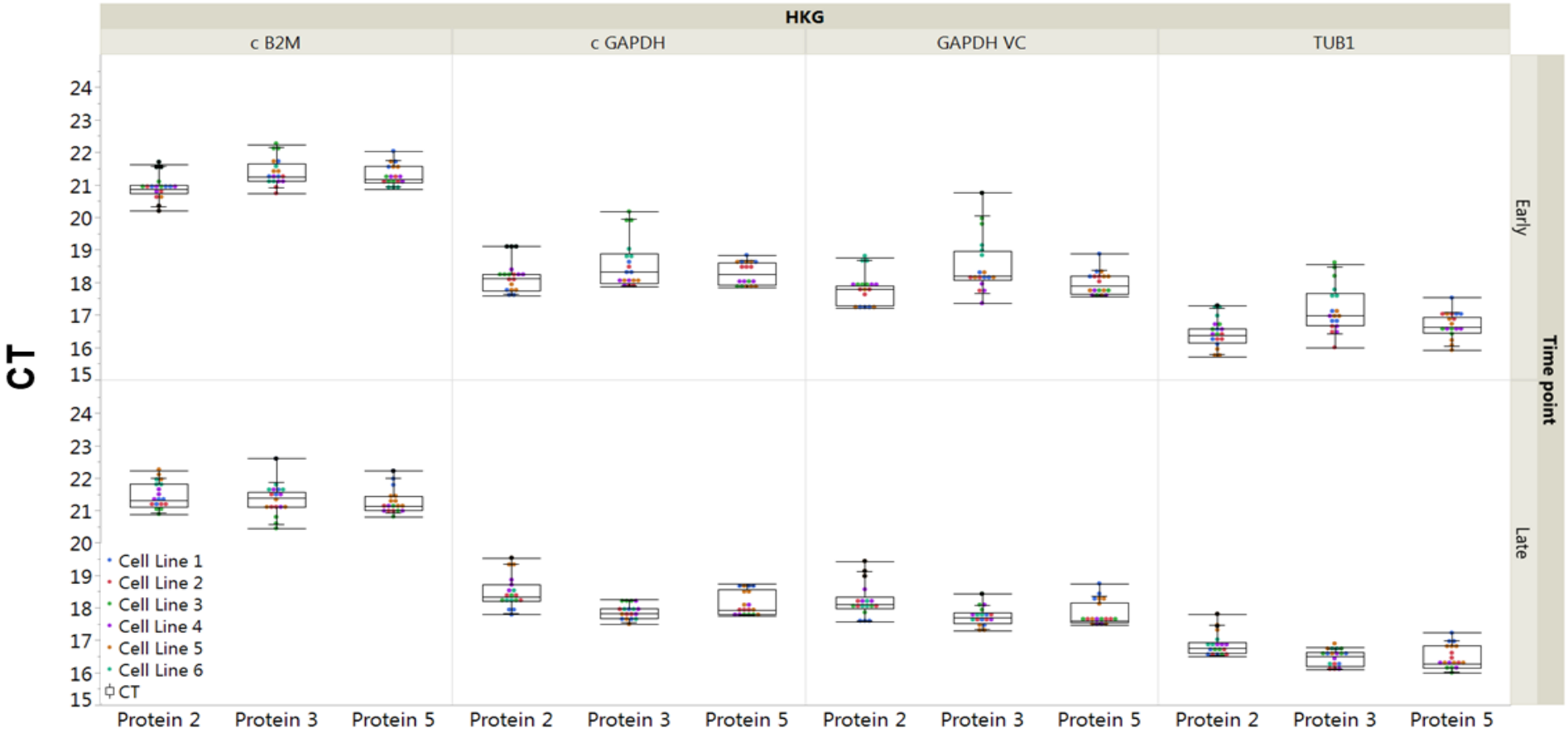


Figure 4.15. Housekeeping gene optimisation experiment to assess the fluctuation of mRNA expression across all therapeutic proteins, cell lines and time points. *Overall %CV are as follows; cB2M = 2.01%, cGAPDH = 2.87%, GAPDH VC = 3.24% and TUB1 = 3.07%. cB2M was selected as the housekeeping gene of choice for the duration of the gene screen* as it obtained the lowest CV% across all samples analysed and was chosen to use as the normaliser for the genetic screen.

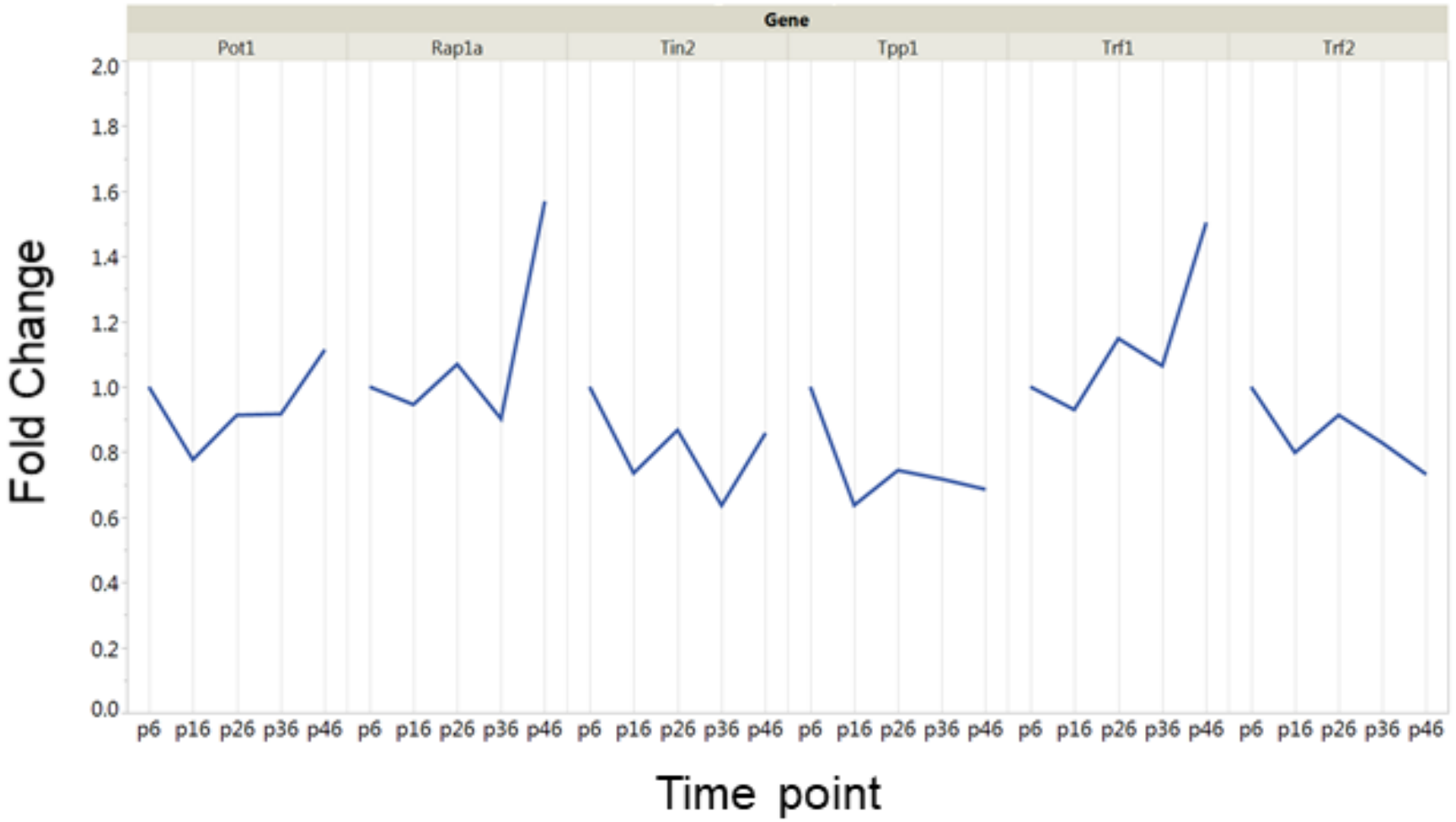


Figure 4.16. Shelterin component expression across a 6-month maintenance culture period. Changes in expression of + or – 2-fold is considered as a significant change in expression. Therefore, shelterin expression was not considered to significantly fluctuate over 6-months of continuous culture.

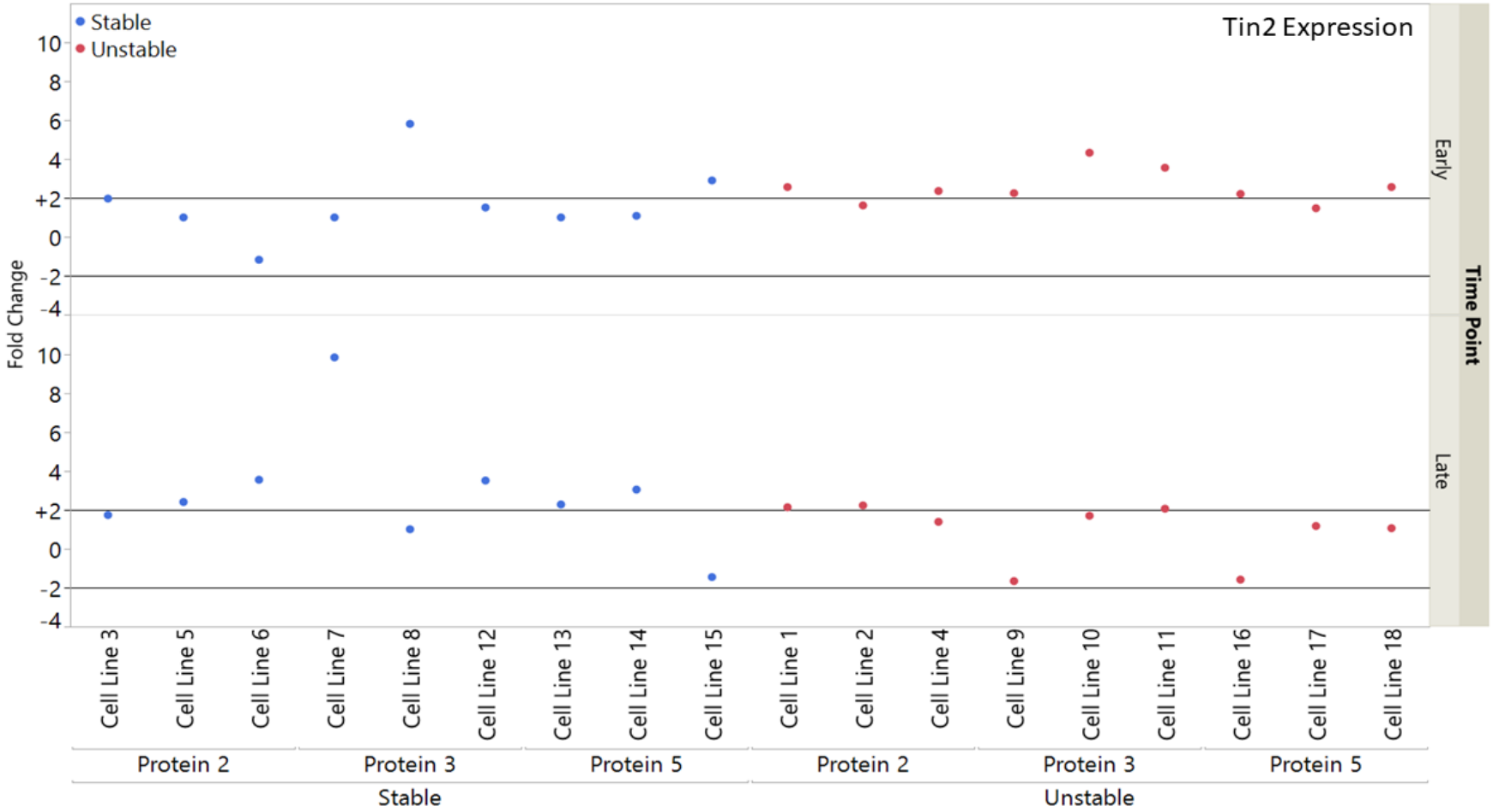


Figure 4.17. Tin2 expression fold change across therapeutic protein 2, 3 and 5 stable and unstable cell lines and early and late time points. Tin2 expression is elevated in early unstable cell lines in 7/9 lines compared to 2/9 early stable lines. Tin2 expression is also elevated in 6/9 late stable cell lines compared to 3/9 late unstable cell lines.

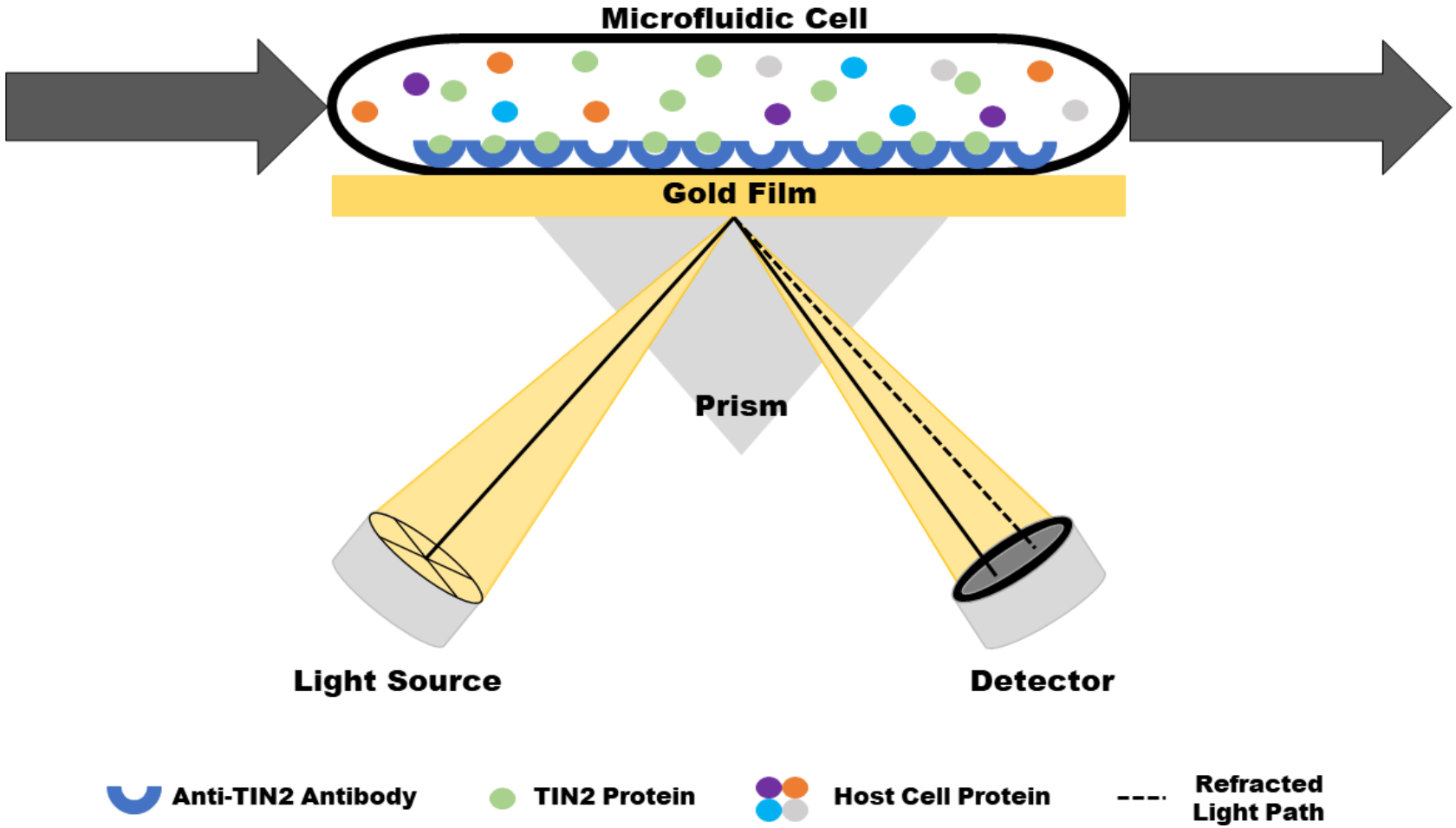


Figure 4.18. Schematic of Surface plasmon resonance protein quantification method. Whole cell lysate is injected into microfluidic channels with immobilised anti-TIN2 antibody that specifically binds to Tin2 protein. A light source is beamed through a prism and reflects from the gold surface and the refracted light path is collected by a downstream detector. The extent of the light reflection reflects the amount of Tin2 protein bound to the surface of the chip.

To ensure there is specific Tin2 binding from CHOK1a-GS-KO whole cell lysates, an optimisation experiment was performed (figure 4.19). CHOK1a-GS-KO host whole cell lysate was extracted from fresh cells and protein quantified using the Bradford reagent assay. A Tin2 overexpression lysate was used to produce a calibration curve at 1, 5 and 10 µg/ml, allowing back calculation of bound Tin2 from whole cell lysates (figure 4.19a). SPR curves of the positive control calibration curve (figure 4.19b) represent the refraction of the light path during sample injection start (blue arrow), injection stop (orange arrow), chip surface wash (black arrow) and surface regeneration (green arrow).

SPR binding curves increased in relation to Tin2 overexpression lysate concentration - 1 (green line), 5 (pink line) and 10 (gold line) µg/ml, show a concentration specific response. CHOK1a-GS-KO whole cell lysate was injected onto the anti-Tin2 immobilised chip at neat, 1 in 5, 1 in 50 and 1 in 500 dilutions (table 2.10). At neat, 1 in 5 and 1 in 50 dilutions, RIPA buffer interference was still present, as reflected by the increase in CV% between replicates. A dilution of 1 in 500 represented the dilution that obtained the lowest CV% between replicates (table 4.9).

After optimising conditions for Tin2 quantification, whole cell lysates were extracted from stable vs unstable panel of cell lines and Tin2 relative protein concentration quantified using a Tin2 overexpression lysate sample to produce a calibration curve. Relative Tin2 protein levels (figure 4.20), normalised to total protein content, did not share the same pattern as observed at the transcriptomic level (figure 4.17), indicating the initial observation was stochastic in nature.

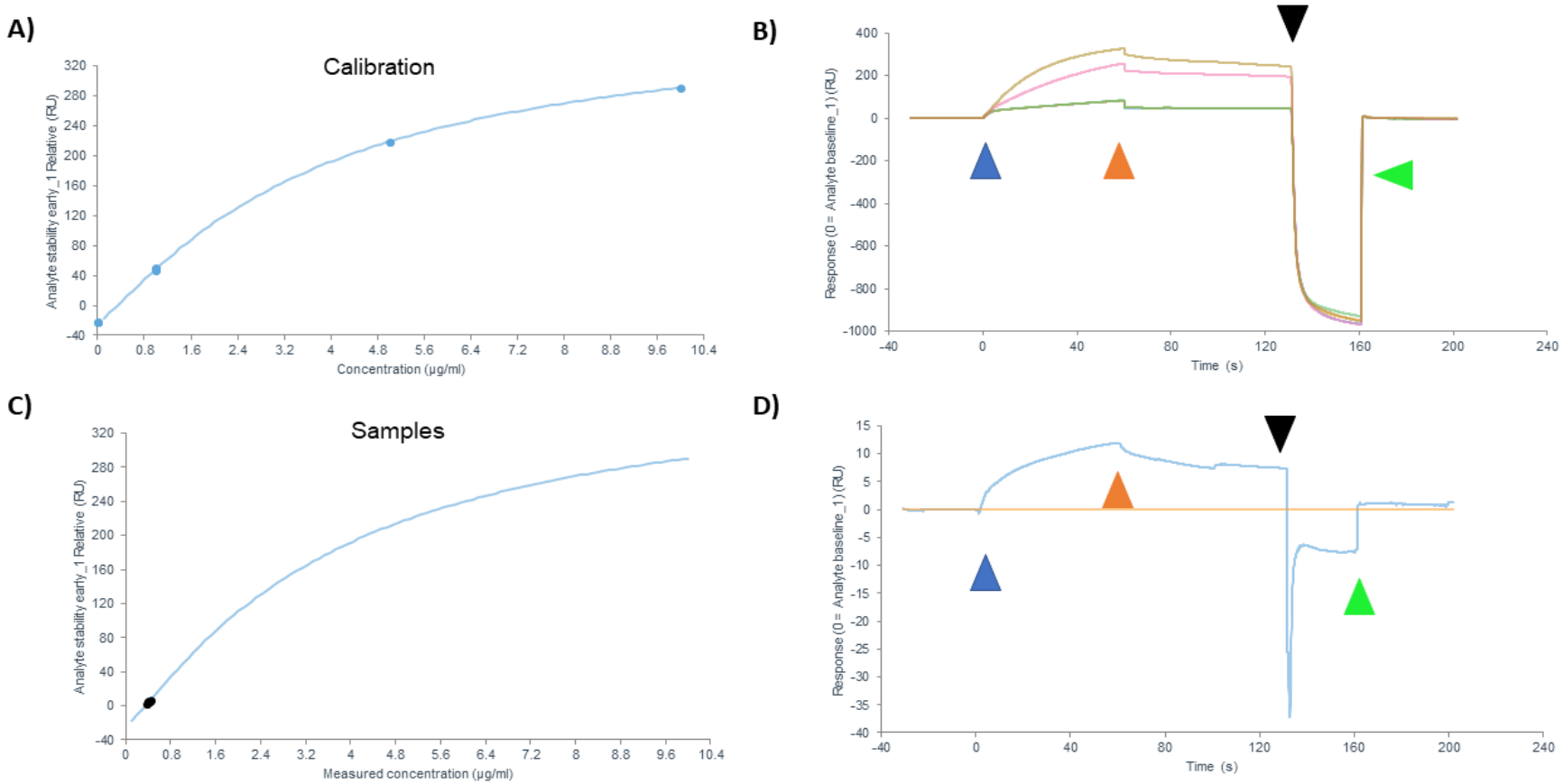


Figure 4.19. Tin2 protein capture optimisation and quantification. A) Calibration curve of Tin2 overexpression lysate at 1, 5 and 10 µg/ml Tin2 protein. B) SPR curves showing injection start (blue arrow), injection stop (orange arrow), chip surface wash (black arrow) and surface regeneration (green arrow) of Tin2 overexpression lysate calibration samples at 1 (green line), 5 (pink line) and 10 (gold line) µg/ml. C) Sample relative units plotted on calibration curve graph. D) SPR curves showing binding, washing and surface regeneration events. Orange line represents the normalisation of the sample to RIPA buffer control sample. RIPA buffer was shown to interfere with SPR curves, therefore RIPA buffer samples are used to normalise CHO whole cell lysate samples.

Table 4.9. Optimisation experiment testing Biacore 8k Tin2 protein binding efficiency. CHOK1a-GS-KO whole cell lysate and Tin2 overexpression lysate was used to test Tin2 capturing ability using a Biacore 8k chip with immobilised anti-Tin2 antibody. Samples were diluted at neat, 1 in 5, 1 in 50 and 1 in 500 dilutions in order to counter RIPA buffer interference.

Cycle	Channel	Sensorgram type	Analysis step purpose	Analysis step name	Calib. curve	Analyte 1 Solution	Response (RU)	Analyte 1 Concentration (µg/ml)	Calculated conc. (µg/ml)	Avg. calc. conc. (µg/ml)	CV (%)	Calc. conc. vs expected (%)
4	1	Reference subtracted	Calibration	Calibration	1	Calib	46.54	1	0.972	1.00	2.44	97.2
5	1	Reference subtracted	Analysis	Samples	1	Sample 1:5	1.727		0.368	0.391	8.25	
4	2	Reference subtracted	Calibration	Calibration	1	Calib	49.39	1	1.01	1.00	2.44	101
5	2	Reference subtracted	Analysis	Samples	1	Sample 1:5	5.159		0.413	0.391	8.25	
4	3	Reference subtracted	Calibration	Calibration	1	Calib	49.64	1	1.02	1.00	2.44	102
5	3	Reference subtracted	Analysis	Samples	1	Sample 1:50	6.632		0.433	0.422	3.79	
4	4	Reference subtracted	Calibration	Calibration	1	Calib	218.6	5	5.00	5.00		100
5	4	Reference subtracted	Analysis	Samples	1	Sample 1:50	4.927		0.410	0.422	3.79	
4	5	Reference subtracted	Calibration	Calibration	1	Calib	290.3	10	10.0	10.0		100.0
5	5	Reference subtracted	Analysis	Samples	1	Sample 1:500	4.383		0.403	0.404	0.47 5	
4	6	Reference subtracted	Calibration	Calibration	1	Calib	-22.86	0				
5	6	Reference subtracted	Analysis	Samples	1	Sample 1:500	4.587		0.406	0.404	0.47 5	

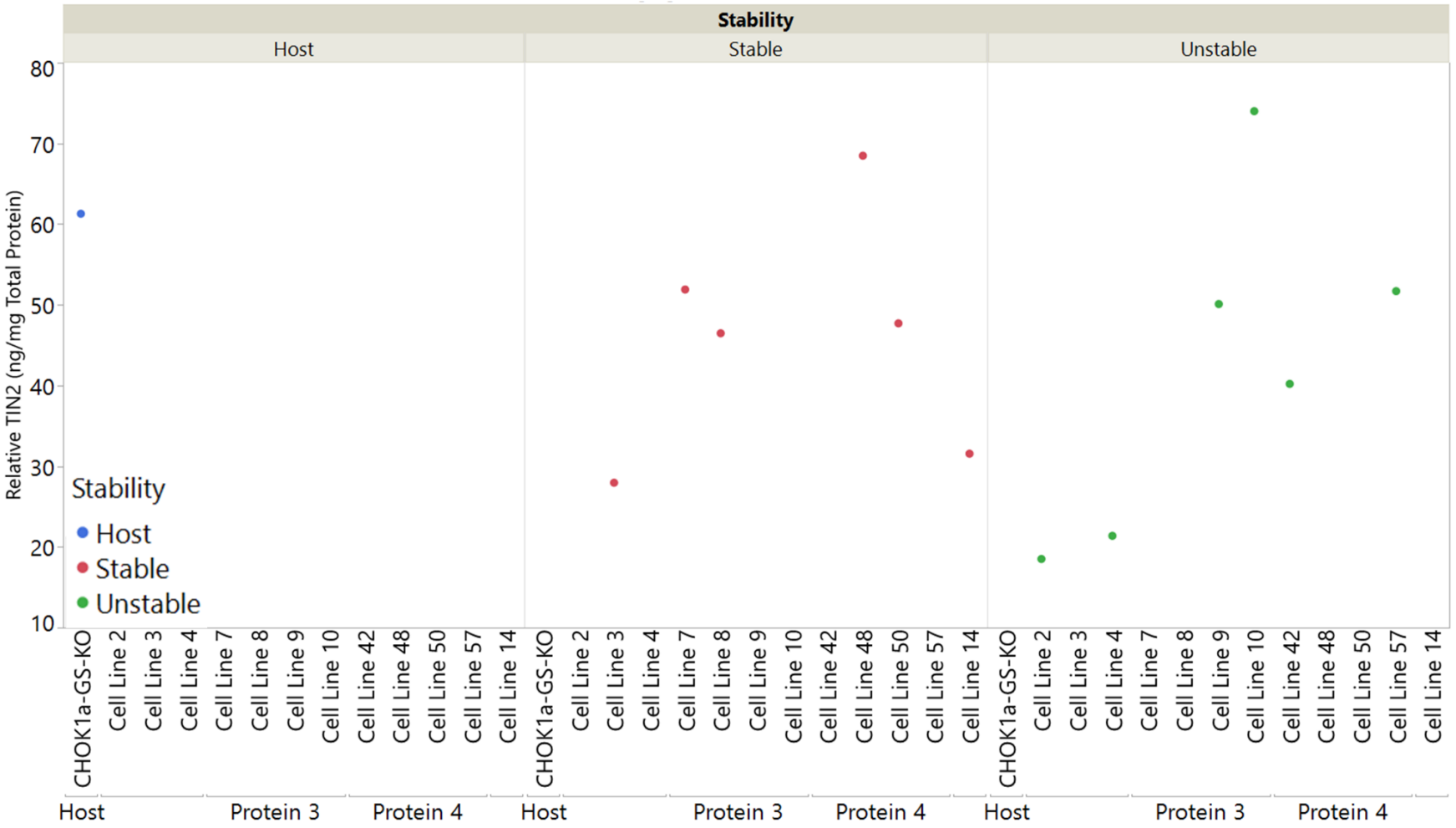


Figure 4.20. SPR calculated relative Tin2 concentrations (ng/mg total protein) of CHOK1a-GS-KO host with stable and unstable producer cell lines. No significant difference was observed between stable and unstable groups (pooled T-test, P=>0.05).

Chapter 5: Phenotypic assessment of production (in)stability. Investigating telomere and DNA damage role in karyotypic heterogeneity and its impact on therapeutic protein production.

Introduction (5.1)

Thus far, CHOK1a-GS-KO host and a panel of stable and unstable therapeutic protein producing cell lines have been characterised to investigate a potential mechanism for the identified karyotypic heterogeneity and its action as a potential key driver for production instability. Karyotypic heterogeneity has been identified through implementing chromosome number distribution analyses and multicolour FISH ("chromosome painting", MFISH), to identify mutational differences between distinct populations within cell cultures, during routine maintenance. Increases in genetically unstable (NCCA) populations has been identified to correlate with production instability at early (~20 generations) and late time points, across a 4-6-month stability assessment window.

Productionally unstable cell lines that obtained greater karyotypic heterogeneity and genetically unstable populations (NCCA) at early time points, gained a greater amount of mutations over prolonged periods of maintenance culture. The increase of diverse populations may represent a potential causative mechanism for the production instability phenotype, due to potential divergent cellular populations. Increases in genetically unstable populations may produce different amounts of therapeutic protein (IgG titre), leading to the fluctuations in peak titre observed during the production stability assessment.

Co-localisation assays, investigating the levels of overall DNA damage (gH2AX foci) and telomere specific damage (TIF) in CHOK1a-GS-KO host and therapeutic protein producing cell lines obtained inconclusive results regarding their impact on production instability. Overall, DNA damage and TIFs were unchanged across production stability groups and the host, suggesting CHO cells during maintenance culture have either acquired the ability to rapidly proliferate whilst maintaining low levels of DNA damage, or the highly optimised media provides a DNA damage protective environment for the cells.

To investigate the extent of telomere protection, co-localisation of TRF2 and telomere signals was assessed. TRF2 is a key component of the Shelterin complex, a 6 membered protein that confers telomeric protection from DNA damage, and its presence on CHOK1a telomeres identifies the potential for shelterin complex assembly within CHOK1a cell lines (Bianchi et al., 1997; Broccoli et al., 1997; de Lange, 2005). Seven stable cell lines across four different therapeutic proteins were found to have elevated TRF2 and telomere co-localisation foci per cell compared to the CHOK1a-GS-KO host.

This may indicate a potential mechanism by which productionally stable cell lines maintain a more homogenous karyotype during therapeutic protein production. However, with an increase in telomere and TRF2 co-localisation, one would expect a concomitant increase in single telomere foci within unstable cell lines, but this was not observed.

As overall DNA damage (highlighted by gH2AX foci quantification) and telomere protection (as highlighted by co-localisation of TRF2 and telomere foci) was deemed inconclusive during routine maintenance, within this chapter I aimed to explore how overall DNA damage and prolonged telomere specific damage might affect therapeutic protein production during a 4-6-month stability assessment. Inducing DNA damage within a production run environment may provide the means to elucidate a causative pathway that links the karyotypic heterogeneity witnessed here and CHOK1a-GS-KO based production instability.

Investigation into the impact of DNA damage within a therapeutic protein production run environment (5.2).

Aims and hypotheses (5.2.1)

Karyotypic heterogeneity of productionally stable and unstable cell lines has thus far been assessed during routine maintenance culture. To understand how CHOK1a-GS-KO based cell line heterogeneity fluctuates within a production environment, which is optimised to promote increased production of therapeutic protein, experiments were designed to assess genomic instability during normal production run conditions, under telomere specific damage stress and in the presence of a DNA damaging agent. Increased karyotypic heterogeneity during the production run may lead to fluctuations in peak titre due to diverse populations producing differing levels of therapeutic protein. If this holds true, it may elucidate a potential pathway that is causative of large fluctuations in peak titre, during multiple production runs performed during a stability assessment.

Results (5.2.2)

To investigate overall DNA damage and telomere specific damage (TIF) role in karyotype heterogeneity a mock stability assessment was designed to induce chronic telomere specific damage utilising Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR)-Cas9 technology, using a telomere guide RNA (gRNA, Telo-Cas9) to elicit continuous telomere specific double strand breaks, as previously described (Mao et al., 2016). Using constitutive expression of Telo-Cas9 within cell lines that produce therapeutic proteins, it provides the means to explore how chronic telomere specific damage may impact karyotypic heterogeneity and therapeutic protein production within the stability assessment environment.

