The use of site-directed integration to study genomic and transcriptional stability of recombinant promoters in CHO cells

A thesis submitted to The University of Manchester
for the degree of Doctor of Philosophy in
The Faculty of Life Sciences

2015

Mário Rui Pimentel Pereira

Contents

List of Figures

List of Tables

Abstract

Transcriptional regulation is a determinant of stability of recombinant protein production in CHO cells. Fundamental studies of recombinant gene transcription in relation to chromatin environment and promoter regulation are important for CHO cell line development and selection. This study has developed a methodology based on a cell/vector system to study recombinant transcription and expression stability of different promoters and/or proteins in the similar genomic environment.

The CHO-FRT mini-pools developed in this project were mini-pools of CHO-S cell lines containing Flp Recombination Target (FRT) sites with ß–galactosidase gene, under the influence of a SV40 promoter. Continuous culture of these mini-pools for 8 weeks using a robotic system demonstrated that 20% of the mini-pools studied revealed an unstable profile (with 30% loss of protein expression). Two of these mini-pools with different characteristics, CHO-FRT 1 (low producer/unstable) and CHO-FRT 108 (high producer/stable), were selected to be used on the study of influence of SV40 and CMV promoters in long-term recombinant expression.

Genes encoding fluorescent proteins were integrated in a site-directed manner under the influence of SV40 or CMV promoters. A sub-clonal population of the top 10% yellow fluorescent protein (YFP) expressing cells of each mini-pool/promoter combination was selected by cell sorting and cultured for 4 weeks. During this period protein expression was monitored by flow cytometry and compared between both promoters. The results revealed that both SV40 and CMV promoters had an unstable expression with different degrees of instability and long-term expressing behaviours. For CMV, instability was considerably high displaying a long-term logarithmic loss of 50-80% of productivity while for SV40 the loss of productivity observed was only 40-45% with a linear behaviour during long-term culture.

The vector system generated contained an MS2-RNA tag sequence cloned 3'- of the recombinant gene to track the recombinant mRNA by using the MS2/MCP-GFP system. This study showed the development of a protocol to measure the transcriptional output of recombinant promoters in CHO cells. The results showed background signal in CHO cells that requires further optimisation studies to allow the direct live cell image quantification of the transcriptional activity of recombinant promoters. Although not yet optimised, the successful combination of site-directed integration with recombinant mRNA tagging method has the potential to become a valuable tool to study the mechanisms of transcriptional activity and stability of transcription driven by different promoters in CHO cells.

Declaration

No portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning;

Copyright Statement

The following four notes on copyright and the ownership of intellectual property rights must be included as written below:

i. The author of this thesis (including any appendices and/or schedules to this thesis) owns certain copyright or related rights in it (the "Copyright") and s/he has given The University of Manchester certain rights to use such Copyright, including for administrative purposes.

ii. Copies of this thesis, either in full or in extracts and whether in hard or electronic copy, may be made only in accordance with the Copyright, Designs and Patents Act 1988 (as amended) and regulations issued under it or, where appropriate, in accordance with licensing agreements which the University has from time to time. This page must form part of any such copies made.

iii. The ownership of certain Copyright, patents, designs, trade marks and other intellectual property (the "Intellectual Property") and any reproductions of copyright works in the thesis, for example graphs and tables ("Reproductions"), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property and/or Reproductions.

iv. Further information on the conditions under which disclosure, publication and commercialisation of this thesis, the Copyright and any Intellectual Property and/or Reproductions described in it may take place is available in the University IP Policy (see http://documents.manchester.ac.uk/DocuInfo.aspx?DocID=487), in any relevant Thesis restriction declarations deposited in the University Library, The University Library's regulations (see http://www.manchester.ac.uk/library/aboutus/regulations) and in The University's policy on Presentation of Theses

Acknowledgments

First and foremost I would like to sincerely thank my supervisor, Alan Dickson, for his advice, guidance and mentoring during this project. It is with no doubt, that without his support this project would not have been possible. Additionally, I also would like to thank him for his huge patience that allow him to answer with great fairness, and/or great humour, the questions that I have been "hhhasking" for the past four years.

I would like to thank all the past and present members of the "Dickson Ducks" that I had the pleasure to work with. Believe me, without you guys all this journey would be much more painful and definitely not so special. I would like to especially thanks Hirra, for the help and patience during the writing of this thesis, Eric, for sharing with me my first entrepreneurial adventures, and above all, Alejandro, because more than being a "lab mate", he became a friend.

I want to thank all my friends especially, Camões, Joana, Miguel, Comprido e Zé for all the support during tough moments. You are the proof that real friendship overcomes any distance.

Finally, I would like to thank Fundação para a Ciência e Tecnologia (FCT) for funding my PhD project.

Dedications

I dedicate this thesis to my family for their unconditional love and support.

Esta tese é especialmente dedicada à minha Mãe e ao meu Pai, por sempre acreditarem em mim e por serem uma inspiração para querer ser sempre melhor, e à Catarina, por continuar a ser o meu pirralho e a alegria da família.

Abbreviations

3C	Chromosome Conformation Capture
4C	Circular Chromosome Conformation Capture
5C	Carbon-Copy Chromosome Conformation Capture
Bre	TFIIB Recognition Element
CDK	Cuclin Dependent Kinase
ChIA-PET	Chromatin Interaction Analysis By Paired-End Tag Sequencing
CHO	Chinese Hamster Ovary
CHO-FRT	CHO-FRT mini-pool
CMV	Cytomegalovirus
CpG	Citosine-Guanine Palindromic
CPRG	Chlorophenolred-ß-D-Galactopyranoside
CPRG	Chlorophenol Red-ß-D-Galactopyranoside
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
CT	Chromossome Territory
CTD	C-Terminus Domain
ddH2O	Double-distilled Water
DHFR	Dihydrofolate
DNA	Deoxyribonucleic Acid
DNMT	Dna Metyltransferase
E. coli	*Escherichia coli*
EPO	Erythropoietin
ER	Endoplasmic Reticulum
FACS	Fluorescence-Activated Cell Sorting
FISH	Fluorescence In Situ Hybridisation
FRAP	Fluorescence Recovery After Photobleaching
FRT	Flp-Recombinase Target
GFP	Green Fluorescence Protein
HAT	Histone Acetyltransferase
HDAC	Histone Deacetylase
HEK293	Human Embryonic Kidney 293 Cells
Hi-C	Genome-Wide Chromosome Conformation Capture
HMT	Histone Methyltransferase
HP1	Heterochormatin-Like Protein 1
Ig	Immunoglobulin

Inr	Transcription Initiation region
K.O.	Knocked-Out
LB	Luria-Bertuni
LCR	Locus Control Region
mAb	Monoclonal Antibody
MALDI-TOF	Matrix-Assisted Laser Desorption/Ionization - Time Of Flight
MCP	MS2 Coat Protein
MS2	MS2 Bacteriophage
MTX	Metrotexate
NLS	Nuclear Localisation Signal
ONPG	O-Nitrophenyl--D-Galactopyranoside
P. Pastoris	*Picchia Pastoris*
PCR	Polymerase Chain Reaction
PIC	Pre-Initiation transcriptional Complex
PP7	PP7 Bacteriophage
PTM	Post-Translational Modification
qPCR	Quantitative Polymerase Chain Reaction
rAAV	Recombinant Adeno-Associated Virus
RNA	Ribonucleic Acid
RNA-TRAP	RNA-FISH Tagging And Recovery Of Associated Proteins
RNAP II	RNA Polymerase II
S. cerevisiae	*Saccharomyces cerevisiae*
S/Mars	Scaffold / Matrix Attachment Regions
STRB	mStrawberry Fluorescent Protein
SV40	Simian Virus 40
TAF	TBP Associated Factors
TALENs	Transcription Activator-Like Effector Nucleases
TBP	TATA Box Binding Protein
TFII	RNAP II Transcription Factor
TFREs	Transcription Factor Regulatory Elements
UCOE	Ubiquitous Opening Elements
UTR	Un-Translated Region
YFP	Yellow Fluorescent Protein
ZFN	Zinc-Finger Nuclease

1. Introduction

1.1. Introductory remarks

The term 'biopharmaceutical' was first used in the 1980s and to date it has not been well defined. Despite the lack of consistent definition, biopharmaceuticals are known as a class of therapeutic products produced by modern biotechnological techniques (Rader 2008; Walsh 2002). The majority of biopharmaceuticals are recombinant proteins and monoclonal antibodies (mAbs) expressed in heterologous systems (Walsh 2002). However the definition also includes gene therapy products and anti-sense oligonucleotides (Walsh 2002).

The first recombinant protein approved for medical use was recombinant insulin expressed in *Escherichia coli* (*E. coli*) (Johnson 1983). Since then, there has been a significant increase in the number of biopharmaceuticals within European Union and United States with over 212 products being approved with a market value of $140 billion in 2013 (Walsh 2014). Table 1.1, ranks the 10 top-selling biopharmaceuticals products in 2013.

Table 1.1. The 10 top-selling biopharmaceuticals in 2013.
(Source of data: Walsh 2014)

Rank	Commercial Name	Active Ingredient	Molecule Type	Expression System	Company	Sales ($ billion)
1	Humira	Adalimumab	mAb	CHO cell line	AbbVie	11.00
2	Enbrel	Etanercept	Fusion protein	CHO cell line	Amgen Pfizer Takeda	8.76
3	Remicade	Infliximab	mAb	Sp2/O cell line	Johnson & Johnson Merck & Co.	8.37
4	Lantus	Insulin Glargine	Hormone	E. coli	Sanofi	7.95
5	Rituxan/ MabThera	Rituximab	mAb	CHO cell line	Roche/Genentech Biogen Idec	7.91
6	Avastin	Bevacizumab	mAb	CHO cell line	Roche/Genentech	6.97
7	Herceptin	Trastuzumab	mAb	Murine cell line	Roche/Genentech	6.91
8	Neulasta	Pegfilgrastim	Growth Factor	E. coli	Amgen	4.39
9	Lucentis	Ranibizumab	mAb	E. coli	Roche/Genentech	4.27
10	Epogen/ Procrit	Epoetin Alfa	Growth Factor	CHO cell line	Amgen Johnson & Johnson	3.35

Most of the approved therapeutic proteins are produced in bacteria, yeast and mammalian cell systems (Walsh 2014). Alternative recombinant expression systems for the production of therapeutic proteins include insect cell lines (Chang et al. 2003), transgenic plants (Goldstein and Thomas 2004) and transgenic animals (Kling 2009). The different expression systems will be discussed in more detail in Section 1.2 with an emphasis on Chinese Hamster Ovary (CHO) cells due to the use of this expression system in this project.

For over 20 years, CHO cell lines have been the most widely used expression systems in the biopharmaceutical industry. Out of the 54 new products approved between 2010-2014 18 were produced in CHO cells (Walsh 2014). This popularity is mainly due to the accumulation of extensive testing and safety data which allows easier approval by the regulatory authorities (Jayapal et al. 2007).

Figure 1.1. Expression systems used to manufacture approved biopharmaceutical products.
Product approvals (1982–2014) in the context of expression systems employed. Each data set is expressed as a percentage of total biopharmaceutical product approvals. (Source of data: Walsh 2014)
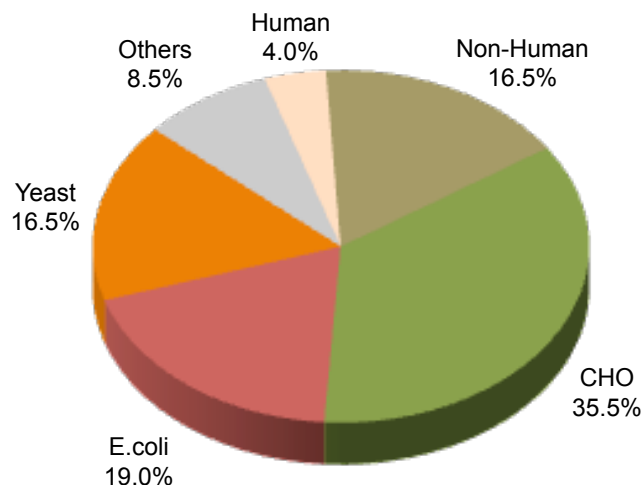
Source: Walsh 2014

One of the main issues of using certain CHO cell lines for the production of therapeutic proteins is loss of constant recombinant protein production (Barnes et al. 2003b). The maintenance or loss of constant production is termed cell line stability or instability, respectively (Bailey et al. 2012; Barnes et al. 2003a). The commercial consequence of this loss of productivity by the selected cell line can result in

problems for process yields, effective usage of time and money, and regulatory approval (Barnes et al. 2003a).

There are many cellular events that can influence the stability of CHO cell lines (Barnes and Dickson 2006; Hussain et al. 2014). However due to the focus of this project, the genomic and transcriptional mechanisms that can influence cell line stability will be discussed in detail (Section 1.3). Furthermore, recent advances in the prediction and prevention of cell line instability at the genomic and transcriptional level will be discussed (Section 1.4).

1.2. Expression Systems

Most biopharmaceuticals approved for human therapy require specific post-translational modifications (PTM) such as glycosylation. PTMs affect protein folding, stability, trafficking, immunogenicity and activity (Walsh and Jefferis 2006). Glycosylation is a well characterised PTM required for production and function of many proteins (Walsh and Jefferis 2006). In eukaryotic cells, glycosylation occurs in the endoplasmic reticulum (ER) and Golgi apparatus as the protein transits through the secretory pathway (Braell et al. 1984).

This Section will describe the relative advantages and disadvantages of bacterial, yeast (Section 1.2.1) and mammalian systems (Section1.2.2) specifically CHO cells (Section 1.2.3).

1.2.1. Bacterial and Yeast systems

Bacterial systems are widely used to express recombinant proteins. The simple cultivation method, short generation period, good scale-up potential and high yields of recombinant protein are the main advantages of these systems (Baneyx 1999). *E. coli* has been used successfully to express a variety of small therapeutic proteins such as insulin, human growth hormone, human tissue type plasminogen activator and growth factors (Jensen and Carken 1990; Johnson 1983; Joly et al. 1998; Qiu et al. 1998). However, a major disadvantage of the *E. coli* system is the inability to perform complex PTMs, especially glycosylation (Walsh and Jefferis 2006).

Yeast systems such as *Saccharomyces cerevisiae* (*S. cerevisiae*) and *Pichia pastoris* (*P. pastoris*) are alternatives to bacterial systems due to their ability to perform some PTMs. However, the oligosaccharide composition in yeast proteins is

different from those in human glycoproteins (Gemmill and Trimble 1999). As a consequence, the high mannose content of yeast-directed PTM oligosaccharides can affect the overall properties of protein (Walsh and Jefferis 2006). In addition, different glycans have been shown to induce inappropriate immunogenicity in therapeutic application (Hopkins et al. 2011; Trimble et al. 2004).

In 2006, a genetically engineered *P. Pastoris* strain (GlycoFi) was developed that was capable of performing glycosylation identical to that in mammalian cells (Hamilton et al. 2006; Li et al. 2006). Although this strain increased the amount of sialylated glycans, which lead to an increase of activity and half-life of some therapeutic proteins, it failed to produce other types of PTMs (Hamilton et al. 2006; Li et al. 2006). In 2009, a PEGylated form of erythropoietin produced in this system was reported to be in phase 3 clinical trials (Ratner 2009). However, to date, no product using this genetically engineered *P. Pastoris* has been launched.

1.2.2. Mammalian systems

In contrast to bacterial and yeast systems, the production of the therapeutic proteins using mammalian cells is highly complex and expensive (Walsh and Jefferis 2006). Although this can be considered a disadvantage compared with bacterial and yeast systems, the ability of mammalian cells to perform PTMs similar to humans makes mammalian cell systems the major systems used by industry to produce therapeutic proteins (Butler and Spearman 2014). This Section will focus on the human (Section 1.2.2.1) and non-human (Section 1.2.2.2) mammalian cell lines used to produce therapeutic proteins.

1.2.2.1. Human cell lines

The advantage of using human cell lines for protein expression is the ability to produce an authentic human glycan profile (Butler and Spearman 2014). Human embryonic kidney cells (HEK293) and human embryonic retinal cells (PER.C6) are two mammalian cell lines derived from primary culture. Both cell lines were immortalized by transfection with an early region 1 (E1) of adenovirus type 5 (Ad5) (Fallaux et al. 1998; Graham 1987; Graham et al. 1977).

HEK293 cells are well characterised, mainly due to their extensive use for research grade recombinant proteins (Pham et al. 2006). The fact that this cell type can grow in suspension in serum-free media, and can be easily transfected at a large scale,

make them desirable vehicles for the production of therapeutic proteins (Pham et al. 2006). Nevertheless, to date, only 3 marketed products have been manufactured using this cell line (Dietmair et al. 2012; Walsh 2014).

The PER.C6 cell line was originally used for the production of clinical-grade influenza vaccines (Pau et al. 2001). However, the ability to grow in suspension, achieve high cell densities (up to 1×10^7 cell/ml) and perform human PTM makes this cell line a very desirable cell line for protein production (Butler and Spearman 2014; Pau et al. 2001). Furthermore, it has the potential to achieve very high productivities. One of the first reports of the use this cell line to produce a therapeutic product reported a productivity of 300-500mg/L of an IgG mAb without medium feed addition (Jones et al. 2003). Since then a few reports have shown higher productivities with one recent report indicating a record of productivity titre of 30g/L when expressing one antibody (Butler and Spearman 2014)

1.2.2.2. Non-human cell lines

Other than CHO cells (discussed in Section 1.2.3), the most commonly used non-human cell line for the production of therapeutic proteins are baby hamster kidney cells (BHK-21) and murine myeloma cell lines (NS0 and Sp2/0). NS0 and Sp2/0 are mostly used in the production of mAbs due to their protein yield and capability of producing highly sialylated glycan products (Byrne et al. 2007; Yoo et al. 2002). Furthermore, NS0 cells have been shown to give good and stable expression over an extended period of batch culture (Barnes et al. 2003a).

BHK-21 cells are mainly used in the production of vaccines (Durocher and Butler 2009). However, this cell line is also used to produce complicated coagulation factors (Butler and Spearman 2014). Factor VIIa and Factor VIII are highly complex molecules that required specific PTMs that can be provided by this cell line (Ishaque et al. 2007; Nivitchanyong et al. 2007; Soukharev et al. 2002)

1.2.3 CHO cells

Since the commercialization of the first approved recombinant therapeutic produced from a mammalian cell line, (human tissue plasminogen activator-tPA from CHO cells) CHO cell lines have dominated the production of therapeutic proteins (Deschênes et al. 1998; Wurm 2004). The ability to grow at high cell densities in suspension culture with unparalleled adaptability to different conditions (such as

serum-free medium) has made this cell line the mammalian "workhorse" for protein production (Bandaranayake and Almo 2014).

1.2.3.1. CHO cell History

Chinese Hamster (*Cricetulus griseus)* was initially used as animal and cell model for genetic analysis. This was due to its low chromosome number (2n=22) and easy identification of each chromosome by size (Ono and Sonta 2001; Tjio and Puck 1958; Yang et al. 2000).

CHO cells were cultivated and immortalised by mutagenesis in 1957 by Tijo and Puck in order to study the chromosomal composition of cells in tissue culture (Tjio and Puck 1958). The continuous subculture of the original CHO cells lead to the generation of a proline-dependent sub-clone (CHO-Pro$^-$) (Kao and Puck 1967) (Figure 1.2). One of the proline-dependant sub-clones was named as the CHO-K1 cell line, the parental cell line for most of the cell lines used in industry today (Hacker et al. 2009; Kao and Puck 1968). Since then, different cell lines have been generated either by sub-clone isolation or mutagenesis selection.