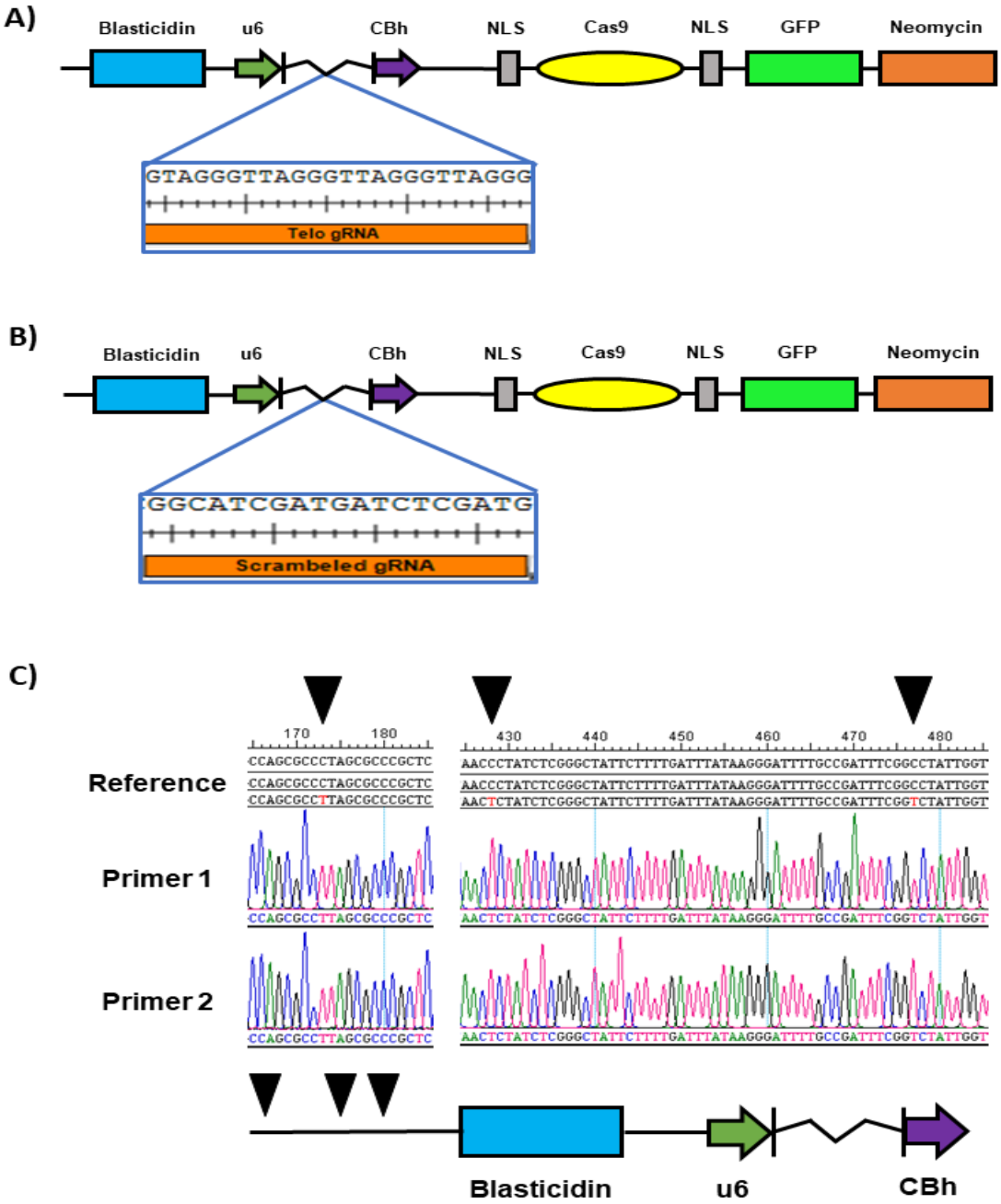


Figure 5.1. Telomere specific DNA damage Cas9 plasmid construction. A and B) Plasmid construct components of modified PX458 (Ran et al., 2013b). A neomycin selection marker has been cloned into the original plasmid with telomere gRNA (A), targeting telomeric sequences, and a scrambled gRNA (B) control plasmid. C) Full plasmid sequencing revealed three-point mutations located upstream from the Blasticidin resistance gene cassette. As the point mutations were located in a non-critical part of the plasmid construct, plasmids were used for transfection.

Telo-Cas9 plasmid was constructed using pSpCas9(BB)-2A-GFP (PX458) (Ran et al., 2013b) that obtains a single gRNA insertion site directly linked to a chimeric gRNA scaffold and a Cas9 protein flanked by two nuclear localisation signal linked to an enhanced GFP by 2A peptide from Thosea asigna virus (figure 5.1a). PX458 was transformed into bacteria, amplified and sequence confirmed before modifications of the plasmid was performed (methods, 2.2.4).

Primers were designed to produce a DNA fragment containing full neomycin resistance gene cassette, from a therapeutic protein expression plasmid, and homology arms that allow for homologous recombination into the linearised PX458 plasmid (figure 1 a and b, in fusion cloning methods, 2.2.9). After cloning the neomycin cassette, PX458-Neomycin plasmid was sequence confirmed and amplified. Telomere gRNA (Mao et al., 2016) and scrambled gRNA was designed to contain complementary sequence overhangs, created when linearising PX458-Neomycin using Bbsi restriction enzyme (Ran et al., 2013b).

Scrambled gRNA sequence was designed and blasted against the CHOK1 genome to assess any off-target hits. Upon confirmation of no sequence complementarity with CHOK1 genome, scrambled gRNA was cloned into PX458-Neomycin, as previously described. Scrambled gRNA PX458-neomycin plasmid was used as a control to assess if phenotypic effects witnessed, during expression of Telo-Cas9 plasmid in a production run environment, can be attributed to the telomere gRNA itself or if observations are due to the transfection and constitutive production of the Cas9 protein.

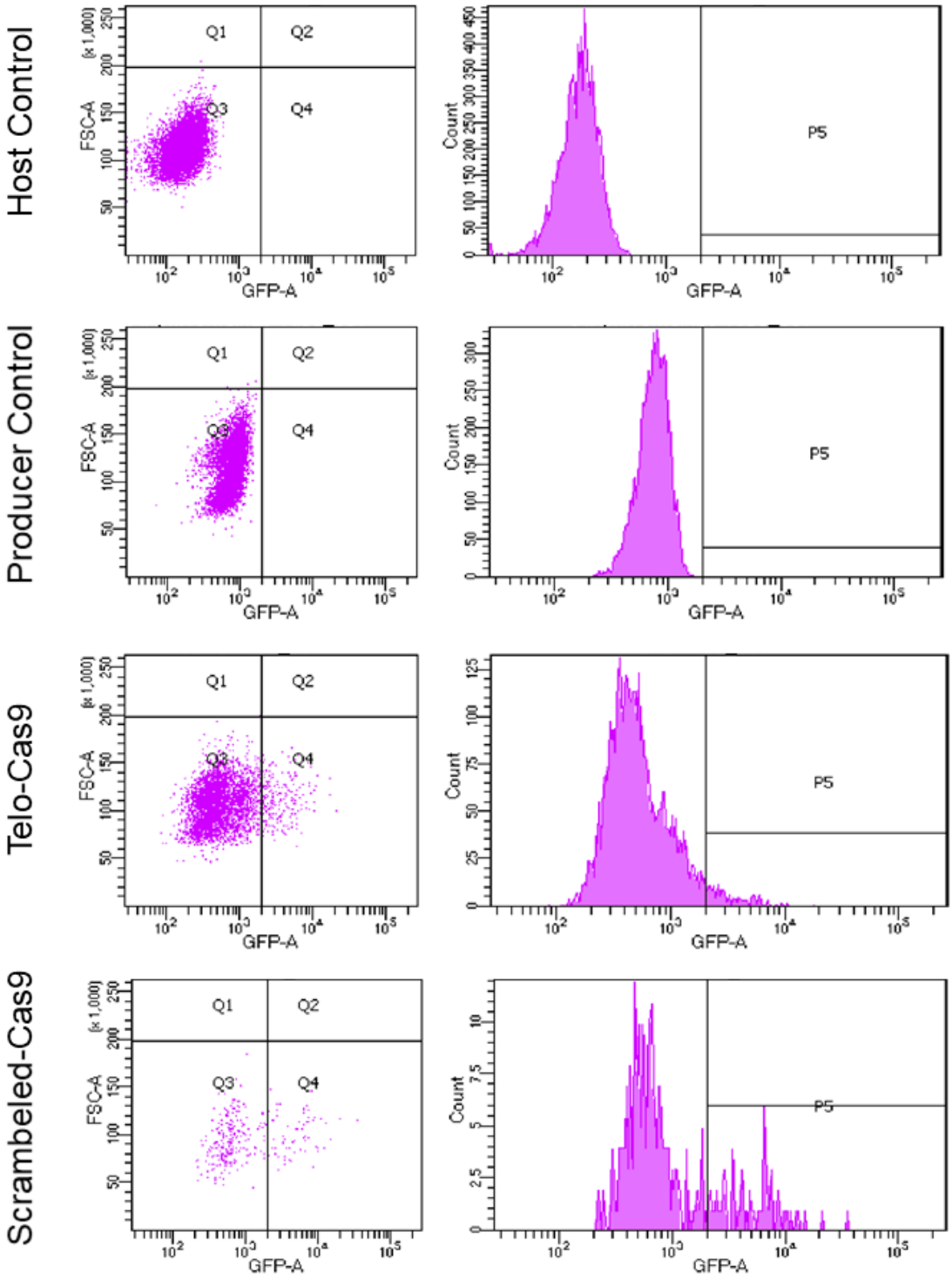


Figure 5.2. Flow cytometry dot plots and histograms with associated GFP positive (Q4 and P5) and negative (Q3) gates. CHOK1a-GS-KO host and a therapeutic protein producer cell line were used as negative controls for GFP expression. Interestingly there seems to be a shift in GFP expression when analysing cell lines that produce therapeutic proteins, hence Q4 and P5 gates were set against the producing cell line control.

10µg of Telo-Cas9 and Scrambled-Cas9 plasmid was individually transfected into a cell line that was previously confirmed as productionally stable, using Amaxa 4D nucleofector (methods, 2.1.6). Transfected cells were incubated in static plates at 37°C and 5% CO_2 for two weeks. Viability was assessed, and cell lines scaled up consecutively (~3-4 weeks) to 125ml shake flasks. Cell lines were analysed for GFP expression, using CHOK1a-GS-KO host and a producer cell line as GFP gating controls. As Cas9 is linked to GFP by a 2A peptide that cleaves after translation (Donnelly et al., 2001), the amount of GFP observed during flow cytometry analysis is proportional to Cas9 production. Interestingly, the therapeutic protein producing cell line obtained a positive shift of GFP expression compared to the host (figure 5.2), indicating mAb production within cells leads to some autofluorescence at the 488nm wave length.

Positive (Q4 and P5) and negative (Q3) gates were set based on the producer cell line control. Initially single GFP positive cells that resided within the P5 gate were sorted per well in a 96 well plate in conditioned media (media from 3-day old cell cultures, methods, 2.1.7). Outgrowth was monitored over a 14-day period; however, no colonies were present in the 96 well plates after three weeks post transfection. Subsequently, 5,000 cells were sorted per well into 24 well plates, containing conditioned media. After two weeks, cells were pooled into T25s corresponding to their row location (A, B, C and D) and expanded consecutively into 125ml shake flask in duplicates (A1, A2, B1, B2, etc).

Upon cell viability reaching >98%, a production run experiment was performed to assess the effect of telomere specific DNA damage in a production environment. After completing the 14-day protocol (methods, 2.1.9), cell lines were re-analysed for their GFP (Cas9) expression (figure 5.3 and 5.4). GFP analysis was performed using the same voltage settings and gates of the previous fluorescent-activated cell sorting (FACS).

A producer cell line control was utilised to ensure the negative control remained in the previously determined negative gate. After 35 days of maintenance culture cell lines expressing Telo-Cas9 obtained a bimodal distribution of GFP fluorescence peaks. Indicating that there are two cell populations with differing GFP expression (figure 5.3). Scrambled-Cas9 cell lines retained a uniform GFP expression peak, however this was slightly shifted to the left, indicating GFP loss over 35 days of routine culture (figure 5.4).

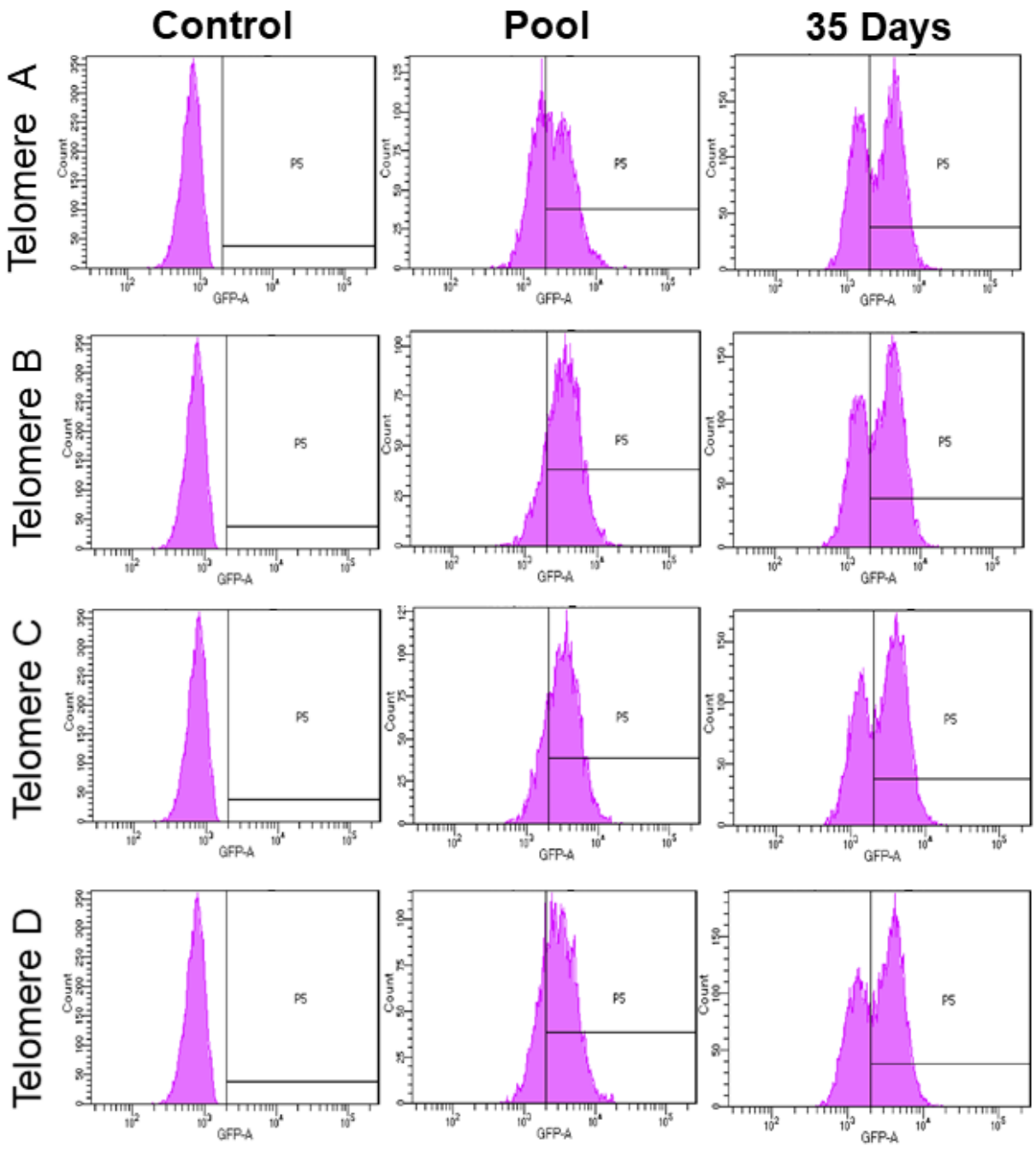


Figure 5.3. GFP expression shift comparison between recovered pools derived from sorting of Telo-Cas9 transfection bulks and cell lines after 35 days in routine maintenance culture. After 35 days, there appears to be a bimodal distribution of GFP expression, indicating a loss of GFP expression over the culturing period.

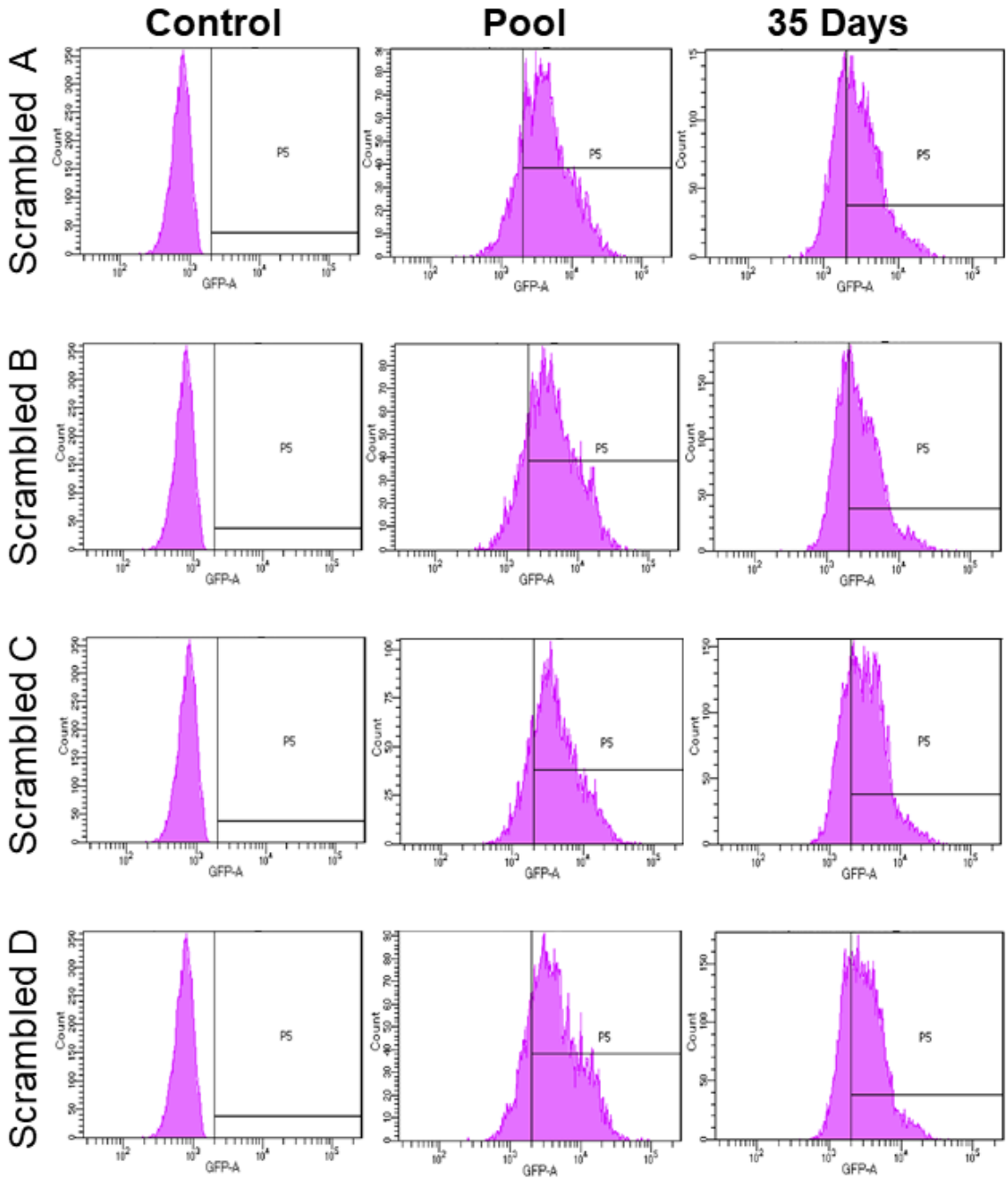


Figure 5.4. GFP expression shift comparison between Scrambled-Cas9 pools and 35 days of culture in a maintenance environment. Although there is a uniform peak, it appears that the peak has been shifted to the left, indicating a slight loss of GFP expression.

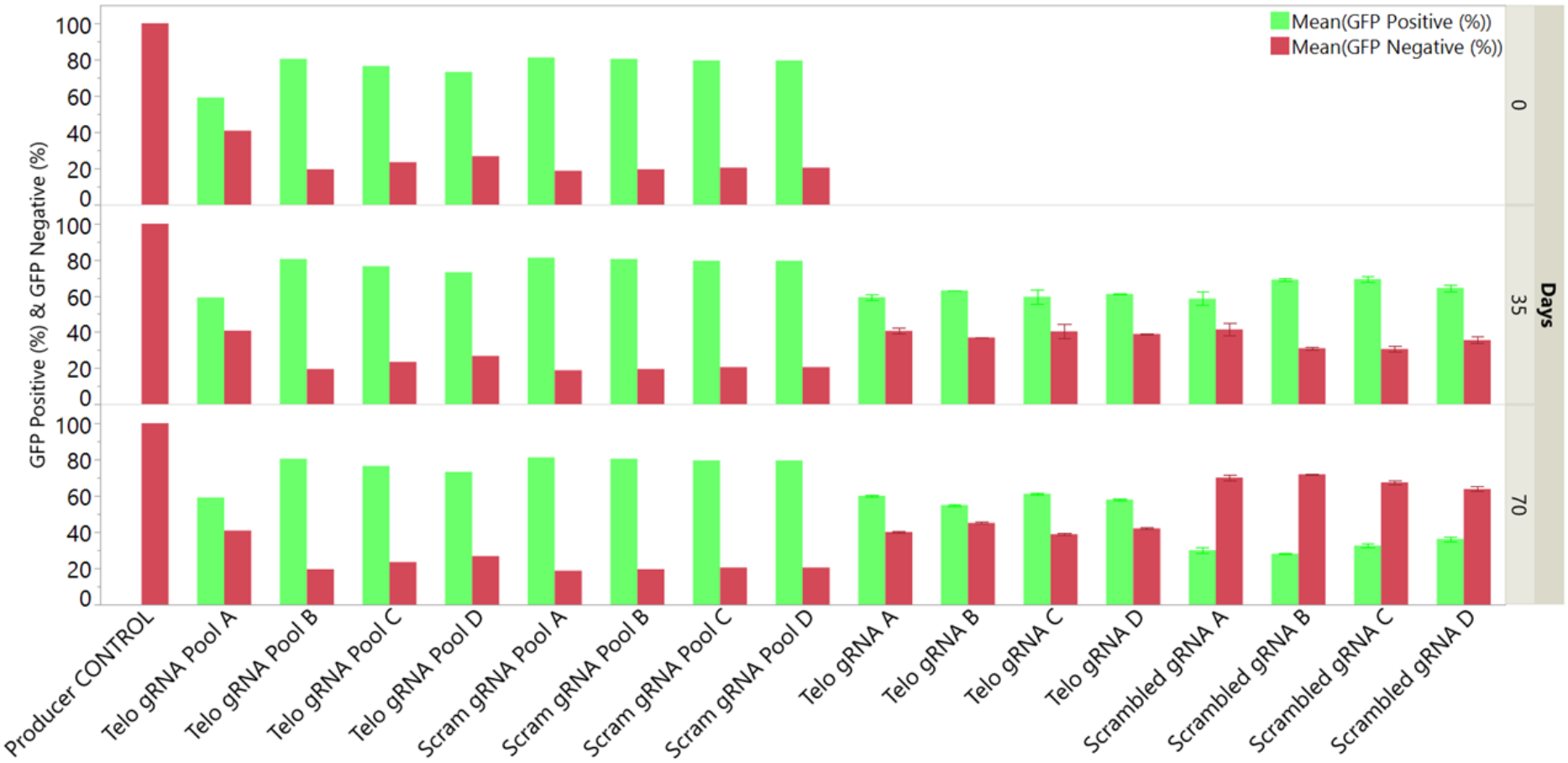


Figure 5.5. GFP positive and negative expression (%), comparing recovered sorted pools across 35 and 70 days in maintenance culture. Both Telo-Cas9 and Scrambled-Cas9 GFP expression decreases after 35 days in routine culture. The loss of GFP expression is exacerbated in Scrambled-Cas9 cell lines after 70 days (Error bars = Standard Deviation).

Cell lines were maintained for a further month (70 days, figure 5.5) to assess whether GFP fluorescence loss exacerbated over time. Telo-Cas9 cell lines maintained a similar proportion of GFP positive and GFP negative cells within the cell line cultures, whereas Scrambled-GFP cell lines GFP negative proportion increased (figure 5.5). Due to the heterogeneity in GFP (Cas9) expression the experiment was not pursued further as it was deemed results obtained from an heterogenous environment would be difficult to analyse and specifically attribute effects observed to telomere specific damage. In light of this, an experiment was designed to assess overall DNA damage effect within a production environment, using Neocarzinostatin as a DNA damaging agent. 6 productionally stable and 6 unstable stables were selected from previously analysed cell lines in the initial stable vs unstable and blinded validation panel of cell lines (methods, 2.1.1). Cell line production cultures were set up in duplicate and contained two groups of non-treated cell lines and treated with 1ng/ml Neocarzinostatin at day 0 only, using the 24 deep well production run method (2.1.9). Viable cell counts (VCC) and IgG titre were quantified on days 6, 8 and 15 and chromosomes were harvested on day 8 to assess karyotype population heterogeneity.

VCC and IgG titre levels were quantified across sampling days 6, 8 and 15 (figure 5.6 and 5.7). Neocarzinostatin had a profound effect on VCC across all sampling days, compared to their cell line counter parts in the un-treated group. This may be due to the acute genetic insult initiating apoptotic pathways, leading to cell death, or DNA damage has prevented cells from proliferating as rapidly. Protein 4, cell line 42 (P4.C2, gH2AX) retained the highest VCC of the treatment group across all sampling days (Day 6 – 6.6x10^6 cells/ml, Day 8 – 7.8 x10^6 cells/ml and Day15 – 7.8 x10^6 cells/ml). Additionally, P4.C2 obtained the highest VCC in normal conditions on days 6 and 8, indicating the cell line could have acquired a favourable mutation (before entering the production run) that promotes proliferation, compared to other cell lines in the panel.

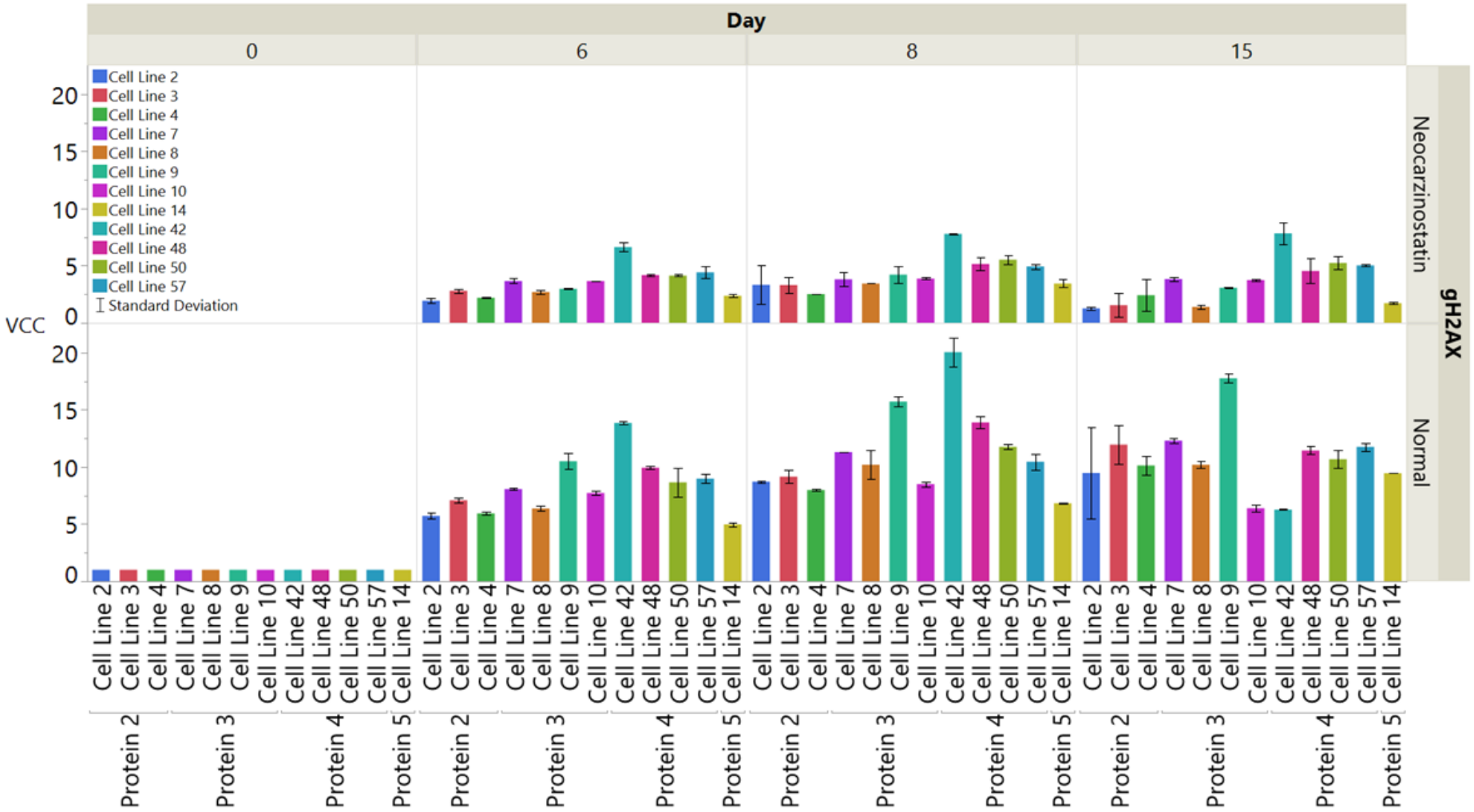


Figure 5.6. Viable cell count (VCC) of production run sample days 0, 6, 8 and 15 for non-treated and Neocarzinostatin treatment groups. 1ng/ml Neocarzinostatin had a profound effect on VCC across all sampling days, decreasing VCC (error bars = standard deviation).

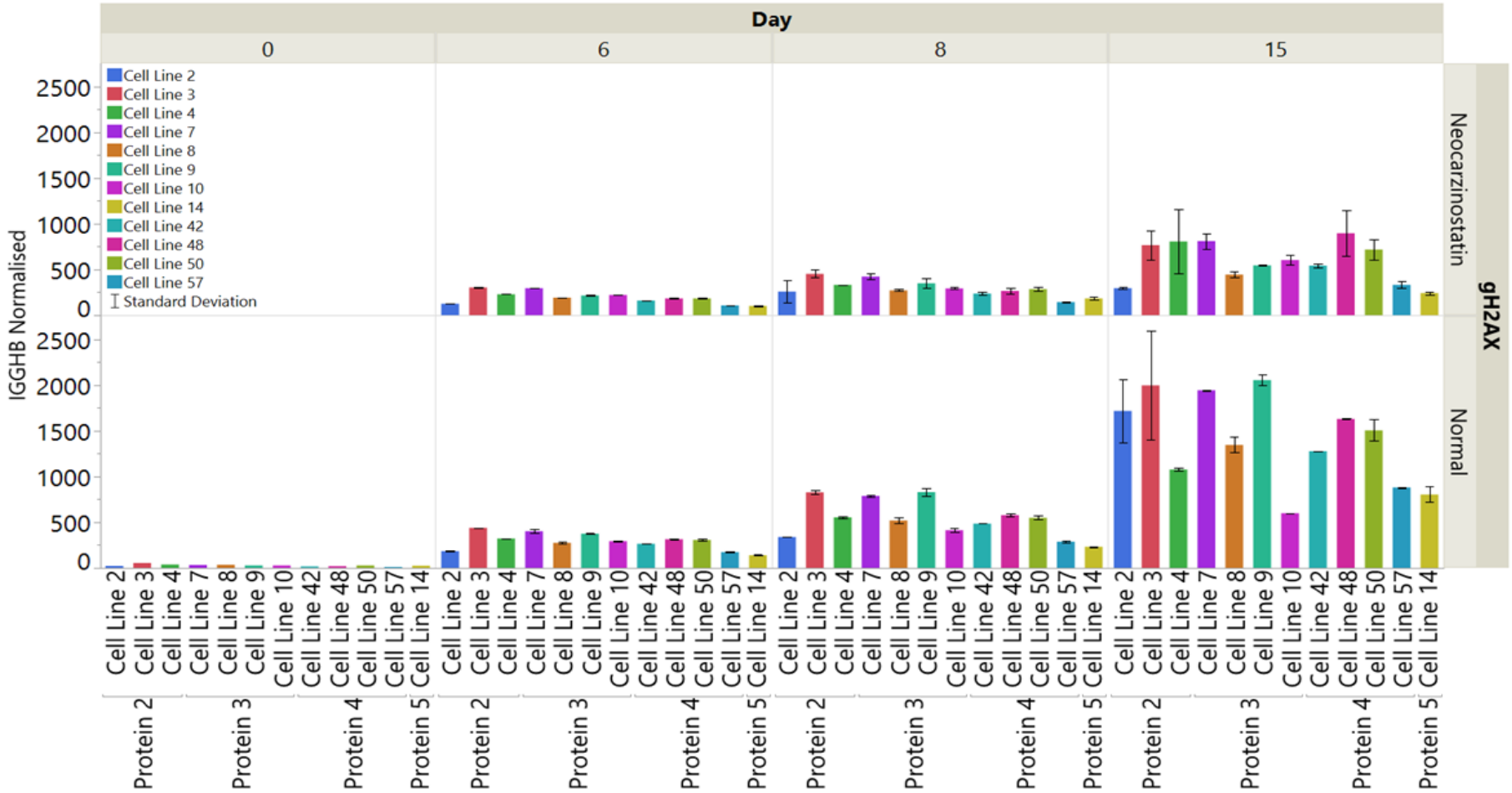


Figure 5.7. IgG titre (mg/L) quantified on sampling days 6, 8 and 15. Day 15 obtained the highest IgG titre, as expected, for both treatment groups. Despite the decrease of VCC in Neocarzinostatin treatment group, IgG titre levels remained similar to un-treated cell lines. Day 15 titre sampling showed the largest difference in IgG between the two treatment groups with untreated cell lines being able to produce higher titres to their relevant counterpart in the treated group (error bars = standard deviation

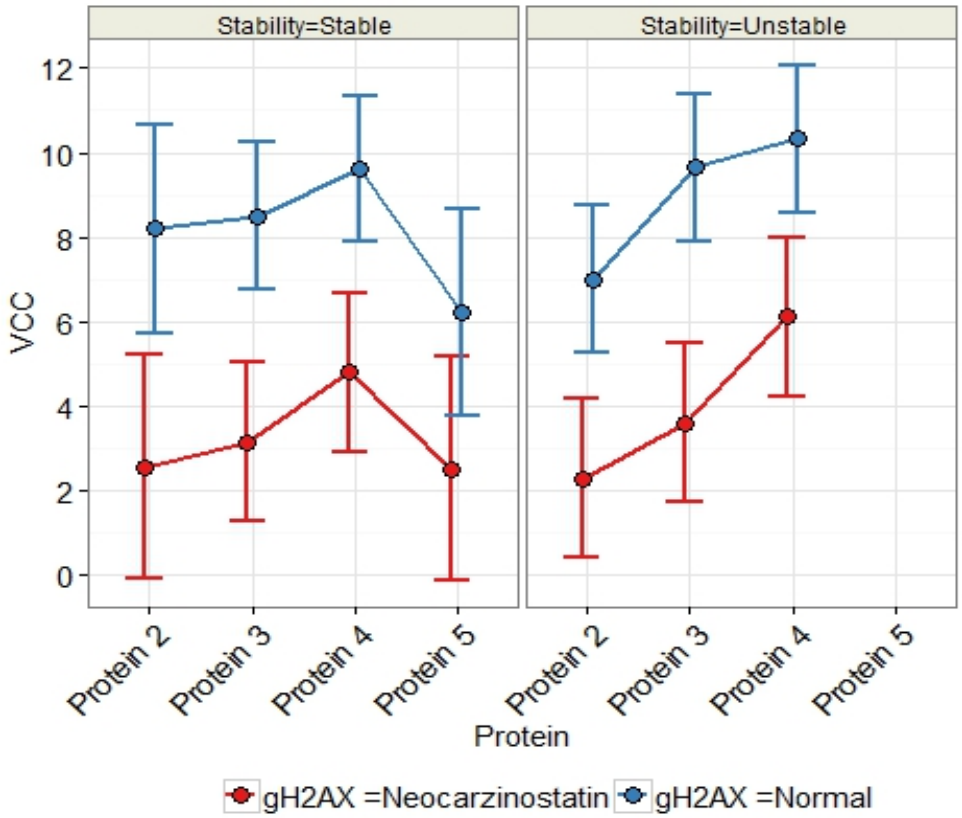


Figure 5.8. Mean VCC of cell lines belonging to each therapeutic protein within treated (red) and untreated (blue) groups.

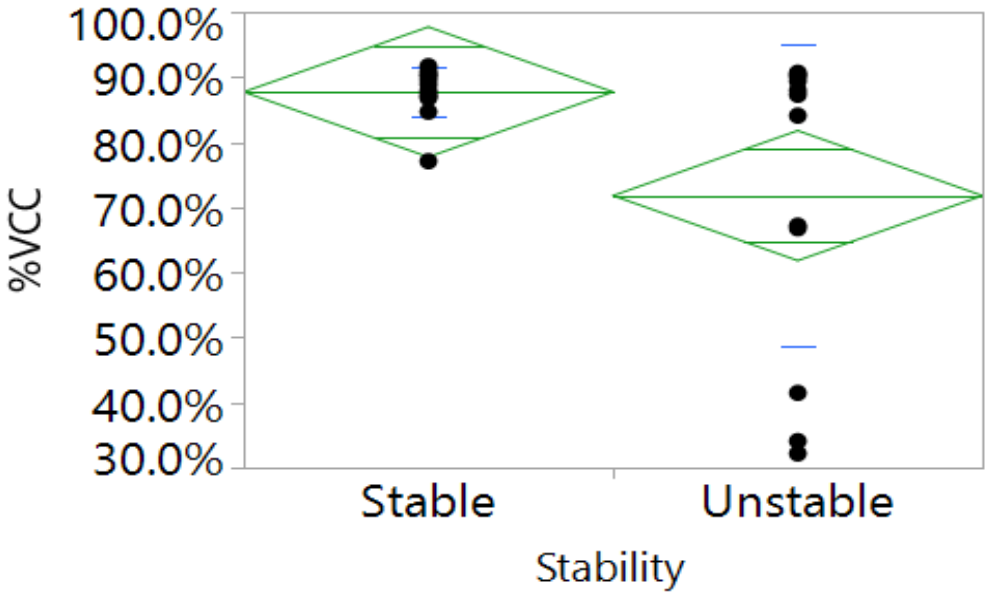


Figure 5.9. %VCC comparison of Day 15 untreated cell lines, grouped by production stability. P=0.03.

Overall, cell line VCC did not fluctuate massively across a 9-day culturing window within the treated group**,** indicating a potential inhibition of proliferation by Neocarzinostatin after the early induction of cell death by day 6. VCC decrease observed between un-treated and treated groups was statistically significant for most therapeutic proteins, except protein 5. (figure 5.8 and table 5.1, two-way ANOVA, Hochberg multiple comparison adjusted P-values). No significant difference was observed when comparing productionally stable and unstable cell lines in the un-treated group across day 6 and day 8 sampling time points. Interestingly, at day 15 there was a significant difference in %VCC between stability groups, indicating production stability may be influenced by VCC due to an increase in apoptosis, which impacts max titre for unstable cell lines (figure 5.9, pooled T-test, P=0.03).

IgG titre remained low within the Neocarzinostatin treatment group, compared to the non-treated, across all sampling days (figure 5.7). Although there was still a gradual increase in IgG titre the treatment group was unable to produce the same level of IgG titre compared to the untreated (figure 5.10). There were no significant titre differences when comparing day 6 and 8 between treatment groups (figure 5.10, four-way ANOVA, Hochberg's procedure, P=>0.05). However, there was a significant difference in titre at day 15 (figure 5.10, four-way ANOVA, Hochberg's procedure, P=<0.0001). The increase in titre would most likely be due to the increase in VCC in the un-treated group, resulting in a greater number of cells that have the ability to produce the therapeutic protein, compared to the Neocarzinostatin treated group (figure 5.6, 5.7 and 50).

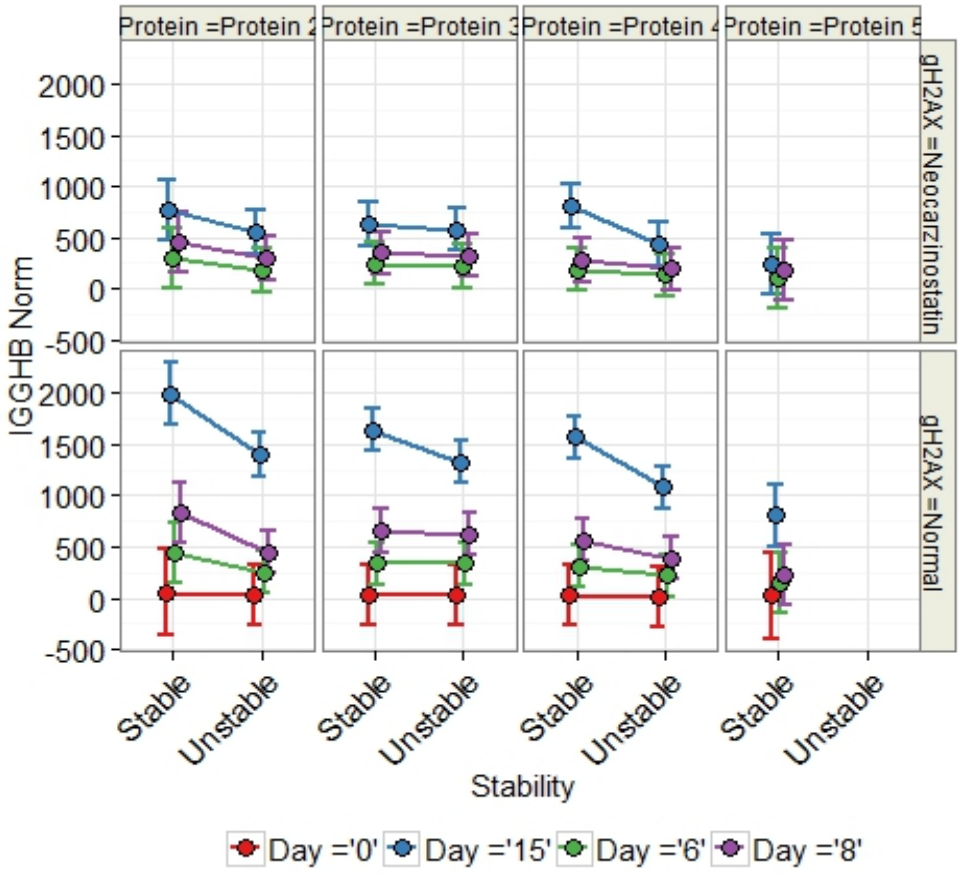


Figure 5.10. Mean normalised IgG titre for therapeutic protein cell lines across day 0, 6, 8 and 15.

To assess whether increases in DNA damage affected specific productivity rate (SPR, pg protein/cell/day), SPR was compared across sampling days, treatment groups, therapeutic protein and production stability status (figure 5.11). SPR remained consistent across sampling days and did not fluctuate between treatment groups (four-way ANOVA, Hochberg's all pair-wise comparisons, P=>0.05). SPR may remain unchanged between treatment groups due to the decrease in viability of treated cell lines, leading to greater cell death. SPR is invariably influenced by VCC number, therefore a decrease in %VCC could lead to a concomitant increase in SPR (figure 5.11 and 52), despite having less cells to produce the therapeutic protein.

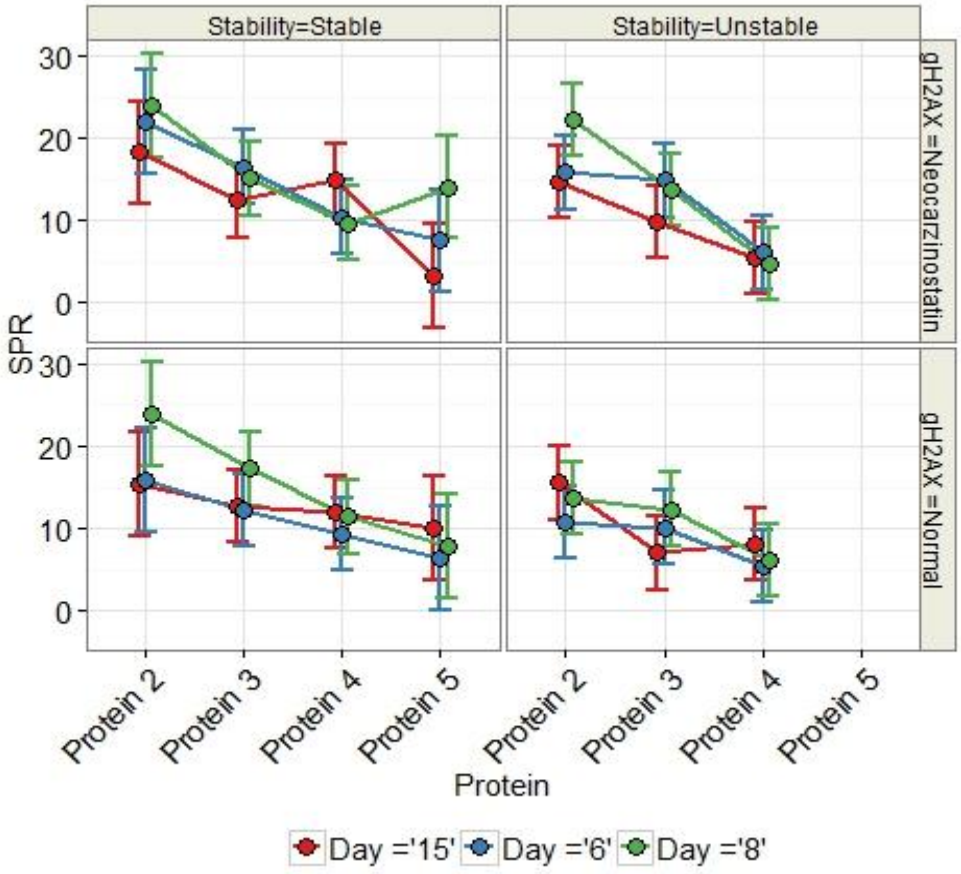


Figure 5.11. Mean specific productivity rate (SPR, pg/cell/day) for cell lines within each therapeutic protein. SPR remained consistent regardless of days in production media and treatment group.

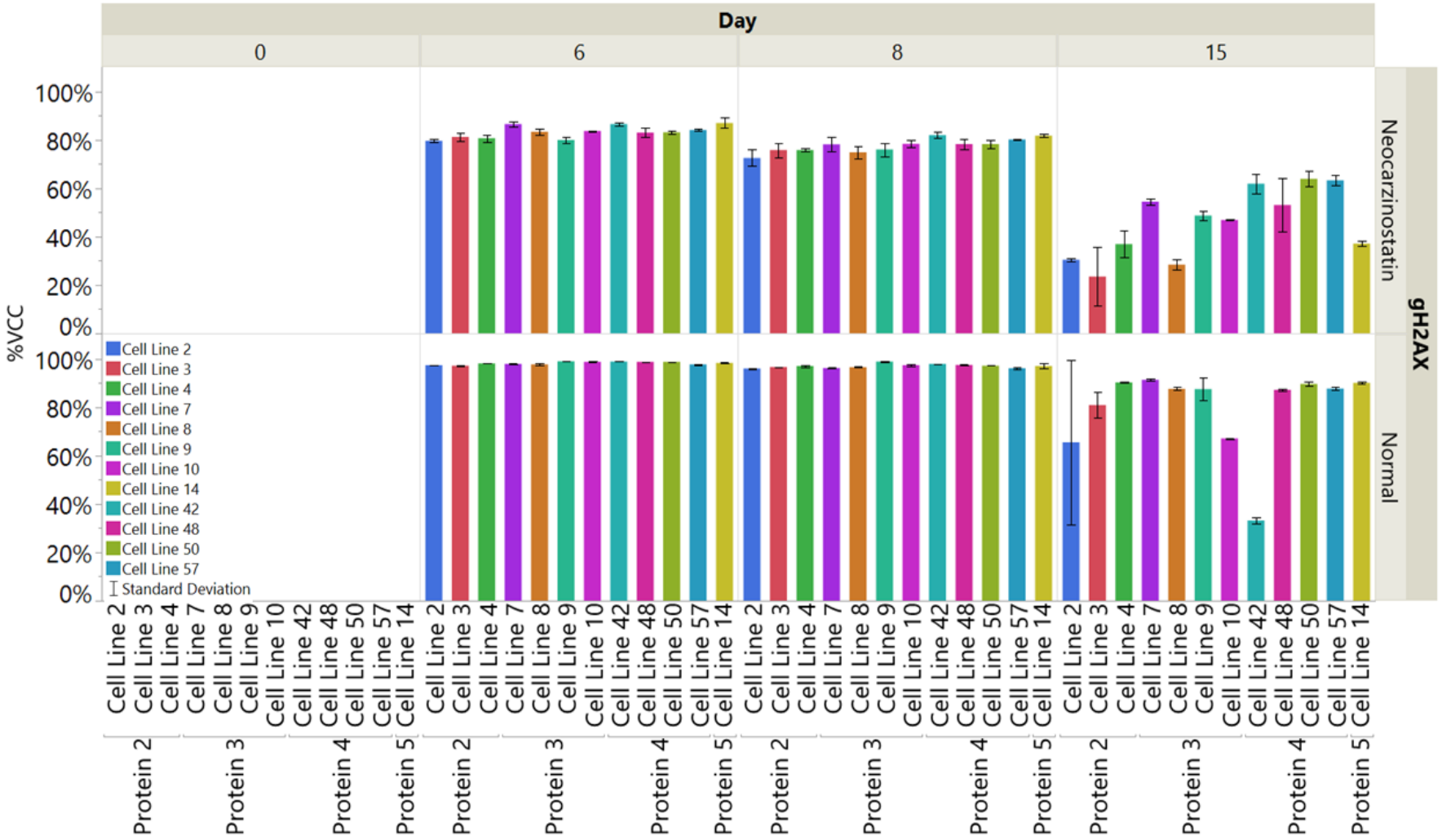


Figure 5.12. % Viability (%VCC) of cell lines across sampling days and treatment groups. Note; day 0 VCC is used to normalise VCC across samples for SPR calculations.

Data presented here highlights the effect of acute DNA damage on a cell within the production run environment. Although there was a drastic increase in IgG titre for untreated day 15 cell lines, this increase in titre can be attributed to the greater cell numbers present in the mini-bioreactor vessel (figure 5.6, 5.7, 5.8 and 50). This notion is corroborated by similar SPR values, that may be influenced by a release of IgG from apoptotic cells.

To investigate whether there is an effect on the genomic instability of cell lines, chromosomes were harvested from samples at day 8 of the production run. Day 8 was selected as a potential timepoint that would allow the stress of the production environment to elicit any potential effects, whilst maintaining a high enough %VCC to allow for appropriate sampling for analysis (figure 5.12, comparing day 8 and day 15 %VCC).

Karyotype heterogeneity was assessed using MFISH as previously described (methods, 2.3.4 and 2.7.1). Karyotype populations were assigned CCA (>5%) or NCCA (=<5%) designations based on their frequency of occurrence. Day 0 represents the baseline karyotypic heterogeneity that the cell line obtained before going through the production run protocol, which is designed to push the cells to produce as much therapeutic protein as possible. As observed in the previous stable and unstable cell line panel, productionally stable cell lines obtained a greater proportion of CCA populations compared to their productionally unstable counter parts by ~29% (figure 5.13a, b and c, table 5.1 and 5.2, P=0.004).

After 8 days within the production run environment, %CCA decreased by 32% in stable cell lines and ~17% in unstable cell lines (figure 5.13b and c, table 5.1 and 5.2, P=<0.0001*** and P=0.07n.s, respectively). This suggests the environmental stress of the production run has an impact on genetic stability as there are increases in NCCA populations (~32% and ~17%) compared to a less stressful maintenance environment at day 0. The addition of a DNA damaging agent exacerbated NCCA population increase, compared to day 8, by ~26% for stable cell lines and 23% for unstable cell lines (figure 5.13b and c, P=0.006 and P=0.014, respectively).

Table 5.1. Mean and 95% confidence limits of %CCA across production stability for sample day 0, day 8 and day 8 gH2AX.

	Sample	Stability	Mean	Lower 95% CI	Upper 95% CI
1	Day 0	Stable	0.866	0.760	0.971
2	Day 8	Stable	0.544	0.438	0.649
3	Day 8 gH2AX	Stable	0.282	0.176	0.388
4	Day 0	Unstable	0.588	0.482	0.694
5	Day 8	Unstable	0.415	0.309	0.521
6	Day 8 gH2AX	Unstable	0.184	0.078	0.290

Table 5.2. Two-way ANOVA with Hochberg's adjust P-values for pre-determined comparisons for %CCA comparisons across sample and production stability.

	Comparison	Unadjusted p-value	Adjusted p-value
1	Day 8 gH2AX Stable vs. Day 0 Stable	6.72617428421063e-09	< 0.0001
2	Day 8 gH2AX Unstable vs. Day 0 Unstable	5.33622199627715e-06	< 0.0001
3	Day 8 Stable vs. Day 0 Stable	0.000125955500982267	0.0009
4	Day 0 Unstable vs. Day 0 Stable	0.000675965734175499	0.0041
5	Day 8 Stable vs. Day 8 gH2AX Stable	0.00121681933485585	0.0061
6	Day 8 Unstable vs. Day 8 gH2AX Unstable	0.00362759781038302	0.0145
7	Day 8 Unstable vs. Day 0 Unstable	0.0246282525808061	0.0739
8	Day 8 Unstable vs. Day 8 Stable	0.089086173693866	0.1782
9	Day 8 gH2AX Unstable vs. Day 8 gH2AX Stable	0.190188840671827	0.1902

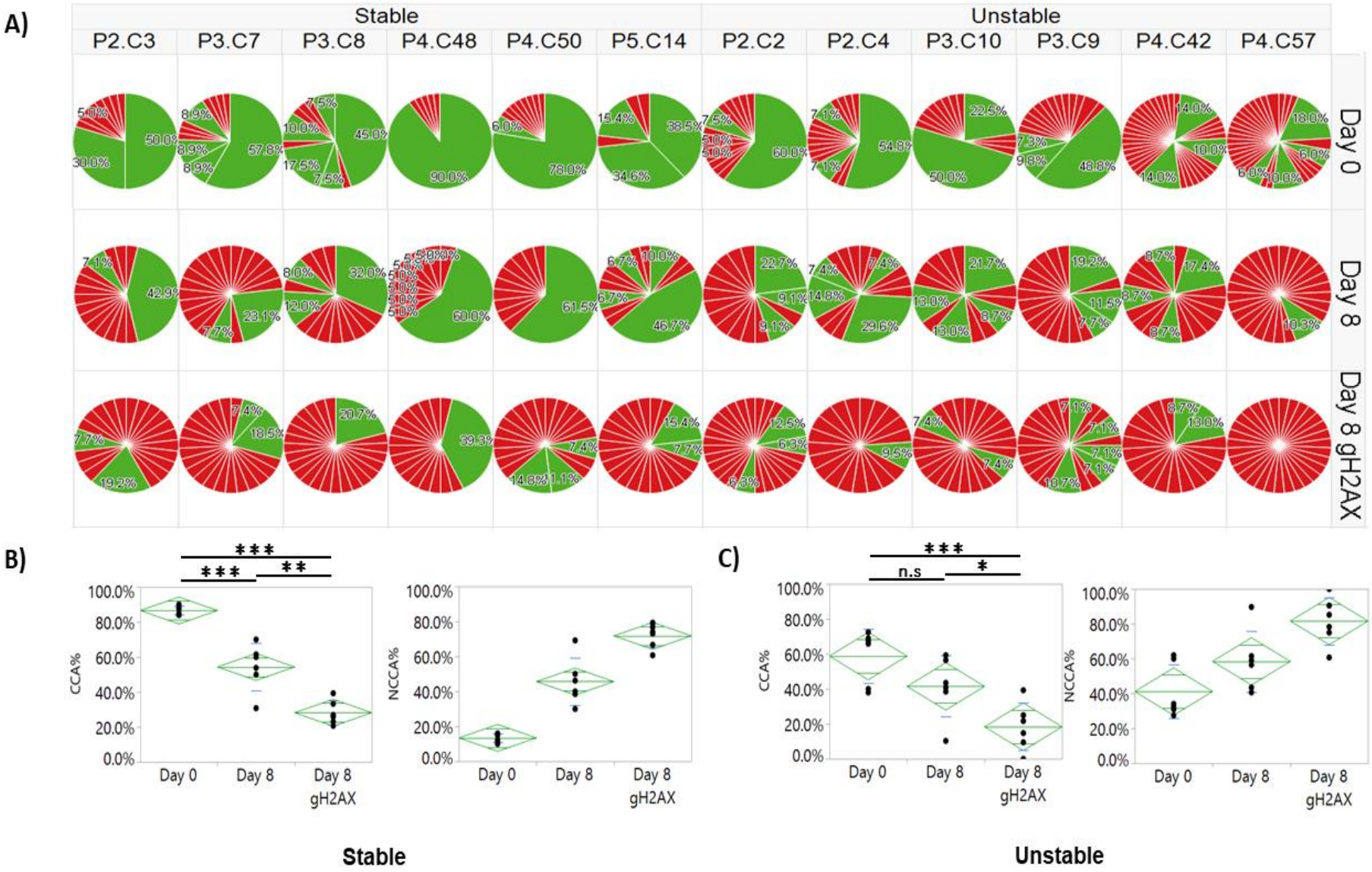

A)
Stable
Unstable
P2.C3
P3.C7
P3.C8
P4.C48
P4.C50
P5.C14
P2.C2
P2.C4
P3.C10
P3.C9
P4.C42
P4.C57
Day 0
Day 8
Day 8 gH2AX
B)
C)
CCA%
NCCA%
Day 0
Day 8
Day 8 gH2AX
***
**
*
n.s
Stable
Unstable

Figure 5.13. Karyotypic heterogeneity comparison between stable, unstable and DNA damage induced cell lines in a production environment. A) CCA (green) and NCCA (red) populations of productionally stable and unstable cell lines that have been sampled on day 8 during a production run. Day 0 time point reflects the cell lines baseline heterogeneity before entering the production run environment. Increases in NCCA populations was observed after 8 days within the production environment. Day 8 gH2AX represents the same cell lines that have been treated with 1ng/ml Neocarzinostatin for the duration of the production run. Addition of Neocarzinostatin has increased NCCA populations further (red segments). B) %CCA and %NCCA of stable cell lines across day 0, day 8 and day8 gH2AX (treated with Neocarzinostatin). Stable cell lines obtained a decrease in CCA populations after 8 days within a production run environment (two-way ANOVA, Hochberg's adjusted P-value, P=<0.001***). CCA population decrease was exacerbated further, compared to day 0 and day 8, due to the addition of DNA damaging agent (P=<0.0001*** and P=<0.01**, respectively). C) %CCA and %NCCA of unstable cell lines across day 0, day 8 and day 8 gH2AX. %CCA decreased by 17.5% between day 0 and day 8, however this was insignificant (P=0.07n.s). CCA populations decreased in the presence of Neocarzinostatin leading to ~40% decrease compared to day 0 (P=<0.0001***) and ~23% decrease compared to day 8 (P=0.015*).

An increase in NCCA populations, upon the addition of a DNA damaging agent, provides evidence that increases in DNA damage within the cell lead to the genomic instability (increase in NCCA populations) witnessed. To assess whether there is an increase in DNA damage and intracellular IgG profiles, duplicate cell lines for each condition were sampled at day 8. As previously identified, DNA damage levels were found to be low in cell lines undergoing routine maintenance using two orthogonal methods (gH2AX immunofluorescence foci counting and flow cytometry analysis).

To understand whether DNA damage levels are elevated in a production run environment, cell lines were analysed for gH2AX by flow cytometry. Cell gating strategy was performed using side scatter (SSC-A) and forward scatter (FSC-A)) to identify cells within the suspension. Subsequently the axes are changed to FSC-A and FSC-H to eliminate cell doublets (figure 5.14a). Single cell gates in combination with positive and negative stained controls were used to set gH2AX positive (CHOK1a-GS-KO cells treated with 1ng/ml Neocarzinostatin) and untreated CHOK1a-GS-KO cell suspension (figure 5.14b).

Duplicate stable and unstable cell lines in un-treated (blue and green peaks) and DNA damage treated groups (orange and red peaks) were analysed for gH2AX by flow cytometry, as previously described (methods, 2.5.1). Despite cells being in a more stressful environment, gH2AX levels remained low within cells at day 8 of the production run protocol. Additionally, gH2AX treated samples also had similarly low levels of gH2AX positive cells compared to day 0 and day 8 untreated samples (figure 5.15). Although day 8 and day 8 gH2AX LSM values look elevated compared to day 0, standard deviation (error bars) indicate a wide spread of data that overlaps between each of the sampling days. Therefore, no significant difference was found between stability groups and sampling days (figure 5.15, P=>0.05).

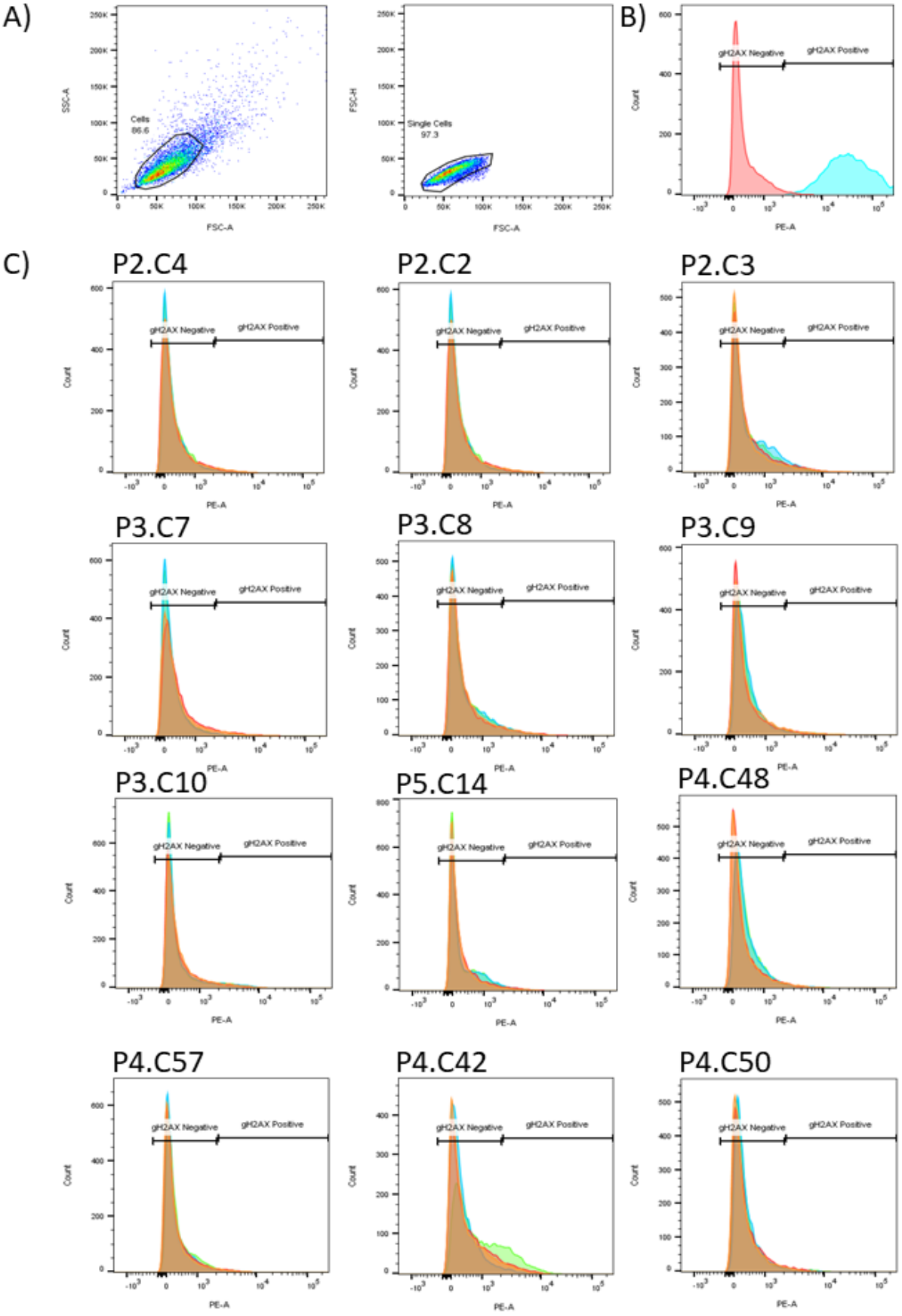

A)
SSC-A
FSC-A
Cells
86.6
FSC-H
FSC-A
Single Cells
97.3
B)
gH2AX Negative
gH2AX Positive
Count
PE-A
C)
P2.C4
P2.C2
P2.C3
P3.C7
P3.C8
P3.C9
P3.C10
P5.C14
P4.C48
P4.C57
P4.C42
P4.C50

Figure 5.14. DNA damage quantification by flow cytometry of Stable and Unstable cell lines in a production environment. A) representative flow cytometry dot plots used for cell gating strategy, to eliminate cell doublets. B) Negative and positive gH2AX gates set using untreated CHOK1a-GS-KO and CHOK1a-GS-KO treated with 1ng/ml Neocarzinostatin, respectively. C) gH2AX flow cytometry analysis of duplicate stable and unstable cell lines in a production run environment with un-treated (blue and green peaks) and DNA damage treated groups (red and orange peaks).

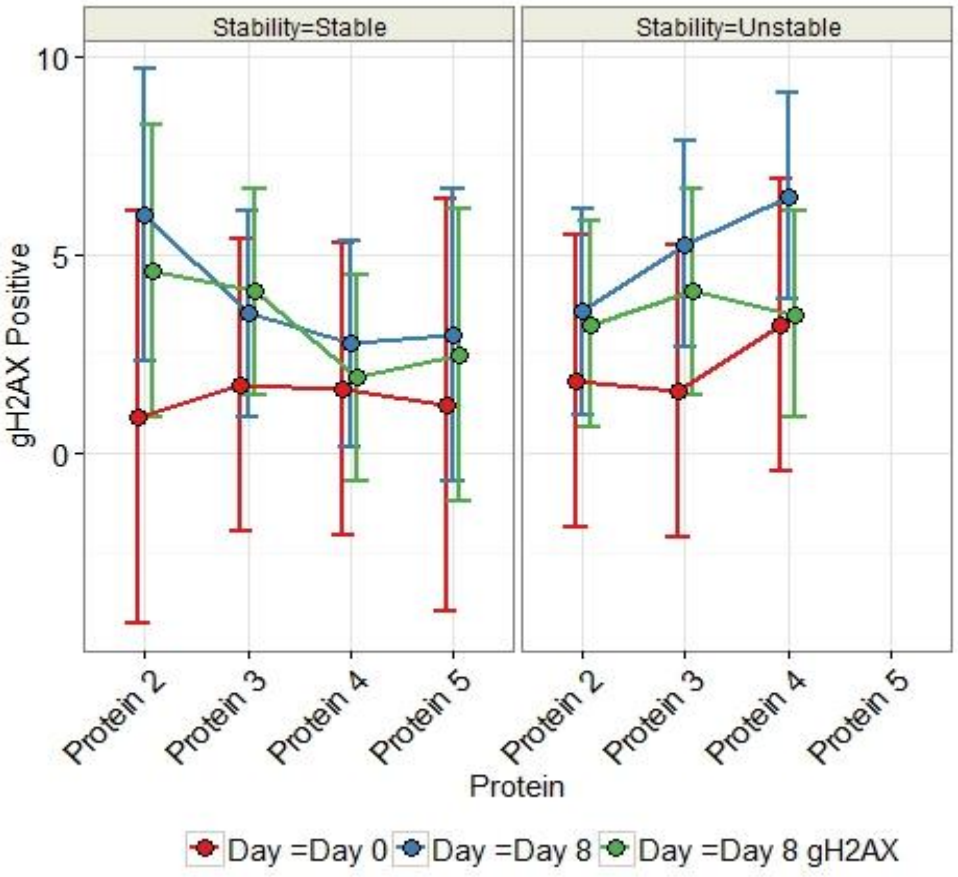


Figure 5.15. Least square mean plot of mean gH2AX positive cells from day 0, day 8 and day 8 DNA damage treated samples within a production run environment. No significant difference was found (P=>0.05, three-way ANOVA, Hochberg's approach).

Although DNA damage is presumably occurring during the cell's nature cycle within the production run environment, either CHOK1a-GS-KO cells have a great propensity to resolve DNA damage quickly or the production media provides the components required to keep DNA damage to a minimum. Although the DNA damaging agent initially elevates DNA damage, as signified by the positive control, where gH2AX-PE fluorescence is increased upon incubating cells with 1ng/ml Neocarzinostatin for 1 hour. No elevated gH2AX was observed within the cell lines treated with the DNA damaging agent. Conversely, %viability of cells within treated samples was decreased compared to untreated cell lines, indicating that Neocarzinostatin had a positive effect on apoptotic pathways during the production run even though gH2AX levels remained low (day 15, figure 5.12, 54 and 55).