Figure 1.2, shows the genealogy of CHO cell lines derived from the original CHO cells cultured by Puck in 1957 to the cell lines that are commonly used in industry today such as CHOK1SV, CHO-S, CHO DG44 and CHO-DXB11.

Figure 1.2. CHO cell line lineage.

Bold fonts represent the most widely used CHO cell lines for recombinant protein expression studies; Blue lines indicate the process of cell line generation.

Cell isolation and immortalisation

CHO cell
(Tjio & Puck et al 1958)

Subcloning

CHO pro⁻
(Kao & Puck 1966)

Subcloning

Subcloning

CHO-K1
(Kao & Puck 1968)

CHO-pro3⁻
(Flintoff et al 1976)

Metrotexate treatment and clone selection

CHO-pro3⁻ MtxRIII
(Flintoff et al 1976)

Subcloning

[³H]deoxyuridine mutagenesis and clone selection

[³H]deoxyuridine mutagenesis and clone selection

Subcloning

CHOK1SV

CHO-DXB11
(Urlaub & Chasin, 1980)

CHO-DG44
(Urlaub et al 1983)

CHO-S
(Deaven & Petersen 1973; D'Anna 1996)

1.2.3.2. CHO genome and karyotype

The karyotype of different CHO cell lines has been extensively characterised. CHO cells have an unstable karyotype as observed by Puck and Tijo in 1958. Original CHO cells had 22 chromosomes with a variation between 21 or 23 chromosomes (Tjio and Puck 1958). This natural karyotype instability of CHO cells was later confirmed in several karyotype characterisations of different CHO cell lines (Deaven and Petersen 1973; Derouazi et al. 2006; Kao and Puck 1969). By comparing the karyotype of CHO cell lines with the karyotype of Chinese Hamster, these studies showed CHO cell lines to have undergone extensive chromosomal re-arrangement, including translocations, homologous recombination and deletions. Table 1.2 shows the results obtained from karyotype profiling.

Table 1.2. Karyotype characterisation of different CHO cell lines.
Chromosome nomenclature was defined in conjunction with Chinese hamster karyotype (Conserved chromosome – chromosomes observed in Chinese hamster karyotype; Z chromosome – chromosomes that are not observed in Chinese hamster karyotype; Derivative chromosome - structurally altered chromosomes by one or more rearrangements of different chromosomes; Marker chromosome - structurally abnormal with no recognised and defined landmark).

	Number of Chromosomes	Conserved Chromosomes	Z chromosomes	Derivative chromosomes	Marker chromosomes	
Chinese Hamster	22					(Tijo and Puck 1958)
CHO	22 (21 – 23)					(Tijo and Puck 1958)
CHO-pro⁻	21	12	9			(Kao and Puck 1969)
		8	13			(Deaven and Petersen 1973)
CHO-K1	20	12	8			(Kao and Puck 1969)
CHO DG44	20	7	4	7	2	(Derouzi et al. 2006)

Although the CHO karyotype has been extensively characterised over the years using different cell lines, the first CHO cell genome was only fully sequenced in 2011 (Xu et al. 2011). The CHO cell line sequenced was CHO-K1 and initial analysis showed an assembly of 2.45Gb genomic sequence with 24,383 predicted genes (Xu et al. 2011). In 2013, two more groups published the sequence of a total of 9 different CHO cell lines from the lineage tree of CHO-K1, CHO-DG44 and CHO-S (Brinkrolf et al. 2013; Lewis et al. 2013). From their studies, both groups observed that the CHO-K1 genome published by Xu et al 2011 was not fully suitable for use as a reference genome for other CHO cell lines due to differences in chromosome numbers and sequence variations (Derouazi et al. 2006). Both studies recommended the use of the genome of female Chinese Hamsters (*Cricetulus griseus*) as a genome reference (Brinkrolf et al. 2013; Lewis et al. 2013).

Despite the activity of several groups with published sequences of different CHO genomes (plus information publicly available at www.chogenome.org) there is not yet a full detailed and organised set of information to fully understand the relevant genes that give each CHO cell line variant its specific characteristics (Baik and Lee 2014).

1.2.3.3. Gene amplification

One of the main strategies used to improve recombinant protein productivity is gene amplification. The generation of DHFR-deficient cell lines CHO DG44 and CHO-DXB11 made possible to use the DHFR gene as a selection and amplification marker in the expression of recombinant genes. DHFR is responsible for reduction of folate

to dihydrofolate and then to tetrahydrofolate (Kaufman et al. 1983). Tetrahydrofolate is the precursor in the generation of purines and some amino acids, such as glycine, hypoxanthine and thymidine (Kaufman et al. 1983). Due to this fact, the culture medium "normal" CHO DG44 and CHO-DXB11 is supplemented with glycine, hypoxanthine and thymidine (Pallavicini et al. 1990). The re-introduction of the DHFR gene into these cell lines restores glycine, hypoxanthine and thymidine production, eliminating the dependence of these nutrients (Florin et al. 2011). The co-cloning of the DHFR gene in the same vector as the biopharmaceutical for production, under the influence of the same or different promoter, leads to the co-integration of both genes into the cell genome bar. This co-integration allows clone selection by growing cells in medium deficient in glycine hypoxanthine and thymidine (Kaufman et al. 1983).

The major advantage of use of DHFR as a selective marker is that amplification of the DHFR gene can be achieved by use of the DHFR inhibitor methotrexate (MTX) (Kaufman et al. 1983). MTX is a folic acid analogue that inhibits DHFR by competing with the natural substrate (Kaufman et al. 1983). By treating the cell with increasing concentrations of MTX, survivor cells express increased amounts of DHFR ("amplification" the DHFR gene) (Kaufman et al. 1983). Since the DHFR gene and the gene of interest were cloned together, during the DHFR gene amplification process, the gene of interest is co-amplified (Pallavicini et al. 1990).

The influence of gene integration and amplification in CHO DHFR-deficient cell lines has been extensively investigated (Pallavicini et al. 1990). In all studies, after gene amplification, chromosome abnormalities were observed, especially the appearance of additional genetic material in chromosomal arms. The appearance of additional genetic material in chromosomes accounts for the most common chromosomal abnormalities observed in CHO cells (Derouazi et al. 2006; Kaufman et al. 1983; Kim and Lee 1999; Lattenmayer et al. 2006; Yoshikawa et al. 2000b). Nevertheless other genomic rearrangement process, such as translocations and deletions have also been observed (Ruiz and Wahl 1990).

There is no general consensus about the best location for amplification to be effective. Kim et al (1999) and Yoshikawa et al (2000) observed that integration and amplification of recombinant genes near to telomeres lead to a better stability of recombinant protein expression (Kim and Lee 1999; Yoshikawa et al. 2000a). In contrast, Lettenmeyer et al (2006) and Derouzi et al (2006) observed that stability of recombinant expression was greater when amplification occurred in non-telomeric regions (Derouazi et al. 2006; Lattenmayer et al. 2006). Also, in both of the latter

studies, recombinant genes were localised to chromosomes 1 and 2 (the two largest chromosomes) (Lattenmayer et al. 2006), Consequently, the authors suggested that the integration of recombinant gene into larger chromosomes was preferential to obtain a stable cell line (Lattenmayer et al. 2006).

1.2.3.4. CHO cell line stability

The word stable in relation to CHO cell lines can be used to describe two different situations. According to Barnes et al., stable can be used to describe either the type of recombinant expression used in expression system (stable expression) or to describe the ability of cell line to maintain a constant protein production in long-term culture (stable cell line) (Barnes et al. 2003a). The stable integration of the recombinant gene into the CHO genome does not guarantee stable recombinant protein expression. Decreased protein yields in CHO cells have been reported to occur in long-term culture (Fann et al. 2000; Jun et al. 2006; Kim et al. 1998; Pallavicini et al. 1990).

The production of recombinant proteins is a complex process that involves several steps, including gene integration, transcription, posttranscriptional processing, translation, posttranslational processing and, for most of therapeutic proteins, secretion (Figure 1.3) (Barnes et al 2003).

Figure 1.3 Illustration of the recombinant protein expression pathway.
Each step has several regulation sites that can potentially affect protein expression. (Adapted from Barnes et al. 2003)
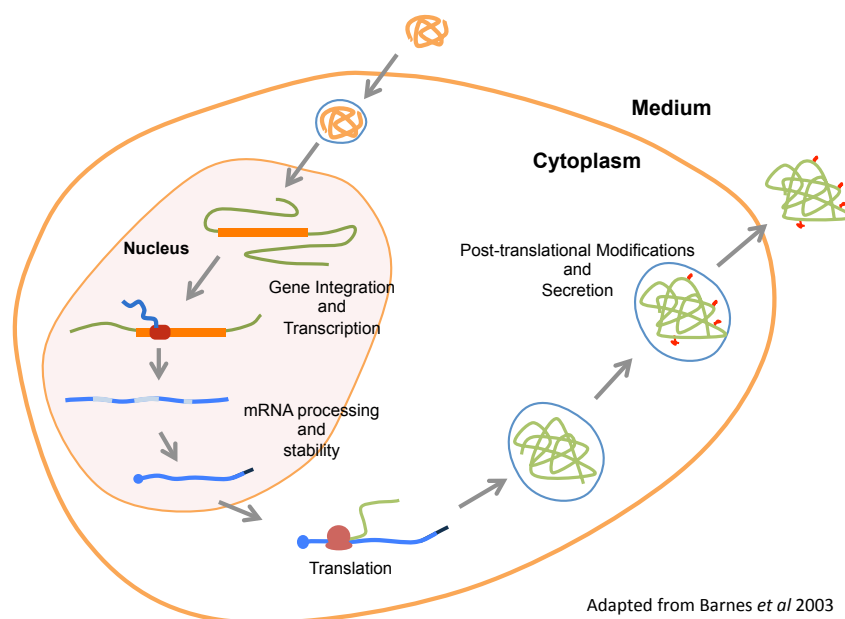
Traditionally transcription has been considered to be the dominant factor that controls long-term protein expression (Barnes and Dickson 2006). However, because the definition of stability is measured at the protein level, any loss of protein production could also arise by the efficiency of each of the individual points of control of the stages mentioned above (Barnes and Dickson 2006).

One of the main reasons for instability in CHO cells is the loss of recombinant genes from the host genome, particularly after amplification processes (Fann et al. 2000; Jun et al. 2006; Kim et al. 1998). Furthermore, the down-regulation of transcription can also occur by transcriptional gene silencing (Holliday and Ho 1998; Paulin et al. 1998). A description of transcriptional gene silencing and an understanding of the causes that can cause this phenomenon are described further in Section 1.3.

1.3. Transcriptional Gene Silencing

Transcriptional gene silencing, observed without the loss of the gene from the genome, can be controlled by regulatory mechanisms that respond to the chromatin environment around the transgene (Holliday and Ho 1998; Paulin et al. 1998). The control mechanisms of the chromatin environment are related to epigenetic regulation that defines the heritable changes in gene expression without a change in

DNA sequence (Wolffe and Matzke 1999). Epigenetic is a term that describes how cells regulate gene expression in response to the environment (Wolffe and Matzke 1999). As transcription occurs in the nucleus (a three-dimensional, compartmentalised structure) it is important to understand the relationships between nuclear structures and recombinant gene expression (Cook 2002). For that reason, Section 1.3.1 will describe chromatin structure focusing on the epigenetic events that lead to gene silencing. Then, Section 1.3.2 will focus on the nuclear organization and its influence on gene transcription. Finally, Section 1.3.3 will focus on the influence of compartmentalisation of transcription in transcription factories. While CHO cells are the system used in this project, this Section will describe regulatory mechanisms elucidated from general mammalian studies due to the shortage of these descriptive studies in CHO cells.

1.3.1. Chromatin Structure and Nuclear organisation

Eukaryotic DNA is compacted and organised in a complex of DNA protein called chromatin (Kornberg 1974). Chromatin is a dynamic structure that not only helps the package of the genomic DNA into the nucleus as also regulates the DNA accessibility to the transcription, recombination, reparation and replication machinery (Razin et al 2007). The chromatin structure is based on nucleosome core subunit, linker DNA and a linker histone that repeats every 200±40 bp throughout the eukaryotic genome (Luger et al 1997).

The nucleosome core subunit is formed of an octamer of histone proteins that consists of two copies of each histone, H2A, H2B, H3 and H4, wrapped by the 145-147 base pairs of DNA (Figure 1.4-A) (Finch et al 1977, Kornberg 1974, Luger et al 1997). The histone octamer core is formed by three histone subunits, a main tetramer subunit composed of histone H3 and H4, $(H3/H4)_2$ and two smaller dimer subunits composed of histone H2A and H2B (H2A/H2B) (Eickbush and Moudrianakis 1978). Histones are a small basic protein, approximately 11-15 kDa, rich in lysine and arginine residues in the N-termini (Razin et al. 2007). These basic residues are important for the histone-DNA and histone-histone interactions (Axel et al. 1974; Eickbush and Moudrianakis 1978). The H2A/H2B dimers are responsible for the interaction with DNA at the entry and exit regions of the nucleosome while the $(H3/H4)_2$ tetramer interacts with the central part of the DNA fragment wound around the globule (Razin et al. 2007). On the other hand, the C-terminal domain (CTD) of the histone is responsible for the histone-histone interactions (Arents et al 1991). The

CTD fold is identical between the four histones that form the core nucleosome. This domain is constituted by three α-helices domains, two small ones (α1 and α3) of 9 and 14 residues and a relative long (α2) of 29 residues, connected by short loop/ß-segments (Figure 1.14-B) (Arents et al 1991). Additional α-helices are also found outside of this CTD of some of the histones and are important for the direction of the dimerization of H2A with H2B and H3 with H4 (Figure 1.4-B) (Arents et al 1991; Cutter and Hayes 2015).

The linker histones (H1) are a group of proteins that are present on the nucleosomes of higher eukaryotic organisms and have been related with various functions (Cutter and Hayes 2015). Among various functions, H1 stabilises the wrapped DNA, protects DNA from nuclease digestion, promotes the folding and assembly of higher order of chromatin structures and regulates gene expression (Cutter and Hayes 2015).

The existence of higher order of chromatin structure in nuclei still remains to be elucidated. *In vitro*, this structure is very well described as the 30-nm fiber (Tremethick 2007) with two main models, solenoid (Robinson et al 2006) and zig-zag (Schalch et al 2005). In vivo no higher order structures has been observed, however due to highly dynamic composition of chromatin *in vivo* and the fact different structures are observed in different biophysical conditions *in vitro*, is possible that the formation of this 30-nm fibre is variable and highly dynamic *in vivo*, varying between the different cell states during the cell cycle (Maeshima et al 2010, Song et al 2014).

Figure 1.4. Nucleosome core and chromatin fiber.

A – Top: Schematic representations of the histone protein organisation in the nucleosome core; Bottom: linear representation of the primary contacts between the core histones protein. Core histone dimerisation partners separated by ━ and dimer-dimer interactions represented by : ;B - Schematic representation of the secondary structure of the core histones proteins. Each α-helix represented by columns. The 3-helix histone fold domain are inside of the grey box area; C – Schematic representation of nucleosome organization along the chromatin. (Yellow – histone H1, dark red – histone H2A; light red – histone H2B; green – histone H3; blue – histone H4; DNA backbone – orange) (based on: Cutter and Hayes 2015; Venkatesh and Workman 2015)
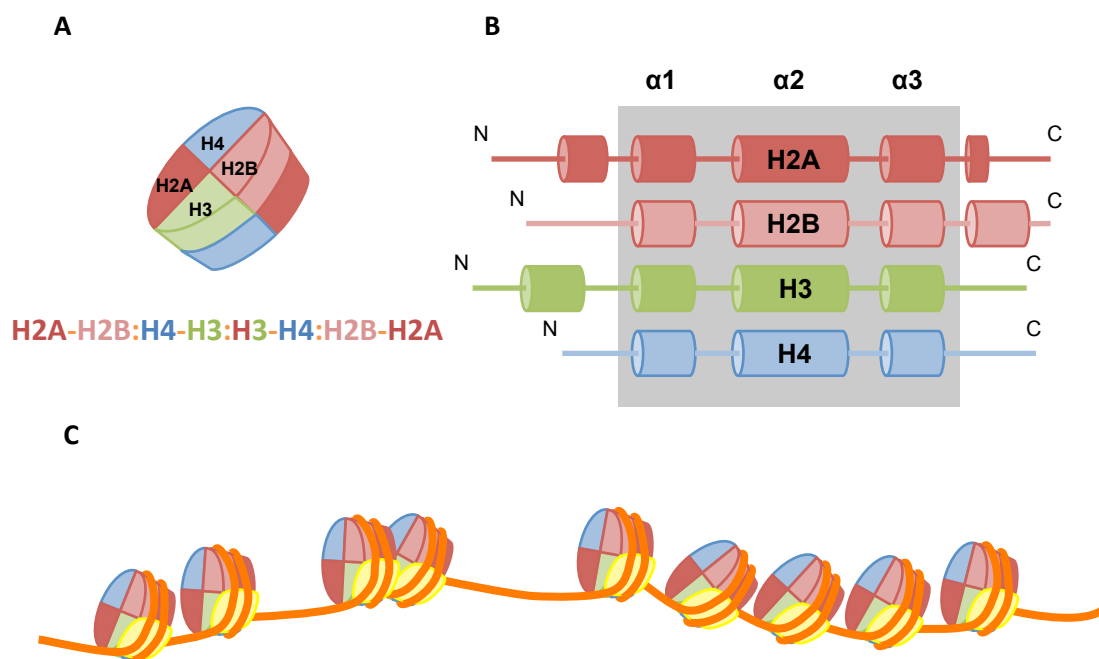
1.3.1.1. Heterochromatin, Euchromatin and Position effect

The distribution of chromatin within the nucleus is not homogeneous (Carlberg and Molnár 2013). Chromatin can be divided into two distinct regions depending on the chromatin density. Heterochromatin is described as the condensed chromatin and is normally correlated with gene-poor, transcriptionally inactive and late replication regions (Grewal and Moazed 2003; Razin et al. 2007). On the other end, euchromatin is described as a less dense chromatin and it is normally related with the gene-rich and transcriptionally active regions (Grewal and Moazed 2003; Razin et al. 2007). Furthermore, gene-poor heterochromatin is associated with the nuclear

periphery, while the gene-rich euchromatin is localised in the internal areas of the nucleus (Boyle et al. 2001).

Most of heterochromatin, called constitutive heterochromatin is transcriptionally inactive and is found next to the functional and gene-poor chromosome structures such as centromeres and telomeres (Sieger et al. 1971). However, smaller heterochromatin regions (facultative heterochromatin) are able to encode genes that are physically condensed and transcriptionally repressed (Berlowitz 1974). This phenomenon is described as position effect variegation (Wilson et al. 1990). Position effect variegation is caused by a variable packaging degree of normal euchromatic regions into a transcriptionally inactive heterochromatin-like conformation and is typically related to histone hypoacetylation and DNA methylation (Section 1.3.1.2).

Heterochromatin and euchromatin have been characterised by covalent modifications that affect histones and DNA. These covalent modifications, such as methylation and acetylation are the basis of epigenetic phenomenon that affect the gene transcription and will be discussed in Sections 1.3.1.2 and 1.3.1.3.