Diminished gH2AX levels, in the presence of a DNA damaging agent, has been shown to increase within an 8-hour window but decreases after 24-48 hours (Mao et al., 2016). Therefore, DNA damage could elicit mutations within cells, which leads to the vast heterogeneity witnessed in day 8 treated and untreated cells and become undetectable by day 8 analysis (192 hours after initial exposure). During routine maintenance, different intracellular IgG peak patterns were observed for stable (more uniform) and unstable (wider distribution of peaks) cell lines. To understand whether these expression peak profiles are maintained in a production run environment, anti-human IgG-647 intracellular flow cytometry analysis was performed, as previously described. Single cells (figure 5.16a, gating strategy) were analysed for IgG-647 expression using CHOK1a-GS-KO host cells and a producer cell line to set negative and positive IgG gates, respectively (figure 5.16b). Cell lines were analysed for intracellular IgG in un-treated (green and blue peaks) and treated groups (orange and red peaks). IgG production remained consistent across stability group and treatment groups (figure 5.17, P=>0.05), indicating that additional DNA damage did not have an effect on the amount of IgG that is produced internally in the cell.

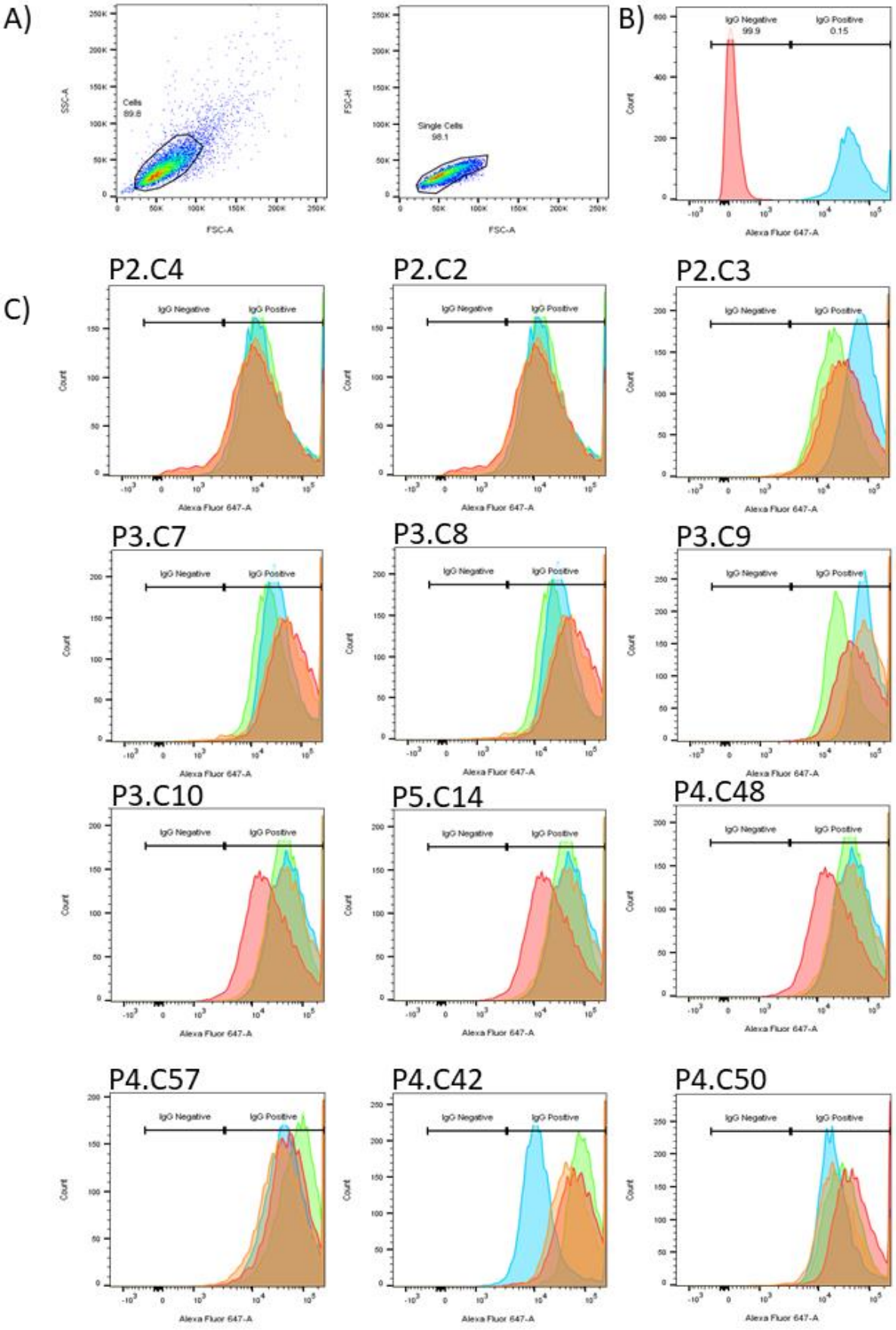

A)
SSC-A
FSC-A
Cells
89.8
FSC-H
FSC-A
Single Cells
98.1
B)
IgG Negative
99.9
IgG Positive
0.15
Count
Alexa Fluor 647-A
C)
P2.C4
P2.C2
P2.C3
P3.C7
P3.C8
P3.C9
P3.C10
P5.C14
P4.C48
P4.C57
P4.C42
P4.C50
IgG Negative
IgG Positive
Count
Alexa Fluor 647-A

Figure 5.16. Quantification of intracellular IgG expression in Stable, Unstable and DNA damage induced cell lines, within a production environment. A) representative flow cytometry dot plots used for cell gating strategy, to eliminate cell doublets. B) Negative and positive IgG gates set using CHOK1a-GS-KO host and producer cell line, respectively. C) Intracellular IgG flow cytometry analysis of duplicate stable and unstable cell lines in a production run environment with un-treated (blue and green peaks) and DNA damage treated groups (red and orange peaks).

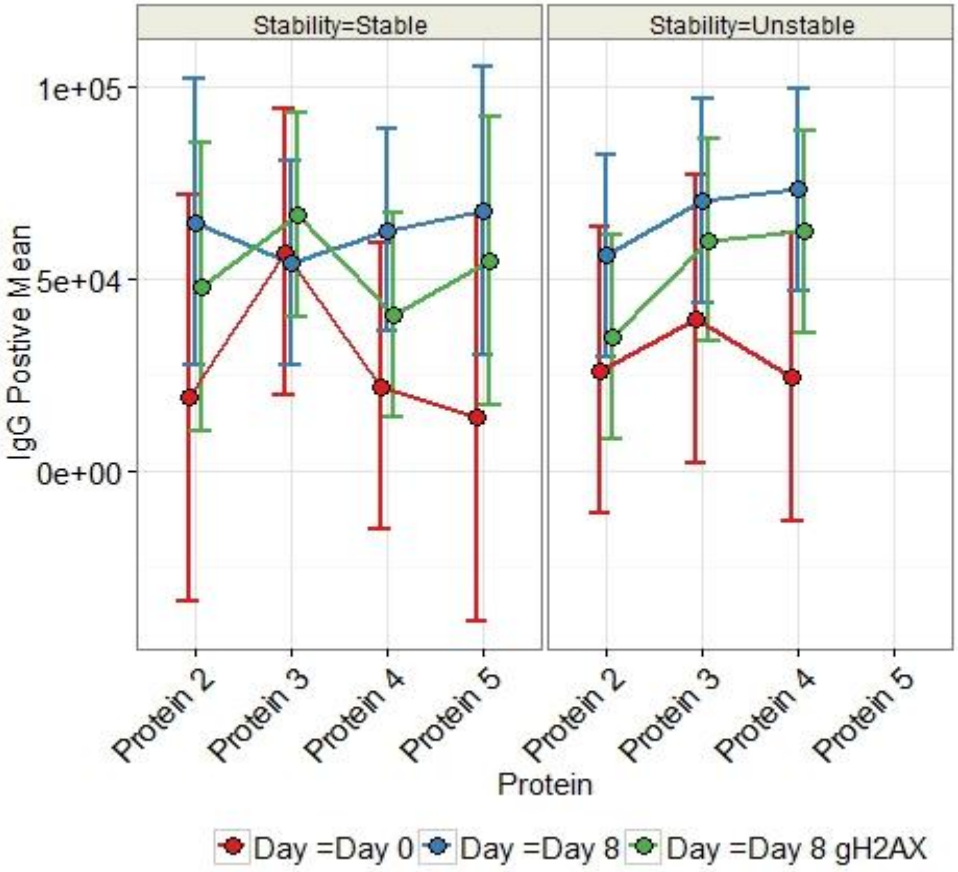


Figure 5.17. LSM plot of IgG positive cells in stable and unstable cell lines across treatment groups. No significant difference found between groups (three-way ANOVA, Hochberg's procedure, P=>0.05).

Chapter 6: Application of automation strategies to enable the industrialisation of MFISH prediction method and assays to characterise CHOK1a-GS-KO based producer cell lines.

Introduction (6.1)

Throughout this thesis, the majority of assays have been performed manually which is receptive to certain caveats such as increasing the likelihood of experimental errors, quantifications being open to unintentional bias / subjectivity and routine experiments being largely laborious. With this in mind, there is a strong focus on automating routine assays in order to minimise scientists time at the bench and focus their time and effort on experimental planning and data analysis.

There are numerous hardware options available that optimise routine tasks such as liquid handling, single cell sorting and counting, fluorescent in-situ hybridisation (FISH) and molecular cloning tasks, such as DNA extraction. However, automated data analysis pipelines are severely lacking, representing a major bottleneck in decreasing lead times for cellular characterisation compared to other highly automated and optimised processes performed downstream to cell line selection and development.

Historically, biologic pharmaceutical companies have invested time and money into highly optimising cell culture conditions, with the majority of companies using their own proprietary growth and production

media (Wurm, 2004). Mini-bioreactors are now common place within industry and are implemented to replicate culturing conditions at manufacturing scale, covering cell suspension volumes from 15ml (e.g. Ambr15) to 50,000L scale. Processes surrounding transfection, single cell sorting, scale up and cell culturing, have all undergone several iterations of standards operating procedure (SOP) to ensure processes can be completed in the shortest time whilst maintaining quality of cell lines, harvested therapeutic protein and process data.

With all of the improvements that have been achieved through process optimisation and hardware automation, there are several bottlenecks in CLD that biopharmaceutical companies have encountered, which have not been resolved through the use of traditional process optimisation. Two major bottlenecks encountered are the length of stability assessment and the productivity of therapeutic protein producing cell lines. With media optimisation alone, max titres have increased to ~4g/L (Wurm, 2004), but companies are voraciously chasing cell lines that produce >10g/L. Doubling the overall titre during a single manufacturing run has obvious positive effects on a company's balance sheet, allowing for the production of twice the therapeutic protein for half of the cost.

To achieve such a feat, companies and research laboratories are pursuing academic centric methodologies to characterise CHO cell lines at every CHO'mic level, in a bid to identify bioengineering targets that are amenable to manipulation. Although outsourcing options are available, such as CRISPR knock out screens (e.g. Horizon Discovery), many of these relatively newer techniques (new in the biopharmaceutical space) require the same concepts of process optimisation and automation strategies to meet industry timeline demands.

Within this thesis, there has been a strong focus on fluorescent imaging techniques, utilised to provide a better understanding of CHO cell biology. To ensure such assays could provide real world value to

biopharmaceutical cell line development, multiple automation efforts were pursued for the duration of this project. This chapter will describe the automation methodologies devised to ensure experimental procedures performed within this thesis are industry scalable and, if required, meet timeline schedules for cell line progression in the CLD critical path.

Aims and hypotheses (6.1.1)

Throughout this thesis the majority of the assays performed have been through manual means. Within an ever-changing pharmaceutical environment, cell characterisation and analysis must be industrially scalable, and data rapidly generated to provide a greater depth of CHOK1a-GS-KO cell characterisation, without impacting project timelines during cell line development. Primarily image analysis and liquid handling for genetic screens were identified as major bottlenecks for these types of analyses. This chapter will outline the solutions conceptualised and implemented during this project, to allow for the industrialisation of image analysis and genetic screens.

Genetic screen automation using the *epMotion 5075 (Eppendorf)* liquid handler (6.2).

Aims (6.2.1)

To establish, test and optimise an automated workflow on the *epMotion 5075* (Eppendorf) for genetic screens performed in 384-well plates.

Results (6.2.2)

Manual genetic screens have been utilised widely to assess differential expression patterns between different cell lines or treatment groups in a bid to better understand the underlying biology that culminates in a phenotype observed by researchers. Here, a genetic screen was designed to target pathways involved in DNA damage repair, resolution and telomere homeostasis, in order to identify a potential bioengineerable target that may provide an inherently more robust CHOK1a-GS-KO host. To ensure the genetic screen could be performed within a reasonable time frame, a 384-well plate format was selected. Performing a genetic screen manually using 384-well plate format represents its own issues in terms of reproducibility and accuracy, due to potential human pipetting error and day to day variance, in addition to its highly laborious nature. In light of this, a workflow was devised on the Epmotion liquid handler (figure 6.1), which has multiple pipetting options and single use filter tips, required to eliminate cross contamination risk.

Figure 6.1. Image of liquid handler Epmotion 5075.

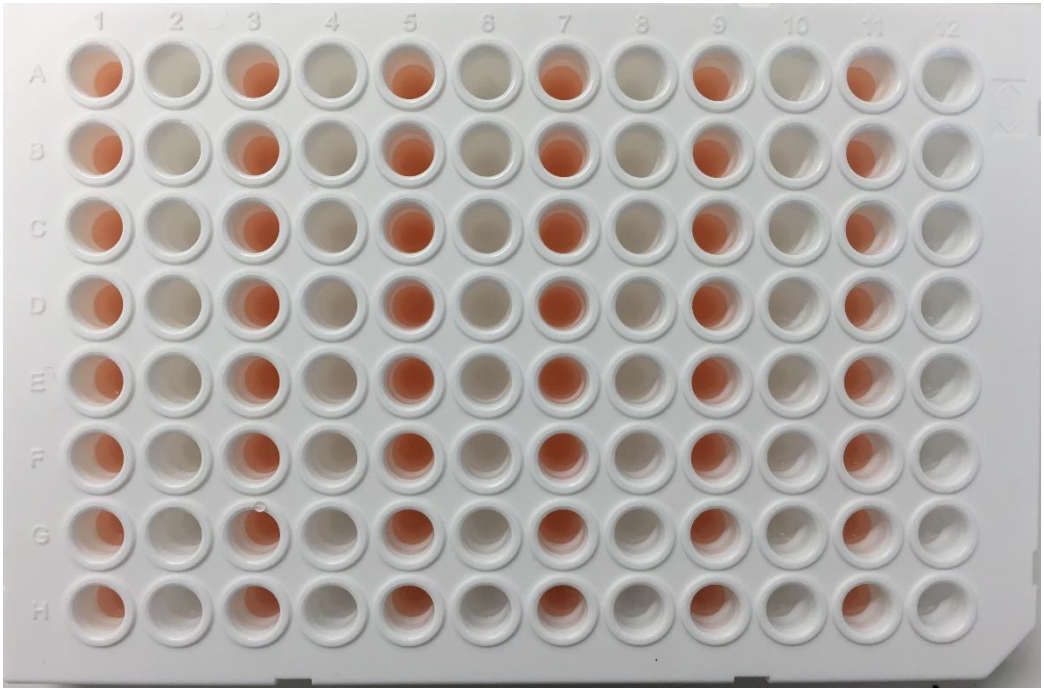


Figure 6.2. Image of initial carry over test using water and TrypLE with phenol red.

To ensure there was no well to well carry over from moving tips, a dummy experiment was performed in a 96-well plate containing alternating rows of water and TrypLE with phenol red. Liquid handling protocol was performed by pipetting water into each sample well to assess cross contamination. No carry over was visually observed (figure 6.2).

As PCR assays are highly sensitive to minute cross contamination, another experiment was performed to using CHOK1a-GS-KO cDNA (6ng/µl) and a house keeping gene (HKG) Tubulin in a SYBR green based master mix (methods, 2.4.4). HKG master mix was decanted into each well and sample template decanted into alternating rows (odd rows). The resulting plate was analysed on a Light Cycler 480 to assess any cross contamination between wells, quantified by cycle thresholds. Within wells containing master mix only, no cycle threshold (CT) was reached after 35 cycles, indicating the liquid handling protocol does not produce cross contamination across wells as no template was carried over (figure 6.3).

Next, a full plate was produced using host cDNA and HKG SYBR green master mix, to assess coefficience of variance (CV%) of CT across a whole 384-well plate. CTs obtained minimal variation across the whole plate, with an overall 1% CV, showing consistency in pipetting from the liquid handling protocol (figure 6.4). After optimisation of the liquid handling protocol, an HKG determination experiment was performed, testing each HKG across each sample and timepoint, as previously described (figure 4.15).

Having an optimised liquid handling protocol for 384-well plate-based gene screens provided the means to produce 6 full 384-well plates in four hours, with minimal intervention from the scientist, allowing their time to be focussed on another task. Additionally, due to the small CV% between wells, there is guaranteed consistency of pipetting across all 384-well plates produced for the genetic screen, which may not be achievable when performing manually. Overall, 48 384-well plates were produced using the automated liquid handling protocol, amounting to a total time of 32 hours that was spent on performing other experiments rather than performing the gene screen manually.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
A	16.44	>35	16.25	>35	16.32	>35	16.54	>35	16.63	>35	16.46	>35	16.53	>35	16.63	>35	16.24	>35	16.42	>35	16.72	>35	16.34	>35
B	16.32	>35	16.48	>35	16.84	>35	16.19	>35	16.34	>35	16.69	>35	16.26	>35	16.67	>35	16.34	>35	16.89	>35	16.57	>35	16.56	>35
C	16.45	>35	16.33	>35	16.44	>35	16.6	>35	16.45	>35	16.27	>35	16.54	>35	16.99	>35	16.28	>35	16.23	>35	17	>35	16.34	>35
D	16.43	>35	16.59	>35	16.64	>35	16.41	>35	16.44	>35	16.51	>35	16.57	>35	16.45	>35	16.21	>35	17.08	>35	16.43	>35	16.55	>35
E	16.5	>35	16.41	>35	16.34	>35	16.56	>35	16.21	>35	16.42	>35	16.46	>35	16.84	>35	16.33	>35	16.48	>35	16.46	>35	16.57	>35
F	16.29	>35	16.6	>35	16.87	>35	16.5	>35	16.45	>35	16.62	>35	16.56	>35	16.89	>35	16.56	>35	16.51	>35	16.49	>35	16.5	>35
G	16.57	>35	16.57	>35	16.49	>35	16.7	>35	16.49	>35	16.75	>35	16.32	>35	16.82	>35	16.2	>35	16.46	>35	16.45	>35	16.22	>35
H	16.33	>35	16.58	>35	16.76	>35	16.58	>35	16.59	>35	16.53	>35	16.3	>35	16.45	>35	16.51	>35	16.47	>35	16.63	>35	16.57	>35
I	16.73	>35	16.42	>35	16.42	>35	16.5	>35	16.73	>35	16.52	>35	16.57	>35	16.7	>35	16.46	>35	16.41	>35	16.33	>35	16.48	>35
J	16.51	>35	16.57	>35	16.45	>35	16.49	>35	16.44	>35	16.51	>35	16.46	>35	16.54	>35	16.51	>35	16.8	>35	16.66	>35	16.51	>35
K	16.56	>35	16.45	>35	16.55	>35	16.46	>35	16.43	>35	16.54	>35	16.57	>35	16.58	>35	16.56	>35	16.46	>35	16.44	>35	16.46	>35
L	16.53	>35	16.56	>35	16.42	>35	16.57	>35	16.57	>35	16.58	>35	16.46	>35	16.51	>35	16.53	>35	16.8	>35	16.64	>35	16.47	>35
M	16.66	>35	16.47	>35	16.58	>35	16.5	>35	16.49	>35	16.52	>35	16.56	>35	16.7	>35	16.66	>35	16.44	>35	16.28	>35	16.41	>35
N	16.52	>35	16.59	>35	16.5	>35	16.71	>35	16.59	>35	16.55	>35	16.32	>35	16.42	>35	16.52	>35	16.66	>35	16.57	>35	16.47	>35
O	16.62	>35	16.47	>35	16.48	>35	16.45	>35	16.49	>35	16.44	>35	16.53	>35	16.69	>35	16.27	>35	16.44	>35	16.67	>35	16.31	>35
P	16.49	>35	16.55	>35	16.69	>35	16.33	>35	16.61	>35	16.54	>35	16.3	>35	16.59	>35	16.66	>35	16.73	>35	16.31	>35	16.67	>35

Figure 6.3. 384-well plate CT values of alternating columns containing HKG master mix and template (odd numbers) and master mix only (even numbers) to assess any template carryover using the automated liquid handling method.

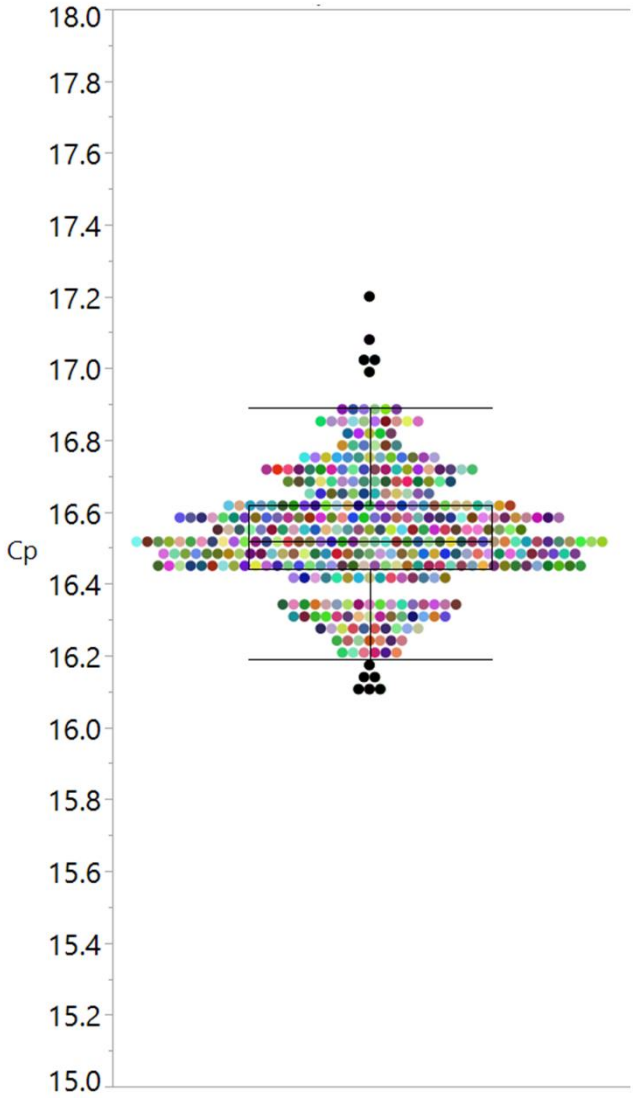


Figure 6.4. CT values for Tubulin (HKG) across a full 384-well plate. Each dot represents a single well CT value. Master mix and template was decanted into wells using the automated workflow on the Epmotion. 1% CV was observed across all 384 well plates, showing consistency in pipetting from the automated liquid handling protocol.

Image analysis automation workflows devised to streamline MFISH production stability prediction analysis timelines and provide industry scalable image analysis tools (6.3).

Aims (6.3.1)

Image analysis is often performed using software that allows characterisation of fluorescent images, but often in a manual and subjective manor (e.g. ImageJ). In a bid to remove this subjectivity from analyses and decrease analysis time lines, image analysis workflows were created on Cell profiler (http://cellprofiler.org/) using their built-in image analysis modules to confirm mutations observed. This section will describe said workflows and how they could be applied on the CLD critical path. Additionally, results from a successful collaboration with GlaxoSmithKline's Artificial Intelligence and Machine Learning (AI/ML) group, which allowed the industrialisation of the production stability prediction method, will be described.

Results (6.3.2)

Fluorescent based image analysis represents an important tool for cellular characterisation that is used across all walks of science. It provides the ability to visualise any protein or DNA sequence (when there are appropriate antibodies and probes are available) within the cell that aids a better description of the underlying cellular biology when investigating a desired phenotype. However, image analyses have historically been analysed manually, opening up the analyses to unintentional bias and subjectivity that may impact the output of results.

Within this thesis, analysis of MFISH karyotypes of CHOK1a-GS-KO host, productionally stable and unstable cell lines was performed manually. To remove potential subjectivity and bias in mutation identification, a cell profiler workflow was created to extract the fluorescent intensities from 5 separate colour channels, from each individual chromosome. Single channel images are extracted from the Metafer software (Metasystems, V5.7.4) and undergo a series of threshold corrections to remove background fluorescence (methods, 2.7.2).

Chromosome masks are identified through the identify primary objects module using the DAPI channel (figure 6.5). Automated masks are manually edited to remove any artefacts (e.g. cells or debris) within the image. Additionally, chromosomes that are in close proximity can be split into individual masks to faithfully replicate the original image. Semi-automated chromosome segmentation allows for the extraction of fluorescent intensity values of pixels in each colour channel contained within the mask.

Expressing the fluorescent pixel intensities from each channel within a single chromosome mask, as a percentage of each other, provides a chromosome colour profile (figure 6.6 and 6.7, showing CHOK1a-GS-KO host conserved chromosomes) that is utilised to confirm chromosome mutations that are visually identified (figure 6.8). This allows the analyst to have a colour profile of the mutation in question providing further evidence that the mutation observed by the analyst is reflected at the fluorescent pixel intensity level.

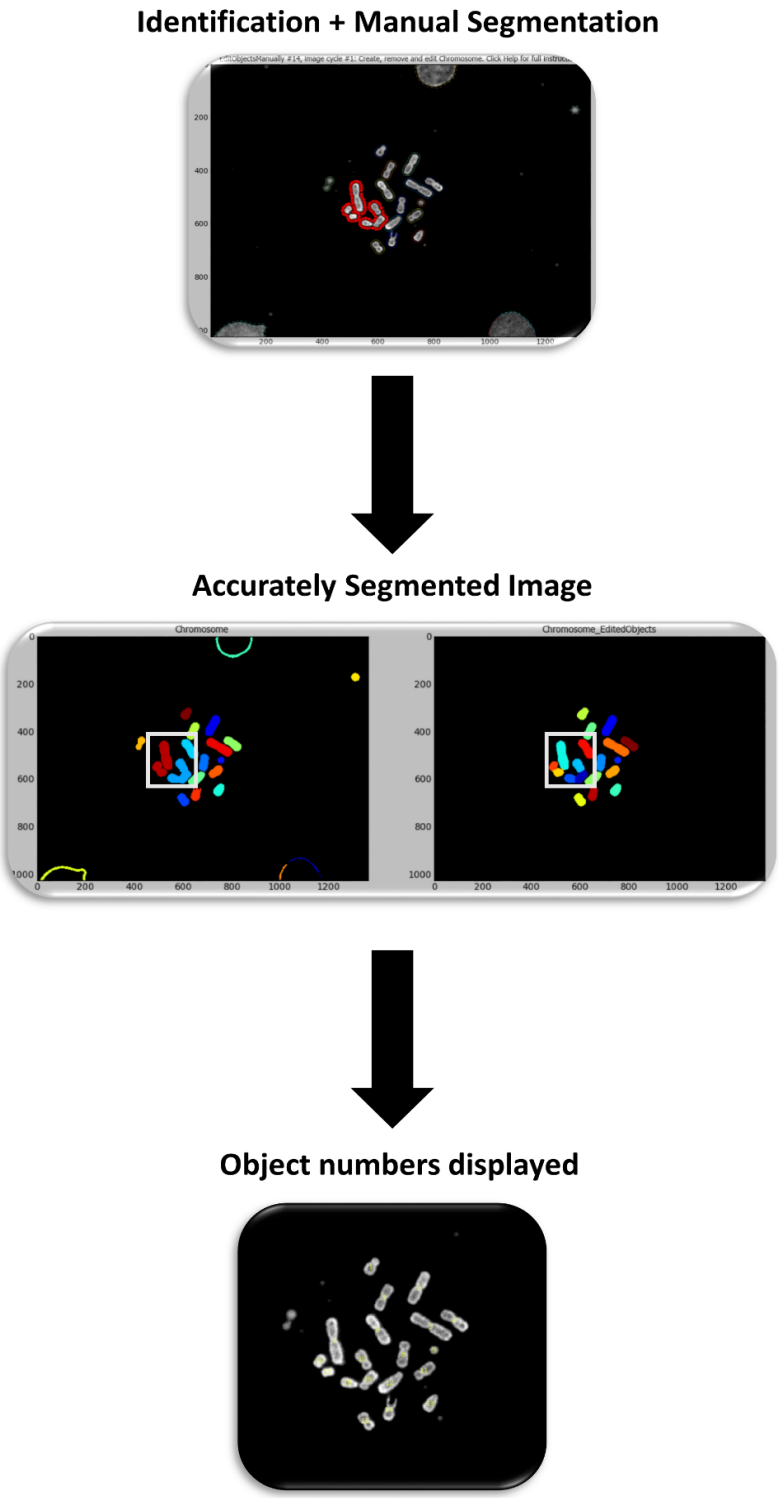


Figure 6.5. Cell profiler workflow showing chromosome segmentation steps using a semi-automated approach. Object numbers (identifying chromosomes in excel output) allow analysts to correlate fluorescent pixel intensities to individual chromosomes within the image.

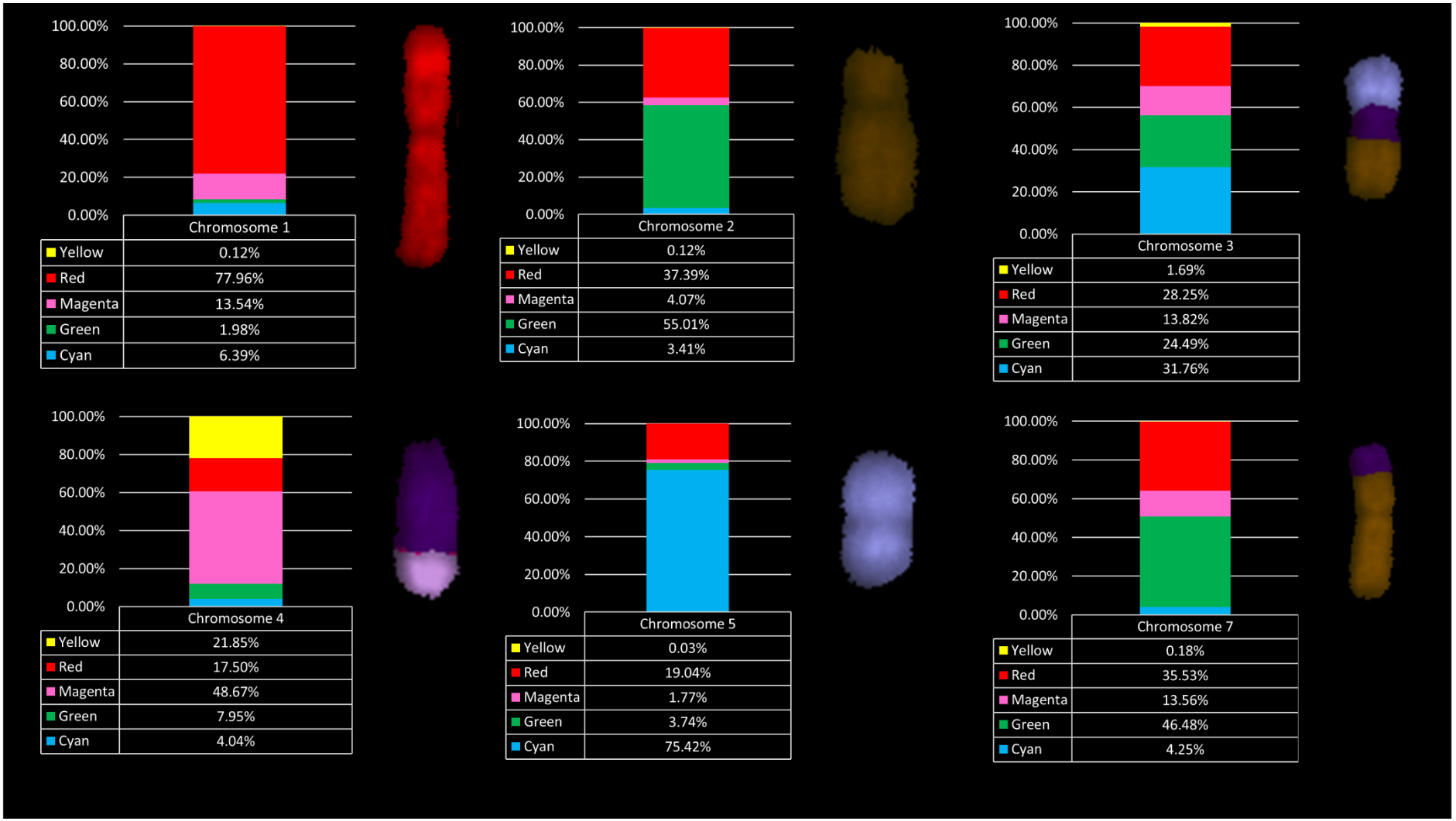


Figure 6.6. Fluorescent intensity colour profiles of CHOK1a-GS-KO chromosomes 1, 2, 3, 4, 5 and 7.