1.3.1.2. Histone modification

Histone covalent modifications play an important role on chromatin organisation. Most of these modifications are targeted to the N-termini of histones and ultimately regulate gene transcription by affecting the nucleosome stability or chromatin condensation (Kouzarides 2007). Traditionally histone modifications have been described as acting as platforms that recruit specific transcription factors and remodelling complexes (Tessarz and Kouzarides 2014). However, more recently it has been shown that these modifications can also alter significantly the function of the histone itself by direct affecting the normal histone-histone and histone-DNA interactions (Tessarz and Kouzarides 2014).

There are several types of histone modification including acetylation, methylation, ubiquitination, phosphorylation and SUMOlation that affect the gene activation or repression (Kouzarides 2007) This Section will focus on acetylation and methylation, of lysine of H3 and H4 as at the present they are best described modifications and also are considered to have greatest influence on gene expression (Tessarz and Kouzarides 2014).

Histone acetylation is normally associated with transcriptional activation (Lee et al. 1993). The acetylation of lysine residues in histones H3 and H4 is mediated by

histone acetyltransferase (HAT) (Figure 1.5-A) (Jenuwin ans Allis 2001, Lee at al. 1993).

Histone lysines residues are positively charged and can form a salt bridge with the negatively charged DNA backbone. The acetylation of the lysines in histones not only introduces a bulkier side chain to this amino acid but also removes the positive charge to this amino acid (Figure 1.5-A) (Tessarz and Kouzarides 2014). This alteration on lysine charge and structure disrupts the salt-bridges interactions between histone and DNA molecules and also the histone-histone interactions. These alterations diminish the binding affinity of the histone with the DNA backbone and disrupts of the nucleosome structure (Tessarz and Kouzarides 2014). The disruption of the chromatin structures ultimately leads to a more open and transcriptionally active DNA (Mellor et al. 2008). Gene activation mediated by histone acetylation was observed in mouse embryonic fibroblast, when the acetylation of lysine 9 of histone H3 was correlated with to transcriptional gene activation (Mutskov and Felsenfeld 2004). In contrast, histone deacetylation, mediated by histone deacetylase (HDAC) increases the positive charge of the histone N-termini, increasing the affinity of the histone-DNA interaction and stabilise histone interactions (Hong et al. 1993, Rundlett et al. 1998). Histone acetylation in heterochromatin and euchromatin arises from the balance between the activity of HAT and HDAC (Peserico and Simone 2011).

Histone methylation is normally associated with transcriptional repression, however in particular cases it can lead to gene activation (Mellor et al. 2008). In mouse embryonic fibroblasts, methylation of lysine 9, 20 and 27 in histone H3, was strongly related with transcriptional inactivation of a cluster of genes (Regha et al. 2007). Nevertheless, the same study observed that methylation of lysine 4 of histone H3 lead to gene activation (Regha et al. 2007).

The mechanisms by which methylation leads to gene activation or repression are different to those observed for acetylation (Berger 2007). Whilst acetylation directly affects the histone-histone and histone-DNA interactions, histone methylation recruits non-histones proteins that affect gene transcription (Berger 2007). For example, methylation of lysine 9 of histone H3 can act as a binding site for heterochromatin-like protein 1 (HP1) (Mutskov and Felsenfeld 2004). HP1 mediates the formation of heterochromatin by associating histones methyltransferases (HMT) and DNA methyltransferases (DNMT), ultimately leading to gene inactivation (Figure 1.5-A) (Smallwood et al. 2007).

1.3.1.3. DNA Methylation

DNA methylation is strongly associated with gene silencing. Methylation of DNA normally occurs in cytosine-guanine palindromic (CpG) dinucleotide sequences termed as CpG island. (Razin 1998). This CpG islands are often found in the promoter regions, leading to gene inhibition if methylated (Berger 2007; Carvalho et al. 2010).

DNA methylation is mediated by DNMT which introduces a methyl group onto the carbon-5 of cytosine leading to the formation of 5-methylcytosine (Carvalho et al 2010, Osterlehner et al. 2011). Methylated DNA leads to gene silencing either by direct inhibition of the promoter to the transcription factors or by recruiting methylated DNA-binding proteins, such as methylated DNA complex 2 (MeCP2) (Pradhan et al. 1999). MeCP2 is a protein that contains a methyl binding domain and this mediates the formation of transcriptional repressive chromatin by interacting with HDAC and an adapter protein, Sin 3 (Figure 1.5-B) (Vaissière et al. 2008). Furthermore, MeCP2 also interacts with DNMT leading an increase of DNA methylation and, consequently, increased gene silencing (Figure 1.5-B) (Vaissière et al. 2008)

DNA methylation has been incriminated as one of the major causes for gene silencing during recombinant protein expression in CHO cells (Yang 2010; Kim et al. 2011; Osterlehner et al. 2011). In most of the cases where methylation was the cause of gene loss it was observed that methylation promoter regions exhibited greatest correlation with transcriptional status (Yang 2010; Kim et al. 2011; Osterlehner et al. 2011). The sequence of human CMV (hCMV) promoter, widely used in therapeutic production, contains 33 CpG islands that are significantly targeted for methylation (Osterlehner et al. 2011). However, methylation in intragenic CpG islands can also occur and such modification is also correlated with gene silencing (Bauer et al. 2010).

Figure 1.5. Histone modifications and DNA methylation.
A – Lysine charge and structure changes due to methylation and acetylation; B - Histone Acetylation, and Histone and DNA methylation mediated by HP1; C – Histone Dacetylation and DNA methylation mediated by MeCP2. (based in Tessarz and Kouzarides 2014, Vaissière et al. 2008)
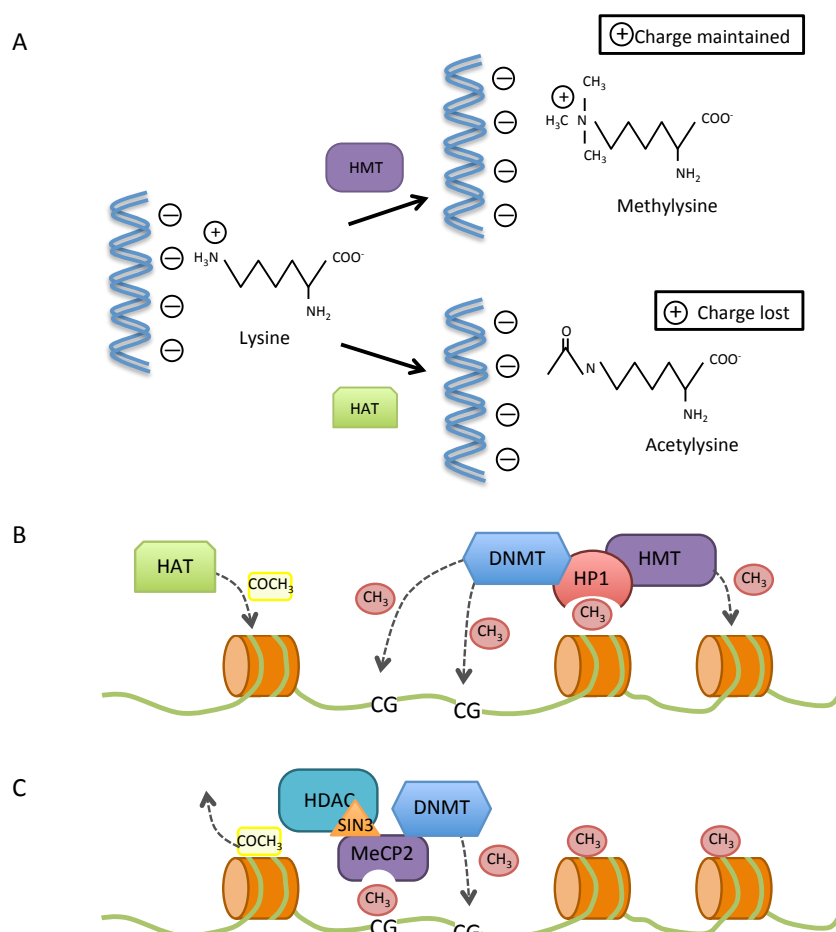
1.3.2. Nuclear organization

As described in Section 1.3.1, eukaryotic genomes are organised into chromatin structures. During mitosis, chromatin is rearranged and compacted into chromosomes, and throughout the rest of the cell division cycle, is organised as a thread-like structure termed as chromosomes territories (CTs) (Cremer et al. 2001). CT distribution along the nucleoplasm is not random and the spatial positioning of CTs relative to each other differs between different tissues and cell types. (Cremer et al. 2001; Parada et al. 2004). The spatial organisation of CTs has important roles in the regulation of genome function and gene expression as this organisation

influences the interactions between genes, regulatory elements and transcription factories (Fraser and Bickmore 2007).

Using fluorescence microscopy techniques, especially Fluorescence *in situ* Hybridisation (FISH), chromatin was observed to be heterogeneously distributed along the CT and that CT had a preferred distribution in the nucleus (Dostie and Bickmore 2011). These observations suggested that gene availability could be regulated by three-dimensional chromatin interactions rather than solely on local chromatin interactions. (Dostie and Bickmore 2011). The development of techniques such as Chromosome Conformation Capture (3C) and all its derivatives (Circular Chromosome Conformation Capture (4C), Carbon-Copy Chromosome Conformation Capture (5C), Genome-wide Chromosome Conformation Capture (Hi-C) and Chromatin Interaction Analysis by Paired-End Tag Sequencing (ChIA-PET)) has provided better tools to study chromosome structure and transcription regulation through chromatin interactions (Dostie and Bickmore 2011; Sanyal et al. 2011). Combinations of FISH and one or more 3C based techniques have been widely used to study the gene regulatory actions of CT organisation and chromatin distribution.

Most active genes are associated in clusters localised in euchromatin (Boyle et al. 2001; Shopland et al. 2003). Active genes have the ability to relocate substantial distance either inside or outside their CT which subsequently leads to the formation of chromatin loops (Fraser and Bickmore 2007). Although not absolute, chromatin loop formation is often related to transcription activity (Branco and Pombo 2006). Branco and Pombo (2006) observed that loop formation activity was decreased by inhibiting the transcriptional machinery (Branco and Pombo 2006).

1.3.2.1 Transcription activity and chromatin loop formation

The exact role of loop formation during transcription activation has not been determined. Evidence revealed that loop formation can either be a direct response to gene activation and targets active genes towards the transcriptional machinery, or can be part of the regulation mechanism that promotes the interactions between genes and gene regulators (Fraser and Bickmore 2007).

While studying gene co-localisation in transcription factories (Section 1.3.3), Osborne et al (2004) detected that active genes relocated by looping in order to be transcribed (Osborne et al. 2004). Later, Osborne et al (2007) used FISH and 3C techniques to study the dynamics of gene induction and observed that genes quickly relocated in transcription factories upon transcription activation (Osborne et al. 2007).

Chromatin relocation may also have an important role in gene regulation by promotion of interactions between genes and chromatin regulator elements. One of the best well-characterised examples of gene regulation by long-range interactions is the interaction between the ß-globin locus control region (LCR) and the ß-globin genes in mouse erythrocyte cells (Carter et al. 2002; Tolhuis et al. 2002). In mouse, ß-globin gene locus is located on chromosome 7 and contains four functional ß-globin genes. The LCR which regulates the ß-globin activation consists of six DNase I–hypersensitive sites (HS) and is localised 30-60 kb upstream of the first ß-globin gene. Using a modified RNA FISH method called RNA-TRAP (Tagging and Recovery of Associated Proteins) (Carter et al. 2002) and 3C (Tolhuis et al. 2002), it was observed that LCR and active ß-globin genes were in close interaction by looping the 50kb sequence (Carter et al. 2002; Tolhuis et al. 2002).

Another example of complex long-range chromatin interaction and loop formation was observed in the activation of *Hoxd13* gene during the late limb patterning in mouse (Montavon et al. 2011). In mammals the *HoxD* gene cluster contains 13 genes (*Hoxd1* to *Hoxd13)* that undergo sequential transcriptional activation, in both time and space, following their respective positions within the cluster (Tarchini and Duboule 2006). Transcriptional activation of *Hoxd* genes is mediated by the direct interaction with two enhancer elements, situated within global control region (GCR) and Prox region (Gonzalez et al. 2007). The combined effects of these two enhancers situated 180 kb and 50 kb upstream of the *hoxd13* has proven to be essential for the activation of *Hoxd* gene (Montavon et al. 2011)

Supporting these two cases, other long-range chromatin regulatory interactions, promoted by chromatin relocation, have also been described between genes and enhancers (Li et al. 2012), LCR (Spilianakis and Flavell 2004; Spilianakis et al. 2005), insulators (Ling et al. 2006) and silencers (Horike et al. 2004; Tiwari et al. 2008).

1.3.3. Transcription and Transcription factories

Transcription mediated by RNA polymerase II (RNAP II) is divided in four separate phases, pre-initiation, initiation, elongation and termination (Pandit et al. 2008). RNAP II is a multi-unit enzyme that despite its structural complexity is not able to recognise the target promoter sequence (Sainsbury et al. 2015). To start the transcription process and to recognise the promoter sequence, RNAP II interacts with several general initiation factors (Trancription factor [TF] IIA, TFIIB, TFIID, TFIIE,

TFIIF and TFIIH) at the CTD (Heidemann et al. 2013, Sainsbury et al. 2015). The CTD is a very conserved sequence, across fungi, plants and animals, and it comprises 25 to 52 tandem repeats of a consensus 7 amino acids ($Tyr_1Ser_2Pro_3Thr_4Ser_5Pro_6Ser_7$) (Corden 1990). CTD is not essential for the activity of the RNAP II in vitro, however it is essential for transcription in vivo as is the phosphorylation patterns of the CTD that dictate which factors associate with RNAPII and consequently the stage of the transcription (Phatnani and Greenleaf 2006, Heidemann et al 2013). Furthermore, the phosphorylation state of CTD regulates the interactions with the RNA processing machinery during the transcription process (Pandit et al. 2008). CTD can be predominantly phosphorylated on Ser2 and Ser5 (Dahmus 1995, Dhamus 2996) and is the CTD phosphorylation pattern that dictates which factors associate with RNAPII.

TFIID is a conserved, multifunctional general transcription factor involved in promoter recognition, assembly of the pre-initiation transcriptional complex (PIC) and chromatin remodelling (Sainsbury et al 2015). This transcription factor is a central component of the transcription machinery and is composed by TBP (TATA box binding protein) and 13-14 different TAFs (TBP associated factors). The different TAFs that compose the TFIID are the key factors for the identification and binding of the TFIID to a specific promoter (Sainsbury et al 2015). Each TAF recognise and bind to different specific elements of the core promoter such as the motif ten element (MTE), downstream promoter element (DPE) and downstream core elements (DCE) regulating the binding of the transcription machinery to specific promoters (Figure 1.6–1) (Sainsbury et al 2015). TFIID also interacts with TFIIA and TFIIB at the promoter in the Bre (TFIIB recognition element) and TATA box motifs present in promoter sequence to form the upstream promoter complex that will recruit RNAP II and TFIIF at the Inr (transcription initiation site) to form the core PIC (Figure 1.6–2, 3) (Pandit et al. 2008, Sainsbury et al 2015).

During initiation, further transcription factors are recruited including TFIIE and TFIIH (Figure 1.6 – 4) that are responsible for the opening of the DNA and the formation of the transcription bubble. TFIIH is also responsible for phosphorylation of the Ser5 of the CTD to form the initiating RNAP II-Ser5P and consequently, the initiation of the transcription process (Nikolov and Burley 1997; Sarah et al. 2015).

At the end of the initiation stage, the initially transcribing complex pauses and P-TEFb (positive transcription elongation factor) is recruited into the assembled transcriptional machinery (Figure 1.6 – 7) (Nikolov and Burley 1997). P-TEFb is a complex of elongation factors that in which its composition has a cyclin dependent

kinase 9 (CDK9) that is responsible to phosphorylate Ser2 of the CTD. The phosphorylation of the Ser2 of the CTD leads to dissociation of the initially transcribing complex and the formation of the elongating RNAP II-Ser2P. Further elongation factors are recruited to the RNAP II leading to the start of the elongation stage (Price 2000, Jonkers and Lis 2015).

Contrary to initiation and elongation, the termination process is not so well understood with two models being proposed for this last step (Pandit et al. 2008). The allosteric model suggests that the nascent RNA is cleaved downstream of the polyadenylation site that could trigger a conformational change of the RNAP II-Sser2P induction pausing and eventually releasing the mRNA from the complex (Luo et al. 2006; West et al. 2008). Alternatively, the torpedo model suggests that the nascent RNA becomes exposed to 5' exonucleases that eventually release the RNA from the RNAP II-Ser2P complex (Figure 1.6 – 8) (Luo et al. 2006; West et al. 2008).

Figure 1.6. Schematic representation of transcription pre-initiation, initiation, elongation and termination phases

Pre-initiation complex (PIC) assembly from general transcription factors and RNA polymerase (identified in image) on promoter DNA (Inr – initiator motif; Bre^U and Bre^D – Upstream and downstream TFIIB recognition element; TATA – TATA box motif; MTE – motif ten element, DPE – downstream promoter element; DCE – downstream core elements). 1 - The TATA box-binding protein (TBP) subunit of the TFIID binds to promoter DNA, inducing a bend; 2 - The TBP–DNA complex is then stabilized by TFIIB and TFIIA, which flank TBP on both sides; 3 - The resulting upstream promoter complex is joined by the Pol II–TFIIF complex, leading to the formation of the core PIC; 4 - Subsequent binding of TFIIE and TFIIH complete the PIC; 5, 6 - In the presence of ATP, the DNA is opened (forming the 'transcription bubble') and RNA synthesis initiates; 7- The dissociation of initiation factors and the recruit of P-TEFb and other elongation factors enables the formation of the Pol II elongation complex, which is associated with transcription elongation factors (different coloured circles); 8 – At the end of the transcription process, it is prosed that termination can either occur due to conformational change of RNAP II-Ser2P that eventually releases the nascent RNA (Allosteric Model) or due to 5' exonuclease cleavage (Torpedo Model). (Adapted from Sainsbury et al. 2015; Pandit et al. 2008)

1. Unbound promoter

TBP
TAFs

-30
Bre^U TATA Bre^D Inr DCE -30
MTE DPE
Promoter DNA
+1

2. Upstream promoter complex

TFIIA TFIIB

3. Core PIC

TFIIF Pol II
Upstream DNA Downstream DNA

4. Closed PIC

TFIIE TFIIH kinase
TFIIH core

ATP
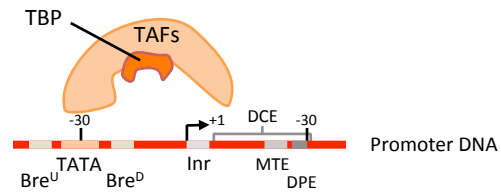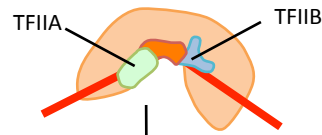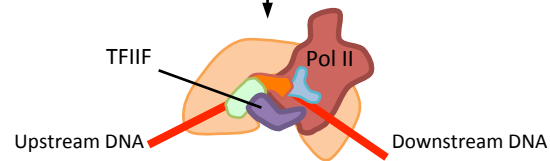
5. Open PIC
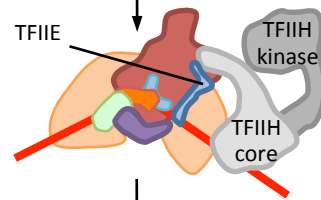
Transcription bubble
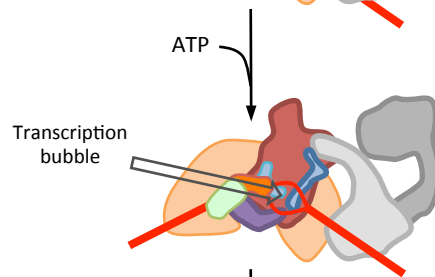
6. Initiating transcribing complex

Nascent mRNA

Initiation factors P-TEFb and Elongation factors
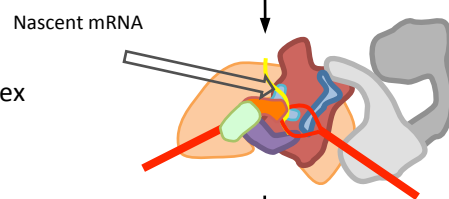
7. Elongation complex

8. Termination

Allosteric Model Torpedo Model
5' exo
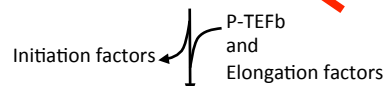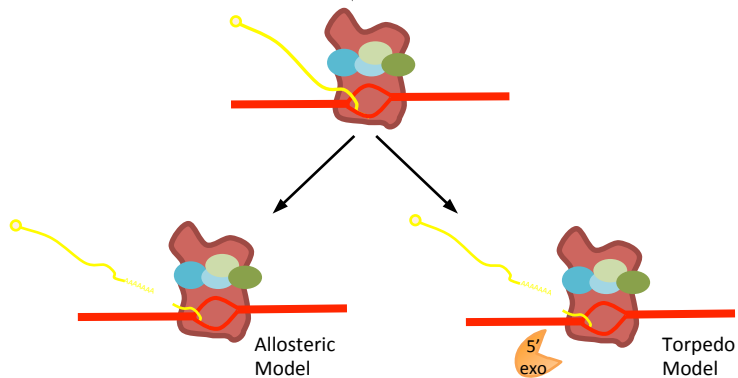
1.3.3.1. Transcription factories

In eukaryotes, most of the protein encoding genes are transcribed by RNAP II. In 1993, Jackson et al (1993) observed that the nascent transcripts were co-localised within the nuclear space, suggesting that the transcription units were organised into factories where transcripts are synthesised and processed (Jackson et al. 1993). Iborra et al (1996) observed that the distribution of RNAP II through the nuclear space was not uniform, and most RNAP II molecules were concentrated in transcription factories (Iborra et al. 1996). The composition of transcription factories remains uncharacterised but it has been described to be a complex of transcription factors, splicing factors and multi active RNAP II which anchor with an, as yet, unidentified nuclear scaffold (Iborra et al. 1996; Jackson et al. 1998; Kimura et al. 1999). Initially Iborra et al (1996) projected the existence of approximately 2,100 transcription factories in each Hela cell (Iborra et al. 1996). This calculation was later confirmed by Jackson et al (1998), that estimated the number of active RNAP II in Hela cells was approximately 75,000 per cell (Jackson et al. 1998). The authors estimated that active RNAP II were concentrated in approximately 2,400 sites with 80 nm containing approximately 30 active RNAP II each (Jackson et al. 1998).

1.3.3.2. Transcription factories and transcription initiation

Although transcription factories have been traditionally described as pre-assembled RNAP II clusters, two recent studies have shown some evidence against this idea ((Cisse et al. 2013; Ghamari et al. 2013). Both studies revealed a dynamic formation of such clusters during the transcription process and hypothesised that their formation was linked to transcription initiation (Cisse et al. 2013; Ghamari et al. 2013).

Ghamari et al 2013, showed that transcription initiation and elongation occurred in different cell compartments and only the initiation occurred in transcription factories (Ghamari et al. 2013). Ghamari and colleagues indirectly marked transcription factories by tracking CDK9, a transcription elongation factor that regulates RNAP II activity following transcriptional initiation. The authors observed that CDK9 initially co-localised in the promoter region with initiating RNAP II-ser5P at the end of the initiation process (Ghamari et al. 2013). Using an mCherry tagged CDK9 in combination with immunostaining against different proteins involved in transcription initiation, elongation and splicing, the authors observed that CDK9 co-localised in the promoter region with initiating RNAP II-ser5P but not with elongating RNAP II-ser2P

(Ghamari et al. 2013). As Ghamari and colleagues only observed the formation of transcription factories during the initiation step, the authors suggested that initiation and elongation took place in different nuclear compartments (Ghamari et al. 2013).

Similar observations were made by Cisse et al 2013. By expressing an α-amanitin-resistant form of the RNAP II fused to a photoswitchable fluorescent protein (Dendra2) in human osteosarcoma cell line U2OS, the authors demonstrated, by single molecule light microscopy, that RNAP II clusters (of average 100nm diameter) were transiently formed with an average lifetime of 5.1(±0.4) seconds (Cisse et al. 2013). They also observed that, when transcription was stimulated, both size and lifetime of these clusters was considerably increased. By blocking transcription elongation using a P-TEFb inhibitor, the authors not only observed the formation of RNAP II clusters, but also, that these clusters were considerably more stable (Cisse et al. 2013). With this observations the authors demonstrated that RNAP II cluster formation was highly correlated to transcription initiation and that it was likely that cluster dissolution was linked to the transition from transcription initiation into elongation (Cisse et al. 2013).

Based on the interpretations from these new observations, Buckely and Lis (2014) recently proposed a second dynamic model for the origin of transcription factories (Figure 1.7) (Buckely and Lis 2014).

Figure 1.7. Schematic representation of models for the origins of transcription factories.
A- Static pre-assembled transcription factory model. Upon activation, the gene loops out of its chromosome to a preformed transcriptional factory for its expression; B- Dynamic self-assembling transcriptional factory model. The inactive gene is surrounded by free nucleoplasmics RNAP II. Upon on activation, RNAP II dynamic cluster and self-assembled in a transcription factory around the active gene leading to its expression. (Adapted from Buckley and Lis 2004)
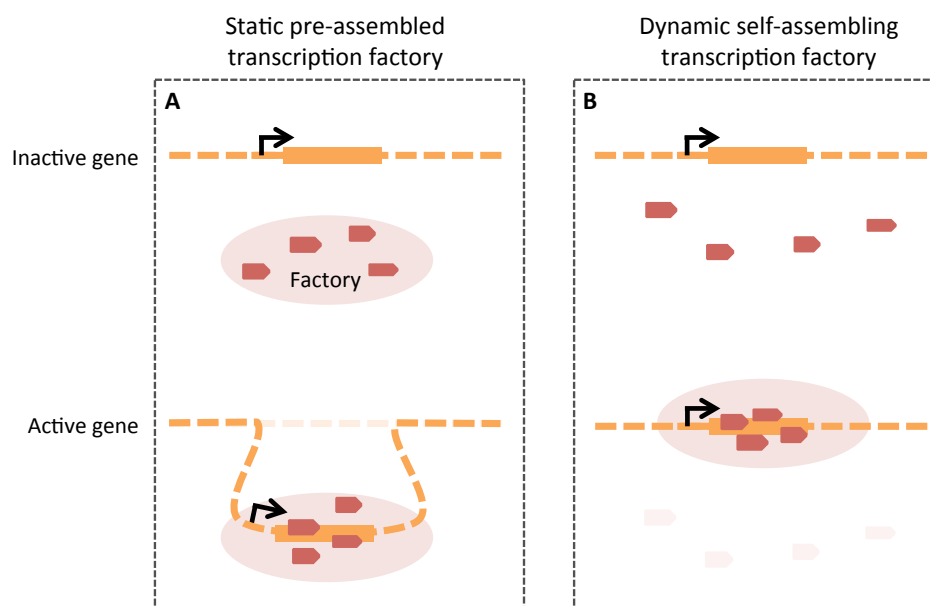
Adapted from Buckley and Lis 2014

1.3.3.3. Specialisation of genes for specific transcription factories

The relative number of transcription factories per cell is substantially lower than the number of active genes, suggesting that different genes share the same transcription factory (Kimura et al. 1999; Martin and Pombo 2003). This co-localisation of transcription of different genes within the same transcription factories has been shown for genes localised in the same and different chromosomes (Osborne et al. 2004; Park et al. 2014; Schoenfelder et al. 2010).

Using different variants of FISH (RNA, DNA and Immuno FISH) and 3C techniques, Osborne et al (2004) observed that in erythroid cells, *Hbb-1* (b-like globin) gene dynamically shared the same transcription factory with four other genes (*Eraf* - α-hemoglobin-stabilizing protein, *Uros* - uroporphyrinogen III synthetase, *Igf2* - insulin-like growth factor 2; Kcnq1ot1 - long QT intronic transcript) localised in a 40 Mb region of the same chromosome (Osborne et al. 2004)). The authors observed that

43

the co-localisation of the genes during transcription occurred at high frequency, ranging from 42-60%. (Osborne et al. 2004). Although with much lower frequency (7%), Osborne and colleagues also observed that *Hbb-1* shared the same transcription machinery with *Hba* (α-globin) that is localised in a different chromosome (Osborne et al. 2004).

Later, further studies using similar techniques revealed evidence that the co-localisation was probably due to a specialisation of the genes for particular transcription factories (Osborne et al. 2007; Xu and Cook 2008, Schoenfelder et al 2010). Schoenfelder et al. 2010 observed that genes co-localised to the same transcription factories shared similar transcription regulatory mechanisms (Schoenfelder et al. 2010). The authors evaluated the transcription of *Hba* and *Hbb* with a set of eight different genes and observed that genes with promoters that had conserved motifs for the specific transcription factor Klf1 (Kruppel-like factor 1) were observed to share the same transcription factories (Schoenfelder et al. 2010). The authors suggested that this observation could be due to different composition of transcription factors between the different transcription factories (Schoenfelder et al. 2010).

Furthermore, Park et al (2014) observed that the co-localisation and specialisation of particular genes in the same transcriptional machinery could have an important role the regulation of their expression (Park et al. 2014). The authors studied the transcriptional activity of three active Ig genes that form IgM in plasma cells. These three genes (*Igk, IgH and IgJ* gene) are localised in three different human chromosomes (chromosome 6. 12 and 5, respectively) (Park et al. 2014). Park et al (2014) observed that when antibody synthesis was maximal, the transcription of these three genes co-localised in the same transcription factories, often near the nuclear periphery (Park et al. 2014). Additionally, the authors observed that after transcription, the transcripts frequently shared interchromatin trafficking channels suggesting that the predominant peripheral localisation of the transcription of these genes was to facilitate the nuclear transcript export (Park et al. 2014). Figure 1.8, shows a schematic representation proposed by Park et al (2014) of the transcription of these genes in plasma cells.

Figure 1.8. Schematic representation of the transcription of plasma cells IgK, IgH and IgJ genes in transcription factories.

IgJ, *IgK* and *IgH* genes from chromosomes 5, 6 and 12 sharing a transcription factory. Functionally rearranged Ig genes loop out from their specific chromosome territories (Park et al. 2014). The resulting Ig gene transcripts that are generated at the boundaries on an interchromatin channel traffic for export to the cytoplasm through the nuclear pores. (Adapted from Park et al. 2014)

Plasma Cell Nucleus

Adapted from Park et al. 2014
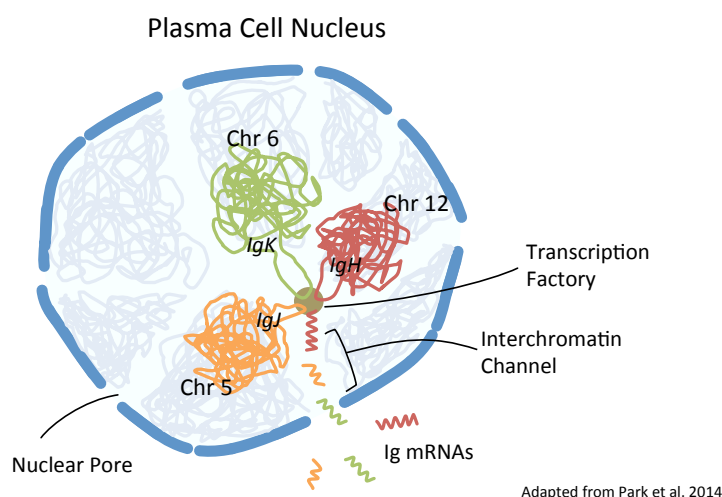
1.3.3.3. Visualisation of transcription in real time in living cells

Imaging of mRNA in vivo using MS2-MCP systems presents a pioneering and revolutionary approach, originally developed to track the movement in real-time of ASH1 mRNA in yeast cells into the daughter cell or bud during cell division (Bertrand et al. 2000). More recently, this system has been used to study the kinetics of the transcriptional machinery in living bacteria, yeast, Drosophila and mammalian cells (Chubb et al. 2006; Golding et al. 2005; Yao et al. 2006; Yunger et al. 2010).

Yunger et al 2010 used a combination of Flp-recombinanse site-directed and MS2-tag/MCP-GFP methods to visualise transcription kinetics in HEK 293 (Yunger et al. 2010). The author cloned a sequence of tandem bacteriophage sequence repeats (MS2) downstream of a recombinant cyclin d1 gene and the sequences underwent site-directed integration to a pre-selected genomic location (Yunger et al. 2010). When transcribed, the MS2 mRNA sequence formed a loop structure that can be specifically recognised and bound by the MS2 coat protein (MCP) (Darzacq et al. 2007). The presence of the MS2-loop structures upstream the 3' un-translated region (UTR) allows the detection of the recombinant mRNA by the MCP protein fused with GFP containing a Nuclear Signal Sequence (MCP-NLS) (Figure 1.9). The fluorescently-labelled recombinant mRNA allows visualisation of mRNA at single-

molecule resolution using live-imaging fluorescent microscopy. Using this approach it is possible to identify the number of RNAP II recruited and, using FRAP, to measure the rate of transcription of the recombinant gene in single living cells (Yunger et al. 2010).

Figure 1.9. Schematic representation of the use of MS2/MCP method to visualise transcriptional in living cells.
A- Schematic representation of the recombinant mStrawberry production in a cell; B- Representation of the final genomic region after site-directed integration of the recombinant-promoter/mStrawberry cassette; C- Schematic representation of the detection of the nascent recombinant mRNA tagged with the MS2 loop tag by the MCP-GFP protein. (Based on Yunger et al 2010)
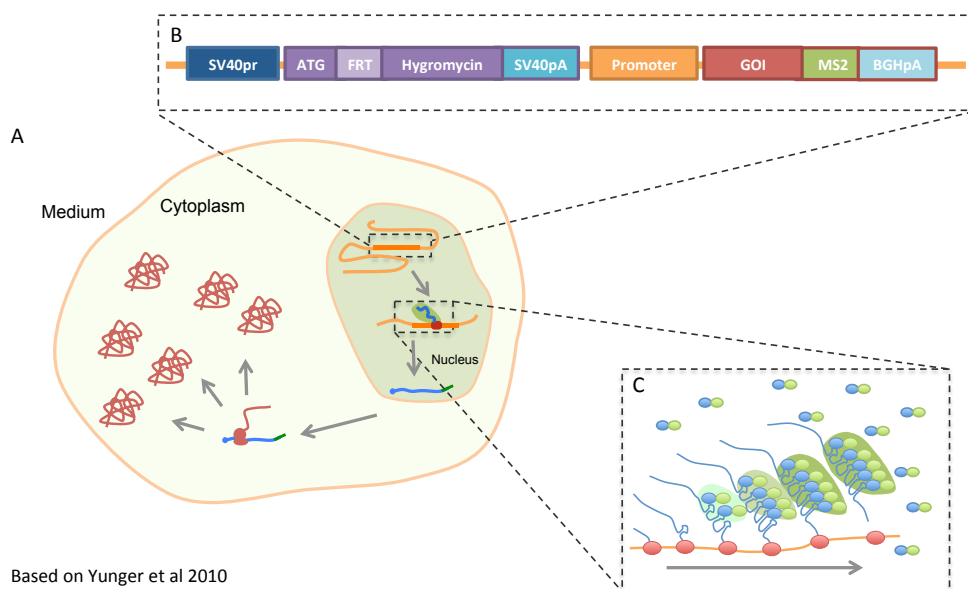
1.4. Prevention of transcriptional instability in CHO cells

Multiple strategies have been developed to improve CHO cell line stability due to commercial pressures for predictability of recombinant protein productivity. The next Section will describe some of those strategies used in CHO cell systems to prevent the effects of transcriptional silencing.

1.4.1. Early detection of CHO cell line instability

The identification of stable/unstable cell lines in the early stages of cell line development is a major challenge for industry. One of the strategies used to predict unstable cell lines is the evaluation of early methylation of the recombinant promoter (Osterlehner et al. 2011). The authors quantified the methylation of cytosine 179 (C-179) of the hCMV-MEI promoter/enhancer at early stages of cell line development and observed that the cell lines that contained methylated C-179 later had a unstable profile (Osterlehner et al. 2011). Ostherlehner et al (2011) suggested a quantification method based on a methylation-specific qPCR to predict cell line instability (Osterlehner et al. 2011).

Apoptosis was identified as a possible cause or effect of cell line instability and is being related as a natural response of the host cell against the stress induced by the high expression of the recombinant gene (Dorai et al.2012). Apoptosis can be measured in the cell by quantification of apoptotic markers such as caspase-3 and Annexin V (Elmore 2007). Dorai et al (2012) observed that unstable cell lines had a higher caspase-3 activity when compared to stable cell lines (Dorai et al. 2012). This difference was observed in the early stage of cell line development and the authors suggested that this apoptotic marker can be used to predict the production instability of CHO cell lines (Dorai et al. 2012).

The use of mass spectroscopy analysis by Matrix-Assisted Laser Desorption/Ionization-Time of Flight (MALDI-TOF) has also been reported as a strategy to identify high producer cell lines (Feng et al. 2011). The identification of high producers and stable cell lines was performed based on the analysis of recombinant protein on the cell surface. The authors suggested, despite lack of quantitation in MALDI-TOF, this method could be used to identify stable producing cell lines in relation to the intensity of specific proteins on the cell surface (Feng et al. 2011).

1.4.2. New Hosts

As discussed in Section 1.2.3.2, CHO cells have a natural ability for genomic rearrangements. This phenomenon ultimately leads to the formation of heterogeneous populations of cells with different functional properties within the same cell line (Davies et al. 2013). This was recently demonstrated by Davies et al (2013) when the authors cloned 100 cell lines from CHOK1SV parental cell line (Davies et al. 2013). Davies and colleagues observed that the selected cell lines

47

were highly heterogeneous in terms of growth and productivity (Davies et al. 2013). Based on this high heterogeneity of cells within a cell population it was possible to select for new cell lines that have the desirable productivity. This was the approach that Pybus et al (2014) used to produce a CHO-DG44 host library (Pybus et al. 2014). The authors used a selective pressure approach to generate a library of cells with the desirable attributes for manufacturing therapeutic proteins (Pybus et al. 2014). Pybus and colleagues described that the new cell lines had a decreased doubling time during subculture and increased productivity when compared with the CHO-DG44 parental cell line (Pybus et al. 2014). These cell lines currently commercially available as Apollo™ expressing system (FUJIFILM Diosynth Biotechnologies).

Besides selecting for desirable clones from a pool of parental cells, other approaches to improve the host cells involve the potential to use genome editing tools such as zinc-fingers nucleases (ZFNs), Transcription activator-like effector nucleases (TALENs), rAAV vectors and clustered regularly interspaced short palindromic repeats (CRISPR-Cas 9) (Cheng and Alper 2014). The ability to edit the genome of CHO cells at any specific genomic location will allow further diversification of CHO hosts cell available for production, and with the increasing availability of precise genome data this approach is now fully possible (Cheng and Alper 2014; Kuystermans and Al-Rubeai 2015). Ultimately, editing the genome of CHO could allow the generation of tailor-made CHO cell hosts for specific biological purposes e.g. glycosylation, optimal metabolic properties, capacity to handle specific products (Cheng and Alper 2014; Kuystermans and Al-Rubeai 2015).