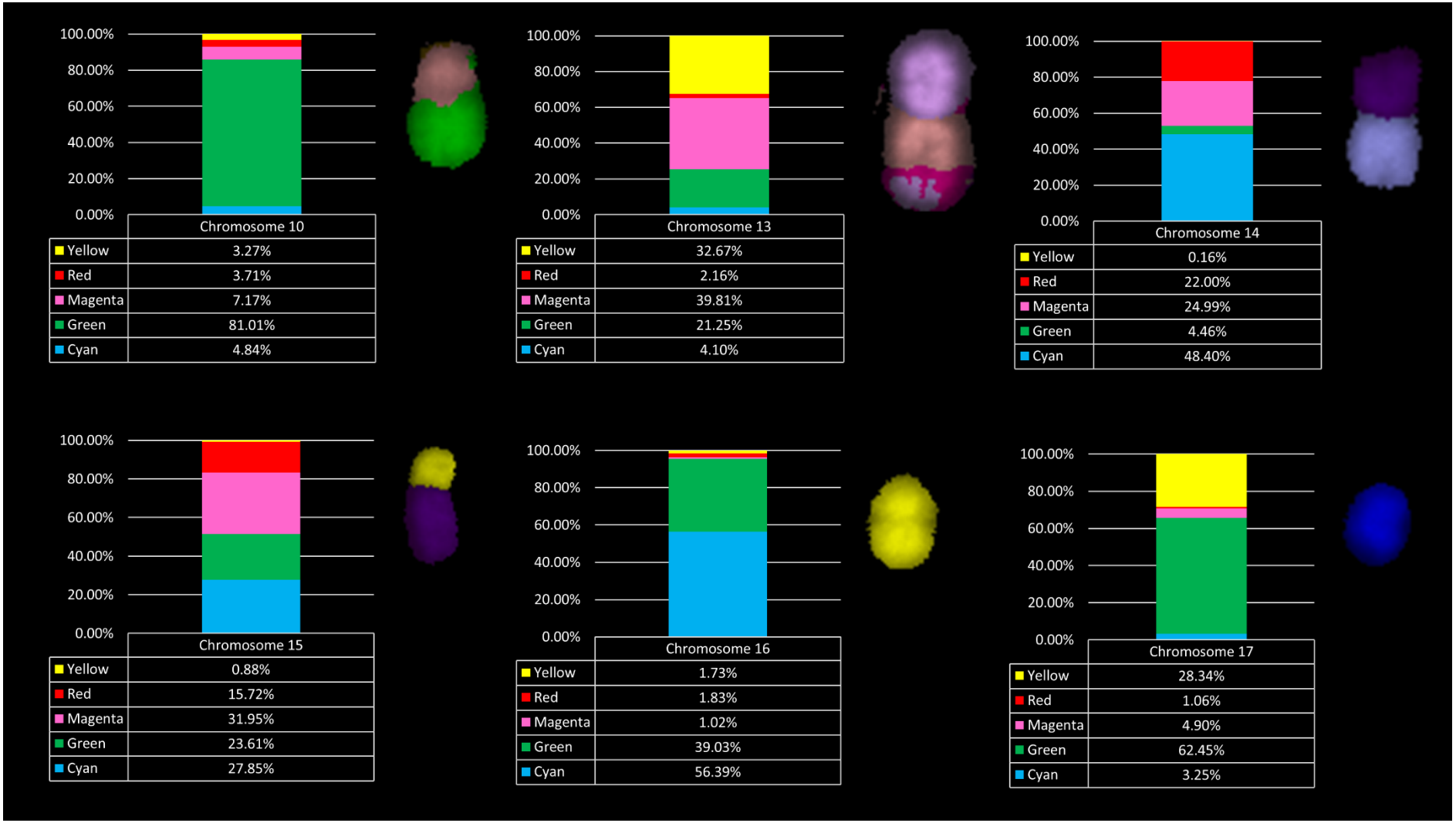


Figure 6.7. Fluorescent intensity colour profiles of CHOK1a-GS-KO chromosomes 10, 13, 14, 15, 16 and 17.

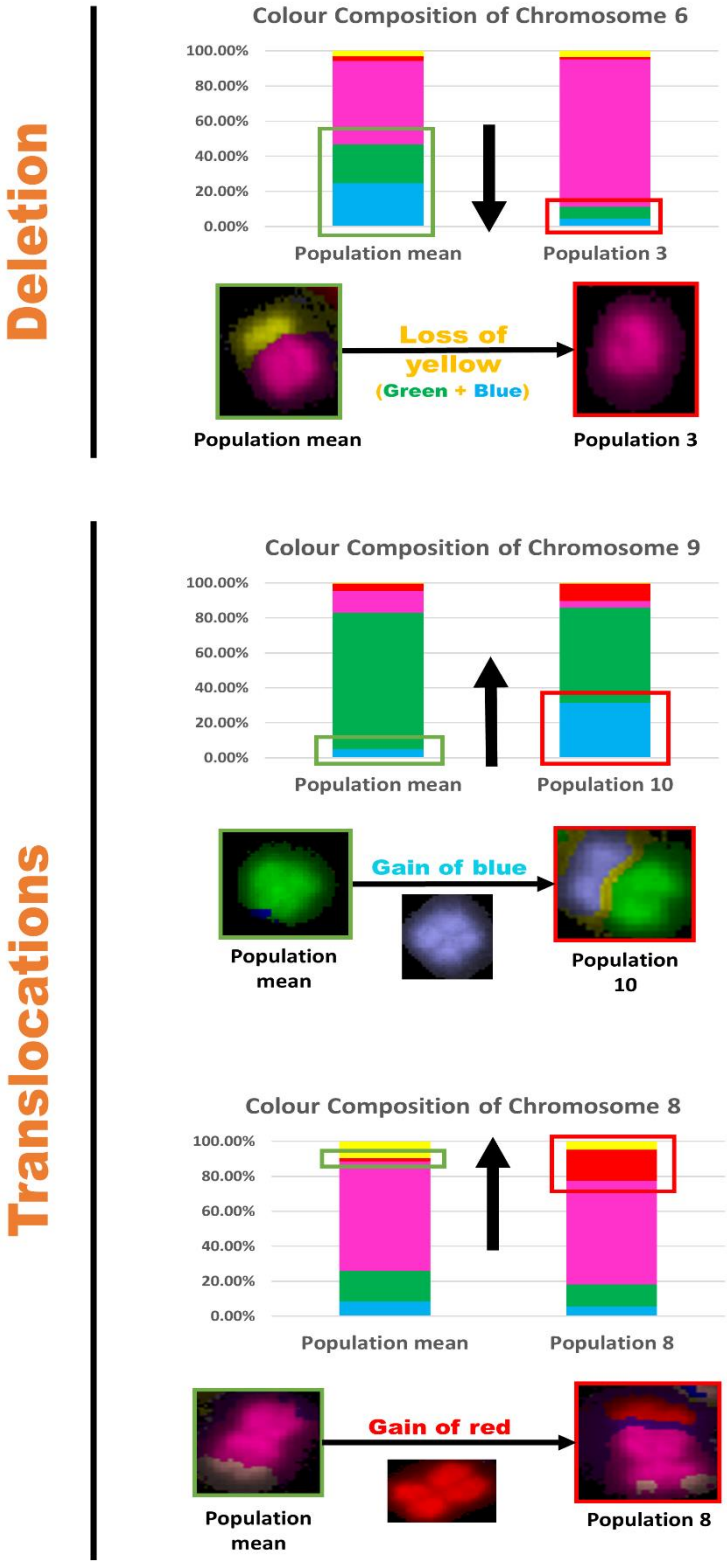


Figure 6.8. Examples of different mutation confirmations using cell profiler extracted fluorescent intensity profiles. Loss (deletion) or gain (translocation) of genetic material is reflected in the decrease or increase of corresponding percent fluorescent intensity, respectively.

Although the semi-automated cell profiler workflow provides an objective means to profile a chromosomal mutation observed during MFISH karyotype analysis, the workflow can still be laborious due to manual editing of each individual image and post analysis processing of fluorescent intensity data. With the current rise in interest in artificial intelligence and machine learning algorithms, most famously known for Alan Turing's 'learning machine' implemented during the second world war, a collaboration was established with GSK's Artificial Intelligence and Machine Learning (AI/ML) group in a bid to fully automate the whole stability prediction manual process, removing potential analyst bias, subjectivity and to decrease overall analysis timelines.

Faithfully segmenting chromosomes within images is the critical step of the whole automated pipeline. Including any artefacts within the image or obtaining masks that contain two or more chromosomes will provide misleading output results. To this end, a uNet machine learning algorithm (https://arxiv.org/abs/1505.04597), which was first created to learn disease state image characteristics on a small set of patient images, was applied.

A machine learning (ML) algorithm produces its optimum results when there are large annotated training data sets available, to refine the algorithm upon. Annotating a large enough data set for standard ML algorithms can often be highly laborious. To decrease time spent on annotating, uNet takes each annotated image and applies a contracting and expansive path of image edits (such as rotating the image 90^{o}). This provides a greater number of images that can be trained on compared to the original data set. Upon training the uNet algorithm with semi-automated and manually annotated masks of DAPI images of chromosomes (using the cell profiler workflow), masks of images were generated automatically without analyst intervention and artefacts (such as cells) were ignored naturally by the algorithm.

However, the algorithm had trouble segmenting chromosomes that were in close proximity and there were several instances of groups of two or more chromosomes being considered as a single entity within the mask. To resolve this issue, masks were edited using the edit objects manually module (cell profiler) and the algorithm retrained on these newly edited images. Additionally, a weight map was created within the ML algorithm to aid its learning of small separation boarders between chromosomes in close proximity. Upon the second iteration, chromosomes were faithfully segmented in the majority of images analysed (figure 6.9a1 and a2, insert image).

Chromosomes are 'painted' using a colour coding system that is built into the proprietary Metasystems software. As there can be different binding efficiencies of fluorescent probes, when comparing across different slides or different probe batches, it was noticed that the fluorescent intensities for each channel were not identical for each set of images, thus having an impact downstream of the segmentation protocol.

As there was no access to Metasystems code for pseudo colouring ('painting') chromosomes, a 12-component gaussian mixture model was applied to analyse fluorescent pixel intensities of each channel, based on metasystems colour coding system (figure 6.10). Chromosomes are clustered by their fluorescent intensity pixel values in a 5-dimension greyscale space and based on the pixel profile, described by an 11-dimension (DAPI channel not considered) vector where the *i*th element is the proportion of the chromosome with pseudo colour *i*. This culminates in a pseudo colour being assigned to each chromosome, based on the relative pixel colour intensities of one chromosome compared to the others within the image (figure 6.9b1 and b2). Our pseudo colouring model replicates the pseudo colouring performed by metasystems, which can be observed by comparing pseudo colours in figure 6.9b1 and b2 to figures 63 and 64.

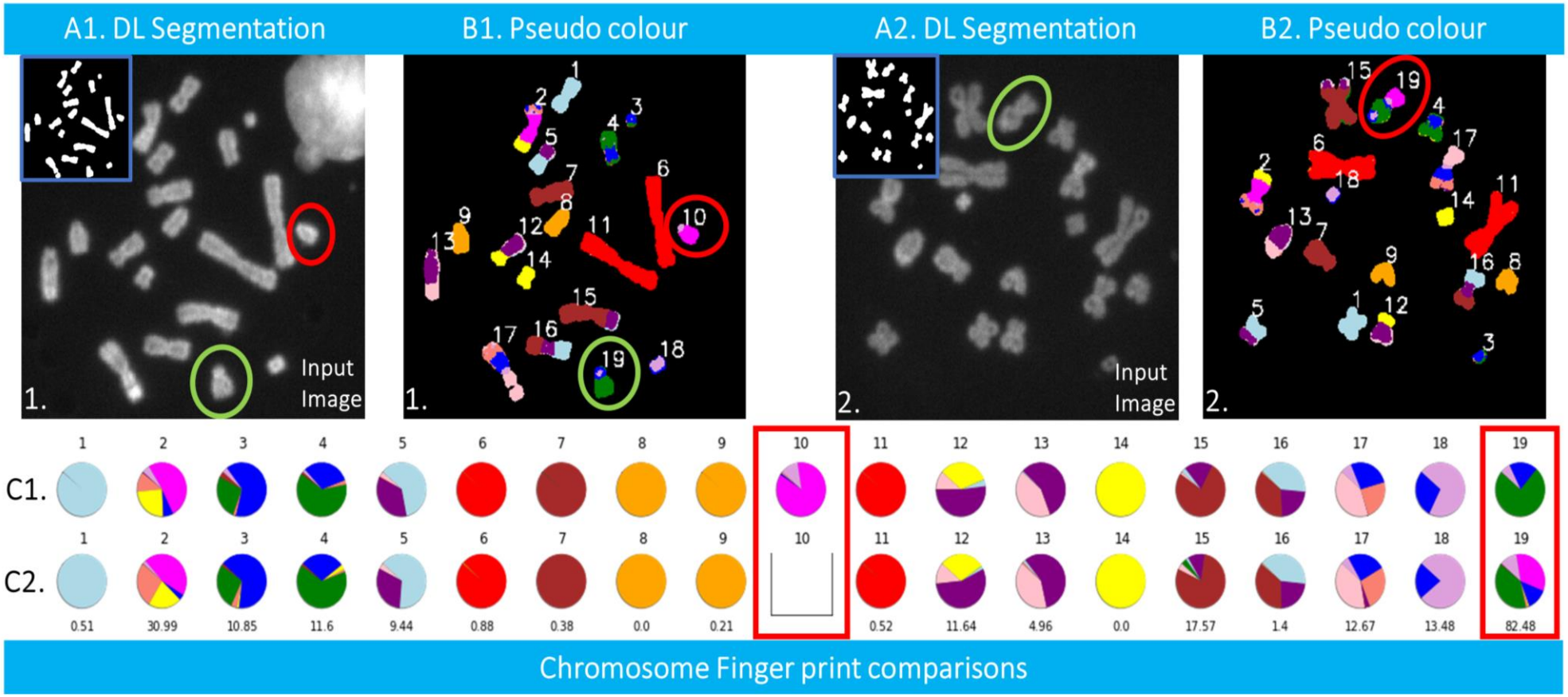


Figure 6.9. Schematic overview of automated prediction workflow. A1 and A2) Automated image segmentation using uNet ML algorithm. Faithful segmentation of chromosomes allows for robust pseudo colouring using a gaussian mixture model (B1 and B2). C1 and C2) Pairwise linear assignment of chromosomes, a translocation of 10 and 19 is detected by the algorithm, returning a large matching cost.

After the images have been segmented and pseudo colouring applied, individual chromosomes undergo a pairwise-linear assignment, where a single chromosome of image 1 is compared to all chromosomes of image 2 (figure 6.9c1 and c2 and figure 6.11). This is repeated for all chromosomes in image 1 and generates a 19x19 matching cost matrix. A matching cost simply describes the value of subtracting one chromosome from another, a small matching cost represents similarity in the colour profile (genomic similarity) and a large matching cost represents genetic dissimilarity (figure 6.11). The algorithm returns the smallest cost matching value from the cost matrix, equating to the chromosomes most likely match.

An example of automated mutation detection is described in figure 6.9. Chromosomes assigned number 10 and 19 are shown to be separate within image 1 (a1 and b1, green and red circle). Within image 2, these chromosomes have undergone a translocation event, which can be confirmed using DAPI channel and pseudo coloured image (a2 and b2, green and red circle). Upon performing the pairwise linear assignment (C1 = image 1 and C2 = image 2), no match can be found for chromosome 10 (as it is not present in image 2) and chromosome 19 has been matched to the mutated chromosome, however with a large matching cost of 82.48. To put this value into context, two chromosomes with genetic similarity (number 6, red chromosome) has a matching cost of 0.88. Therefore, a matching cost threshold can be applied to quickly identify mutations in large image sets (e.g. >50 matching cost = a mutation).

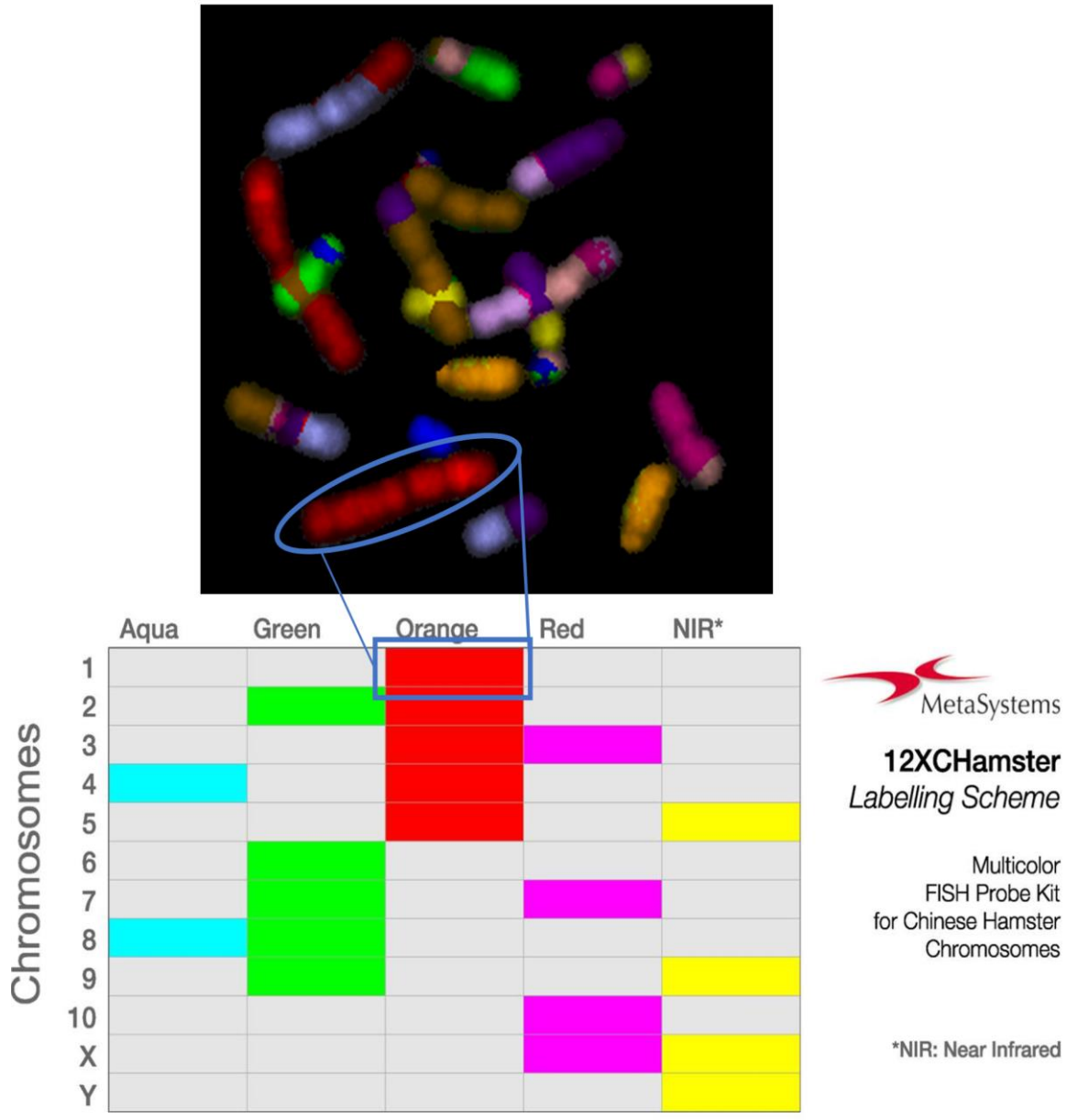


Figure 6.10. Metasystems 12XCHamster labelling scheme, used to dictate 12D gaussian mixture model for pseudo colouring chromosomes within the stability prediction workflow.

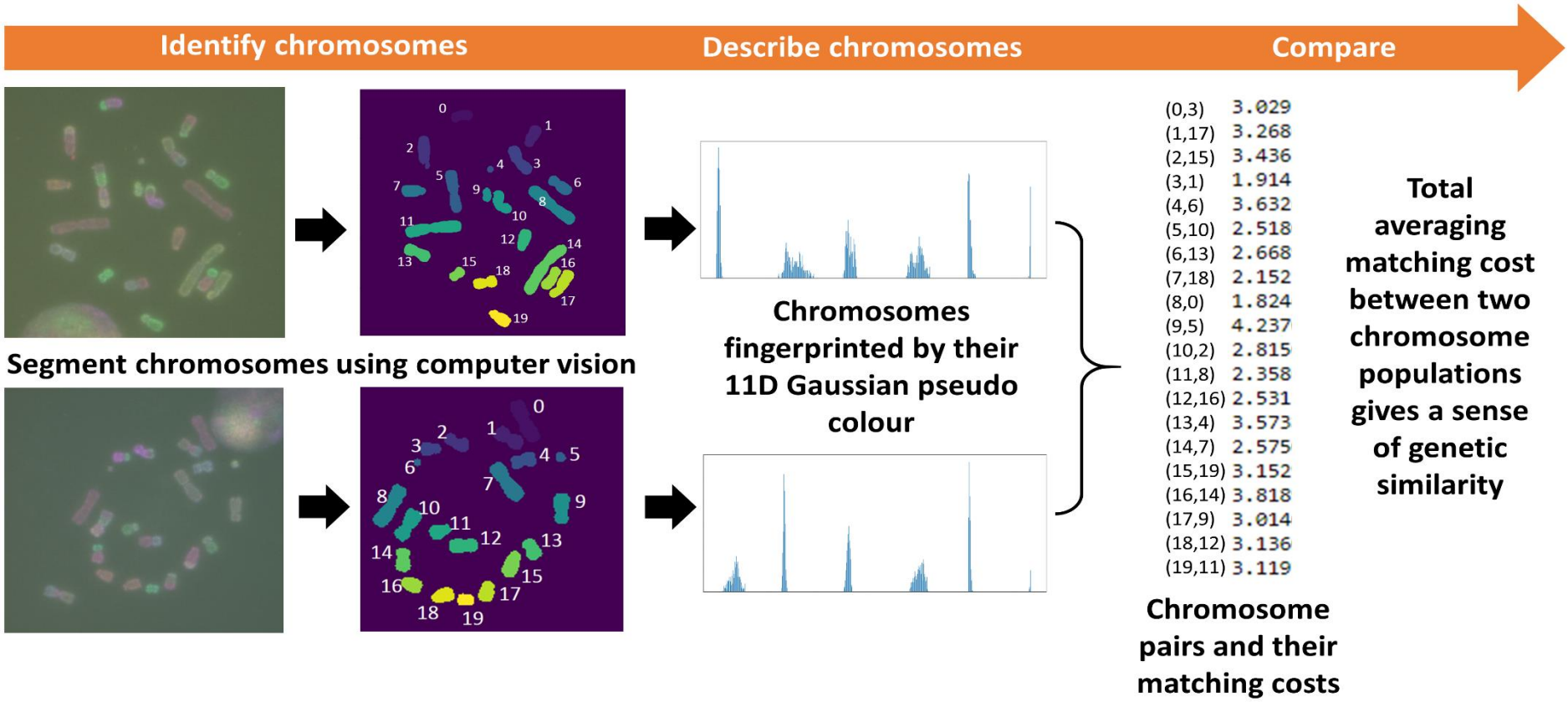


Figure 6.11. Pair-wise linear assignment workflow. Chromosomes are identified using segmentation algorithm. After pseudo colouring, chromosome pixel fluorescent intensity profile of each image is subtracted from one another (describe chromosomes), creating a cost matching matrix of the smallest returnable value (i.e. the closest chromosome match or genetic similarity). Comparing the overall matching costs within a cell line provides an overall sense of the number of mutations which are present within the sample, allowing comparison of CCA and NCCA populations for production stability prediction.

To validate the automated prediction workflow (APW), images used in manual MFISH analyses were analysed through the APW algorithm and data was compared against manual methods. APW identification of CCA and NCCA populations were largely consistent compared to manual method (figure 6.12a). Unstable cell lines obtained a larger proportion of NCCA populations compared to their stable counterparts, as observed with manual analysis. Comparing CCA and NCCA frequency showed a significant difference between stable and unstable groups, as observed in manual analyses (figure 6.12b, pooled T-test, P=<0.05). Upon comparing CCA% and standard deviation (SD) of matching cost, it was observed there was a distinct separation between stable and unstable cell lines based on their matching cost SD, indicating SD could be used as another genomic instability metric analogous to CCA and NCCA designated populations (figure 6.12c).

Sample preparation and analysis timelines for manual and automated methods are outlined in table 6.1. Performing the manual MFISH karyotyping 40 images per cell line for 48 cell lines culminates in a total analysis time (minus sample preparations) of 159 hours. In contrast to APW that can complete the same analysis in 1.3 hours, a time saving of ~157 hours for researchers.

Due to the highly laborious nature of the manual analysis, 40 images per sample were analysed. Although throughout this thesis the same pattern of CCA and NCCA frequencies for stable and unstable cell lines has been observed, there remains a question whether 40 images (40 cells) truly reflects a culture flask that contains millions of cells. APW analysis time savings provides the means to increase images analysed from 40 to 200-400 images per cell line, providing a greater in-depth characterisation of the cell culture flask. APW provides an upscaled (200 images per cell line, 48 cell lines) analysis time saving of 32.9 days, providing an industrialised algorithm that could be integrated into CLD's critical path, without impacting project timeline

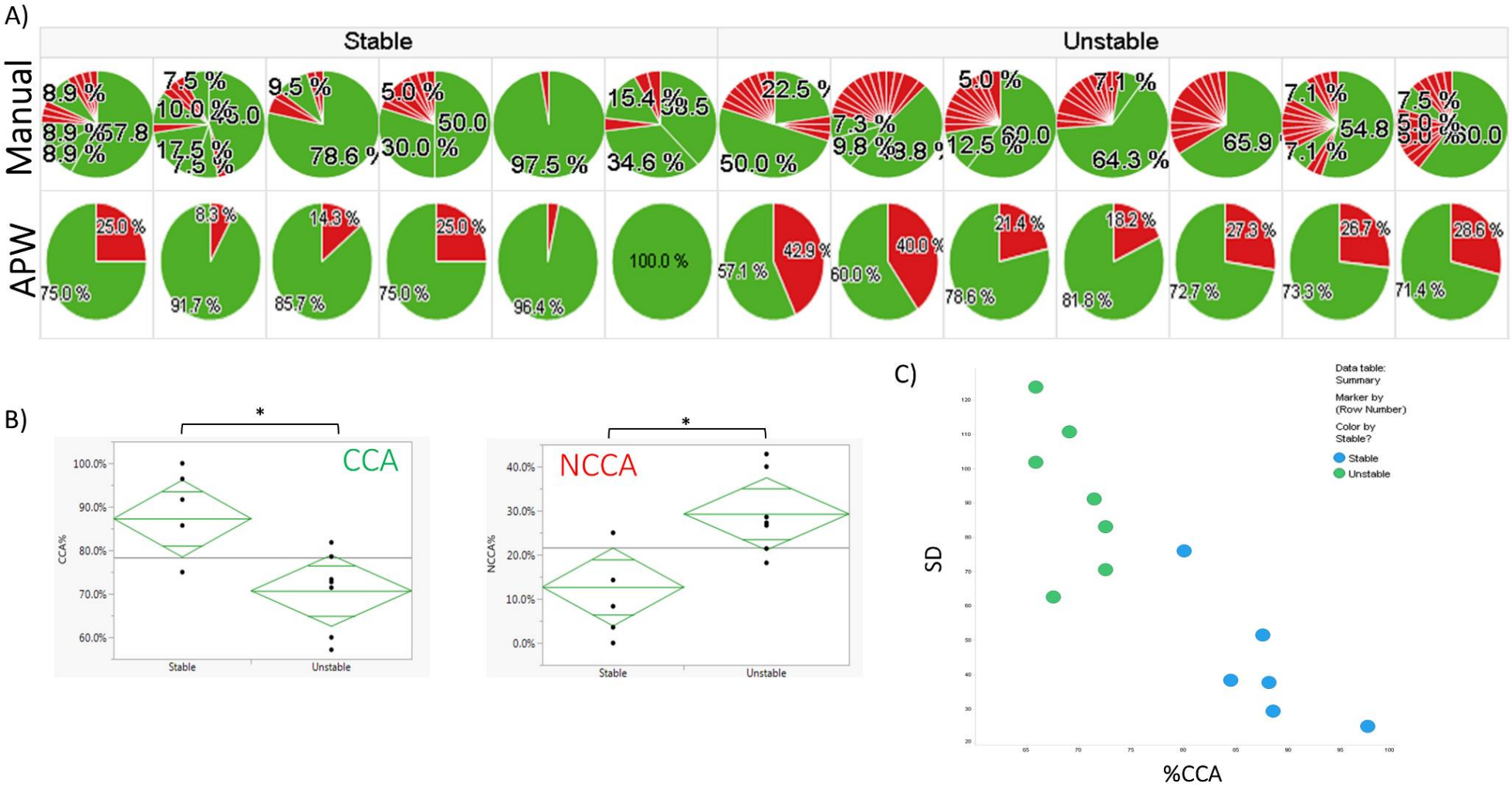

A)
Stable
Unstable
Manual
APW
B)
CCA
NCCA
CCA%
NCCA%
Stable
Unstable
C)
Data table: Summary
Marker by (Row Number)
Color by Stable?
Stable
Unstable
SD
%CCA

Figure 6.12. Comparison of manual and automated workflow karyotype heterogeneity identification. A) Comparison of manual and automated (APW) calculated CCA and NCCA subpopulations showing similar profiles in CCA and NCCA proportions within each cell line. B) Comparison of %CCA and %NCCA generated by the automated prediction workflow showing a distinct separation between stable and unstable cell lines, as observed in manual analyses (P=<0.05). C) Dot plot depicting correlation between cost matching standard deviation indicating SD could be used as a genetic instability marker (Large SD = increased instances of large matching costs = greater number of mutations).

Table 6.1. Comparison of manual and automated analysis timelines. Automation of cell profiler semi-automated workflow provides a potential 32.9 days analysis time saving for upscaled image analysis.

Stage	Manual			Automated		
	Per Round (Hours)	Total for 48 Cell Lines (Hours, 40 images per cell line)	Total for 48 Cell Lines (Upscaled analysis – 200 images per cell line)	Per Round (Hours)	Total for 48 Cell Lines (Hours, 40 images)	Total for 48 Cell Lines (Upscaled analysis – 200 images per cell line)
Slide Prep & Start of M-FISH	24	144	144	24	144	144
M-FISH Completion & Acquisition	24	144	144	24	144	144
Select Images & Cell-Profiler semi-automated segmentation	16	96	480	0.22	1.3	6.5
Pseudo Colour & Export	4	24	120			
Population Determination (including CCA/NCCA calculations)	6	36	180			
Mutation Confirmation	0.5	3	15			
Total	74.5	447	1083	48.2	289.3	294.5
Total analysis (minus slide preparation & imaging)	26.5	159	795	0.22	1.3	6.5
	Total saving (hours)			26.3	157.7	788.5
	Total saving (days)			1.1	6.6	32.9

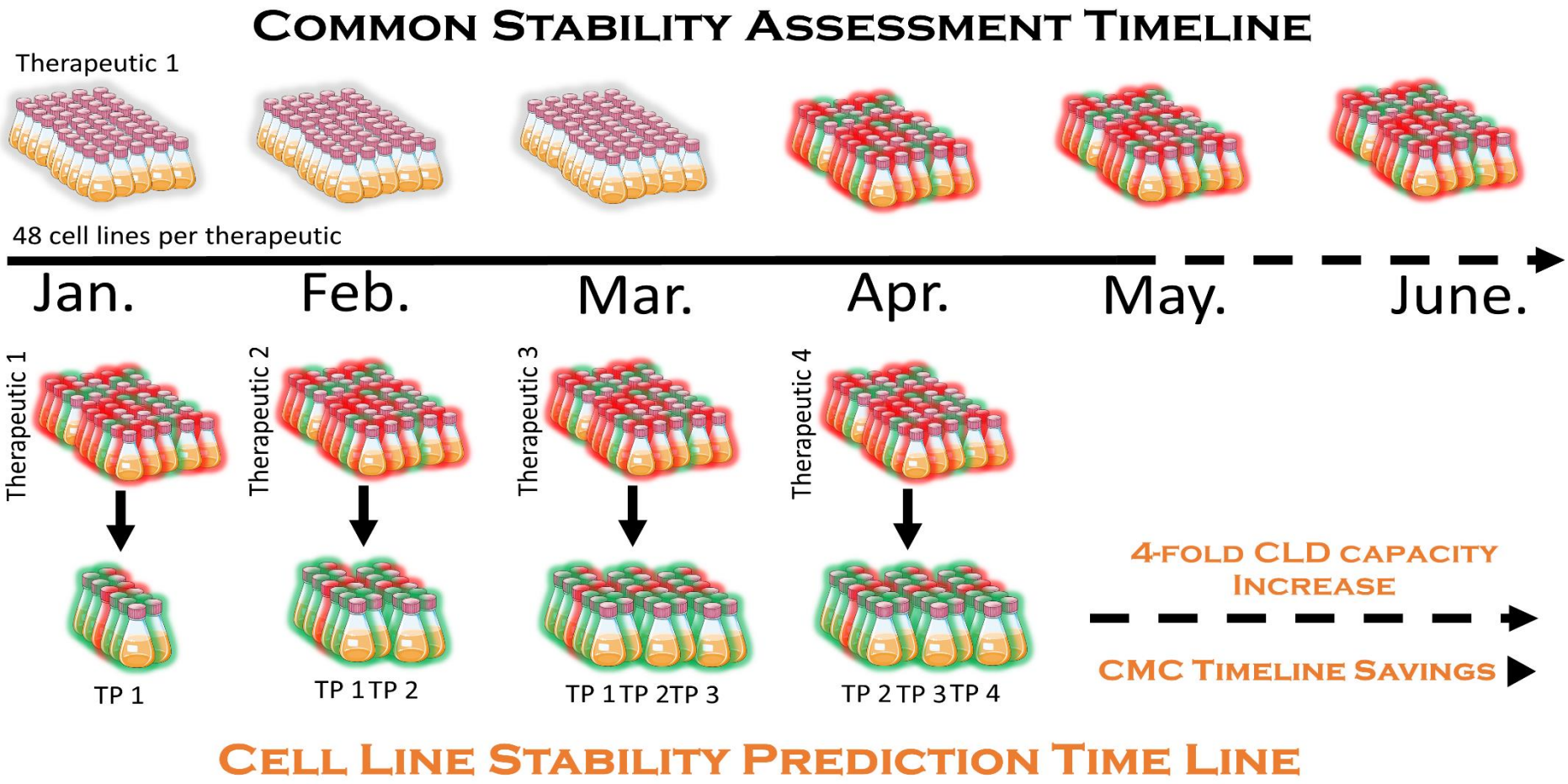


Figure 6.13. Potential cell line development stability assessment standard of procedure after integration of automated cell line stability prediction workflow into CLD critical path. Early triaging of cell lines allows for multiple therapeutic proteins to be assessed at one time, allowing four therapeutic proteins to be assessed within a 7-month period compared to 16 months in the current sequential stability assessment format.