Use of genome editing tools is still in early stages for CHO cells, however three different types of genetic engineered CHO hosts are commercially available; CHO GS knocked-out (K.O.) (Sigma-Aldrich, Lonza and Horizon Discovery); CHO DHFR K.O. (Sigma-Aldrich); CHO Fut-8 KO cell line (Lonza).

1.4.3. Synthetic promoters

Recombinant promoters are essential for the initiation and regulation of the transgene expression in CHO cells (Ho and Yang 2014). Developing novel recombinant promoters for specific, active and stable recombinant gene expression in CHO cells has become an important approach for the optimisation of the productivity and stability in CHO cells. Traditionally, therapeutic proteins are expressed using viral promoters such as SV40, RSV and CMV promoters (Pontiller

et al. 2008). Besides these viral promoters, mammalian elongation factor 1 alpha promoters (EF1α) have also been used to express recombinant products in CHO cells (Ho et al. 2015; Kim et al. 2002). Although these promoters usually lead to a constitutive high expression of the recombinant protein, a number of problems such as gene silencing and induction of cellular stress reactions have been associated with them (Pontiller et al. 2008).

Recently a series of reports developing novel synthetic recombinant promoters for CHO cells have been published using different approaches from generating libraries of modified versions of existing promoters (Brown et al. 2014; Lianchun et al. 2013; Mariati et al. 2014a; Mariati et al. 2014b) to the identification of novel genomic regulatory elements from CHO (Chen et al. 2013; Pontiller et al. 2008; Pontiller et al. 2010).

Brown et al 2014, generated a library of 140 synthetic promoters using bioinformatic tools to identify transcription factor regulatory elements (TFREs) from 10 common viral promoter sequences (Brown et al. 2014). The authors used the identified TFREs as building blocks to randomly produce promoters of multiple sequences of TFREs. These sequences were cloned into a promoter-less vector that contained the minimal hCMV-IE1 core promoter and screened for promoter strength by quantifying the expression of SEAP and GFP proteins (Brown et al. 2014). The authors identified that 57% of the new promoters developed had higher expression than the "standard" CMV promoter, with the strongest promoter exhibiting 2.2 fold increase of productivity (Brown et al. 2014)

On the other hand, the identification of genomic regulatory elements has not yet been so successful. The strategy used by Pontlier et al was to clone random genomic sequences from a CHO genomic library into promoter-less vectors and screen for producing cells (Pontiller et al. 2008; Pontiller et al. 2010). Towards the same end, Chen et al (2013) used a prompter trap (Chen et al. 2013). The authors randomly integrated a promoter-less vector with a reporter gene into the genome of CHO cells and then selected for expressing clones (Chen et al. 2013). Both strategies succeeded in identifying regulatory elements from the CHO genome that could regulate recombinant expression. However, both strategies failed to identify a strong promoter (Chen et al. 2013; Pontiller et al. 2008; Pontiller et al. 2010). In both studies, none of the generated promoters were as effective as SV40 promoter (Chen et al. 2013; Pontiller et al. 2008; Pontiller et al. 2010).

The fact that the CHO genome has just recently become available opens new opportunities to identify regulatory CHO specific regulatory elements. Analysing the CHO genome information using a bioinformatics approaches similar to those used by Brown et al (2014) will possibly allow the identification of strong and stable regulatory elements specifically to be used in CHO cells.

1.4.4. Modification of gene integration site

The inclusion of chromatin regulatory elements such as Ubiquitous Opening Elements (UCOEs) and Scaffold/Matrix Attachment Regions (SMARs) in the expression vectors is one of the strategies used to improve cell line stability (Barnes and Dickson 2006). UCOEs are transcriptional active chromatin elements that increase histone acetylation, leading to gene activation (Zhang et al. 2010). Nair et al (2011) evaluated the influence of one human and one mouse UCOEs on the production of Factor VIII in BHK11 cells driven by different promoters (Nair et al. 2011). The authors observed the influence of each UCOE in the maintenance of long-term expression was dependent on the UCOE/Promoter combination (Nair et al. 2011). In CHO cells, the use of UCOE elements upstream to the recombinant promoter leads to an increase of the selection for high expressing clones (Benton et al. 2002; Betts and Dickson 2015; de Poorter et al. 2007; Hou et al. 2014; Ye et al. 2010). The ability of UCOEs to maintain long-term expression has recently been shown by Betts and Dickson (2015) where the authors observed that the long-term productivity of recombinant CHO-DG44 cell lines was retained in the presence of the selection marker (Betts and Dickson 2015).

An increase in the number of selected high producing cells was also observed when a ß-globin Matrix-Attachment Region (MAR) element was incorporated into expression vectors (Kim et al. 2004). S/MARs are chromatin regulatory elements that enhance the formation of chromatin loops, acting as an anchor to the nuclear matrix and stabilising the loop (Mirkovitch et al. 1984). Recently, Ho et al 2015 evaluated the influence of two MARs, iMAR (human interferon ß MAR element) and cMAR (chicken lysonzyme MAR element) in recombinant expression regulated by SV40, EF1α and CMV promoters (Ho et al. 2015). The author observed that the influence of each MAR was different on the expression recombinant driven by each promoter (Ho et al. 2015). For SV40 promoter, both iMAR and cMAR led to an increase of expression, while no significant increase of expression was observed for EF1α and CMV promoters. Furthermore, Ho and colleagues also suggested that the influence

of MARs in maintaining long-term expression was also promoter-dependent. Both MARS maintained the stable long-term stable expression observed for SV40 promoter but failed to prevent the loss of productivity observed for EF1α and CMV promoters (Ho et al. 2015).

Another strategy to modify the gene integration site utilised Artificial Chromosome Expression technology (Kennard et al. 2009). This technology involves in the generation of a new CHO cell host that contains a pre-engineered artificial chromosome, which contains multiple recombination acceptor sites. This artificial chromosome allows the targeted integration of single or multiple genes for specific locations that were designed to improve the epigenetic environment. This strategy is reported to eliminate the random integration into the native CHO genome and is claimed to increase protein expression yields (Omasa et al. 2009)

1.4.5. Targeting gene integration site

Targeting recombinant gene sequences to genomic transcriptionally-active loci offers a powerful approach for obtaining good expressing cell lines. One of the strategies recently described to target the recombinant gene is the use of *Piggy*Bac transposon which mediates transgene integration. (Matasci et al. 2011). *Piggy*Bac transposons are class II transposable elements derived from insect *Trichoplusia ni* and specifically targets the TTAA tetranucleotides sites in the genome (Fraser et al. 1996). Although it is not directly targeting to any specific active region in CHO cells, the use of this strategy for gene integration leads to the generation of cell lines with greater expression cell lines and greater stability (Matasci et al. 2011). The authors suggested that this is may be due to the ability of the method to integrate the functional transgene through transposition, compared to the traditional random integration of the plasmid that often results in transgene disruption and inactivation (Matasci et al. 2011).

The major strategy used to target recombinant genes to specific loci is through the use of site-specific recombinant systems (Kim et al. 2012). The most promising site-specific systems used in mammalian cells are the Cre/loxP system from bacteria (Kito et al. 2002) and the Flp/FRT system from yeasts (O'Gorman et al. 1991; Sauer 1994). Both systems are based in homologous recombination between two DNA fragments catalysed by a specific enzyme. To target the gene, the enzyme recognises a small homologous sequence present in both host genome and gene sequence (Sauer 1994). Because this project used the Flp/FRT recombination the

next Section will describe this system in more detail and the strategies that can be used to integrate recombinant genes in a site-directed manner.

1.4.5.1 Flp/FRT recombination system

The Flp/FRT recombination system is a site-specific gene recombination mediated by the Flp-recombinase. Flp-recombinase is an enzyme that recognise and catalyses gene recombinations that are targeted by Flp-Recombinase Target (FRT) sequence (O'Gorman et al. 1991, Sauer 1994).

The minimal FRT site consists of a 34 base-pair (bp) sequence, which has two inverted 13 bp sequences separated by an 8 bp asymmetric spacer (O'Gorman et al. 1991). However the typical FRT site also contains a third 13 bp sequence upstream to the minimal FRT site as it improves the recombination process *in vivo* (Lyznik et al. 1993; Schlake and Bode 1994).

The 8 bp spacer is the key essential sequence for the correct recombination as this small sequence is involved in DNA-DNA pairing that occurs during the DNA strand exchange (Schlake and Bode 1994). Furthermore, it is the assymteric sequence of this spacer that determines the correct direction of site alignment, which will consequently lead to either inversion or excision (O'Gorman et al. 1991; Schlake and Bode 1994).

Flp/FRT systems rely on a three vector strategy, consisting of a cloning vector, an expression vector and the pOG44 vector (O'Gorman et al. 1991). The cloning vector that encodes a FRT sequence and a reporter gene, is used to generate the FRT host. The expression vector contains a multiple cloning site for the insertion of the recombinant gene and a FRT sequence to target the vector towards the FRT site on the FRT host cell. The pOG44 vector transiently expresses the Flp-recombinase which mediates gene recombination. Both expression and pOG44 vectors are co-transfected into the CHO FRT host in order to integrate the recombinant gene.

There are two Flp/FRT based strategies to target the insertion of the recombinant gene. The Flp/FRT site-specific gene integration uses one FRT sequence upstream to the gene of interest (Figure 1.10-A) (Huang et al. 2007; Merrihew et al. 1995). In contrast, in the Flp-recombinase Mediated Cassette Exchange (Schlake and Bode 1994) the gene of interest is flanked with two FRT sites which allows the exchange of recombinant gene in the same genome location (Figure 1.10-B).

The Flp/FRT recombination system has been successfully used in different mammalian cell lines. In CHO cells, there are reports using both Flp/FRT strategies to express different types of recombinant genes such as, polyclonal antibodies (Huang et al. 2007, Nehlsen et al. 2009, Wilberg et al 2006), erythropoietin (EPO) (Soo and Lee 2008; Zhou et al. 2010), human tissue-type plasmid activator (tPA) (Zhou et al. 2010) and secreted alkaline phosphatase (SEAP) (Zhou et al. 2010)

Figure 1.10. Flp/FRT site-specific recombination methods.
A – Flp/FRT site-specific gene integration [based on (Zhou et al. 2010)]; B – Flp-Recombinase Mediated Cassette Exchange [based on Schlake 1994)] (Orange line – cloning vector; Red – FRT site; Green line – chromatin; Blue line – expression vector; Brown circle – Flp-recombinase).
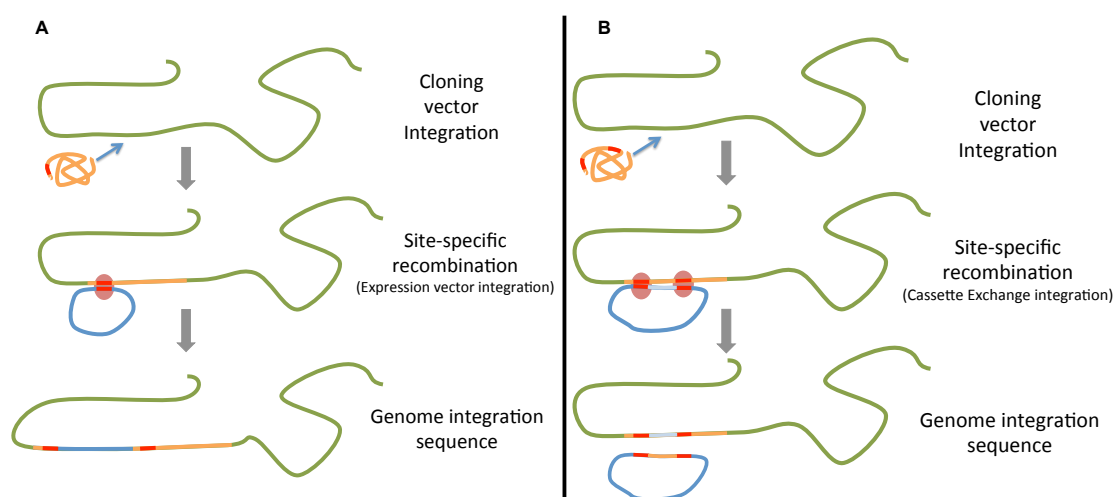
The site-directed method used in this project is based on the FLP-In™ system (Life Technologies). Flp-In™ is a commercially available site-directed system that follows a Flp/FRT site-specific gene integration approach To assess protein expression after cell line selection, the cloning vector pFRT/LacZeo encodes ß-galactosidase as a reporter gene. The gene of interest is integrated under the influence of CMV promoter by the expression vector (pcDNA5/FRT). The last main component, pOG44 vector transiently expresses Flp-recombinase that recognise both FRT sites and mediates the gene integration (Flp-In™ system manual).

To complement this system, Life Technologies also developed different mammalian cell lines (CHO, HEK293 and BHK) that allow the integration of the gene of interest to a pre-determined genomic location. Life Technologies claims that their proprietary

CHO host cell line encodes a single FRT integrated into a stable active locus that ensures a high and stable expression of recombinant gene (CHO Flp-In™ Manual).

1.5. Summary and project aims

The expression of recombinant protein involves the co-ordination of many cellular events, which together can effect production of functional protein. These cellular processes include transcriptional and translational regulation, protein folding and protein secretion. Whilst high productivity cell lines are an important consideration from an industrial perspective, it is also vital that this high productivity is maintained during the production process.

Although the selection and development of new CHO cell hosts, the optimisation of vectors, selection methods and cell culture medium have lead to major improvements on protein yield, there are still fundamental questions that arise in order to optimise the overall productivity.

One of the biggest concerns of the biopharmaceutical industry is the loss of recombinant gene expression in long-term cultures, described as cell line instability. This phenomenon arises from regulatory problems of gene expression control mechanism during the expression of recombinant genes. Although the fundamental mechanism that regulates gene transcription has been largely described for mammalian cells, the regulation of these mechanisms during recombinant gene expression is not fully described. Understanding the fundamental biology behind the regulation of recombinant gene transcription in CHO is one of the important steps on the pursuit of high and constant productivity.

Based on the position effect theory (Wilson et al. 1990), it is believed that different recombinant promoters will have different transcriptional activity in the same genomic locus. Also, that these promoters will vary in stability of expression during long-term culture when under the influence of the same genomic environment. Following these two principles, the overall aim of this project is to study transcriptional activity and stability of recombinant promoters in CHO cells.

Therefore, the aims for this project are:

a) Generate a cell/vector system to enable the study of recombinant promoters in CHO cells;

b) Generate a methodology to evaluate recombinant promoter activity in CHO cells when promoters were localised in the same genomic environment;

c) Characterise the activity of two widely-used recombinant promoters in CHO cells, SV40 and CMV promoters, in the same genomic environment;

d) Study the long-term expression of both promoters by evaluating the influence of each promoter in long-term expression of recombinant proteins in CHO cells;

e) Determine the transcriptional activity of each promoter in CHO cells *in vivo.*

To achieve the objectives in this project, Chapter 3 shows the generation and characterisation of a library of CHO cells that contain FRT sequences to target different promoters/recombinant gene cassettes. These cells were generated by integration of the Flp site-directed integration (based on the Flp-In™ system) in CHO-S cells and express ß-galactosidase under the influence of SV40.

Due to the possible impact that this project could have in the bioprocessing industry, the methodology used to generate these cell lines was intended to be as close as possible to the industry standards within the academic resources. Because of the nature of the industry protocols for the achievement of high expression cell lines, it is possible that the cell lines generated have multiple FRT sequences within their genome. The possible presence of multiple FRT sequences, represent a possible caveat of these cell lines as the ideal platform for the target of promoters/genes to the same genomic location. For that reason it was imperative that in Chapter 3, a full cell line stability characterisation was performed to understand the nature of each cell line and to support the analyses of the results Chapters 4 and 5.

Chapter 4 shows the development of the methodology to study and characterise the recombinant promoters in CHO cells. Fluorescent proteins were integrated into the CHO-FRT cells developed in Chapter 3, in a site-directed manner, under the influence of SV40 and CMV promoters. The activity of the promoters was evaluated by quantifying the expression of the fluorescent proteins whilst the influence of the each promoter in the long-term expression, in these cell lines, was evaluated by monitoring the expression of the fluorescent proteins throughout a long-term culture. Based on the nature and previous characterisation of both SV40 and CMV promoters it is expected that the expression driven by SV40 promoter is lower than CMV promoter but possibly more stable (Du et al. 2013; Kim et al. 2011; Ho et al. 2015; Mariati et al. 2014b; Osterlehner et al. 2011; Spencer et al. 2015; Williams et al. 2005; Yang et al. 2010). However, due to the possible multiple presence of FRT sequences, it is possible that the expected differences between promoters could be diluted by the generation of heterogeneous populations on the resulting cell lines.

Finally, Chapter 5 shows the strategy used to quantify the transcriptional activity of both promoters in CHO cells *in vivo*. The vector system that was developed contained a MS2 RNA tag sequence cloned downstream of the recombinant gene that can be used to quantify the transcriptional output of the recombinant gene in single living cells using fluorescent microscopy. Because there are no reports of the use of MS2 RNA tag technology in CHO cells, it is imperative to firstly certify that this technology can be successfully used in CHO cell. If successfully implemented, it would be expected that the transcription rate of CMV promoter *in vivo* is considerably higher than SV40 promoter in the similar conditions.

2. Materials and Methods

2.1. Materials

2.1.1 Sources of chemicals and reagents

All chemical reagents were of the highest grade and obtained from standard sources. The materials used, and their suppliers, are listed in Appendix 1.

2.1.2. Source of DNA plasmid vectors

2.1.2.1. Flp-In™ system

Complete Flp-In™ system was purchased from Life Technologies. The major components of this system are the pFRT/lacZeo, the pCDNA5/FRT and the pOG44 vectors (Appendix 2).

The pFRT/lacZeo vector encodes a lacZ-Zeocin™ fusion gene, whose expression is under control of a SV40 promoter and a FRT sequence downstream to the ATG initiation codon of the lacZ-Zeocin™ fusion gene.

The pCDNA5/FRT vector contains a FRT sequence to target the insertion of the recombinant gene in a site-directed manner. The vector contains a multicloning site for integration gene of interest under the control of human cytomegalovirus (CMV) promoter and the hygromycin resistance gene for the selection of the positively integrated cell lines.

The pOG44 plasmid encodes the Flp recombinase gene used to mediate the homologous recombination of the FRT sequences.

2.1.2.2. Other vectors

The vectors pRSET/YFP and pRSET/mStrawberry were kindly donated by Dr. Luis Stephano (Faculty of Sciences, The Autonomous University of Baja California). The vector pGEM-MS2(11) was kindly provided by Dr. Mark Ashe (Faculty of Life Sciences, University of Manchester). pMS2-GFP was generated by Dr. Robert Singer and was acquired from Addgene (plasmid # 27121). p901-GFP was generated in the Dickson laboratory by Alexandra Croxford. (Vector maps in Appendix 2)

2.1.3 Preparation and sterilisation of solutions

All solutions were prepared in milliQ water (ddH2O) unless otherwise stated. If necessary, solutions were sterilised by autoclaving in a LTE Scientific Series 250

autoclave or, in cases for which autoclaving was not suitable, by filtration through a 0.22 µm sterile filter.