Upon integration into CLD's critical path, I envisage the APW to be utilised as an early cell line triaging method (figure 6.13). A standard stability assessment requires 48 cell lines, belonging to a single therapeutic protein, which is cultured from 4 to 6 months before the cell lines production stability is identified. Through performing a stability prediction on a blinded panel of cell lines, it was observed that the prediction workflow obtained greater correct prediction results for unstable cell lines (80-100%, figure 4.5).

Using this method to triage unstable cell lines would provide an enrichment of stable cell lines after one month, reducing the number of cell lines that are subjected to the full stability assessment to 12 cell lines per therapeutic protein. Therefore, four therapeutic proteins could have their stability assessed in a single stability run, in a 7-month period. In the current sequential format (1 therapeutic protein, 48 cell lines, 4-6 months per protein), it would take 16 months to assess four therapeutic protein cell lines stabilities. Thus, implementing APW could lead to a 4-fold increase in CLD capacity and savings on Chemistry, Manufacturing and Controls timelines.

Chapter 7: Discussion

Introduction (7.1)

Chinese hamster ovary (CHO) cell lines have gradually become the production system of choice for therapeutic protein production. Originally, they were selected as auxotroph's after their initial identification by Puck and colleagues (Kao and Puck, 1968; Puck et al., 1958) and were adopted for therapeutic protein production due to their relative safety profile, ease of exogenous DNA uptake and their ability to confer mammalian post-translational glycosylation profiles (Bandaranayake and Almo, 2014; Butler, 2005).

Through constant iterations of media and process optimisations CHO cell lines can be grown at great cellular densities (>10x10^6 cells/ml) and can produce ~4g/L of therapeutic protein during a manufacturing run (Wurm, 2004). However, cell line production stability is still a major concern for pharmaceutical companies as roughly 50-60% of cell lines are considered productionally unstable and cannot maintain consistent titre production for the duration of the manufacturing window (in house data, ~4 months).

Due to the inefficiency of identifying stable cell lines, pharmaceutical companies have to produce and analyse a large number of cell lines across a 4-6-month stability assessment window, before confirming cell line stability. This represents a major bottleneck in cell line development critical path timelines and solutions to alleviate these bottlenecks are being voraciously investigated.

There are numerous potential pathways currently being investigated that could contribute to the production instability of a cell. Some areas that have been explored are IgG secretary pathway bottle necks (Kaneyoshi et al., 2019), different plasmid integration strategies (e.g. piggy back transposon mediated integration (Matasci et al., 2011)) and epigenetic regulation of production during batch culture (Hernandez et al., 2019), all showing varied success in alleviating the production instability phenotype, albeit in cell models that do not fully recapitulate CHO cell lines use in industry. Despite having numerous avenues explored, there is still no consensus explanation as to the causative pathway of the production instability phenotype.

CHO cell line Chromosomal heterogeneity, the potential impact of telomere specific DNA damage and protein production instability (7.2).

CHOK1a-GS-KO host cell line karyotypic heterogeneity (7.2.1)

A potential explanation for lack of consensus may be found in the vast heterogeneity observed at the chromosomal and transcriptomic level between CHO host and producer cell lines (Puck, 1964; Singh et al., 2018; Vcelar et al., 2018a; Vcelar et al., 2018b; Worton et al., 1977). Yusufi and colleagues (Yusufi et al., 2017) performed an in-depth characterisation of a single transfection event within a CHO cell line and observed ~49% of predicted genes (Xu et al., 2011) to obtain a mutation due to the transfection event. Whether cells in transfection bulks obtain different profiles of mutations, leading to genetic and transcriptomic dissimilarity upon single cell sorting, remains unknown.

Such heterogeneity witnessed in a single transfection event may be exacerbated further within an industry setting. Numerous transfections and cell line scale ups occur on a regular basis, compounding heterogeneity across all cell lines producing therapeutic proteins. With this in mind, CHOK1a-GS-KO host and CHOK1a-GS-KO based producer cell lines were characterised at a high level, focusing on chromosomal mutation profiles, telomere homeostasis and DNA damage quantification. I hypothesise that heterogeneity observed across different CHO cell lines at the transcriptomic level (Singh et al., 2018) may be a result of chromosomal mutations that occur due to improper protection of interstitial telomeric sequences (ITS) and a resulting elevated DNA damage response at these sites, leads to gross chromosomal rearrangements.

Chromosomal instability has been documented in CHO, through observations of varying chromosomal number and G-banding patterns (Deaven and Petersen, 1973; Yusufi et al., 2017), to gross chromosomal mutations identified by multi-colour fluorescent in-situ hybridisation (MFISH, chromosome 'painting') (Auer et al., 2018; Vcelar et al., 2018a; Vcelar et al., 2018b). MFISH provides the means to visualise large structural variants and balanced translocations and to our knowledge this technique has only been used in CHO cell lines by ourselves and Nicola Borth's group (Auer et al., 2018; Vcelar et al., 2018a; Vcelar et al., 2018b). Their observations coincide with our own observations in that the CHOK1a based host cell lines obtains a modal chromosome number of 19 rather than the original 22 chromosomes that is present in the Chinese hamster (figure 3.2, (Vcelar et al., 2018b)).

Although CHO cell lines obtain similar modal chromosome numbers, suggesting genetic stability, once MFISH probes are applied to said chromosomes, a very different picture emerges. Upon applying MFISH probes to the CHOK1a-GS-KO host, cells within the suspension culture already obtained multiple gross chromosomal rearrangement at an early time point (~20 generations). Vcelar et al (Vcelar et al., 2018b) observations indicate that subcloning in multiple rounds does not prevent cell lines from acquiring heterogenous populations during the subsequent scale up phase. Indicating that routine culturing of our CHOK1a-GS-KO host leads to karyotypic heterogeneity, in spite of the host being single cell sorted.

From our own observations, CHOK1a-GS-KO host cell line obtained 18 distinct karyotypes at an early time point. The number of populations reduced to 16 after a 6-month culturing period, with 6 out of the original 18 populations still remaining present within the culturing flask. A novel karyotype (population 4, figure 3.4) became the second greatest population due to obtaining a favourable mutation in chromosome 6 that allowed it to establish itself as a dominant

population. It should be noted that the decrease in population size may be an artefact of the number of images (cells) analysed per sample, rather than an actual loss of karyotype.

CHOK1a-GS-KO telomere and DNA damage characterisation (7.2.2)

To investigate a potential causative factor for the genomic heterogeneity witnessed within the host, interstitial telomere sequences (ITS), shown to exist in large intra-chromosomal blocks by ourselves and others (Krutilina et al., 2001; Smilenov et al., 1998), were characterised to assess their relative length, expressed as a proportion to DAPI staining (DNA content). Dysfunctional telomeres have been highlighted as a potential source of genome instability in cancer cell lines (O'Sullivan and Karlseder, 2010), which show similar characteristics in proliferative capacity and genomic instability compared to CHO cell lines.

At the time of writing, there have been no publications profiling CHOK1a based cell line telomere length. Therefore, before investigating comparisons of telomere length in a panel of productionally stable and unstable cell lines, the host cell line was profiled for its relative telomere length across a 6-month maintenance culturing period. Relative telomere length increased from 2.8% to 8.9% after 6-months (figure 3.3, P=<0.0001). This indicates potential ITS amplification during routine culture which has been implicated as a result of genetic instability (Aksenova and Mirkin, 2019; Bolzan, 2012; Mondello et al., 2000). Amplifications could be a result of gross chromosomal rearrangements that has been observed by ourselves and others (human cancer models, (Artandi and DePinho, 2010; Murnane, 2010). However, a knock-out experiment of telomere repeats binding factor 1 (TRF1) or TRF2, which has been shown to cause telomere loss in human cellular models

(Smogorzewska et al., 2000), would be required to confirm telomere lengths role in genetic instability within CHOK1a based cell lines.

The existence of TRF1/2 in CHO cell lines was first described by Smilenov et al., (Smilenov et al., 1998), who observed punctate staining of TRF1 during interphase but not during metaphase, without quantification. TRF2 and Telomere co-localisation quantification within CHOK1a-GS-KO host provides a sense of the extent of telomere protection by shelterin, although TRF2 bound to telomere sequences is not indicative of full shelterin complex binding. Here, it was observed that an average of 6 TRF2 and telomere co-localisation foci per cell compared to 1.85 telomere only foci, at an early time point. TRF2 and telomere co-localisation was observed to decrease to 4.75 with an increase of telomere only foci (2.2) after 6 months of culturing. However, these observations were shown to be insignificant (pooled T-test, P=>0.05). It was noticed that there were instances of TRF2 only foci within images, suggesting there is free TRF2 that exists within the nucleus. This has also been observed by another group (Zhang et al., 2004b), who suggest that TRF2 resides within the nucleolus of human cells.

To further explore whether telomere sequences operate as potential DNA damage hot spots within the host, gH2AX, a core histone protein that is phosphorylated at serine 139 upon double strand breaks (Rogakou et al., 1998), was characterised by immunofluorescence foci quantification. Neocarzinostatin, a chromoprotein that creates DNA damage through abstracting hydrogen atoms from the deoxyribose sugar mediated by thiol nucleophilic attack (Kohnlein and Jung, 1982), was used as a positive control to elicit DNA damage and assess whether the foci quantification assay could detect changes in DNA damage. Although gH2AX foci only and telomere specific damage (TIF) was shown to be significantly elevated, the host cell line obtained relatively low levels of

gH2AX and TIF foci that did not fluctuate across 6-months of routine culture.

Do productionally stable and unstable cell lines obtain similar traits to CHOK1a-GS-KO host? (7.2.3)

CHOK1a-GS-KO host has been characterised to establish baseline observations that can be used as a reference point when analysing therapeutic protein producing cell lines. A panel of 9 productionally stable and 9 unstable cell lines were selected across three therapeutic proteins, to replicate industry cell line diversity. Stability is defined here as being able to produce the same level of titre within a +/- 30% max titre loss threshold, across a 4-6-month production window. I hypothesised that cell lines producing therapeutic protein may encounter higher levels of stress due to constitutive production of protein, potentially impacting DNA damage levels and MFISH karyotype heterogeneity.

Chromosome number distribution of CHOK1a-GS-KO producer cell lines was quantified as previously described. An interesting observation identified here, that was not observed within Vcelar et al (Vcelar et al., 2018b), is instances of CHOK1a-GS-KO based producer cell lines obtaining a “diploid” karyotype after single cell sorting. 3 out of 4 cell lines that obtained a “diploid” karyotype were considered productionally stable and in one instance, protein 3 cell line 7 was shown to adopt a “diploid” karyotype (modal chromosome number of 19 to 32) over a 6-month culturing period (figure 4.4).

Interestingly, no population containing 32 chromosomes was observed at the early timepoint, suggesting the modal chromosome population arose from a de novo mutation. Our observations may be more representative of CHO cell lines within an industry setting, as cell lines across three different therapeutic proteins have been assessed, compared to Vcelar et al's analysis of single cell sorted cell lines from the same parent. Thus, indicating that different transfection events may have influenced the cells ability to acquire a "diploid" karyotype or provide a favourable genetic mutation that arms "diploid" cells with a competitive advantage within the culture flask.

Relative telomere proportion was calculated in stable and unstable cell lines across a 6-month culturing period, to assess whether there was a difference in telomere length between stable and unstable groups. This experiment was designed to assess whether unstable cell lines obtained a greater proportion of telomere sequence, that may lead to an increase in TIFs, causing karyotypic heterogeneity. A least square means (LSM) model was applied to relative telomere length to consider covariates such as time point and chromosome number. Stable cell lines obtained an average of telomere proportion of 2% that was shown to be elevated to 2.3% in unstable cell lines. Although this difference was deemed significant (figure 4.2c, page 149, P=<0.0001), it is likely that the large number of analysed images has increased statistical sensitivity to small changes within the mean. One has to question whether there is biological relevance of a 0.3% increase or if this value is simply variation within the data. Interestingly, there was no increase observed over a 6-month culturing period, unlike the host.

Does genomic stability impact production stability? (7.2.4)

Subsequently, stable and unstable panel of cell lines were characterised using chromosome 'painting' to assess each cell lines heterogeneity and to investigate whether there are commonalities across each group that may indicate a potential cause of the production instability phenotype. Karyotype populations were defined by analysing each image and providing newly observed karyotypes, which obtain a novel mutation, a new population ID. Based on the population frequency within the sample they are designated a clonal chromosomal aberration (CCA, >5%) and non-clonal chromosomal aberration (NCCA, =<5%), which describes the rarity of mutation compared to the cells analysed.

Results described here indicate a correlation between high CCA% and low NCCA% frequency to productionally stable cell lines and vice versa for productionally unstable cell lines. Applying this methodology to a blinded panel of cell lines a frequency of >78% CCA, a threshold determined using the previously analysed stable and unstable panel of cell lines (figure 4.5), predicts productionally stable cell lines with 67.5% efficiency. Moreover, <=78% CCA predicts productionally unstable cell lines with 80-100% efficiency.

One set of cell lines producing therapeutic protein 1 did not obtain good prediction results, but this may be due to the relatively low generation number compared to the other cell lines analysed (~10 generations vs. ~20 generations). To confirm this hypothesised explanations, further work has to be performed. MFISH populations should be analysed over a time course series spanning 10-20 generations, to investigate whether an additional 10 generations are an appropriate window for cell lines to adopt their stabilised populational make up. If this does not confirm the initial observation, perhaps the poor

prediction result can be attributed to therapeutic protein specific variation.

In a similar study, Vcelar et al (Vcelar et al., 2018b) profile chromosome number distribution and mutations in subcloned cell lines, generated from the same working cell bank. They make some interesting observations in that modal chromosome number remains consistent at 19 chromosomes across all analysed cell lines. However, mutation profiling using MFISH shows that these cell lines are karyotypically heterogenous, but they do not go further to assess karyotypic heterogeneity and production stability correlation. Their results highlight that subcloning cell lines does not improve downstream production stability and they indicate multiple rounds of enrichment sorting, based on GFP fluorescence co-expressed with cluster of differentiation 4 (CD4) as a means to produce cell lines that maintain production stability. However, their observation was not confirmed using cell lines producing full sized monoclonal antibodies (mAbs) that are around 123 kDa larger (~150kDa mAb – 27kDa GFP), which may not recapitulate the protein production kinetics and resulting cellular stress exerted within mAb producing cell lines.

To our knowledge, results presented here are the first study that correlates a genetic instability marker (CCA/NCCA frequency) with the production instability phenotype. Moreover, to our knowledge this study is the first to test a novel finding across a panel of cell lines that replicates the vast heterogeneity of CHO cell lines in an industry setting. This was achieved by testing the MFISH based prediction method against 60 cell lines across 5 different therapeutic proteins. In light of these results, a patent application has been submitted based on the work presented here.

Although our data show a correlation between production instability and genetic instability, there were observations of cell lines that do not follow the general trend. During analysis, it was observed cell lines that contained a diploid number of chromosomes did not show similar CCA and NCCA frequencies compared to cell lines containing a “haploid” chromosome number (19). In one instance, Protein 3:Cell line 7 (P3:C7), a stable producer cell line, saw a modal chromosome number shift from 19 to 32. Upon analysing CCA and NCCA populations of late diploid P3:C7, not a single CCA population was observed, meaning each population analysed was karyotypically distinct. This indicates that with double the genetic content, CHO producing cell lines have the ability to produce consistent therapeutic protein titres regardless of genetic instability. Although, this may not be attributed to increase in copy number as there are multiple publications that do not observe such trend (Ley et al., 2015; Noh et al., 2018).

Characterisation of pathways that may be associated with genomic heterogeneity in productionally stable and unstable cell lines (7.2.5)

Although karyotypic heterogeneity within CHO producing cell lines has been clearly demonstrated by ourselves and others (Vcelar et al., 2018a; Vcelar et al., 2018b), a definitive link between genetic heterogeneity and production stability is yet to be established. Here, an intracellular IgG staining method was applied to samples to assess cell line fluorescence intensity by flow cytometry. Upon grouping fluorescent peaks into stable and unstable producer cell lines, stable cell lines were observed to have a consistent fluorescent peak, whilst unstable cell lines had a wider spread of fluorescence. However, there was a discrepancy in Bartlett's test of variance, which tests variance statistically in fluorescence peaks between stable and unstable cell lines. Dependent on performing the test on the whole data set (figure 4.7, P=>0.05), or performing the test removing an outlier (~17 SD from stable group mean, P=<0.0001) there was a discrepancy between obtained results.

This observation provides an interesting avenue that could be explored further, as the discrepancy in results is influenced by the relatively small sample size, requiring further samples to be analysed to confirm or disprove the original observation. If confirmed, this could represent a link between genomic instability, correlating karyotypic heterogeneity and heterogenous intracellular heterogeneity of IgG production between cell lines, which cannot be observed from max titre reads during production stability assessments.

To elucidate a potential mechanism that creates the genomic heterogeneity witnessed within CHO cell lines, cellular DNA damage, telomere protection and telomere specific damage was investigated within stable and unstable panel of cell lines. CHO host cell lines have previously been used to elucidate DNA damage pathways upon acute DNA damage through boron neutron capture (Kinashi et al., 2011) and ionizing radiation (Matsuya et al., 2014; Rothkamm et al., 2003). However, there are no publications that look to characterise DNA damage levels during therapeutic protein production.

gH2AX foci and TIF foci quantification was shown to detect increases in DNA damage at DNA and telomeric sites through the use of Neocarzinostatin, however, there was no significant differences between production stability groups and overall gH2AX foci and TIF foci remained at similar levels compared to the host. To confirm these observations, gH2AX quantification was analysed using flow cytometry using a different antibody. Although raw values of each analogous experiment cannot be directly compared, due to different antibody composition, relationships between the host, stable and unstable cell line groups can be assessed. Again, no significant difference was observed between groups and by using flow cytometry analysis, a baseline of 0.92-4.63% gH2AX positive cells observed across all cell lines in routine maintenance culture, regardless of stability, has been established within this thesis.

Low levels of TIF foci may be due to protected interstitial telomeric sequences (ITS), conferred by shelterin complex. To assess telomere protection, TRF2 and telomere co-localisation was performed on stable and unstable cell lines, as performed with CHOK1a-GS-KO host. Interestingly, stable cell lines obtained a significant increase in TRF2 and telomere co-localised foci, indicating protection by shelterin, compared to the host cell line (figure 4.13, P=<0.01). This may suggest that TRF2 confers production stability through telomeric protection. However, the observed increase in co-localisation was not corroborated by an increase

in telomere only foci (unprotected telomeres) in the host, nor was the comparison between stable and unstable cell lines for TRF2 and telomere co-localisation found statistically significant. This indicates that the initial observation may be an artefact of the analysis itself. Additionally, a caveat of this assay is that co-localisation of signals is not indicative of protein interaction (a chromatin precipitation (ChIP) experiment could confirm TRF2 bound to telomere sequences), however, ChIP was not pursued due to observing no significant difference between stable and unstable groups.

The only published study that investigates Chinese hamster TRF1, telomere protection and its potential relation to genomic instability, observed punctate staining of TRF1 during interphase but TRF1 foci was undetectable in metaphase spreads (Smilenov et al., 1998). Interestingly, TRF2 foci has been observed in CHO metaphases by another group (Smogorzewska et al., 2000), however they do not explore the interaction of TRF2 in terms of telomere length within CHO cell lines as the publication is focused on human telomeres.

Smogorzewska et al., report TRF2 as being a negative regulator of telomere length in human cells, however this is characterised at the extreme ends of chromosomes, where TRF2-TRF1 block telomerase mediated telomere elongation through sequestering 3' single strand DNA at the end of telomeres, in the proposed t loop model of telomere homeostasis (Griffith et al., 1999). The t loop model has not been characterised at ITS blocks, so it is unknown whether TRF1 and TRF2 play a similar role at these intrachromosomal locations.

How does the production run environment affect cells at the karyotypic level? (7.2.6)

Characterising DNA damage potential role in genomic and production instability has thus far been performed in maintenance cultures. Such an environment may not reflect cellular stresses that cell lines are subjected to during a production run. To understand DNA damage role in production instability, 6 stable and 6 unstable cell lines were put into production run conditions for two weeks under treated (1ng/ml Neocarzinostatin at day 0) and un-treated (normal) conditions. Neocarzinostatin cleaves DNA by abstracting hydrogen atoms from the deoxyribose sugar through nucleophilic attack by a thiol residue (Gibson et al., 1984; Kuromizu et al., 1986). Neocarzinostatin had a profound effect on VCC (figure 5.6), thus resulting in a drastic decrease in max titre by day 15 (figure 5.7). Interestingly, the resulting decrease of max titre was not observed in specific productivity rate (SPR, pg/cell/day), suggesting the remaining live cells were still producing the same amount of therapeutic protein, compared to untreated cell lines. When comparing day 15 %viability between untreated stability groups, unstable cell lines obtained a greater overall decrease in %viability (figure 5.12, P=0.03), indicating production stability may be influenced by VCC as an increase in apoptosis impacts max titre.

To assess the genomic mutation landscape during a production run, karyotype populations were assessed at day 0 and day 8 time points. Increases in NCCA populations were significant for both stable and unstable cell lines (figure 5.13, P=<0.001), indicating Neocarzinostatin had a significant impact on the cell lines mutation profiles. Untreated stable cell lines obtained a significant increase in NCCA frequency during 8 days of production culturing, however this was not observed in untreated unstable cell lines. Although there was an overall increase in NCCA% between unstable day 0 and day 8, the relatively high NCCA% at day 0 may have resulted in the insignificant difference. These results are

the first to highlight the effect of a production run on a cell lines mutation profile, providing evidence that the environmental stress causes cells to acquire de novo mutations, which may have an impact on fluctuating titre levels across a whole stability assessment (4-6 months).

Further investigations into the ratio of CCA and NCCA (mutation profile) and its impact on day 15 %VCC would be an interesting avenue to explore. As day 8 viability remained consistent across stability groups, meaning increases in apoptosis witnessed at day 15 in unstable cell lines may have been influenced by increases in mutations within these cells. Although day 8 stable and unstable %NCCA comparison was similar (P=>0.05), cell lines were not assessed for their karyotypic populations at day 15. This may have provided an insight into whether the mutations observed at day 8 influenced %VCC decrease in unstable cell lines at day 15. It would be interesting to understand whether stable cell lines could maintain their mutation profile observed at day 8, whilst unstable cell lines may obtain increases in mutations that may lead to an increase in apoptosis witnessed here at day 15.

Conclusion (7.3)

Results presented in this study has provided further evidence of the correlation between genomic and production instability within CHOK1a-GS-KO host and CHOK1a-GS-KO based producer cell lines (section 1.5.2, objective and aims: 1 a and b, 2 a and b). In line with Vcelar et al., (Vcelar et al., 2018a; Vcelar et al., 2018b) observations, I present data that highlights the vast genomic heterogeneity both within CHOK1a-GS-KO host and CHOK1a-GS-KO producing cell lines, regardless of production stability. These findings have been expanded further by applying clonal (CCA) and non-clonal (NCCA) chromosomal aberration designations, used within the cytogenetic field for disease diagnosis, providing a general mutation metric that describes the overall mutation landscape within the cell culture flask. Although a stable cell line may have multiple populations, it is the ratio of CCA (genetically stable mutation) or NCCA (genetically unstable/rare) that defines the overall genomic stability of the cell line. When assessing the mutational profile of therapeutic protein producing cell lines within a manufacturing environment, %NCCA obtained a statistically significant increase in all cell lines compared to maintenance media conditions (section 1.5.2, objective and aims: 4), indicating that manufacturing stress causes greater chromosomal mutations.

Utilising CCA and NCCA % metric, I have established a correlation between increased mutations (high % NCCA) and production instability, which showed a consistent trend in 4 out of 5 therapeutic proteins assessed. Moreover, I have provided solid evidence that this metric could be used for production stability prediction at an early time point (~20 generations), testing the methodology on a blinded panel of cell lines to recapitulate its use in a live Cell Line Development project. To my knowledge, this study is the first to test novel findings in an industry relevant diverse panel of cell lines (60 cell lines across 5 therapeutic

proteins) that produce full sized mAbs, which have been developed by GSK's drug discovery process, thus mimicking a full biopharmaceutical therapeutic protein pipeline, from target identification to candidate selection. The observed prediction power across multiple therapeutic protein producing cell lines indicates the method could be used in a therapeutic protein agnostic manor, which allows it to be utilised in any cell line development project, providing overall CMC timesaving of 3-5 months.

Previous work by Nicola Borth's group (Vcelar et al., 2018a; Vcelar et al., 2018b) has provided genomic instability characterisation of multiple CHOK1 based host cell lines, used within the field, during routine maintenance; tracking genomic heterogeneity over the single cell cloning process in a variety of cell culture conditions. Such evidence suggests that findings presented within this thesis may be utilised for other CHOK1 based cell lines, providing impact within the general field. Although the link between genomic heterogeneity and production instability has now been established within this thesis. A causative pathway of the production instability phenotype is still wanting.

There have been attempts at profiling CHO at the 'omic level within a single cell line (Yusufi et al., 2017), to my knowledge, no study has investigated a causative pathway that could be applicable to multiple cell lines producing multiple therapeutic proteins. To this end, this study has characterised overall cellular DNA damage within producer cell lines during maintenance and production cell culture (section 1.5.2, objectives and aims: 1f, 2f and 4). Although there was no fluctuation of DNA damage between host and producer cell lines in different environments, I have established a baseline level of 0.92-4.63% gH2AX positive cells, assessed by flow cytometry staining, in therapeutic protein producing cell lines during maintenance culture. However, no significant differences were found between productionally stable cell lines during routine maintenance, nor within a production run environment, suggesting

either cells have acquired enhanced DNA repair pathways or media and process optimisation provides a DNA damage resistant environment.

Although overall DNA damage increase, as quantified by gH2.AX foci, was not observed, aberrant interstitial telomere sequences (ITS) were explored as a potential causative factor in CHO cell lines high mutation rates. I investigated how ITS length and deprotection (by shelterin) within the CHO genome may lead to increased localised DNA damage at these sequences (telomere induced foci, TIFs), in a bid to understand a mechanism by which the CHO genome mutates so frequently. However, no link could between these factors could be established (section 1.5.2, objectives and aims: 1 c, d and e, 2 c, d, and e).

A gene panel consisting of ~120 genes was designed and validated to investigate whether there were any differential expression patterns between productionally stable and unstable cell lines in genes involved in telomere homeostasis and DNA damage repair pathways (section 1.5.2, objective and aims: 3). No gene was observed to have robust differential expression between production stability groups. I believe that this does not rule out telomere or DNA damage pathways as a potential cause of CHO chromosomal aberrations as the gene screen was not performed as single cell, which may reduce false positive or false negative signals, as highlighted in section 1.3. Combining the knowledge of CHO genome heterogeneity and the fact that the gene screen was performed from pooled cell samples, distinct differences between karyotypically distinct populations may be lost due to an 'averaging' of the mRNA levels within that sample, due to a cell population having the highest expression. It is the authors belief that future work looking to profile expression patterns in CHO cell lines should be performed in single cell to remove potential caveats associated with CHO's natural heterogeneity issues.

As cytogenetic assays are laborious and amenable to subjectivity, I set out to fully automate the manual MFISH based stability prediction method, through a collaboration with GSK AI/ML group, to create a software that incorporates all the manual / semi-automated steps of the prediction method. Automation was achieved by implementing artificial intelligence and machine learning algorithms (as outlined in section 2.7.2.3). The resulting software allows for rapid objective analysis of samples, with minimal analyst intervention. Comparison of automated and manual population identification were similar, with NCCA % difference being statistically significantly higher in unstable cell lines (as determined by the automated workflow, section 1.5.2, objectives and aims: 5). This provides a fully scalable method that allows greater characterisation (increased number of cells analysed) and rapid analysis to provide output results within an industry project time frame. Work presented here on the production stability prediction method has been submitted to the European patent office for an initial patent filling.

Overall, this study provides further evidence of CHOK1's inherent and vast genomic heterogeneity and its effect on therapeutic protein production stability. Utilising CCA and NCCA % metrics, the method outlined here provides researchers a tool to obtain an understanding of CHO cell line mutation landscape and assess whether mutations have increased over prolonged culture. CCA/NCCA % also may provide a metric for stratifying cell lines into genomically stable and unstable categories, which may prove more robust as a categorisation strategy then using titre threshold, for further investigational experiments. Unfortunately, I could not identify a pathway that causes the genomic heterogeneity that the field witnesses within CHO cell lines, but I have provided further characterisation of DNA damage and relative telomere length within CHO, which may be built upon by others.

Limitations of the study (7.4)

The greatest limitation of this study, which not only applies here, but also applies to the CHO field in general, is the vast heterogeneity of CHO at all 'omic levels (Yusufi et al., 2017). Heterogeneity is compounded further by industry process, where multiple therapeutic protein transfections occur constantly, leading to ~49% of genes acquiring differing mutations in each transfection event. It is difficult to observe 'true' patterns when trying to compare cell lines which obtain similar phenotype characteristics (which in turn are often based on arbitrary values), due to this heterogeneity.

A case in point is the genetic screen performed in this study. Expression of ~100 genes was quantified across 9 stable and 9 unstable cell lines. Work presented here and elsewhere (Vcelar et al., 2018a; Vcelar et al., 2018b) has categorically shown CHO cell lines obtain genomically heterogeneous populations within a cell culture flask. When attempting assays that use a pooled sample of cells (i.e. no single cell), the assays essentially detect an average of expression from a single gene, without understanding the ratio in which different cell populations contribute to the detected expression level.

Therefore, as witnessed here, any 'hit' could either be an artefact of the experiment itself or if the differential gene expression is confirmed within your analysed cell lines, there is no guarantee that this trend could be witnessed in a completely new set of cell lines producing a different therapeutic protein. Perhaps this is why studies often concentrate on a single cell line, or subcloned cell lines from the same parent, in a bid to minimise the aforementioned heterogeneity.

Unfortunately, this methodology simply does not replicate how CHO cell lines are used in industry, so the potential irreproducibility of these findings in an industry setting has to be noted. Routine single cell analysis of different CHO 'omic levels will provide a more robust characterisation of CHO cell lines and may provide enough data to confidently assess whether different cell lines obtain similar expression patterns when data is grouped at the single cell level.

A specific caveat to this study, which is shared by the cytogenetic field, is the small number of images (cells) analysed per sample. Here I have analysed 50 images per sample using the manual method, due to the highly laborious nature of the analysis. Although the analysis is based on proportions of CCA/NCCA, it is yet to be determined whether a sample size of 50 cells truly represents the proportion of CCA/NCCA in a flask of >10x10^6 cells. In terms of CCA/NCCA ratio's role in production stability prediction, confidence in observations outlined in this thesis can be drawn from observing the same pattern across numerous experiments performed on a large panel of cell lines (60 cell lines in total). To ameliorate this caveat, we have automated the full prediction stability process, which provides the means to upscale analysis sample size from 50 images to 200-400 images, providing a greater in-depth characterisation of cell lines.

Although a lot of work had gone into producing a Cas9 construct that directly targets telomere sequences, testing the hypothesis of DNA damage at interstitial telomere sequences leads to genomic heterogeneity, the hypothesis could not be fully investigated due to divergent Cas9 expression within the cells during scale up and routine maintenance. Therefore, I did not pursue the experiment, as it would be impossible to discern any potential results and attribute them to telomere specific damage. Thus, were unable to test this hypothesis in this study.

Future directions (7.5)

To further progress the findings presented in this study, a variety of experiments could be performed. Testing the automated stability prediction workflow in a live project setting is imperative to understand if the methodology can truly predict production stability in unseen cell lines, in a therapeutic protein agnostic manor. Additionally, testing the full workflow whilst adhering to project timelines will highlight any limitations in the workflow that need to be addressed.

As previously mentioned, categorising cell lines into productionally stable and unstable cell lines, based on an arbitrary threshold, does not provide a robust metric that is intrinsically linked to the cell's biology. To address this, utilising the MFISH method outlined in this thesis, one could group a panel of cell lines based on their genomic mutation profile (high %CCA vs low %CCA) and perform single cell RNA sequencing. This may identify a genetic marker(s) that are attributable to increased genetic mutations within a cell line. Such information could be utilised in three ways. Firstly, any genetic markers identified could be characterised further to identify pathways involved in genomic instability, with a view to ameliorate said instability through bioengineering. Secondly, identified genetic markers could be used as an early screening tool to aid cell line selection. Thirdly, if a correlation between the identified genetic marker, MFISH mutation profile and production stability is confirmed, these markers could be used to build robust panels of cell lines, based on intrinsic biology of the cell, for further cell line development experiments.

To understand whether interstitial telomere sequences (ITS) play a role in production stability, knock down or knock out of CHO TRF1 or TRF2, leading to telomere deprotection, could be performed. Testing these cell lines in a production run environment could provide a definitive answer as to the role of shelterin, ITS and DNA damage in production stability.

As there is a lack of publications which define any telomeric or DNA damage pathways within the CHO therapeutic production field, there are a multitude of pathways that could prove causative of the production instability phenotype. Here, I concentrated on ITS and DNA damage role, however future work may investigate telomerase and its role in the break-fusion-bridge cycle (BFB) of chromosomes (Thomas et al., 2018). BFBs occur when broken DNA ends are recombinogenic, or when telomeres at the extreme ends of chromosomes are critically low, leading to a fusion of chromosomes which are then sporadically broken upon mitosis (Selvarajah et al., 2006). This provides an interesting pathway that may explain why gross chromosomal changes are observed within CHOK1a cell lines and could represent bioengineering targets once fully characterised within CHO.

References

Adair, G.M., and Carver, J.H. (1979). Unstable, non-mutational expression of resistance to the thymidine analogue, trifluorothymidine in CHO cells. Mutat Res *60*, 207-213.

Aksenova, A.Y., and Mirkin, S.M. (2019). At the Beginning of the End and in the Middle of the Beginning: Structure and Maintenance of Telomeric DNA Repeats and Interstitial Telomeric Sequences. Genes (Basel) *10*.

Alt, F.W., Kellems, R.E., Bertino, J.R., and Schimke, R.T. (1992). Selective multiplication of dihydrofolate reductase genes in methotrexate-resistant variants of cultured murine cells. 1978. Biotechnology *24*, 397-410.

Alvarez, L., Evans, J.W., Wilks, R., Lucas, J.N., Brown, J.M., and Giaccia, A.J. (1993). Chromosomal radiosensitivity at intrachromosomal telomeric sites. Genes Chromosomes Cancer *8*, 8-14.