2.1.4 pH measurements

Measurements of pH were made using a digital Corning pH meter 120 with a glass electrode. The pH was adjusted using hydrochloric acid or sodium hydroxide, as appropriate, unless otherwise stated.

2.1.5. Bacterial cells and culture medium

E. coli DH5α or *E. coli* XL1-Blue were used for the propagation of plasmids. The medium used to grow these cells was Luria Bertuni broth (1% (w/v) Tryptone; 0,5% (w/v) Yeast Extract; 0.5% (w/v) Sodium Chloride). Solid Luria-Bertuni (LB) medium was prepared using the above constituents supplemented with 1.5% (w/v) of agar. Both media were autoclaved before use and supplemented with 100 µg/ml filter-sterilised ampicillin, when required. For solid LB medium, ampicillin was added when medium temperature (following autoclaving) was below 55ºC.

2.1.6 Mammalian cell lines and culture medium

Chinese Hamster Ovary S (CHO-S) cell line was routinely cultivated in CD CHO medium (Life Technologies) supplemented with 1x GlutaMax (Life Technologies). For the selection of transfected cells, the medium was supplemented with Zeocin (Life Technologies) or Hygromycin (Life Technologies), 0.5mg/ml final concentration for either, depending on the selection marker introduced by the transfected vector. For the maintenance of selected cell lines the medium was supplemented with 0.25 mg/ml Zeocin or Hygromycin, depending on the selection method used.

2.2 Generation and purification of plasmids

This Section focuses on the molecular biology methods used to generate, purify and characterise plasmid DNA used in this project

2.2.1. Generation of competent *E. coli* cells

An aliquot of *E. coli* DH5α cells or XL1-Blue was quickly thawed by hand. A 10 µl sample of *E. coli* was placed in 5 ml of SOB medium (2% (w/v) Tryptone; 0.5% (w/v) Yeast Extract; 10mM Sodium Chloride; 2.5mM Potassium Chloride; 20mM Magnesium Sulphate) and incubated overnight at 37ºC, shaking at 230rpm. The culture was added to 195ml of SOB medium and incubated at 37ºC (again at 230rpm) until the OD_{600nm} reached 0.6. Bacterial cells were harvested by centrifugation of the culture at 1000 g for 15 min. at 4 ºC and then the pellet was gently resuspended in 60ml ice-cold Transformation Buffer (Huang et al. 2007) (10 mM PIPES 15 mM $CaCl_2$ –$2H_2O$ 250 mM KCl, 55 mM $MnCl_2$, pH 6.8) The cells were incubated for 10 min. on ice and then centrifuged again at 1000 g for 15min. at 4ºC. The supernatant was discarded and the pellet was then gently resuspended in 15ml ice-cold TB. 7% (v/v) of DMSO was added to the cell suspension, which was then incubated on ice for 10 min. The resultant cell suspension was aliquoted into ice-cold empty sterile microfuge tubes (100 µl per tube) and snap-frozen in liquid nitrogen. The cell aliquots were stored in -80ºC until use.

2.2.2. Transformation of competent *E. coli* cells

An aliquot containing 100 µl of competent *E. coli* cells (Section 2.2.1) was thawed on ice (about 15 min). 100-300 ng of plasmid was transferred to the aliquot tube and gently swirled to mix. The mixture was incubated on ice for 30 min, heat-shocked at 42 ºC for 40 sec. and then incubated on ice for at 5-10 min. In a sterile environment as described in Section 2.2.1, 700ul LB broth was added to the cell mixture and incubated for 1 hour at 37 ºC shaking at 230 rpm for allow recovery. 50 µl cell culture was spread on a warm LB agar plate containing ampicillin and plates were incubated overnight at 37 ºC. The colonies obtained were analysed and the plate was kept for no more than 3 weeks at 4 ºC.

For the transformation of competent cells with DNA vector resulting from a ligation reaction, 5µl ligation reaction mix was added to each 100 µl aliquot of competent cells. After the heat-shock process, performed as mentioned above, 500µl SOC medium was added to the cell mixture and this was incubated for 2 hour at 37 ºC. 200 µl cell incubation was spread on a warm LB agar plate (containing ampicillin) and plates were incubated (lid up) for 1 hour to allow the absorption of excess recovered cell medium, with subsequent overnight incubation (lid down) at 37 ºC.

2.2.3. Mini and Maxi-preparation of plasmid DNA

Mini-preparation of plasmid DNA was used for analysis of plasmids resulting from ligation reactions. A colony obtained on the LB agar plate after transformation (Section 2.2.2) was added to a 50 ml sterile Falcon tube containing 6 ml LB medium (supplemented with ampicillin, 100mg/ml) and then cultured overnight at 37 ℃ shaking at 230 rpm. The cell culture was centrifuged in a 1.5 ml microcentrifuge tube (4x 1.5 ml) for 1 min at 13000 x g and the plasmid was then purified from the pellet using the Qiagen Plasmid Mini Kit (Quiagen).

To prepare plasmid DNA for transfection, a colony from a LB agar plate was added to 200 ml of LB medium with ampicillin (100mg/ml) and this mixture was cultured overnight at 37 ℃ shaking at 230 rpm. The cell culture was centrifuged in 50 ml sterile Falcon tubes (4x 50 ml) for 15 min at 5000 x g. The plasmid DNA was purified and sterilised from the cell pellet using the Qiagen HiSpeed Plasmid Maxi Kit (Qiagen)

A sample of purified plasmid DNA obtained was used for DNA quantification using a NanoDrop^{TM} 1000 (Thermo Scientific).

2.2.4. Restriction digestion of plasmid DNA

100-300 ng plasmid DNA was digested by 10 U of appropriate restriction enzyme in a reaction mixture with 1X final concentration of appropriate restriction enzyme buffer in a final volume of 20 μl. The reaction mixture was incubated for 2 h at 37 ℃.

To digest plasmid DNA for cloning processes, 1μg plasmid DNA was incubated with 20 U of each restriction enzyme in a 50 μl mixture containing 1X final concentration of the appropriate restriction enzyme buffer. The reaction mixture was incubated for 2 h at 37 ℃

2.2.5. Polymerase Chain reaction (PCR)

The primers used for PCR reactions (Table 2.1) were reconstituted in ddH$_2$O to stock solutions of 100 μM. Prior to use primers were diluted to working solutions with a final concentration of 10 μM.

For each PCR reaction, in a 0.2 ml thin-wall PCR tubes, 100 ng plasmid DNA (Section 2.2.3) was mixed with 5 μl 10X BioTaq polymerase buffer (Bioline), 2 μl MgCl2 solution (Bioline), 1 μl 100 mM dNTP mix (Bioline), 1.5 μl forward primer and

1.5 µl reverse primer working solutions, 1 µl BioTaq polymerase (Bioline) and ddH2O was added up to a final volume of 50 µl. Reaction mixes were made quickly, vortex-mixed and pulse-centrifuged before placing the tubes in a T-3000 Thermal Cycler.

Mixtures were incubated for 5 minutes at 94°C for denaturation followed by 35 cycles of denaturation for 30 seconds at 94°C, annealing for 30 seconds at X°C (where X < 70 ºC and is typically the Tm of the primer pair – 2 ºC; Table 2.1), and elongation for 1 minutes at 72°C. A final elongation step for 10 minutes at 72°C was performed before storing the samples at 4 ºC until further use.

Table 2.1. Primers used in PCR reactions.

Primer	Sequence
SV40 forward	5'-CCCAGATCTGGCAGCTGTGGAATGTGTG-3'
SV40 reverse	5'-CCCGCTAGCTAGCCTCCAAAAAAGCCTCC-3'
GFP forward	5'-CCCGGATCCATATGGCCAGCAAAGGGGAAG5-3'
GFP reverse	5'-CCCCTCGAGTTATTACTTGTACACAGCTCATCC-3'

2.2.6. Ligation of PCR products into pCR2.1 cloning vector

PCR products were cloned into the pCR2.1 vector using the TA Cloning Kit (Invitrogen). The ligation reactions contained 2 µl fresh PCR reaction (Section 2.2.5), 2 µl pCR2.1 vector (Invitrogen), 1 µl 10x T4 Ligation Buffer (Roche), 1 µl (10U) T4 DNA ligase (Roche) and 4 µl ddH$_2$O in a 0.2 ml thin-wall PCR tube. Reactants were incubated overnight (up to 16h) at 14 ºC in a T-3000 Thermal Cycler and then used to transform competent *E. coli* cells (Section 2.2.2).

2.2.7 Purification of linearised DNA fragments or PCR products

DNA linearized by restriction digestion with one restriction enzyme (Section 2.2.4) or PCR reaction products (Section 2.2.5) were purified using the QIAquick PCR Purification Kit (Quiagen) following the manufacturer's protocol. Purified fragments were eluted in ddH2O for use in ligation reactions (Section 2.2.10)

2.2.8. Agarose Gel Electrophoresis of DNA Fragments

Agarose gel electrophoresis was used to separate DNA fragments through their size and consequently analyse restriction digestion results. The agarose gel was prepared by dissolving 1% (w/v) agarose in TBE buffer (0.09 M Tris, 0.09 M OrthoBoric Acid, 0.2 mM EDTA). To help the dissolution of the agarose into the TBE buffer, the mixture was boiled in a microwave. The agarose solution was cooled, until it was comfortable to touch the container against the hand (approximately 50 ºC) and then Ethidium Bromide or Safer Green fluorescence Dye was added to a final concentration of 0.01% (v/v). The agarose mixture was poured into a gel bed mould with the appropriate comb until set.

For electrophoresis, the gel was transferred to a running cell tank and immersed in TBE buffer. Before loading, the samples were mixed with Sample Loading Buffer (Bioline) to a 1X final concentration. DNA Hyperladder^{TM} 1kb (Bioline) and the samples were loaded into wells and electrophoresis was performed at 100V until the blue dye of the Sample Loading Buffer (Bioline) reached the end of the gel (approximately 50-70 min). DNA fragments were visualized and analysed by UV light using a UV transilluminator.

2.2.9. Purification of DNA Fragments from Agarose gels

Purified DNA fragments were isolated following restriction digestion (Section 2.2.4) and agarose gel electrophoresis (Section 2.2.8) for use in ligation reactions (Section 2.2.10). The selected DNA fragments were cut from agarose gels using a razor blade on a UV transilluminator and purified using the QIAquick Gel Extraction Kit (Quiagen) according to the manufacturer's protocol. The purified DNA fragments were eluted in ddH$_2$O.

2.2.10 Ligation of DNA fragments with vectors

All ligation reactions were performed at three different molar ratios 3:1, 1:1 and 1:3 with the amount of vector kept constant (150-250ng). Ligation reactions were performed in a 0.2 ml thin-wall PCR tube by mixing the DNA fragment and vector with the correct molar ratio with 1 µl 10X T4 ligase buffer (Roche), 1 µl (10U) T4 DNA ligase (Roche) and ddH2O up to a final volume of 10 µl. Ligation mixes were incubated overnight (up to 16 hours) at 14ºC in a T-3000 Thermal Cycler then used to transform competent *E. coli* cells (Section 2.2.2).

2.3. Construction of mammalian expression vectors for site-directed integration.

Figures 2.1 - 2.5 summarise strategies and methods used to create mammalian expression vectors used in this project. More detailed plasmid maps of vectors generated are included in Appendix 2.

Figure 2.1. Schematic representation of the strategy used to create pS/FRT plasmid.
SV40 promoter was amplified by PCR as described in Section 2.2.5 using the pFRT/LacZeo
SV40 promoter as template sequence and SV40 Forward and SV40 reverse primers (Table
2.1). The amplified SV40 promoter was ligated into pCR2.1 vector (Section 2.2.6) and
transformed into competent E.coli cells (Section 2.2.2). Positive clones containing SV40
promoters were screened by mini-preparation (Section 2.2.3), restriction digestion with BglII
and NheI (Section 2.2.4) and agarose gel electrophoresis (Section 2.2.5). BglII and NheI
restriction sites were not present in the pFRT/lacZeo vector but were added to the 5' and 3'
(respectively) of the amplified SV40 promoter sequence by the primers used for amplification.
1 µg of each pCR2.1-SV40pr and pcDNA5/FRT were digested with BglII and NheI restriction
enzymes (Section 2.2.4), separated by agarose gel electrophoresis (Section 2.2.5) and the
selected bands were purified from the gel (Section 2.2.6). The SV40 promoter fragment and
the linearised pcDNA5/FRT vector were ligated (Section 2.2.10) and transformed into
competent E.coli cells (Section 2.2.2). Positive clones were screened by mini-preparation
(Section 2.2.3), restriction digestion with BglII and NheI (Section 2.2.4) and agarose gel
electrophoresis (Section 2.2.5). (The size and proportion of the Figures are not represented at
scale)

Figure 2.2. Schematic representation of the strategy used to create pcDNA5/FRT-GFP and pS/FRT-GFP plasmids.

Methods and strategy used to create pcDNA5/FRT-GFP and pS/FRT-GFP vectors are detailed in the legend of the Figure 2.1. The GFP sequence was amplified from vector p901-GFP using the primers GFP forward and GFP reverse (Table 2.1) and inserted into pcDNA5/FRT and pS/FRT vectors. The restriction enzymes used for screening and cloning were BamHI and XhoI and were added to the 5' and 3' (respectively) ends of the amplified GFP sequence by the primers used for amplification. (The size and proportion of the Figures are not represented at scale)
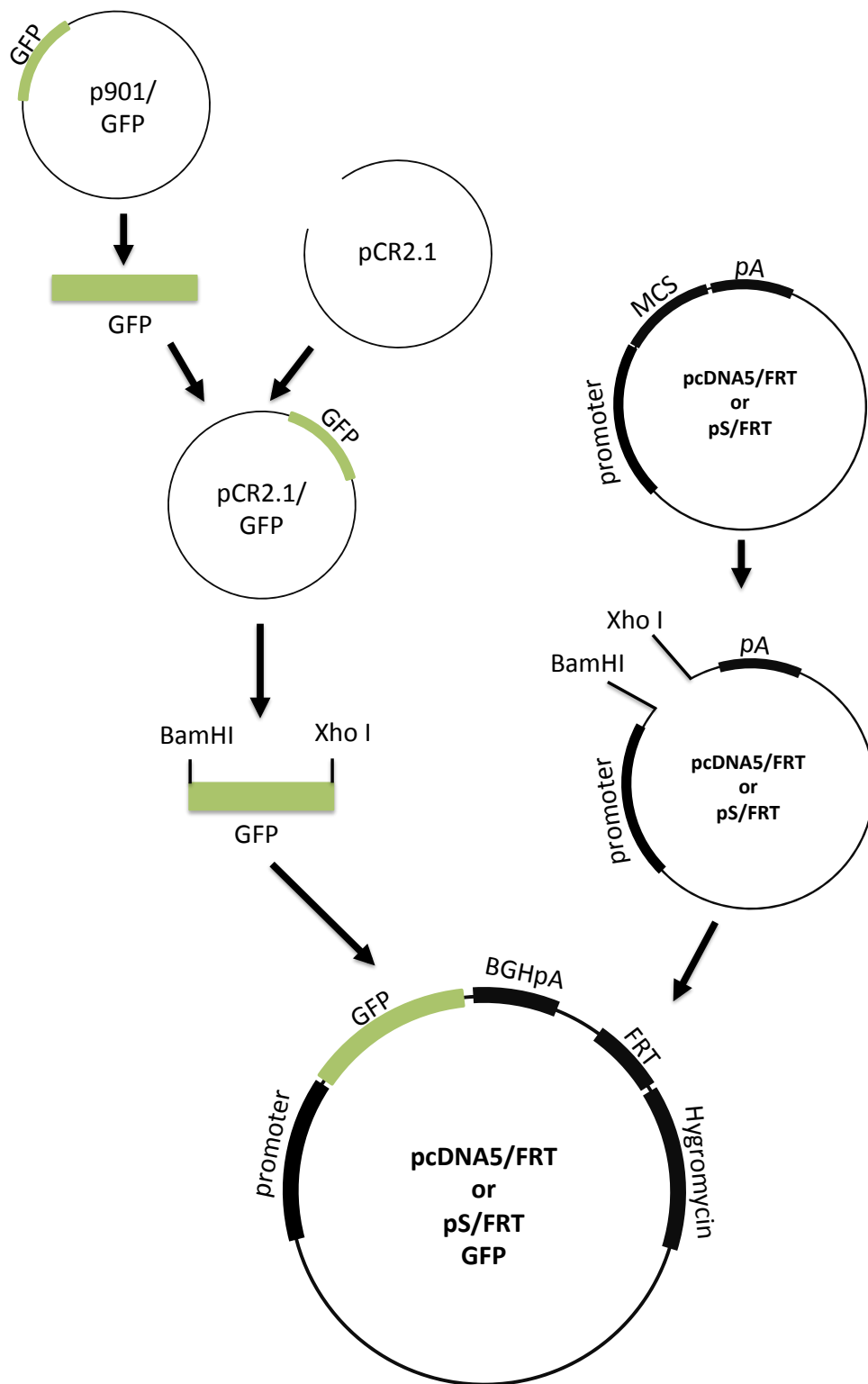
Figure 2.3. Schematic representation of the strategy used to create pcDNA5/FRT/MS2 and pS/FRT/MS2 plasmids.