Antoniou, M., Harland, L., Mustoe, T., Williams, S., Holdstock, J., Yague, E., Mulcahy, T., Griffiths, M., Edwards, S., Ioannou, P.A.*, et al.* (2003). Transgenes encompassing dual-promoter CpG islands from the human TBP and HNRPA2B1 loci are resistant to heterochromatin-mediated silencing. Genomics *82*, 269-279.

Armstrong, M.J., Gara, J.P., Gealy, R., 3rd, Greenwood, S.K., Hilliard, C.A., Laws, G.M., and Galloway, S.M. (2000). Induction of chromosome aberrations in vitro by phenolphthalein: mechanistic studies. Mutat Res *457*, 15-30.

Artandi, S.E., and DePinho, R.A. (2010). Telomeres and telomerase in cancer. Carcinogenesis *31*, 9-18.

Auer, N., Hrdina, A., Hiremath, C., Vcelar, S., Baumann, M., Borth, N., and Jadhav, V. (2018). ChromaWizard: An open source image analysis software for multicolor fluorescence in situ hybridization analysis. Cytometry A *93*, 749-754.

Bahr, S.M., Borgschulte, T., Kayser, K.J., and Lin, N. (2009). Using microarray technology to select housekeeping genes in Chinese hamster ovary cells. Biotechnol Bioeng *104*, 1041-1046.

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.

Balajee, A.S., Bertucci, A., Taveras, M., and Brenner, D.J. (2014). Multicolour FISH analysis of ionising radiation induced micronucleus formation in human lymphocytes. Mutagenesis *29*, 447-455.

Bandaranayake, A.D., and Almo, S.C. (2014). Recent advances in mammalian protein production. FEBS Lett *588*, 253-260.

Barbi, G., Steinbach, P., and Vogel, W. (1984). Nonrandom distribution of methotrexate-induced aberrations on human chromosomes. Detection of further folic acid sensitive fragile sites. Hum Genet *68*, 290-294.

Barnes, L.M., Bentley, C.M., and Dickson, A.J. (2003). Stability of protein production from recombinant mammalian cells. Biotechnol Bioeng *81*, 631-639.

Bartek, J., and Lukas, J. (2007). DNA damage checkpoints: from initiation to recovery or adaptation. Curr Opin Cell Biol *19*, 238-245.

Batista, L.F., Chigancas, V., Brumatti, G., Amarante-Mendes, G.P., and Menck, C.F. (2006). Involvement of DNA replication in ultraviolet-induced apoptosis of mammalian cells. Apoptosis *11*, 1139-1148.

Bauman, J.G., Wiegant, J., Borst, P., and van Duijn, P. (1980). A new method for fluorescence microscopical localization of specific DNA sequences by in situ hybridization of fluorochromelabelled RNA. Exp Cell Res *128*, 485-490.

Bebbington, C.R., Renner, G., Thomson, S., King, D., Abrams, D., and Yarranton, G.T. (1992). High-level expression of a recombinant antibody from myeloma cells using a glutamine synthetase gene as an amplifiable selectable marker. Biotechnology (N Y) *10*, 169-175.

Beers, S.A., and Glennie, M.J. (2013). Neutrophils: "neu players" in antibody therapy? Blood *122*, 3093-3094.

Berk, A.J. (2005). Recent lessons in gene expression, cell cycle control, and cell biology from adenovirus. Oncogene *24*, 7673-7685.

Bertoni, L., Attolini, C., Faravelli, M., Simi, S., and Giulotto, E. (1996). Intrachromosomal telomere-like DNA sequences in Chinese hamster. Mamm Genome *7*, 853-855.

Betts, Z., Croxford, A.S., and Dickson, A.J. (2015). Evaluating the interaction between UCOE and DHFR-linked amplification and stability of recombinant protein expression. Biotechnol Prog *31*, 1014-1025.

Betts, Z., and Dickson, A.J. (2016). Ubiquitous Chromatin Opening Elements (UCOEs) effect on transgene position and expression stability in CHO cells following methotrexate (MTX) amplification. Biotechnol J *11*, 554-564.

Betts, Z., and Dickson, A.J. (2017). Improved CHO Cell Line Stability and Recombinant Protein Expression During Long-Term Culture. Methods Mol Biol *1603*, 119-141.

Beum, P.V., Peek, E.M., Lindorfer, M.A., Beurskens, F.J., Engelberts, P.J., Parren, P.W., van de Winkel, J.G., and Taylor, R.P. (2011). Loss of CD20 and bound CD20 antibody from opsonized B cells occurs more

rapidly because of trogocytosis mediated by Fc receptor-expressing effector cells than direct internalization by the B cells. J Immunol *187*, 3438-3447.

Bianchi, A., Smith, S., Chong, L., Elias, P., and de Lange, T. (1997). TRF1 is a dimer and bends telomeric DNA. EMBO J *16*, 1785-1794.

Bibeau, F., Lopez-Crapez, E., Di Fiore, F., Thezenas, S., Ychou, M., Blanchard, F., Lamy, A., Penault-Llorca, F., Frebourg, T., Michel, P., *et al.* (2009). Impact of Fc{gamma}RIIa-Fc{gamma}RIIIa polymorphisms and KRAS mutations on the clinical outcome of patients with metastatic colorectal cancer treated with cetuximab plus irinotecan. J Clin Oncol *27*, 1122-1129.

Boffa, L.C., Morris, P.L., Carpaneto, E.M., Louissaint, M., and Allfrey, V.G. (1996). Invasion of the CAG triplet repeats by a complementary peptide nucleic acid inhibits transcription of the androgen receptor and TATA-binding protein genes and correlates with refolding of an active nucleosome containing a unique AR gene sequence. J Biol Chem *271*, 13228-13233.

Bolton, E.T., and Mc, C.B. (1962). A general method for the isolation of RNA complementary to DNA. Proc Natl Acad Sci U S A *48*, 1390-1397.

Bolzan, A.D. (2012). Chromosomal aberrations involving telomeres and interstitial telomeric sequences. Mutagenesis *27*, 1-15.

Boscolo, S., Mion, F., Licciulli, M., Macor, P., De Maso, L., Brce, M., Antoniou, M.N., Marzari, R., Santoro, C., and Sblattero, D. (2012). Simple scale-up of recombinant antibody production using an UCOE containing vector. N Biotechnol *29*, 477-484.

Bouffler, S., Silver, A., Papworth, D., Coates, J., and Cox, R. (1993). Murine radiation myeloid leukaemogenesis: relationship between

interstitial telomere-like sequences and chromosome 2 fragile sites. Genes Chromosomes Cancer *6*, 98-106.

Bouffler, S.D., Morgan, W.F., Pandita, T.K., and Slijepcevic, P. (1996). The involvement of telomeric sequences in chromosomal aberrations. Mutat Res *366*, 129-135.

Bowles, J.A., and Weiner, G.J. (2005). CD16 polymorphisms and NK activation induced by monoclonal antibody-coated target cells. J Immunol Methods *304*, 88-99.

Brahmer, J.R., Tykodi, S.S., Chow, L.Q., Hwu, W.J., Topalian, S.L., Hwu, P., Drake, C.G., Camacho, L.H., Kauh, J., Odunsi, K.*, et al.* (2012). Safety and activity of anti-PD-L1 antibody in patients with advanced cancer. N Engl J Med *366*, 2455-2465.

Brinkrolf, K., Rupp, O., Laux, H., Kollin, F., Ernst, W., Linke, B., Kofler, R., Romand, S., Hesse, F., Budach, W.E., *et al.* (2013). Chinese hamster genome sequenced from sorted chromosomes. Nat Biotechnol *31*, 694-695.

Broccoli, D., Smogorzewska, A., Chong, L., and de Lange, T. (1997). Human telomeres contain two distinct Myb-related proteins, TRF1 and TRF2. Nat Genet *17*, 231-235.

Brown, M.E., Renner, G., Field, R.P., and Hassell, T. (1992). Process development for the production of recombinant antibodies using the glutamine synthetase (GS) system. Cytotechnology *9*, 231-236.

Burton, D.R., Barbas, C.F., 3rd, Persson, M.A., Koenig, S., Chanock, R.M., and Lerner, R.A. (1991). A large array of human monoclonal antibodies to type 1 human immunodeficiency virus from combinatorial libraries of asymptomatic seropositive individuals. Proc Natl Acad Sci U S A *88*, 10134-10137.

Butler, M. (2005). Animal cell cultures: recent achievements and perspectives in the production of biopharmaceuticals. Appl Microbiol Biotechnol *68*, 283-291.

Butterfield, D.A., Poon, H.F., St Clair, D., Keller, J.N., Pierce, W.M., Klein, J.B., and Markesbery, W.R. (2006). Redox proteomics identification of oxidatively modified hippocampal proteins in mild cognitive impairment: insights into the development of Alzheimer's disease. Neurobiol Dis *22*, 223-232.

Carlsson, M., and Carlsson, A. (1990). Interactions between glutamatergic and monoaminergic systems within the basal ganglia--implications for schizophrenia and Parkinson's disease. Trends Neurosci *13*, 272-276.

Carrington, W.A., Lynch, R.M., Moore, E.D., Isenberg, G., Fogarty, K.E., and Fay, F.S. (1995). Superresolution three-dimensional images of fluorescence in cells with minimal light exposure. Science *268*, 1483-1487.

Chan, C.E., Chan, A.H., Lim, A.P., and Hanson, B.J. (2011). Comparison of the efficiency of antibody selection from semi-synthetic scFv and non-immune Fab phage display libraries against protein targets for rapid development of diagnostic immunoassays. J Immunol Methods *373*, 79-88.

Chan, V.L., Whitmore, G.F., and Siminovitch, L. (1972). Mammalian cells with altered forms of RNA polymerase II. Proc Natl Acad Sci U S A *69*, 3119-3123.

Chang, T.Y., and Siegel, D.L. (2001). Isolation of an IgG anti-B from a human Fab-phage display library. Transfusion *41*, 6-12.

Chen, C., Wu, B., Wei, T., Egholm, M., and Strauss, W.M. (2000). Unique chromosome identification and sequence-specific structural analysis with short PNA oligomers. Mamm Genome *11*, 384-391.

Cheng, A.W., Wang, H., Yang, H., Shi, L., Katz, Y., Theunissen, T.W., Rangarajan, S., Shivalila, C.S., Dadon, D.B., and Jaenisch, R. (2013). Multiplexed activation of endogenous genes by CRISPR-on, an RNA-guided transcriptional activator system. Cell Res *23*, 1163-1171.

Cimprich, K.A., and Cortez, D. (2008). ATR: an essential regulator of genome integrity. Nat Rev Mol Cell Biol *9*, 616-627.

Clark, M.A., Hawkins, N.J., Papaioannou, A., Fiddes, R.J., and Ward, R.L. (1997). Isolation of human anti-c-erbB-2 Fabs from a lymph node-derived phage display library. Clin Exp Immunol *109*, 166-174.

Clementi, N., Mancini, N., Solforosi, L., Castelli, M., Clementi, M., and Burioni, R. (2012). Phage display-based strategies for cloning and optimization of monoclonal antibodies directed against human pathogens. Int J Mol Sci *13*, 8273-8292.

Clynes, R., Takechi, Y., Moroi, Y., Houghton, A., and Ravetch, J.V. (1998). Fc receptors are required in passive and active immunity to melanoma. Proc Natl Acad Sci U S A *95*, 652-656.

Cockett, M.I., Bebbington, C.R., and Yarranton, G.T. (1990). High level expression of tissue inhibitor of metalloproteinases in Chinese hamster ovary cells using glutamine synthetase gene amplification. Biotechnology (N Y) *8*, 662-667.

Codamo, J., Munro, T.P., Hughes, B.S., Song, M., and Gray, P.P. (2011). Enhanced CHO cell-based transient gene expression with the epi-CHO expression system. Mol Biotechnol *48*, 109-115.

Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., *et al.* (2013). Multiplex genome engineering using CRISPR/Cas systems. Science *339*, 819-823.

Conrad, U., and Scheller, J. (2005). Considerations on antibody-phage display methodology. Comb Chem High Throughput Screen *8*, 117-126.

Cooper, A.J. (2012). The role of glutamine synthetase and glutamate dehydrogenase in cerebral ammonia homeostasis. Neurochem Res *37*, 2439-2455.

Cox, D.M., and Puck, T.T. (1969). Chromosomal nondisjunction: the action of colcemid on Chinese hamster cells in vitro. Cytogenetics *8*, 158-169.

Cuvier, O., Hart, C.M., and Laemmli, U.K. (1998). Identification of a class of chromatin boundary elements. Mol Cell Biol *18*, 7478-7486.

Dahodwala, H., and Lee, K.H. (2019). The fickle CHO: a review of the causes, implications, and potential alleviation of the CHO cell line instability problem. Curr Opin Biotechnol *60*, 128-137.

Dall'Ozzo, S., Tartas, S., Paintaud, G., Cartron, G., Colombat, P., Bardos, P., Watier, H., and Thibault, G. (2004). Rituximab-dependent cytotoxicity by natural killer cells: influence of FCGR3A polymorphism on the concentration-effect relationship. Cancer Res *64*, 4664-4669.

Das, K., Gunasegaran, B., Tan, I.B., Deng, N., Lim, K.H., and Tan, P. (2014). Mutually exclusive FGFR2, HER2, and KRAS gene amplifications in gastric cancer revealed by multicolour FISH. Cancer Lett *353*, 167-175.

Davy, A.M., Kildegaard, H.F., and Andersen, M.R. (2017). Cell Factory Engineering. Cell Syst *4*, 262-275.

de Haard, H.J., van Neer, N., Reurs, A., Hufton, S.E., Roovers, R.C., Henderikx, P., de Bruine, A.P., Arends, J.W., and Hoogenboom, H.R. (1999). A large non-immunized human Fab fragment phage library that permits rapid isolation and kinetic analysis of high affinity antibodies. J Biol Chem *274*, 18218-18230.

de Lange, T. (2002). Protection of mammalian telomeres. Oncogene *21*, 532-540.

de Lange, T. (2005). Shelterin: the protein complex that shapes and safeguards human telomeres. Genes Dev *19*, 2100-2110.

Deaven, L.L., and Petersen, D.F. (1973). The chromosomes of CHO, an aneuploid Chinese hamster cell line: G-band, C-band, and autoradiographic analyses. Chromosoma *41*, 129-144.

Demidov, V., Frank-Kamenetskii, M.D., Egholm, M., Buchardt, O., and Nielsen, P.E. (1993). Sequence selective double strand DNA cleavage by peptide nucleic acid (PNA) targeting using nuclease S1. Nucleic Acids Res *21*, 2103-2107.

Derouazi, M., Martinet, D., Besuchet Schmutz, N., Flaction, R., Wicht, M., Bertschinger, M., Hacker, D.L., Beckmann, J.S., and Wurm, F.M. (2006). Genetic characterization of CHO production host DG44 and derivative recombinant cell lines. Biochem Biophys Res Commun *340*, 1069-1077.

Dianov, G.L., and Hubscher, U. (2013). Mammalian base excision repair: the forgotten archangel. Nucleic Acids Res *41*, 3483-3490.

Donnelly, M.L., Luke, G., Mehrotra, A., Li, X., Hughes, L.E., Gani, D., and Ryan, M.D. (2001). Analysis of the aphthovirus 2A/2B polyprotein 'cleavage' mechanism indicates not a proteolytic reaction, but a novel translational effect: a putative ribosomal 'skip'. J Gen Virol *82*, 1013-1025.

Dorai, H., Corisdeo, S., Ellis, D., Kinney, C., Chomo, M., Hawley-Nelson, P., Moore, G., Betenbaugh, M.J., and Ganguly, S. (2012). Early prediction of instability of Chinese hamster ovary cell lines expressing recombinant antibodies and antibody-fusion proteins. Biotechnol Bioeng *109*, 1016-1030.

Eisenberg, D., Gill, H.S., Pfluegl, G.M., and Rotstein, S.H. (2000). Structure-function relationships of glutamine synthetases. Biochim Biophys Acta *1477*, 122-145.

Fan, L., Kadura, I., Krebs, L.E., Hatfield, C.C., Shaw, M.M., and Frye, C.C. (2012). Improving the efficiency of CHO cell line generation using glutamine synthetase gene knockout cells. Biotechnol Bioeng *109*, 1007-1015.

Fann, C.H., Guirgis, F., Chen, G., Lao, M.S., and Piret, J.M. (2000). Limitations to the amplification and stability of human tissue-type plasminogen activator expression by Chinese hamster ovary cells. Biotechnol Bioeng *69*, 204-212.

Faravelli, M., Azzalin, C.M., Bertoni, L., Chernova, O., Attolini, C., Mondello, C., and Giulotto, E. (2002). Molecular organization of internal telomeric sequences in Chinese hamster chromosomes. Gene *283*, 11-16.

Femino, A.M., Fay, F.S., Fogarty, K., and Singer, R.H. (1998). Visualization of single RNA transcripts in situ. Science *280*, 585-590.

Fernandez, J.L., Gosalvez, J., and Goyanes, V. (1995). High frequency of mutagen-induced chromatid exchanges at interstitial telomere-like DNA sequence blocks of Chinese hamster cells. Chromosome Res *3*, 281-284.

Ferrara, N., Hillan, K.J., Gerber, H.P., and Novotny, W. (2004). Discovery and development of bevacizumab, an anti-VEGF antibody for treating cancer. Nat Rev Drug Discov *3*, 391-400.

Figler, H., Olsson, R.A., and Linden, J. (2003). Allosteric enhancers of A1 adenosine receptors increase receptor-G protein coupling and counteract Guanine nucleotide effects on agonist binding. Mol Pharmacol *64*, 1557-1564.

Fiore, M., Zanier, R., and Degrassi, F. (2002). Reversible G(1) arrest by dimethyl sulfoxide as a new method to synchronize Chinese hamster cells. Mutagenesis *17*, 419-424.

Fliedl, L., Grillari, J., and Grillari-Voglauer, R. (2015). Human cell lines for the production of recombinant proteins: on the horizon. N Biotechnol *32*, 673-679.

Foecking, M.K., and Hofstetter, H. (1986). Powerful and versatile enhancer-promoter unit for mammalian expression vectors. Gene *45*, 101-105.

Gaj, T., Gersbach, C.A., and Barbas, C.F., 3rd (2013). ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. Trends Biotechnol *31*, 397-405.

Gall, J.G., and Pardue, M.L. (1969). Formation and detection of RNA-DNA hybrid molecules in cytological preparations. Proc Natl Acad Sci U S A *63*, 378-383.

Gargiulo, G., Serresi, M., Cesaroni, M., Hulsman, D., and van Lohuizen, M. (2014). In vivo shRNA screens in solid tumors. Nat Protoc *9*, 2880-2902.

Giacomelli, E., Bellin, M., Sala, L., van Meer, B.J., Tertoolen, L.G., Orlova, V.V., and Mummery, C.L. (2017). Three-dimensional cardiac microtissues composed of cardiomyocytes and endothelial cells co-differentiated from human pluripotent stem cells. Development *144*, 1008-1017.

Gibson, B.W., Herlihy, W.C., Samy, T.S., Hahm, K.S., Maeda, H., Meienhofer, J., and Biemann, K. (1984). A revised primary structure for neocarzinostatin based on fast atom bombardment and gas chromatographic-mass spectrometry. J Biol Chem *259*, 10801-10806.

Gisselsson, D., Pettersson, L., Hoglund, M., Heidenblad, M., Gorunova, L., Wiegant, J., Mertens, F., Dal Cin, P., Mitelman, F., and Mandahl, N. (2000). Chromosomal breakage-fusion-bridge events cause genetic intratumor heterogeneity. Proc Natl Acad Sci U S A *97*, 5357-5362.

Goldfarb, P.S., Carritt, B., Hooper, M.L., and Slack, C. (1977). The isolation and characterization of asparagine-requiring mutants of Chinese hamster cells. Exp Cell Res *104*, 357-367.

Gossen, M., Freundlieb, S., Bender, G., Muller, G., Hillen, W., and Bujard, H. (1995). Transcriptional activation by tetracyclines in mammalian cells. Science *268*, 1766-1769.

Goulian, M., Bleile, B., and Tseng, B.Y. (1980). Methotrexate-induced misincorporation of uracil into DNA. Proc Natl Acad Sci U S A *77*, 1956-1960.

Griffith, J.D., Comeau, L., Rosenfield, S., Stansel, R.M., Bianchi, A., Moss, H., and de Lange, T. (1999). Mammalian telomeres end in a large duplex loop. Cell *97*, 503-514.

Grossbard, E.B. (1987). Recombinant tissue plasminogen activator: a brief review. Pharm Res *4*, 375-378.

Hammill, L., Welles, J., and Carson, G.R. (2000). The gel microdrop secretion assay: Identification of a low productivity subpopulation arising during the production of human antibody in CHO cells. Cytotechnology *34*, 27-37.

Han, K., Jeng, E.E., Hess, G.T., Morgens, D.W., Li, A., and Bassik, M.C. (2017). Synergistic drug combinations for cancer identified in a CRISPR screen for pairwise genetic interactions. Nat Biotechnol *35*, 463-474.

Harbers, S.O., Crocker, A., Catalano, G., D'Agati, V., Jung, S., Desai, D.D., and Clynes, R. (2007). Antibody-enhanced cross-presentation of self antigen breaks T cell tolerance. J Clin Invest *117*, 1361-1369.

Harding, F.A., Stickler, M.M., Razo, J., and DuBridge, R.B. (2010). The immunogenicity of humanized and fully human antibodies: residual immunogenicity resides in the CDR regions. MAbs *2*, 256-265.

Hayflick, L. (1965). The Limited in Vitro Lifetime of Human Diploid Cell Strains. Exp Cell Res *37*, 614-636.

Hayflick, L., and Moorhead, P.S. (1961). The serial cultivation of human diploid cell strains. Exp Cell Res *25*, 585-621.

Hayward, B.E., Hussain, A., Wilson, R.H., Lyons, A., Woodcock, V., McIntosh, B., and Harris, T.J. (1986). The cloning and nucleotide sequence of cDNA for an amplified glutamine synthetase gene from the Chinese hamster. Nucleic Acids Res *14*, 999-1008.

Hefzi, H., Ang, K.S., Hanscho, M., Bordbar, A., Ruckerbauer, D., Lakshmanan, M., Orellana, C.A., Baycin-Hizal, D., Huang, Y., Ley, D., *et al.* (2016). A Consensus Genome-scale Reconstruction of Chinese Hamster Ovary Cell Metabolism. Cell Syst *3*, 434-443 e438.

Hernandez-Ilizaliturri, F.J., Jupudy, V., Ostberg, J., Oflazoglu, E., Huberman, A., Repasky, E., and Czuczman, M.S. (2003). Neutrophils contribute to the biological antitumor activity of rituximab in a non-Hodgkin's lymphoma severe combined immunodeficiency mouse model. Clin Cancer Res *9*, 5866-5873.

Hernandez, I., Dhiman, H., Klanert, G., Jadhav, V., Auer, N., Hanscho, M., Baumann, M., Esteve-Codina, A., Dabad, M., Gomez, J., *et al.*

(2019). Epigenetic regulation of gene expression in Chinese Hamster Ovary cells in response to the changing environment of a batch culture. Biotechnol Bioeng *116*, 677-692.

Ho, S.C., Mariati, Yeo, J.H., Fang, S.G., and Yang, Y. (2015). Impact of using different promoters and matrix attachment regions on recombinant protein expression level and stability in stably transfected CHO cells. Mol Biotechnol *57*, 138-144.

Hoffmeyer, F., Witte, K., and Schmidt, R.E. (1997). The high-affinity Fc gamma RI on PMN: regulation of expression and signal transduction. Immunology *92*, 544-552.

Hoogenboom, H.R., and Chames, P. (2000). Natural and designer binding sites made by phage display technology. Immunol Today *21*, 371-378.

Hopman, A.H., Wiegant, J., Raap, A.K., Landegent, J.E., van der Ploeg, M., and van Duijn, P. (1986). Bi-color detection of two target DNAs by non-radioactive in situ hybridization. Histochemistry *85*, 1-4.

Hornigold, D.C., Mistry, R., Raymond, P.D., Blank, J.L., and Challiss, R.A. (2003). Evidence for cross-talk between M2 and M3 muscarinic acetylcholine receptors in the regulation of second messenger and extracellular signal-regulated kinase signalling pathways in Chinese hamster ovary cells. Br J Pharmacol *138*, 1340-1350.

Horowitz, M.C., Friedlaender, G.E., and Qian, H.Y. (1996). The immune response: the efferent arm. Clin Orthop Relat Res, 25-34.

Horrobin, D.F. (2003). Modern biomedical research: an internally self-consistent universe with little contact with medical reality? Nat Rev Drug Discov *2*, 151-154.

Huen, M.S., and Chen, J. (2008). The DNA damage response pathways: at the crossroad of protein modifications. Cell Res *18*, 8-16.

Hunter, M., Yuan, P., Vavilala, D., and Fox, M. (2019). Optimization of Protein Expression in Mammalian Cells. Curr Protoc Protein Sci *95*, e77.

Illingworth, R., Kerr, A., Desousa, D., Jorgensen, H., Ellis, P., Stalker, J., Jackson, D., Clee, C., Plumb, R., Rogers, J., *et al.* (2008). A novel CpG island set identifies tissue-specific methylation at developmental gene loci. PLoS Biol *6*, e22.

Irving, R.A., Kortt, A.A., and Hudson, P.J. (1996). Affinity maturation of recombinant antibodies using E. coli mutator cells. Immunotechnology *2*, 127-143.

Jackson, S.P., and Bartek, J. (2009). The DNA-damage response in human biology and disease. Nature *461*, 1071-1078.

Jadhav, V., Hackl, M., Druz, A., Shridhar, S., Chung, C.Y., Heffner, K.M., Kreil, D.P., Betenbaugh, M., Shiloach, J., Barron, N., *et al.* (2013). CHO microRNA engineering is growing up: recent successes and future challenges. Biotechnol Adv *31*, 1501-1513.

Jensen, K.K., Orum, H., Nielsen, P.E., and Norden, B. (1997). Kinetics for hybridization of peptide nucleic acids (PNA) with DNA and RNA studied with the BIAcore technique. Biochemistry *36*, 5072-5077.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science *337*, 816-821.

Jiricny, J. (2006). The multifaceted mismatch-repair system. Nat Rev Mol Cell Biol *7*, 335-346.

Kaneyoshi, K., Yamano-Adachi, N., Koga, Y., Uchiyama, K., and Omasa, T. (2019). Analysis of the immunoglobulin G (IgG) secretion efficiency in recombinant Chinese hamster ovary (CHO) cells by using Citrine-fusion IgG. Cytotechnology *71*, 193-207.

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.

Kastan, M.B., and Bartek, J. (2004). Cell-cycle checkpoints and cancer. Nature *432*, 316-323.

Katinka, M.D., and Bourgain, F.M. (1992). Interstitial telomeres are hotspots for illegitimate recombination with DNA molecules injected into the macronucleus of Paramecium primaurelia. EMBO J *11*, 725-732.

Kaufman, R.J., Sharp, P.A., and Latt, S.A. (1983). Evolution of chromosomal regions containing transfected and amplified dihydrofolate reductase sequences. Mol Cell Biol *3*, 699-711.

Kelley, B. (2009). Industrialization of mAb production technology: the bioprocessing industry at a crossroads. MAbs *1*, 443-452.

Kim, H., Lee, O.H., Xin, H., Chen, L.Y., Qin, J., Chae, H.K., Lin, S.Y., Safari, A., Liu, D., and Songyang, Z. (2009). TRF2 functions as a protein hub and regulates telomere maintenance by recognizing specific peptide motifs. Nat Struct Mol Biol *16*, 372-379.

Kim, J.M., Kim, J.S., Park, D.H., Kang, H.S., Yoon, J., Baek, K., and Yoon, Y. (2004). Improved recombinant gene expression in CHO cells using matrix attachment regions. J Biotechnol *107*, 95-105.

Kim, Y.G., Cha, J., and Chandrasegaran, S. (1996). Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. Proc Natl Acad Sci U S A *93*, 1156-1160.

Kinashi, Y., Takahashi, S., Kashino, G., Okayasu, R., Masunaga, S., Suzuki, M., and Ono, K. (2011). DNA double-strand break induction in Ku80-deficient CHO cells following boron neutron capture reaction. Radiat Oncol *6*, 106.

Kingston, R.E., Kaufman, R.J., Bebbington, C.R., and Rolfe, M.R. (2002). Amplification using CHO cell expression vectors. Curr Protoc Mol Biol *Chapter 16*, Unit 16 23.

Kislauskis, E.H., Li, Z., Singer, R.H., and Taneja, K.L. (1993). Isoform-specific 3'-untranslated sequences sort alpha-cardiac and beta-cytoplasmic actin messenger RNAs to different cytoplasmic compartments. J Cell Biol *123*, 165-172.

Klinger, K., Landes, G., Shook, D., Harvey, R., Lopez, L., Locke, P., Lerner, T., Osathanondh, R., Leverone, B., Houseal, T.*, et al.* (1992). Rapid detection of chromosome aneuploidies in uncultured amniocytes by using fluorescence in situ hybridization (FISH). Am J Hum Genet *51*, 55-65.

Knappik, A., Ge, L., Honegger, A., Pack, P., Fischer, M., Wellnhofer, G., Hoess, A., Wolle, J., Pluckthun, A., and Virnekas, B. (2000). Fully synthetic human combinatorial antibody libraries (HuCAL) based on modular consensus frameworks and CDRs randomized with trinucleotides. J Mol Biol *296*, 57-86.

Koduri, R.K., Miller, J.T., and Thammana, P. (2001). An efficient homologous recombination vector pTV(I) contains a hot spot for increased recombinant protein expression in Chinese hamster ovary cells. Gene *280*, 87-95.

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

Kohnlein, W., and Jung, G. (1982). [Neocarzinostatin: molecular mechanism of action and prospects for clinical use]. Arzneimittelforschung *32*, 1474-1479.

Koike, M., and Koike, A. (2008). Accumulation of Ku80 proteins at DNA double-strand breaks in living cells. Exp Cell Res *314*, 1061-1070.

Krajewski, W.W., Collins, R., Holmberg-Schiavone, L., Jones, T.A., Karlberg, T., and Mowbray, S.L. (2008). Crystal structures of mammalian glutamine synthetases illustrate substrate-induced conformational changes and provide opportunities for drug and herbicide design. J Mol Biol *375*, 217-228.

Kremkow, B.G., Baik, J.Y., MacDonald, M.L., and Lee, K.H. (2015). CHOgenome.org 2.0: Genome resources and website updates. Biotechnol J *10*, 931-938.

Krutilina, R.I., Oei, S., Buchlow, G., Yau, P.M., Zalensky, A.O., Zalenskaya, I.A., Bradbury, E.M., and Tomilin, N.V. (2001). A negative regulator of telomere-length protein trf1 is associated with interstitial (TTAGGG)n blocks in immortal Chinese hamster ovary cells. Biochem Biophys Res Commun *280*, 471-475.

Kumar, R., Singh, S.K., Koshkin, A.A., Rajwanshi, V.K., Meldgaard, M., and Wengel, J. (1998). The first analogues of LNA (locked nucleic acids): phosphorothioate-LNA and 2'-thio-LNA. Bioorg Med Chem Lett *8*, 2219-2222.

Kuromizu, K., Tsunasawa, S., Maeda, H., Abe, O., and Sakiyama, F. (1986). Reexamination of the primary structure of an antitumor protein, neocarzinostatin. Arch Biochem Biophys *246*, 199-205.

Langer, P.R., Waldrop, A.A., and Ward, D.C. (1981). Enzymatic synthesis of biotin-labeled polynucleotides: novel nucleic acid affinity probes. Proc Natl Acad Sci U S A *78*, 6633-6637.

Lawrence, J.B., Singer, R.H., and Marselle, L.M. (1989). Highly localized tracks of specific transcripts within interphase nuclei visualized by in situ hybridization. Cell *57*, 493-502.

Lawrence, J.B., Singer, R.H., Villnave, C.A., Stein, J.L., and Stein, G.S. (1988). Intracellular distribution of histone mRNAs in human

fibroblasts studied by in situ hybridization. Proc Natl Acad Sci U S A *85*, 463-467.

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

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

Lee, K.J., Saha, J., Sun, J., Fattah, K.R., Wang, S.C., Jakob, B., Chi, L., Wang, S.Y., Taucher-Scholz, G., Davis, A.J., *et al.* (2016). Phosphorylation of Ku dictates DNA double-strand break (DSB) repair pathway choice in S phase. Nucleic Acids Res *44*, 1732-1745.

Lefebvre, M.L., Krause, S.W., Salcedo, M., and Nardin, A. (2006). Ex vivo-activated human macrophages kill chronic lymphocytic leukemia cells in the presence of rituximab: mechanism of antibody-dependent cellular cytotoxicity and impact of human serum. J Immunother *29*, 388-397.

Levsky, J.M., Shenoy, S.M., Pezo, R.C., and Singer, R.H. (2002). Single-cell gene expression profiling. Science *297*, 836-840.

Ley, D., Seresht, A.K., Engmark, M., Magdenoska, O., Nielsen, K.F., Kildegaard, H.F., and Andersen, M.R. (2015). Multi-omic profiling -of EPO-producing Chinese hamster ovary cell panel reveals metabolic adaptation to heterologous protein production. Biotechnol Bioeng *112*, 2373-2387.

Li, F., Vijayasankaran, N., Shen, A.Y., Kiss, R., and Amanullah, A. (2010). Cell culture processes for monoclonal antibody production. MAbs *2*, 466-479.

Li, P., Pomianowski, P., DiMaio, M.S., Florio, J.R., Rossi, M.R., Xiang, B., Xu, F., Yang, H., Geng, Q., Xie, J., *et al.* (2011). Genomic characterization of prenatally detected chromosomal structural abnormalities using oligonucleotide array comparative genomic hybridization. Am J Med Genet A *155A*, 1605-1615.

Liaw, S.H., and Eisenberg, D. (1994). Structural model for the reaction mechanism of glutamine synthetase, based on five crystal structures of enzyme-substrate complexes. Biochemistry *33*, 675-681.

Lieber, M.R. (2008). The mechanism of human nonhomologous DNA end joining. J Biol Chem *283*, 1-5.

Lindsay, M.A. (2003). Target discovery. Nat Rev Drug Discov *2*, 831-838.

Lingg, N., Zhang, P., Song, Z., and Bardor, M. (2012). The sweet tooth of biopharmaceuticals: importance of recombinant protein glycosylation analysis. Biotechnol J *7*, 1462-1472.

Lo, A.W., Sabatier, L., Fouladi, B., Pottier, G., Ricoul, M., and Murnane, J.P. (2002). DNA amplification by breakage/fusion/bridge cycles initiated by spontaneous telomere loss in a human cancer cell line. Neoplasia *4*, 531-538.

LoBuglio, A.F., Wheeler, R.H., Trang, J., Haynes, A., Rogers, K., Harvey, E.B., Sun, L., Ghrayeb, J., and Khazaeli, M.B. (1989). Mouse/human chimeric monoclonal antibody in man: kinetics and immune response. Proc Natl Acad Sci U S A *86*, 4220-4224.

Low, N.M., Holliger, P.H., and Winter, G. (1996). Mimicking somatic hypermutation: affinity maturation of antibodies displayed on bacteriophage using a bacterial mutator strain. J Mol Biol *260*, 359-368.

Lu, W., Zhang, Y., Liu, D., Songyang, Z., and Wan, M. (2013). Telomeres-structure, function, and regulation. Exp Cell Res *319*, 133-141.

Ludwig, D.L., Pereira, D.S., Zhu, Z., Hicklin, D.J., and Bohlen, P. (2003). Monoclonal antibody therapeutics and apoptosis. Oncogene *22*, 9097-9106.

Macaulay, I.C., and Voet, T. (2014). Single cell genomics: advances and future perspectives. PLoS Genet *10*, e1004126.

Maccani, A., Ernst, W., and Grabherr, R. (2013). Whole genome sequencing improves estimation of nuclear DNA content of Chinese hamster ovary cells. Cytometry A *83*, 893-895.

MacDonald, M.L., Hamaker, N., and Lee, K.H. (2018). Bioinformatic analysis of Chinese hamster ovary host cell protein lipases. AIChE J *64*, 4247-4254.

Maffioletti, S.M., Sarcar, S., Henderson, A.B.H., Mannhardt, I., Pinton, L., Moyle, L.A., Steele-Stallard, H., Cappellari, O., Wells, K.E., Ferrari, G., *et al.* (2018). Three-Dimensional Human iPSC-Derived Artificial Skeletal Muscles Model Muscular Dystrophies and Enable Multilineage Tissue Engineering. Cell Rep *23*, 899-908.

Maloney, D.G., Grillo-Lopez, A.J., White, C.A., Bodkin, D., Schilder, R.J., Neidhart, J.A., Janakiraman, N., Foon, K.A., Liles, T.M., Dallaire, B.K., *et al.* (1997). IDEC-C2B8 (Rituximab) anti-CD20 monoclonal antibody therapy in patients with relapsed low-grade non-Hodgkin's lymphoma. Blood *90*, 2188-2195.

Manuelidis, L., Langer-Safer, P.R., and Ward, D.C. (1982). High-resolution mapping of satellite DNA using biotin-labeled DNA probes. J Cell Biol *95*, 619-625.

Mao, P., Liu, J., Zhang, Z., Zhang, H., Liu, H., Gao, S., Rong, Y.S., and Zhao, Y. (2016). Homologous recombination-dependent repair of telomeric DSBs in proliferating human cells. Nat Commun *7*, 12154.

Mao, S., Gao, C., Lo, C.H., Wirsching, P., Wong, C.H., and Janda, K.D. (1999). Phage-display library selection of high-affinity human single-chain antibodies to tumor-associated carbohydrate antigens sialyl Lewisx and Lewisx. Proc Natl Acad Sci U S A *96*, 6953-6958.

Marceau, C.D., Puschnik, A.S., Majzoub, K., Ooi, Y.S., Brewer, S.M., Fuchs, G., Swaminathan, K., Mata, M.A., Elias, J.E., Sarnow, P.*, et al.* (2016). Genetic dissection of Flaviviridae host factors through genome-scale CRISPR screens. Nature *535*, 159-163.

Markkanen, A., Heinonen, K., Knuutila, S., and de la Chapelle, A. (1982). Methotrexate-induced increase in gap formation in human chromosome band 3p14. Hereditas *96*, 317-319.

Marotta, M., Chen, X., Watanabe, T., Faber, P.W., Diede, S.J., Tapscott, S., Tubbs, R., Kondratova, A., Stephens, R., and Tanaka, H. (2013). Homology-mediated end-capping as a primary step of sister chromatid fusion in the breakage-fusion-bridge cycles. Nucleic Acids Res *41*, 9732-9740.

Matasci, M., Baldi, L., Hacker, D.L., and Wurm, F.M. (2011). The PiggyBac transposon enhances the frequency of CHO stable cell line generation and yields recombinant lines with superior productivity and stability. Biotechnol Bioeng *108*, 2141-2150.

Mather, J.P. (1998). Laboratory scaleup of cell cultures (0.5-50 liters). Methods Cell Biol *57*, 219-227.

Matsuya, Y., Ohtsubo, Y., Tsutsumi, K., Sasaki, K., Yamazaki, R., and Date, H. (2014). Quantitative estimation of DNA damage by photon

irradiation based on the microdosimetric-kinetic model. J Radiat Res *55*, 484-493.

McClintock, B. (1941). The Stability of Broken Ends of Chromosomes in Zea Mays. Genetics *26*, 234-282.

McVey, M., and Lee, S.E. (2008). MMEJ repair of double-strand breaks (director's cut): deleted sequences and alternative endings. Trends Genet *24*, 529-538.

Meeker, T.C., Lowder, J., Maloney, D.G., Miller, R.A., Thielemans, K., Warnke, R., and Levy, R. (1985). A clinical trial of anti-idiotype therapy for B cell malignancy. Blood *65*, 1349-1363.

Mender, I., and Shay, J.W. (2015). Telomere Dysfunction Induced Foci (TIF) Analysis. Bio Protoc *5*.

Mendez, M.J., Green, L.L., Corvalan, J.R., Jia, X.C., Maynard-Currie, C.E., Yang, X.D., Gallo, M.L., Louie, D.M., Lee, D.V., Erickson, K.L., *et al.* (1997). Functional transplant of megabase human immunoglobulin loci recapitulates human antibody response in mice. Nat Genet *15*, 146-156.

Milstein, C. (1985). From the structure of antibodies to the diversification of the immune response. Nobel lecture, 8 December 1984. Biosci Rep *5*, 275-297.

Mondello, C., Pirzio, L., Azzalin, C.M., and Giulotto, E. (2000). Instability of interstitial telomeric sequences in the human genome. Genomics *68*, 111-117.

Montpetit, A.J., Alhareeri, A.A., Montpetit, M., Starkweather, A.R., Elmore, L.W., Filler, K., Mohanraj, L., Burton, C.W., Menzies, V.S., Lyon, D.E., *et al.* (2014). Telomere length: a review of methods for measurement. Nurs Res *63*, 289-299.

Moore, G.L., Chen, H., Karki, S., and Lazar, G.A. (2010). Engineered Fc variant antibodies with enhanced ability to recruit complement and mediate effector functions. MAbs *2*, 181-189.

Morrison, L.E., Ramakrishnan, R., Ruffalo, T.M., and Wilber, K.A. (2002). Labeling fluorescence in situ hybridization probes for genomic targets. Methods Mol Biol *204*, 21-40.

Mullis, K., Faloona, F., Scharf, S., Saiki, R., Horn, G., and Erlich, H. (1986). Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. Cold Spring Harb Symp Quant Biol *51 Pt 1*, 263-273.

Murnane, J.P. (2010). Telomere loss as a mechanism for chromosome instability in human cancer. Cancer Res *70*, 4255-4259.

Nagorsen, D., Bargou, R., Ruttinger, D., Kufer, P., Baeuerle, P.A., and Zugmaier, G. (2009). Immunotherapy of lymphoma and leukemia with T-cell engaging BiTE antibody blinatumomab. Leuk Lymphoma *50*, 886-891.

Nederlof, P.M., Robinson, D., Abuknesha, R., Wiegant, J., Hopman, A.H., Tanke, H.J., and Raap, A.K. (1989). Three-color fluorescence in situ hybridization for the simultaneous detection of multiple nucleic acid sequences. Cytometry *10*, 20-27.

Nederlof, P.M., van der Flier, S., Vrolijk, J., Tanke, H.J., and Raap, A.K. (1992). Fluorescence ratio measurements of double-labeled probes for multiple in situ hybridization by digital imaging microscopy. Cytometry *13*, 839-845.

Nederlof, P.M., van der Flier, S., Wiegant, J., Raap, A.K., Tanke, H.J., Ploem, J.S., and van der Ploeg, M. (1990). Multiple fluorescence in situ hybridization. Cytometry *11*, 126-131.

Neermann, J., and Wagner, R. (1996). Comparative analysis of glucose and glutamine metabolism in transformed mammalian cell lines, insect and primary liver cells. J Cell Physiol *166*, 152-169.

Ngo, V.N., Davis, R.E., Lamy, L., Yu, X., Zhao, H., Lenz, G., Lam, L.T., Dave, S., Yang, L., Powell, J., *et al.* (2006). A loss-of-function RNA interference screen for molecular targets in cancer. Nature *441*, 106-110.

Nielsen, P.E., Egholm, M., Berg, R.H., and Buchardt, O. (1991). Sequence-selective recognition of DNA by strand displacement with a thymine-substituted polyamide. Science *254*, 1497-1500.

Noh, S.M., Shin, S., and Lee, G.M. (2018). Comprehensive characterization of glutamine synthetase-mediated selection for the establishment of recombinant CHO cells producing monoclonal antibodies. Sci Rep *8*, 5361.

O'Sullivan, R.J., and Karlseder, J. (2010). Telomeres: protecting chromosomes against genome instability. Nat Rev Mol Cell Biol *11*, 171-181.

Osterlehner, A., Simmeth, S., and Gopfert, U. (2011). Promoter methylation and transgene copy numbers predict unstable protein production in recombinant Chinese hamster ovary cell lines. Biotechnol Bioeng *108*, 2670-2681.

Page, D.B., Postow, M.A., Callahan, M.K., Allison, J.P., and Wolchok, J.D. (2014). Immune modulation in cancer with antibodies. Annu Rev Med *65*, 185-202.

Pandita, T.K., and DeRubeis, D. (1995). Spontaneous amplification of interstitial telomeric bands in Chinese hamster ovary cells. Cytogenet Cell Genet *68*, 95-101.

Pardoll, D.M. (2012). The blockade of immune checkpoints in cancer immunotherapy. Nat Rev Cancer *12*, 252-264.

Patel, N.A., Anderson, C.R., Terkildsen, S.E., Davis, R.C., Pack, L.D., Bhargava, S., and Clarke, H.R.G. (2018). Antibody expression stability in CHO clonally derived cell lines and their subclones: Role of methylation in phenotypic and epigenetic heterogeneity. Biotechnol Prog *34*, 635-649.

Pawluczkowycz, A.W., Beurskens, F.J., Beum, P.V., Lindorfer, M.A., van de Winkel, J.G., Parren, P.W., and Taylor, R.P. (2009). Binding of submaximal C1q promotes complement-dependent cytotoxicity (CDC) of B cells opsonized with anti-CD20 mAbs ofatumumab (OFA) or rituximab (RTX): considerably higher levels of CDC are induced by OFA than by RTX. J Immunol *183*, 749-758.

Peggs, K.S., Quezada, S.A., Korman, A.J., and Allison, J.P. (2006). Principles and use of anti-CTLA4 antibody in human cancer immunotherapy. Curr Opin Immunol *18*, 206-213.

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

Pinero, J., Furlong, L.I., and Sanz, F. (2018). In silico models in drug development: where we are. Curr Opin Pharmacol *42*, 111-121.

Pinkel, D., Straume, T., and Gray, J.W. (1986). Cytogenetic analysis using quantitative, high-sensitivity, fluorescence hybridization. Proc Natl Acad Sci U S A *83*, 2934-2938.

Piper, J., Poggensee, M., Hill, W., Jensen, R., Ji, L., Poole, I., Stark, M., and Sudar, D. (1994). Automatic fluorescence metaphase finder speeds translocation scoring in FISH painted chromosomes. Cytometry *16*, 7-16.

Pontiller, J., Gross, S., Thaisuchat, H., Hesse, F., and Ernst, W. (2008). Identification of CHO endogenous promoter elements based on a genomic library approach. Mol Biotechnol *39*, 135-139.

Pontiller, J., Maccani, A., Baumann, M., Klancnik, I., and Ernst, W. (2010). Identification of CHO endogenous gene regulatory elements. Mol Biotechnol *45*, 235-240.

Puck, T.T. (1964). Studies of the Life Cycle of Mammalian Cells. Cold Spring Harb Symp Quant Biol *29*, 167-176.

Puck, T.T. (1971). Biochemical and genetic studies on mammalian cells. In Vitro *7*, 115-119.

Puck, T.T. (1979). Studies on cell transformation. Somatic Cell Genet *5*, 973-990.

Puck, T.T., Cieciura, S.J., and Robinson, A. (1958). Genetics of somatic mammalian cells. III. Long-term cultivation of euploid cells from human and animal subjects. J Exp Med *108*, 945-956.

Puck, T.T., Sanders, P., and Petersen, D. (1964). Life Cycle Analysis of Mammalian Cells. Ii. Cells from the Chinese Hamster Ovary Grown in Suspension Culture. Biophys J *4*, 441-450.

Rajendra, Y., Hougland, M.D., Alam, R., Morehead, T.A., and Barnard, G.C. (2015). A high cell density transient transfection system for therapeutic protein expression based on a CHO GS-knockout cell line: process development and product quality assessment. Biotechnol Bioeng *112*, 977-986.

Ran, F.A., Hsu, P.D., Lin, C.Y., Gootenberg, J.S., Konermann, S., Trevino, A.E., Scott, D.A., Inoue, A., Matoba, S., Zhang, Y., *et al.* (2013a). Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. Cell *154*, 1380-1389.

Ran, F.A., Hsu, P.D., Wright, J., Agarwala, V., Scott, D.A., and Zhang, F. (2013b). Genome engineering using the CRISPR-Cas9 system. Nat Protoc *8*, 2281-2308.

Ried, T., Landes, G., Dackowski, W., Klinger, K., and Ward, D.C. (1992). Multicolor fluorescence in situ hybridization for the simultaneous detection of probe sets for chromosomes 13, 18, 21, X and Y in uncultured amniotic fluid cells. Hum Mol Genet *1*, 307-313.

Riley, T., Sontag, E., Chen, P., and Levine, A. (2008). Transcriptional control of human p53-regulated genes. Nat Rev Mol Cell Biol *9*, 402-412.

Ritter, A., Nuciforo, S., Schulze, A., Oertli, M., Rauschert, T., Voedisch, B., Geisse, S., Jostock, T., and Laux, H. (2017). Fam60A plays a role for production stabilities of recombinant CHO cell lines. Biotechnol Bioeng *114*, 701-704.

Ritter, A., Rauschert, T., Oertli, M., Piehlmaier, D., Mantas, P., Kuntzelmann, G., Lageyre, N., Brannetti, B., Voedisch, B., Geisse, S., *et al.* (2016a). Disruption of the gene C12orf35 leads to increased productivities in recombinant CHO cell lines. Biotechnol Bioeng *113*, 2433-2442.

Ritter, A., Voedisch, B., Wienberg, J., Wilms, B., Geisse, S., Jostock, T., and Laux, H. (2016b). Deletion of a telomeric region on chromosome 8 correlates with higher productivity and stability of CHO cell lines. Biotechnol Bioeng *113*, 1084-1093.

Rivero, M.T., Mosquera, A., Goyanes, V., Slijepcevic, P., and Fernandez, J.L. (2004). Differences in repair profiles of interstitial telomeric sites between normal and DNA double-strand break repair deficient Chinese hamster cells. Exp Cell Res *295*, 161-172.

Rogakou, E.P., Pilch, D.R., Orr, A.H., Ivanova, V.S., and Bonner, W.M. (1998). DNA double-stranded breaks induce histone H2AX phosphorylation on serine 139. J Biol Chem *273*, 5858-5868.

Rogers, L.M., Veeramani, S., and Weiner, G.J. (2014). Complement in monoclonal antibody therapy of cancer. Immunol Res *59*, 203-210.

Ronda, C., Pedersen, L.E., Hansen, H.G., Kallehauge, T.B., Betenbaugh, M.J., Nielsen, A.T., and Kildegaard, H.F. (2014). Accelerating genome editing in CHO cells using CRISPR Cas9 and CRISPy, a web-based target finding tool. Biotechnol Bioeng *111*, 1604-1616.

Roos, W.P., and Kaina, B. (2006). DNA damage-induced cell death by apoptosis. Trends Mol Med *12*, 440-450.

Rothe, C., Urlinger, S., Lohning, C., Prassler, J., Stark, Y., Jager, U., Hubner, B., Bardroff, M., Pradel, I., Boss, M., *et al.* (2008). The human combinatorial antibody library HuCAL GOLD combines diversification of all six CDRs according to the natural immune system with a novel display method for efficient selection of high-affinity antibodies. J Mol Biol *376*, 1182-1200.

Rothkamm, K., Kruger, I., Thompson, L.H., and Lobrich, M. (2003). Pathways of DNA double-strand break repair during the mammalian cell cycle. Mol Cell Biol *23*, 5706-5715.

Running Deer, J., and Allison, D.S. (2004). High-level expression of proteins in mammalian cells using transcription regulatory sequences from the Chinese hamster EF-1alpha gene. Biotechnol Prog *20*, 880-889.

Saliba, A.E., Westermann, A.J., Gorski, S.A., and Vogel, J. (2014). Single-cell RNA-seq: advances and future challenges. Nucleic Acids Res *42*, 8845-8860.

San Filippo, J., Sung, P., and Klein, H. (2008). Mechanism of eukaryotic homologous recombination. Annu Rev Biochem *77*, 229-257.

Sanders, P.G., and Wilson, R.H. (1984). Amplification and cloning of the Chinese hamster glutamine synthetase gene. EMBO J *3*, 65-71.

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

Schier, R., McCall, A., Adams, G.P., Marshall, K.W., Merritt, H., Yim, M., Crawford, R.S., Weiner, L.M., Marks, C., and Marks, J.D. (1996). Isolation of picomolar affinity anti-c-erbB-2 single-chain Fv by molecular evolution of the complementarity determining regions in the center of the antibody binding site. J Mol Biol *263*, 551-567.

Schimke, R.T. (1988). Gene amplification in cultured cells. J Biol Chem *263*, 5989-5992.

Schmittgen, T.D., and Livak, K.J. (2008). Analyzing real-time PCR data by the comparative C(T) method. Nat Protoc *3*, 1101-1108.

Schmutz, I., and de Lange, T. (2016). Shelterin. Curr Biol *26*, R397-399.

Schroeder, H.W., Jr., and Cavacini, L. (2010). Structure and function of immunoglobulins. J Allergy Clin Immunol *125*, S41-52.

Schulte, G., and Fredholm, B.B. (2003). The G(s)-coupled adenosine A(2B) receptor recruits divergent pathways to regulate ERK1/2 and p38. Exp Cell Res *290*, 168-176.

Scott, A.M., Wolchok, J.D., and Old, L.J. (2012). Antibody therapy of cancer. Nat Rev Cancer *12*, 278-287.

Segal, D.M., Weiner, G.J., and Weiner, L.M. (1999). Bispecific antibodies in cancer therapy. Curr Opin Immunol *11*, 558-562.

Seidel, U.J., Schlegel, P., and Lang, P. (2013). Natural killer cell mediated antibody-dependent cellular cytotoxicity in tumor immunotherapy with therapeutic antibodies. Front Immunol *4*, 76.

Selvarajah, S., Yoshimoto, M., Park, P.C., Maire, G., Paderova, J., Bayani, J., Lim, G., Al-Romaih, K., Squire, J.A., and Zielenska, M. (2006). The breakage-fusion-bridge (BFB) cycle as a mechanism for generating genetic heterogeneity in osteosarcoma. Chromosoma *115*, 459-467.

Sennino, B., and McDonald, D.M. (2012). Controlling escape from angiogenesis inhibitors. Nat Rev Cancer *12*, 699-709.

Sha, J., Ghosh, M.K., Zhang, K., and Harter, M.L. (2010). E1A interacts with two opposing transcriptional pathways to induce quiescent cells into S phase. J Virol *84*, 4050-4059.

Shalem, O., Sanjana, N.E., Hartenian, E., Shi, X., Scott, D.A., Mikkelson, T., Heckl, D., Ebert, B.L., Root, D.E., Doench, J.G., *et al.* (2014). Genome-scale CRISPR-Cas9 knockout screening in human cells. Science *343*, 84-87.

Shan, G., Huang, W., Gee, S.J., Buchholz, B.A., Vogel, J.S., and Hammock, B.D. (2000). Isotope-labeled immunoassays without radiation waste. Proc Natl Acad Sci U S A *97*, 2445-2449.

Shapiro, B.M. (1970). Regulation of glutamine synthetase by enzyme catalyzed structural modification. Angew Chem Int Ed Engl *9*, 670-678.

Shiloh, Y. (2003). ATM and related protein kinases: safeguarding genome integrity. Nat Rev Cancer *3*, 155-168.

Singer, R.H., and Ward, D.C. (1982). Actin gene expression visualized in chicken muscle tissue culture by using in situ hybridization with a biotinated nucleotide analog. Proc Natl Acad Sci U S A *79*, 7331-7335.

Singh, A., Kildegaard, H.F., and Andersen, M.R. (2018). An Online Compendium of CHO RNA-Seq Data Allows Identification of CHO Cell Line-Specific Transcriptomic Signatures. Biotechnol J *13*, e1800070.

Slijepcevic, P., and Hande, M.P. (1999). Chinese hamster telomeres are comparable in size to mouse telomeres. Cytogenet Cell Genet *85*, 196-199.

Slijepcevic, P., Xiao, Y., Dominguez, I., and Natarajan, A.T. (1996). Spontaneous and radiation-induced chromosomal breakage at interstitial telomeric sites. Chromosoma *104*, 596-604.

Smilenov, L.B., Mellado, W., Rao, P.H., Sawant, S.G., Umbricht, C.B., Sukumar, S., and Pandita, T.K. (1998). Molecular cloning and chromosomal localization of Chinese hamster telomeric protein chTRF1. Its potential role in chromosomal instability. Oncogene *17*, 2137-2142.

Smith, G.P. (1985). Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. Science *228*, 1315-1317.

Smogorzewska, A., van Steensel, B., Bianchi, A., Oelmann, S., Schaefer, M.R., Schnapp, G., and de Lange, T. (2000). Control of human telomere length by TRF1 and TRF2. Mol Cell Biol *20*, 1659-1668.

Stahlberg, A., Andersson, D., Aurelius, J., Faiz, M., Pekna, M., Kubista, M., and Pekny, M. (2011). Defining cell populations with single-cell gene expression profiling: correlations and identification of astrocyte subpopulations. Nucleic Acids Res *39*, e24.

Stahlberg, A., Rusnakova, V., and Kubista, M. (2013). The added value of single-cell gene expression profiling. Brief Funct Genomics *12*, 81-89.

Takai, H., Smogorzewska, A., and de Lange, T. (2003). DNA damage foci at dysfunctional telomeres. Curr Biol *13*, 1549-1556.

Taylor, M.W., Pipkorn, J.H., Tokito, M.K., and Pozzatti, R.O., Jr. (1977). Purine mutants of mammalian cell lines: III. Control of purine biosynthesis in adenine phosphoribosyl transferase mutants of CHO cells. Somatic Cell Genet *3*, 195-206.

Taylor, R.P. (2006). Of mice and mechanisms: identifying the role of complement in monoclonal antibody-based immunotherapy. Haematologica *91*, 146a.

Teeling, J.L., Mackus, W.J., Wiegman, L.J., van den Brakel, J.H., Beers, S.A., French, R.R., van Meerten, T., Ebeling, S., Vink, T., Slootstra, J.W.*, et al.* (2006). The biological activity of human CD20 monoclonal antibodies is linked to unique epitopes on CD20. J Immunol *177*, 362-371.

Tharmalingam, T., Barkhordarian, H., Tejeda, N., Daris, K., Yaghmour, S., Yam, P., Lu, F., Goudar, C., Munro, T., and Stevens, J. (2018). Characterization of phenotypic and genotypic diversity in subclones derived from a clonal cell line. Biotechnol Prog *34*, 613-623.

Thie, H., Voedisch, B., Dubel, S., Hust, M., and Schirrmann, T. (2009). Affinity maturation by phage display. Methods Mol Biol *525*, 309-322, xv.

Thomas, R., Marks, D.H., Chin, Y., and Benezra, R. (2018). Whole chromosome loss and associated breakage-fusion-bridge cycles transform mouse tetraploid cells. EMBO J *37*, 201-218.

Topp, M.S., Kufer, P., Gokbuget, N., Goebeler, M., Klinger, M., Neumann, S., Horst, H.A., Raff, T., Viardot, A., Schmid, M.*, et al.* (2011). Targeted therapy with the T-cell-engaging antibody blinatumomab of chemotherapy-refractory minimal residual disease in B-lineage acute lymphoblastic leukemia patients results in high response rate and prolonged leukemia-free survival. J Clin Oncol *29*, 2493-2498.

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

Vaickus, L., and Foon, K.A. (1991). Overview of monoclonal antibodies in the diagnosis and therapy of cancer. Cancer Invest *9*, 195-209.

Vcelar, S., Jadhav, V., Melcher, M., Auer, N., Hrdina, A., Sagmeister, R., Heffner, K., Puklowski, A., Betenbaugh, M., Wenger, T., *et al.* (2018a). Karyotype variation of CHO host cell lines over time in culture characterized by chromosome counting and chromosome painting. Biotechnol Bioeng *115*, 165-173.

Vcelar, S., Melcher, M., Auer, N., Hrdina, A., Puklowski, A., Leisch, F., Jadhav, V., Wenger, T., Baumann, M., and Borth, N. (2018b). Changes in Chromosome Counts and Patterns in CHO Cell Lines upon Generation of Recombinant Cell Lines and Subcloning. Biotechnol J *13*, e1700495.

Venkitaraman, A.R. (2003). A growing network of cancer-susceptibility genes. N Engl J Med *348*, 1917-1919.

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

Wang, F., Podell, E.R., Zaug, A.J., Yang, Y., Baciu, P., Cech, T.R., and Lei, M. (2007). The POT1-TPP1 telomere complex is a telomerase processivity factor. Nature *445*, 506-510.

Wang, S.Y., Racila, E., Taylor, R.P., and Weiner, G.J. (2008). NK-cell activation and antibody-dependent cellular cytotoxicity induced by rituximab-coated target cells is inhibited by the C3b component of complement. Blood *111*, 1456-1463.

Wang, S.Y., Veeramani, S., Racila, E., Cagley, J., Fritzinger, D.C., Vogel, C.W., St John, W., and Weiner, G.J. (2009). Depletion of the C3

component of complement enhances the ability of rituximab-coated target cells to activate human NK cells and improves the efficacy of monoclonal antibody therapy in an in vivo model. Blood *114*, 5322-5330.

Wang, S.Y., and Weiner, G. (2008). Complement and cellular cytotoxicity in antibody therapy of cancer. Expert Opin Biol Ther *8*, 759-768.

Wang, T., Birsoy, K., Hughes, N.W., Krupczak, K.M., Post, Y., Wei, J.J., Lander, E.S., and Sabatini, D.M. (2015). Identification and characterization of essential genes in the human genome. Science *350*, 1096-1101.

Wang, W., Jia, Y.L., Li, Y.C., Jing, C.Q., Guo, X., Shang, X.F., Zhao, C.P., and Wang, T.Y. (2018a). Retraction Note: Impact of different promoters, promoter mutation, and an enhancer on recombinant protein expression in CHO cells. Sci Rep *8*, 13482.

Wang, X., Mathieu, M., and Brezski, R.J. (2018b). IgG Fc engineering to modulate antibody effector functions. Protein Cell *9*, 63-73.

Wang, X.Y., Zhang, J.H., Zhang, X., Sun, Q.L., Zhao, C.P., and Wang, T.Y. (2016). Impact of Different Promoters on Episomal Vectors Harbouring Characteristic Motifs of Matrix Attachment Regions. Sci Rep *6*, 26446.

Ward, B.E., Gersen, S.L., Carelli, M.P., McGuire, N.M., Dackowski, W.R., Weinstein, M., Sandlin, C., Warren, R., and Klinger, K.W. (1993). Rapid prenatal diagnosis of chromosomal aneuploidies by fluorescence in situ hybridization: clinical experience with 4,500 specimens. Am J Hum Genet *52*, 854-865.

Waye, M.M., and Stanners, C.P. (1979). Isolation and characterization of CHO cell mutants with altered asparagine synthetase. Somatic Cell Genet *5*, 625-639.

Weiner, G.J. (2010). Rituximab: mechanism of action. Semin Hematol *47*, 115-123.

Weiner, G.J. (2015). Building better monoclonal antibody-based therapeutics. Nat Rev Cancer *15*, 361-370.

Westbrook, T.F., Martin, E.S., Schlabach, M.R., Leng, Y., Liang, A.C., Feng, B., Zhao, J.J., Roberts, T.M., Mandel, G., Hannon, G.J., *et al.* (2005). A genetic screen for candidate tumor suppressors identifies REST. Cell *121*, 837-848.

Williams, B., Stender, H., and Coull, J.M. (2002). PNA fluorescent in situ hybridization for rapid microbiology and cytogenetic analysis. Methods Mol Biol *208*, 181-193.

Williams, S., Mustoe, T., Mulcahy, T., Griffiths, M., Simpson, D., Antoniou, M., Irvine, A., Mountain, A., and Crombie, R. (2005). CpG-island fragments from the HNRPA2B1/CBX3 genomic locus reduce silencing and enhance transgene expression from the hCMV promoter/enhancer in mammalian cells. BMC Biotechnol *5*, 17.

Williamson, R.A., Burioni, R., Sanna, P.P., Partridge, L.J., Barbas, C.F., 3rd, and Burton, D.R. (1993). Human monoclonal antibodies against a plethora of viral pathogens from single combinatorial libraries. Proc Natl Acad Sci U S A *90*, 4141-4145.

Worton, R.G., Ho, C.C., and Duff, C. (1977). Chromosome stability in CHO cells. Somatic Cell Genet *3*, 27-45.

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

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

Yamane-Ohnuki, N., Kinoshita, S., Inoue-Urakubo, M., Kusunoki, M., Iida, S., Nakano, R., Wakitani, M., Niwa, R., Sakurada, M., Uchida, K., *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.

Yang, Y., Mariati, Chusainow, J., and Yap, M.G. (2010). DNA methylation contributes to loss in productivity of monoclonal antibody-producing CHO cell lines. J Biotechnol *147*, 180-185.

Ye, J.Z., and de Lange, T. (2004). TIN2 is a tankyrase 1 PARP modulator in the TRF1 telomere length control complex. Nat Genet *36*, 618-623.

Ye, J.Z., Donigian, J.R., van Overbeek, M., Loayza, D., Luo, Y., Krutchinsky, A.N., Chait, B.T., and de Lange, T. (2004a). TIN2 binds TRF1 and TRF2 simultaneously and stabilizes the TRF2 complex on telomeres. J Biol Chem *279*, 47264-47271.

Ye, J.Z., Hockemeyer, D., Krutchinsky, A.N., Loayza, D., Hooper, S.M., Chait, B.T., and de Lange, T. (2004b). POT1-interacting protein PIP1: a telomere length regulator that recruits POT1 to the TIN2/TRF1 complex. Genes Dev *18*, 1649-1654.

Yerganian, G. (1972). History and cytogenetics of hamsters. Prog Exp Tumor Res *16*, 2-34.

Young, A.B., Greenamyre, J.T., Hollingsworth, Z., Albin, R., D'Amato, C., Shoulson, I., and Penney, J.B. (1988). NMDA receptor losses in

putamen from patients with Huntington's disease. Science *241*, 981-983.

Yusufi, F.N.K., Lakshmanan, M., Ho, Y.S., Loo, B.L.W., Ariyaratne, P., Yang, Y., Ng, S.K., Tan, T.R.M., Yeo, H.C., Lim, H.L.*, et al.* (2017). Mammalian Systems Biotechnology Reveals Global Cellular Adaptations in a Recombinant CHO Cell Line. Cell Syst *4*, 530-542 e536.

Zeng, J., Nagrajan, H.K., and Yi, S.V. (2014). Fundamental diversity of human CpG islands at multiple biological levels. Epigenetics *9*, 483-491.

Zhang, M.Y., Xiao, X., Sidorov, I.A., Choudhry, V., Cham, F., Zhang, P.F., Bouma, P., Zwick, M., Choudhary, A., Montefiori, D.C.*, et al.* (2004a). Identification and characterization of a new cross-reactive human immunodeficiency virus type 1-neutralizing human monoclonal antibody. J Virol *78*, 9233-9242.

Zhang, S., Hemmerich, P., and Grosse, F. (2004b). Nucleolar localization of the human telomeric repeat binding factor 2 (TRF2). J Cell Sci *117*, 3935-3945.

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

Zhu, J., Tsai, H.J., Gordon, M.R., and Li, R. (2018). Cellular Stress Associated with Aneuploidy. Dev Cell *44*, 420-431.

Book references

Alexandru Dan Corlan (2012). Medline trend: automated yearly statistics of PubMed results for any query, 2004. Web resource at

URL:http://dan.corlan.net/medline-trend.html. Accessed: 2012-02-14. (Archived by WebCite at http://www.webcitation.org/65RkD48SV)

Bate, S., & Clark, R. (2014). *The Design and Statistical Analysis of Animal Experiments*. Cambridge: Cambridge University Press. doi:10.1017/CBO9781139344319

de Souza, N (2012). Single-cell methods. *Nat Methods* **9,** 35.

Heller-Harrison RA, Crowe K, Leonard M, Cooley C, Hone M, Mccarthy K (2009). Managing cell line instability and its impact during cell line development. BioPharm Int 2009, 4:1-4

Janeway CA Jr, Travers P, Walport M, et al. Immunobiology: The Immune System in Health and Disease. 5th edition. New York: Garland Science; 2001. The structure of a typical antibody molecule.

James P. Allison – Facts (2018). NobelPrize.org. Nobel Media AB 2019. Sat. 22 Jun 2019