1 μg each of pGEM-MS2(11), pcDNA5/FRT and pS/FRT were digested with BamH I and EcoR V restriction enzymes (Section 2.2.4), separated by agarose gel electrophoresis (Section 2.2.5) and the required bands were purified from the gel (Section 2.2.6). The 11x MS2-loops (MS2(11)) DNA fragment and the linearised pcDNA5/FRT and pS/FRT vectors were ligated (Section 2.2.10) and transformed into E.coli competent cells (Section 2.2.2). Positive clones containing MS2(11) sequence were screened by mini-preparation (Section 2.2.3), restriction digestion with BamHI and EcoRV (Section 2.2.4) and agarose gel electrophoresis (Section 2.2.5). The same process was repeated to clone a second MS2(11) sequence into the pcDNA5/FRT/MS2(11) and pS/FRT/MS2(11) vectors with the difference that Eco53KI and NotI restriction enzymes were used for the cloning process. For the screening process for positive clone containing 22x MS2-loops (MS2(22)) sequence, BamH I and Not I restriction enzymes were used. (The size and proportion of the Figures are not represented at scale)
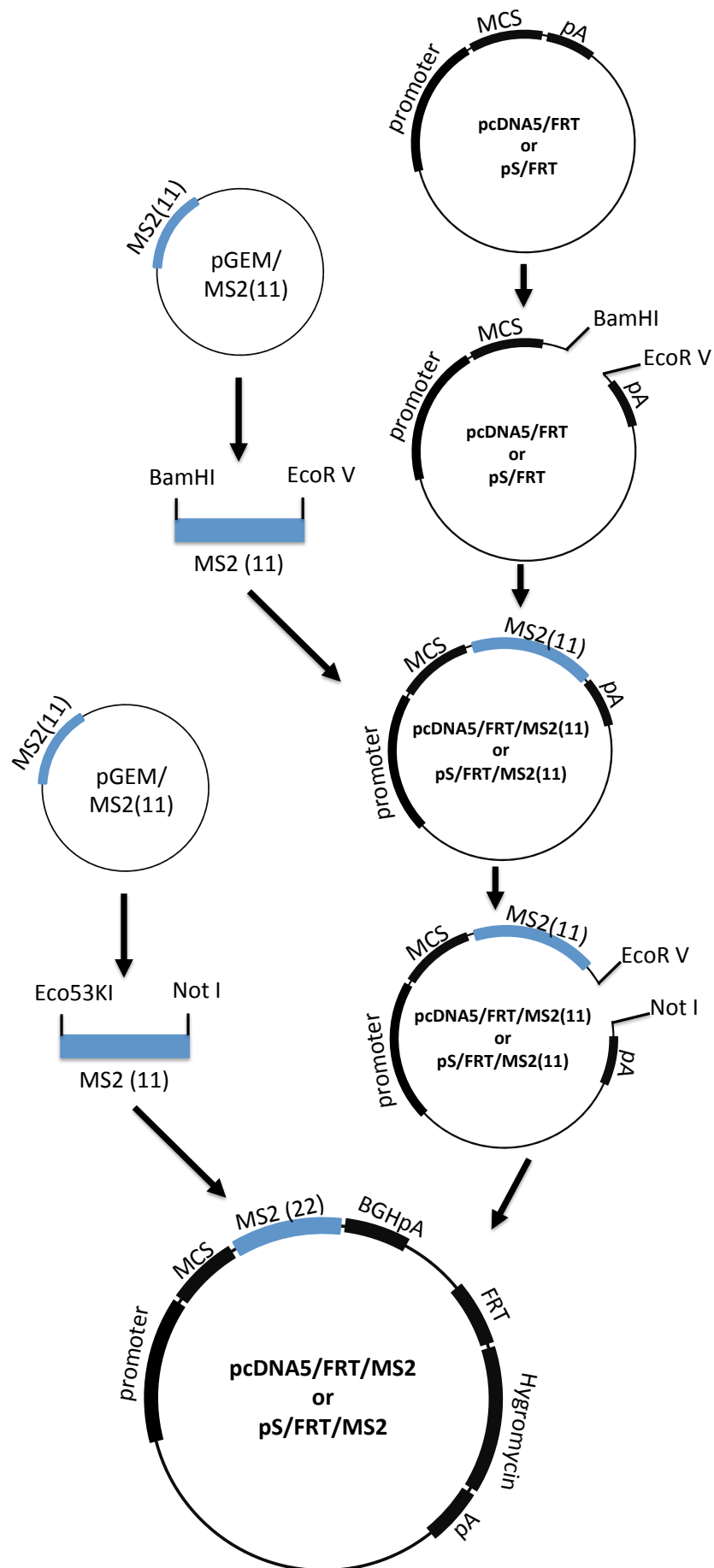
Figure 2.4. Schematic representation of the strategy used to create pcDNA5/FRT/MS2-YFP and pS/FRT/MS2-YFP plasmids.

1 μg of each pRESET/YFP, pcDNA5/FRT/MS2 and pS/FRT/MS2 were digested with Nhe I and Hind III restriction enzymes (Section 2.2.4), separated by agarose gel electrophoresis (Section 2.2.5) and the required bands were purified from the gel (Section 2.2.6). The YFP fragment and the linearised pcDNA5/FRT/MS2 and pS/FRT/MS2 vectors were ligated (Section 2.2.10) and transformed into E.coli competent cells (Section 2.2.2). Positive clones containing YFP sequence were screened by mini-preparation (Section 2.2.3), restriction digestion with NheI and HindIII (Section 2.2.4) and agarose gel electrophoresis (Section 2.2.5). (The size and proportion of the Figures are not represented at scale)
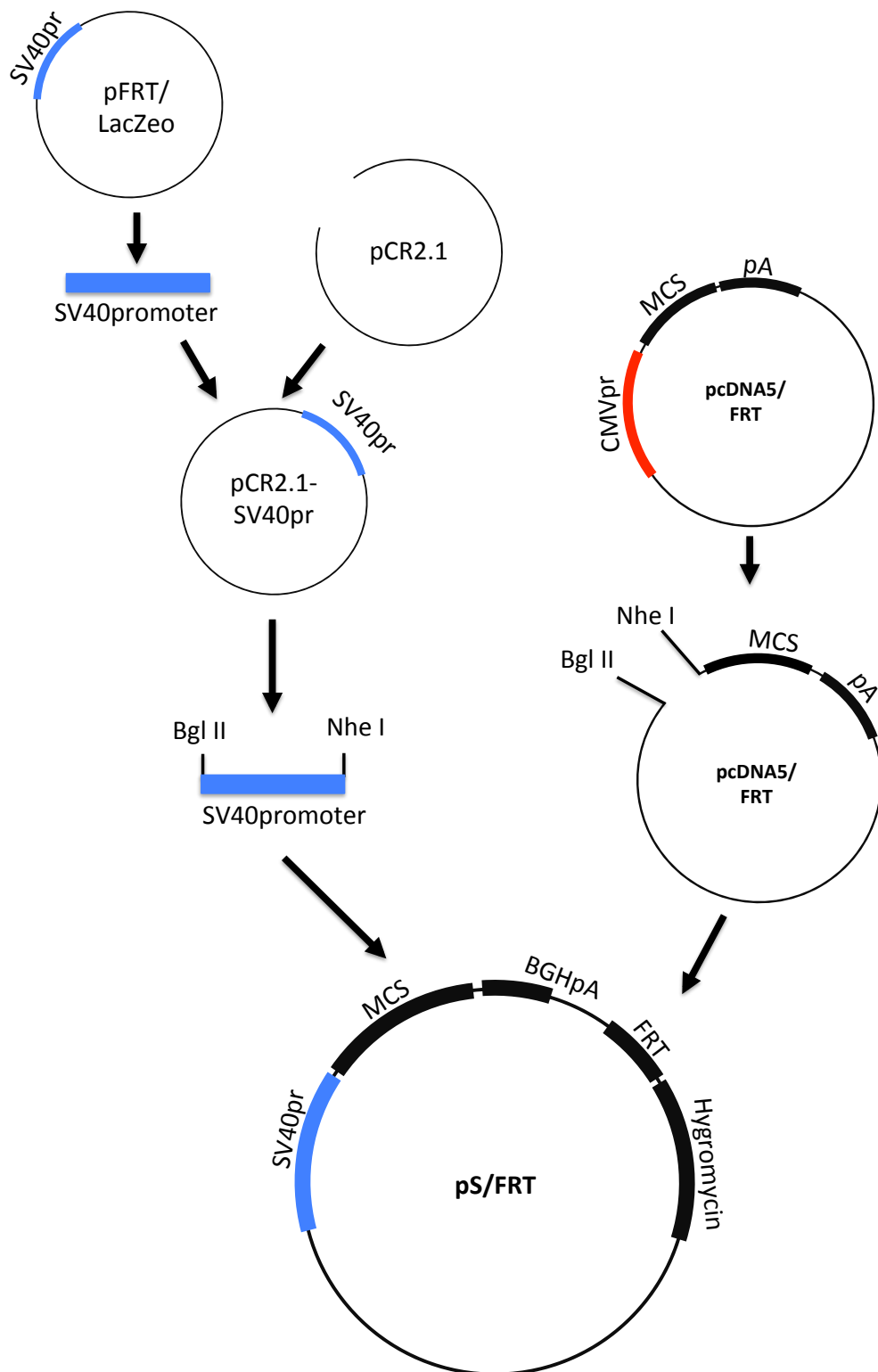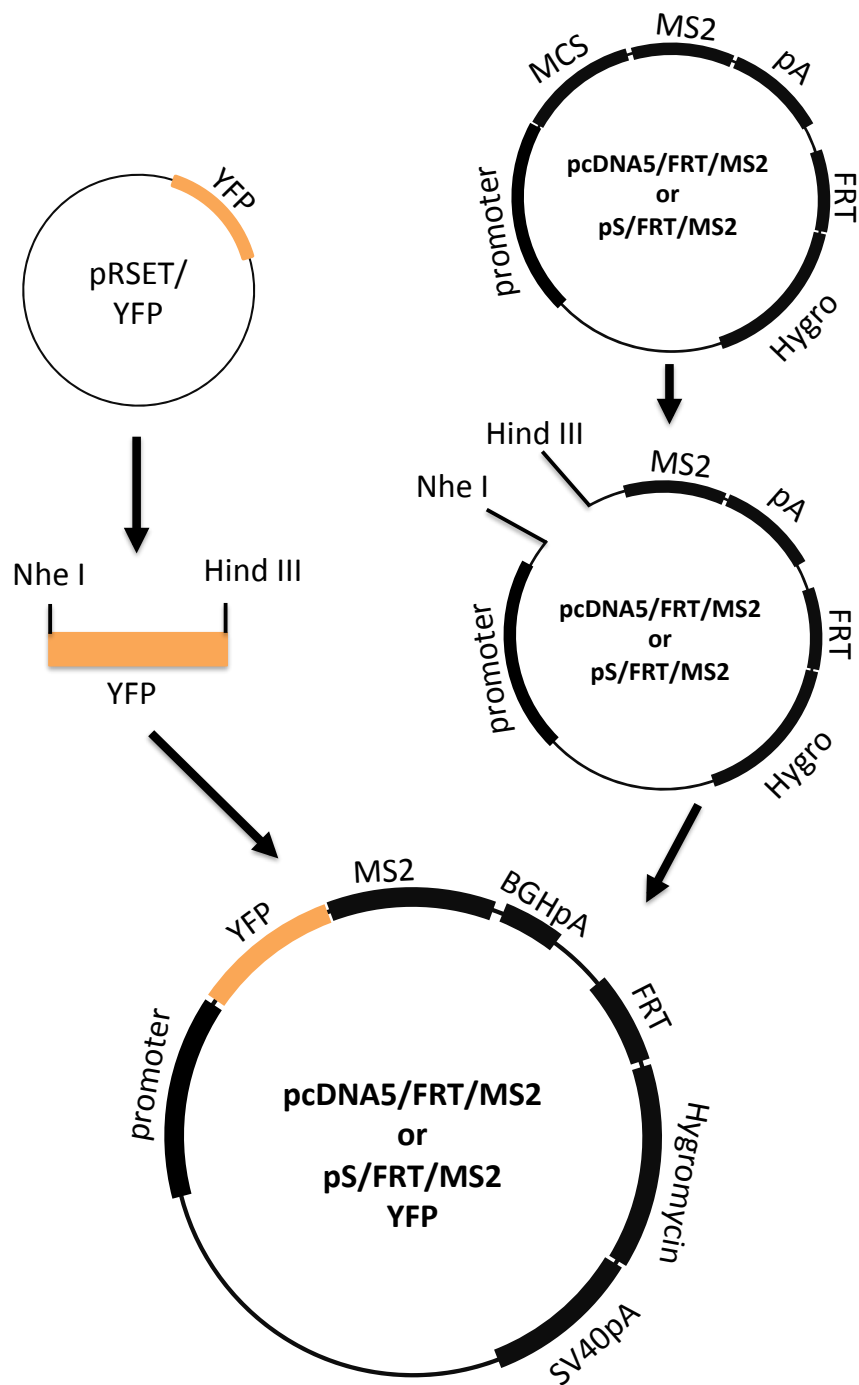
Figure 2.5. Schematic representation of the strategy used to create pcDNA5/FRT/MS2-YFP and pS/FRT/MS2-YFP plasmids.

The methods and process used to clone mStrawberry into pcDNA5/FRT/MS2 and pS/FRT/MS2 vectors are described in the legend of Figure 2.4. (The size and proportion of the Figures are not represented at scale)
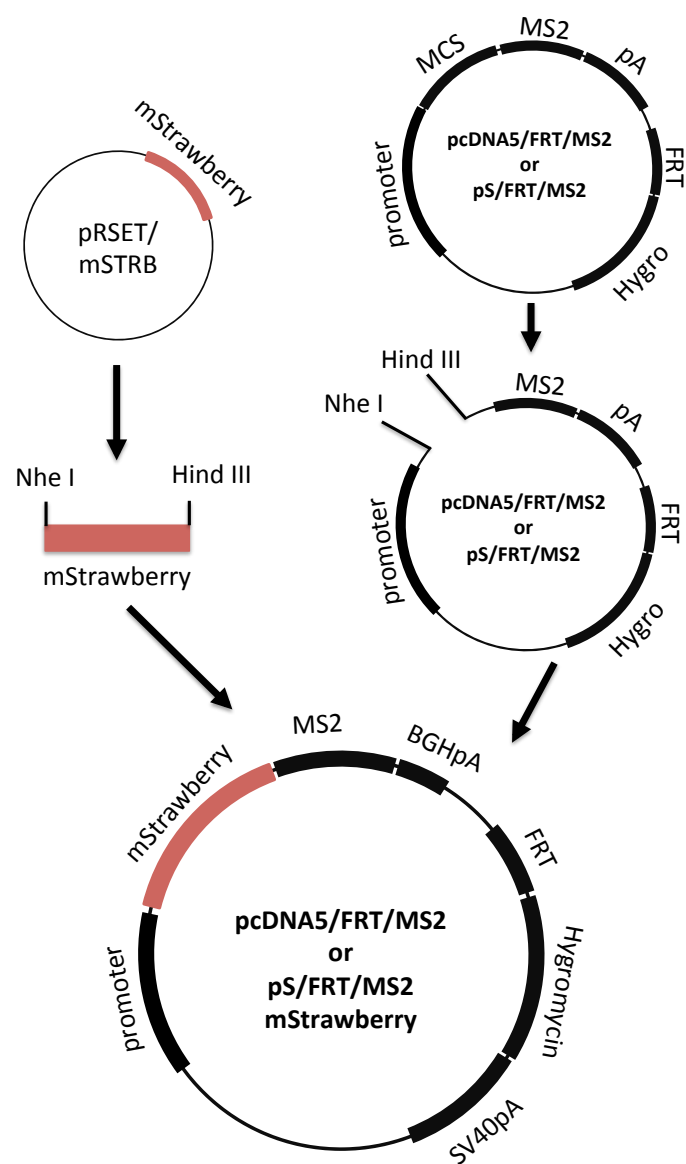
2.4. Mammalian Cell Culture

All cell culture material, media and additives were sterilised either by the manufacturer prior to purchase or by filtration through a 0.22 µm sterile filter. To guarantee a sterile environment the entire cell culture procedure was performed within a sterile laminar flow cell culture cabinet.

2.4.1. CHO Cell Host Revival

A vial of cryopreserved CHO cells was removed from liquid nitrogen and thawed quickly in a water bath at 37 ºC with gently swirling. Immediately, the vial contents were transferred to a 50 ml sterile Falcon tube containing 20 ml CD CHO medium supplemented with 1X Glutamax and this mixture was centrifuged for 5 minutes at 160 x g. The supernatant was removed by gently pouring to a container and the pellet of cells was resuspended in 10 ml fresh warm CD CHO medium supplemented with 1X Glutamax.

For CHO-S cell lines, the 10 ml of resuspended cells were transferred to a 125 ml vented capped Erlenmeyer flasks containing 10 ml CD CHO medium supplemeneted with 1X Glutamax. The cell culture was then incubated at 37 ºC with 5 % CO_2 and shaking (130 rpm).

For ß-galactosidase- and fluorescent protein-expressing cell lines, 10 ml of cells were transferred to a T75 cm^2 flask containing 10 ml CD CHO medium supplemented with 1X Glutamax and incubated in at 37 ºC with 5% CO^2 for 48 hours. The cells were then transferred to a 125 ml vented capped Erlenmeyer flasks containing 10 ml of CD CHO medium supplemented with 1X Glutamax and incubated at 37 ºC with 5 % CO_2 and shaking (130 rpm).

2.4.2. CHO Cell Culture Maintenance

CHO-S cells were routinely cultured in 125 ml vented capped Erlenmeyer flasks at 37 ºC with 5 % CO_2 ,shaking at 130 rpm. Cells were sub-cultured at an initial cell density of $0.2x10^6$ cells/ml every 3 or 4 days into a final volume of 30 ml CD CHO medium supplemented with 1X Glutamax. At each sub-culture, cell density and cell viability was determined by light microscopy and trypan blue exclusion (Section 2.4.3), for seeding the new culture at the required initial cell density.

For ß-galactosidase- and fluorescent protein-expressing CHO cell lines, the cells were routinely cultured in a similar process to CHO-S cell lines but with medium supplemented with 0.25 mg/ml of the appropriate selective marker (Zeocin for ß-galactosidase-expressing cell lines and Hygromycin for fluorescent protein-expressing cell lines).

2.4.3. Measurement of CHO Cell Number and Viability

Measurement of Cell Number and viability was performed either by light microscopy and trypan blue exclusion or by using the Countess® Automated Cell Counter (Life Technologies).

For light microscopy, the cell suspension sample was diluted 1:1 with 0,5% (v/v) trypan blue in PBS and placed into a Neubauer haemocytometer. Under light microscopopic observation, live cells, with intact plasma membranes, were observed as bright white, while dead cells had dark blue nuclei due to the uptake of trypan blue through their compromised plasma membrane. Both live and dead cells were counted in 4 fields of 0.1 mm^3 and then used to calculate the total and viable cell density.

The Countess® Automated Cell Counter (Invitrogen) using the same trypan blue exclusion method to analyse the cell viability with the difference being that the cells are counted by the Countess® Automated Cell Counter. For this method, the cell suspension was diluted 1:1 with 0,4% (v/v) trypan blue in PBS and placed into a Countess® Cell Counting Chamber Slide (Invitrogen). The slide was then introduced into the Countess® Automated Cell Counter and density and viability readings were recorded.

2.4.4. Cryopreservation of CHO Cells

To prepare cells for cryopreservation, a sample of culture containing approximately 1.0x10^7 cells in exponential phase of culture was centrifuged for 5 min. at 160 x g. The supernatant was removed and the pellet was gently resuspended in 1 ml of freezing medium (90% [v/v] CD CHO + 1X Glutamax, 10% [v/v] DMSO). The cell suspension was quickly transferred to 1.6 ml CryoPure Tubes (Sarstedt) and immediately frozen slowly within a polystyrene box or within a 5100 Cryo 1°C Freezing Container at -80 ºC for at least 48 h before transfer to liquid nitrogen for long-term storage.

2.4.5. Batch Culture Analysis

To perform a batch culture, cells were seeded at an initial cell density of 0.2×10^6 cells/ml a total volume 30 ml culture medium in a 125 ml vented capped Erlenmeyer flask. The cells were incubated at 37 ºC with 5 % CO_2 and shaking (130 rpm) and total cell density and cell viability was monitored daily using the Countess® Automated Cell Counter (Section 2.4.3). The cell culture was continued until cell viability reach <40%.

2.5. Generation and Characterisation of CHO-FRT mini-pools

2.5.1. Generation of CHO-FRT mini-pools

2.5.1.1. CHO-S Zeocin resistance test

CHO-S cells in the exponential phase of culture were diluted to a final cell density of 5.0×10^3 cells/ml in 6x 1 ml of CD CHO supplemented with 1X GlutaMax. Zeocin was added to generate final concentrations from 0-1.0 mg/ml. 200 µl of each preparation was added to 3 wells of a 96 well plate and incubated for 10 days at 37 ºC with 5% CO_2 in a static incubator. CHO-S cell growth in different culture conditions was analysed and compared by light-microscopy in a Cellomics® ArrayScan® VTI HCS Reader (Thermo Scientific).

2.5.1.2. pFRT7lacZeo vector linearisation

Prior to use in transfections, pFRT/lacZeo vector was linearised with *Sca* I restriction enzyme as recommended by the manufacturer. 15 µg maxi-prep of pFRT/lacZeo plasmid DNA was incubated with 50 U of ScaI – HF restriction enzyme (New England Biolabs), in a 250 µl mixture containing 1X final concentration CutSmart buffer and ddH$_2$O. The reaction mixture was incubated for 2 h at 37 ºC

2.5.1.3. Ethanol precipitation of linearised DNA

Before transfection, the linearised DNA was purified and sterilised. The linearised DNA (Section 2.5.1) was incubated with a 1/10 reaction volume of sodium acetate (3M, pH 5.5) and 2.5 vol. reaction volume of ice-cold ethanol for 2 h at -80 ºC. The mixture was centrifuged for 30 min. at 4 ºC in a microcentrifuge at 16000 g and the

supernatant was discarded. The pellet was rinsed with 300 µl 70% (v/v) ethanol and then incubated for 10 minutes on ice. The mixture was re-centrifuged using the same conditions as above. Under sterile conditions in a laminar flow cell culture cabinet, the supernatant was removed and the pellet was air-dried for 10-20 min. The final DNA pellet was resuspended in 110 µl autoclaved ddH$_2$0. A 10 µl sample was quantified using a NanoDrop^{TM} 1000 (Thermo Scientific) and analysed in an agarose gel (Section 2.2.8).

2.5.1.4. Transfection of CHO cells by electroporation

CHO-S cells were transfected with linearised plasmid prepared as described in Section 2.4. For a transfection reaction, $1.0x10^7$ cells were harvested from mid-exponential culture by centrifugation at 160 g for 5 min. The supernatant was removed and the cells were resuspended in 700 µl fresh CD CHO supplemented with 1X Glutamax. The resuspended cells were transferred to a 0.4 cm path length electroporation cuvette (Bio-Rad) and 3 µg of plasmid DNA was added to the cuvette and mixed by gentle swirling. The cells were electroporated using a Gene Pulser Xcell Electroporation System (Bio-Rad) at 0.3 kV and 950 µF for a time of 16-18 ms. The cell mixture was removed from the cuvette and gently resuspended in 20 ml warm CD CHO supplemented with 1X Glutamax. Cells were transferred to a T75 cm^2 flask and incubated overnight at 37ºC with 5% CO$_2$. A negative control was performed in parallel, in which the transfection was performed with sterile water instead of plasmid.

2.5.1.5. Clone selection and amplification

On the day following electroporation, cells from plasmid and negative control transfections were diluted to a final concentration of $5.0x10^3$, $1.0x10^3$ and $0.25x10^3$ cell/ml in selective medium (CD CHO supplemented with 1x GlutaMax and 0.5 mg/ml Zeocin). 200 µl portions of cell suspension were added to each well of 96 well plates (14 plates for the plasmid transfections [3 plates for $5.0x10^3$ cell/ml, 4 plates for $1.0X10^3$ cell/ml and 7plates for $0.25x10^3$ cell/ml dilution] and 2 plates for the control). The plates were wrapped in foil and incubated in static conditions at 37ºC with 5% CO$_2$ for 10 days. After 10 days of incubation, the wells were examined for the presence of single colonies by light microscopy. Colony growth, as well as the appearance of new colonies, was monitored daily (for at least 15 days). When the well confluency was 60-100%, cells were transferred to a 24 well plate with a final

volume of 1 ml selective medium. The mini-pools were progressively scaled-up from 24 to 12 to 6 well plate and then from a 6 well plate to T25 cm^2 to T75 cm^2 flask, when confluence of each stage was 65-90% to final volumes of 2, 4, 7, and 15 ml of selective medium. When the T75 cm^2 flask stage was reached, clone-derived cells were maintained for a maximum of 4 weeks by passaging into freshT75 cm2 flasks with a final volume of 20 ml of selective medium every 5-6 days. At this stage each mini-pools was frozen (at least 2 vials per clone) as described in Section 2.4.4 and a 1.5 ml sample of cell culture sample was lysed as described in Section 2.5.2.1.1 for analysis of β-galactosidase expression (Section 2.5.2.1).

2.5.1.6. Mini-pool adaptation to suspension culture

The cell density and viability of cell cultures in T75 cm^2 flask was calculated using the Countess® Automated Cell Counter (Section 2.3.3). When viable cell density was greater than 1.0x10^6 cells/ml and cell viability was more than 85%, 15 ml of cell culture was transferred to a 125 ml vented capped Erlenmeyer flask containing 15 ml of fresh warm selective medium and incubated at 37 ºC with 5 % CO_2 and 130 rpm. Viable cell density and cell viability were monitored daily and when a cell density and viability of above 1.3x10^6 cells/ml and 90%, respectively, was reached, cells were passaged and maintained as described in Section 2.4.2. Two vials of two consecutive early passages were frozen as described in Section 2.3.4 and 1,5 ml cell culture sample was lysed as described in Section 2.6.1

2.5.2. Initial characterisation of CHO-FRT mini-pools

To select the suspension-adapted CHO-FRT mini-pools to be used for the long-term stability study (Section 2.5.3) CHO-FRT mini-pools were initially characterised in terms recombinant protein expression (Section 2.5.2.1) and cell growth (2.5.2.2).

2.5.2.1. Determination of ß-galactosidase protein expression

The method used to determine the expression of ß-galactosidase was an enzymatic assay based on the ß-galactosidase activity assay for mammalian cells described by Eustice et al. 1991 that uses chlorophenolred-ß-D-galactopyranoside (CPRG) as the enzyme substrate.

The ß-galactosidase activity was normalized between mini-pools either to cell number, to mini-pools in the T75 cm^2 flask stage, or to total cell protein in mini-pool extracts for the mini-pools adapted to suspension culture.

2.5.2.1.1 Cell extract preparation

A 1.5 ml cell culture sample, for which viable cell density and viability were previously determined using the Countess® Automated Cell Counter (Section 2.4.3.), was transferred into a 1.5 ml microfuge tube and centrifuged in a microcentrifuge at 13000 x g for 1 min. The supernatant was discarded and the pellet was resuspended in 250 µl of Triton Lysis Buffer (TLB) (100 mM K_2HPO_4, 10 mM KH_2PO_4 and 0.2 % [v/v] Triton X-100, pH 7-8). The cell pellet was then frozen at -80 ºC until use. To maximise cell lysis, before use, the extracts were thawed at room temperature and then subjected to two rounds of snap-freeze/thawing in liquid nitrogen.

2.5.2.1.2. ß-galactosidase activity quantification

The β-galactosidase activity was measured kinetically using CPRG as substrate. The degradation of CPRG by the β-galactosidase leads to the formation of galactose and chlorophenol red. The formation of chlorophenol red was monitored by measuring the rate of increase of absorbance of the reaction solution at 575nm.

Before the assay the substrate was freshly prepared by mixing 200 µl of 5x Z buffer (350 mM Na_2HPO_4, 158 mM NaH_2PO_4, 50 mM KCl, 10 mM $MgSO_4$, pH 7-8), 1 µl 0.5 M CPRG, 1.4 µl β-mercaptoethanol and 800 µl ddH_2O for each ml of assay substrate needed. The assay substrate was then warmed at 37 ºC before use.

Cell extracts were thawed at room temperature after freeze/thawing as described in Section 2.6.1. A 10 µl sample of each cell extract was added to a well in a 96 well plate and this was mixed with 190 µl warmed assay substrate. The plate was immediately introduced into the BioTek Powerwave 340 plate reader that was previously warmed at 37 ºC and the measurement of the absorbance at 575 nm of each well was initiated immediately. The kinetic assay was completed by incubation of the plate at 37 ºC for 1 hour, into the plate reader, and the absorbance of each well was measured every minute at 575 nm, with shaking of the plate between each read. A negative control for ß-galactosidase activity of CHO-S cell line and an assay negative control using the TLB instead of cell lysate were performed in each plate.

The ß-galactosidase activity was calculated in units (U) (1 U = 1 µmol of substrate degraded per minute). At 575nm Chlorophenol Red formed has an extinction coefficient (ε_{575nm}) of 75000 $M^{-1}cm^{-1}$. The Beer-Lambert Law was used to calculate the amount of chlorophenol red formed during the enzymatic reaction and consequently the amount of substrate degraded. The amount of chlorophenol red was calculated as follows.

$$A = \varepsilon l c$$

A - Absrobance (at 575 nm)

ε - Extinction coefficient Chlorophenol red (ε_{575nm} = 75000 $M^{-1}cm^{-1}$)

l – laser path length (The path length for 200 µl of reaction volume in the 96 well plate is 0.6 cm)

c – Concentration of Chlorophenol Red in solution (M)

2.5.2.1.3. Cell protein quantification

Protein quantification was performed using the Bio-Rad Protein Assay. This protein assay based on the Bradford method and was performed in 96 well plates.

Bovine serum albumin (BSA, 100 µg/ml stock) was used to generate a range of serial dilutions of BSA standards The cell extract samples (Section 2.5.2.1.1) were diluted 20x in ddH$_2$O and 50 µl of diluted cell extract and BSA standards were transferred to 96 well plates. 200 µl Bio-Rad solution, previous prepared according to the manual, was added to each well and incubation was performed for 5 min. The absorbance at 570 nm was measured in a BioTek Powerwave 340 plate reader.

2.5.2.2 Cell growth characterisation

CHO-FRT cell growth were characterised by performing a batch culture analysis (Section 2.3.5). The cell growth rate (µ) during exponential phase (day 2-5) was calculated as follows.

$$\mu = (\ln[X] - \ln[Y])/T$$

µ - Cell growth rate (day^{-1})
X- Viable cell density at day 5 (cell/ml)
Y- Viable cell density at day 2 (cell/ml)
T - Time in culture (day)

2.5.3. Characterisation of the stability of CHO-FRT mini-pools

The stability of CHO-FRT mini-pools was characterised throughout a continuous culture the CHO-FRT mini-pools for a period of over 60 generation (Bayle 2012) (Section 2.5.3.1). Batch cultures analyses (Section 2.5.3.2) were performed before (early stage) and after (late stage) the period of culture for the comparison of CHO-FRT mini-pool characteristics. Day 4 cell samples from each batch culture study were collected for analysis and comparison of ß-galactosidase activity (Section 2.5.2.2), gene copy number (Section 2.5.3.3) and mRNA (Section 2.5.3.4).

2.5.3.1 Long-Term continuous culture of CHO-FRT mini-pools

CHO-FRT mini-pools were continuously cultured for 62 days using the Sonata Robot System (TAP Biosciences). Cells were sub-cultured every 4 days to an initial cell density of 0.2×10^6 cells/ml into a final volume of 30 ml of appropriate maintenance medium (CD CHO, 1X Glutamax, 0.25 mg/ml Zeocin) in 125 ml vented capped Erlenmeyer flasks at 37 °C with 5 % CO_2. At the time of each sub-culture, cell density, cell viability and cell diameter were determined and calculated using the Cedex Automated Cell Counter (Roche).

2.5.3.2 Early and late stage batch culture

Early and late batch culture were performed as described in Section 2.4.5, with cells initially seeded using the automated system Sonata Robot (Tap Biosystems) and cell density and cell viability were monitored daily using the Cedex Automated Cell Counter (Roche).

2.5.3.3 Determination of ß-galactosidase gene copy number

The ß-galacosidase gene copy number was determined by quantitative PCR (qPCR) (Section 2.5.3.3.3). Genomic DNA was isolated from CHO-FRT mini-pools (Section 2.5.3.3.1) and qPCR reaction standards and samples were prepared as described in Section 2.5.3.3.2.

2.5.3.3.1. Isolation of genomic DNA

CHO-FRT genomic DNA was isolated using TRIzol[®] Reagent (Life Tecnhologies), following the protocol suggested by the manufacturer. 5-10x 10^6 CHO cells from day 4 (exponential phase) of batch culture (Section 2.5.3.2) were harvested by centrifugation for 4 minutes at 200 x g at room temperature. The supernatant was removed and the pellet was resuspended in 1 ml TRIzol[®] Reagent by pipetting until fully homogeneous. Homogenized solutions were stored at -80°C until further processing and analysis.

When required homogenized extracts were thawed and incubated for 5 minutes at room temperature. 200 **μ**l of chloroform was added to each extract and tubes were vigorously shaken by hand for 15 seconds. After 3 minutes incubation at room temperature, mixtures were centrifuged for 15 minutes at 13000 x g at 4°C. The upper aqueous phase was transferred to a sterile 1.5 ml microcentrifuge tube for RNA extraction (Section 2.5.3.4.1) and the remaining lower interphase and organic phase containing cell protein and genomic DNA were used for genomic DNA isolation.

The genomic DNA was separated from the cell protein by precipitation. 300 μl of 100% ethanol were added to the lower interphase and organic phases and mixtures were mixed by inversion of tubes several times (at least 5x). After 3 minutes incubation at room temperature, the mixtures were centrifuged at 2000 x g for 5 minutes at room temperature to pellet the DNA. The supernatant containing the cell protein was discarded and the genomic DNA pellet was washed with 1 ml of sodium citrate/ethanol solution (0.1 M sodium citrate, 10% [v/v] ethanol, pH 8.5). The mixture was incubated for 30 minutes at room temperature during which time, it was mixed by inversion every 5 minutes. The mixture was centrifuged at 2000 x g for 5 minutes at 4ºC, the supernatant was discarded and the genomic DNA pellet was washed again with 1 ml of sodium citrate/ethanol solution by repeating the same process. 1.5 ml of 75% (v/v) ethanol were added to the final genomic DNA pellet and this was incubated for 15 minutes at room temperature. After incubation, the tubes were centrifuged at 2000 x g for 5 minutes at 4 ºC, the supernatant was discarded and the pellet was air-dried for 15 minutes at room temperature. Finally, the genomic DNA pellet was resuspended in 100 μl of autoclaved ddH$_2$O and kept at -20 ºC until needed.

2.5.3.3.2. Preparation of Standard curve and DNA Samples

The pFRT/LacZeo vector was diluted to a final concentration of 1,000,000 – 457 copies per 5µl reaction, in a background (10ng/µl) of genomic DNA isolated from non-transfected CHO-S cells following the process described in Section 2.5.3.3.1. Background DNA was used to ensure the efficiency of the PCR reaction was the same for all samples. Stock dilutions were made, aliquoted and stored at -80 ºC, for use for future assays.

Before the qPCR reaction (Section 2.5.3.3.3.) CHO-FRT genomic DNA was quantified using a NanoDrop^{TM} 1000 (Thermo Scientific) and diluted to a final concentration of 10 ng/µl in ddH$_2$O

2.5.3.3.3 Quantitative PCR (qPCR)

5 µl of each standard solution, CHO-S genomic DNA and diluted sample were added in triplicate wells to a MJ white 96 well plate. 15 µl of a master mix (10 µl 2x SensiMixTM SYBR No ROX, 2.5 µl 10 µM pFRT/LacZeo forward primer [Table 2.2], 2.5 µl 10 µM pFRT/lacZeo reverse primer [Table 2.2]) was added to each well and the plate was sealed with clear plastic caps. Before the PCR reaction, the plates and centrifuged for 1 minute at 900 x g. To monitor possible abnormalities between PCR reactions a standard curve was performed in every plate assayed.

PCR reactions were performed using a Chromo 4 thermocycler with the following settings: 10 minutes at 95°C, followed by 35 cycles of denaturation for 10 seconds at 95°C, annealing for 10 seconds at 56°C, elongation for 20 seconds at 72°C, denaturation (of any primer dimers) for 1 second at 76°C. A final elongation step was performed for 10 minutes at 72°C, and a melting curve was generated to check the quality of the amplified product.

Table 2.2. Primers used in qPCR for CHO-FRT gene copy-number analysis
Primers were designed using the primer3web software (http://primer3.ut.ee)

Target gene	Forward	Reverse
pFRT/lacZeo	5'-TGGAGGCTACCATGGAGAAG-3'	5'-GTGCTGCAAGGCGATTAAGT-3'
ß-actin	5'-ACTGCTCTGGCTCCTAGCAC-3'	5'-CATCGTACTCCTGCTTGCTG-3'

2.5.3.3.4 Analysis of qPCR results

Data from qPCR products (Section 2.5.3.3.3) was quantified using Opticon Monitor analysis software according to the manufacturer's instructions. The fluorescence threshold was manually set at 0.05 and the cycle at which samples reached this fluorescence was recorded (Ct value). The standard curve was generated by ploting log [gene copies] against Ct value. Gene copy number was calculated compared to the standard curve and was normalised against the expression of ß-actin in each sample. The melting curve was assessed to check the quality of the PCR product; a pure product produced a single peak at 80-90°C.

2.5.3.4. ß-galactosidase mRNA analysis

ß-galacosidase mRNA in CHO-FRT mini-pools was determined and analysed by qPCR reaction (Sections 2.5.3.3.3 and 2.5.3.3.4). Table 2.3 shows the primers used for the quantification of ß-galactosidase mRNA by qPCR. The RNA was extracted from CHO-FRT mini-pools and, before qPCR reaction, samples were treated and prepared as follows

Table 2.3. Primers used in qPCR for CHO-FRT mRNA analysis.
Primers were designed using the primer3web software (http://primer3.ut.ee)

Target mRNA	Forward	Reverse
pFRT/lacZeo	5'-CTGGCGTAATAGCGAAGAGG-3'	5'-GACAGTATCGGCCTCAGGAA-3'
ß-actin	5'-TGTGACGTTGACATCCGTAAA-3'	5'-GCAATGATCTTGATCTTCATG-3'

2.5.3.4.1. RNA extraction

The RNA was extracted from the upper aqueous (Section 2.5.3.3.1) phase containing RNA, described in Section 2.5.3.2.1. The RNA was precipitated by adding 500 µl isopropanol to the aqueous phase. This mixture was incubated for 10 minutes at room temperature, and then centrifuged for 10 minutes at 13000 x g at 4°C. The supernatant was aspirated and discarded and the RNA pellet was washed with 1 ml 75% (v/v) ethanol by vortex. After centrifugation for 5 minutes at 5000 x g at 4°C, the pellet was air-dried for 5-10 minutes. The pellet was dissolved in 30 µl of RNase-free

dd.H$_2$O (Bioline) and incubated for 10 minutes at 55-60°C to improve solubility. The resuspended RNA was stored at -80°C until required.

2.5.3.4.2. DNase treatment of RNA

RNA was treated with DNase I to remove any possible genomic DNA in the sample. RNA content of samples (Section 2.5.3.2.1) was quantified using a NanoDrop® ND-1000 UV-Vis Spectrophotometer. 1 µg RNA, 1µl 10x reaction buffer (Kwaksa et al. 2005), 1 µl DNase I (Sigma-Aldrich) and RNase-free ddH2O (Bioline) up to a total mixture volume of 10 µl, were mixed in a 1.5 ml microcentrifuge tube. The mixture was incubated for 15 minutes at room temperature after which, 1 µl stop solution (Sigma-Aldrich) was added to inactivate DNase I. Samples were heated for 10 minutes at 70°C to denature the DNase I and unfold the RNA. Samples were kept at 4°C and immediately used for the cDNA synthesis (Section 2.4.3.4.3).

2.5.3.4.3. cDNA synthesis

cDNA was generated using a cDNA synthesis kit (Bioline). Samples of quantified DNase-treated RNA (Section 2.5.3.4.2) were mixed with 1 µl oligo (dT) 18 primer and 1 µl 10 mM dNTP. The mixtures were incubated for 10 minutes at 65°C followed by 2 minutes incubation on ice (whilst the subsequent reaction mix was prepared). 4 µl 5x RT buffer, 1 µl RNAse inhibitor, 0.25 µl reverse transcriptase and 4.75 µl DEPC-treated water were added to each sample and the resulting mixtures were incubated for 60 minutes at 42°C. The reaction was terminated by incubation of the mixture for 15 minutes at 70°C, followed by a incubation on ice for at least 5 minutes. The resulting cDNA was aliquoted and stored at -20°C until used in further analytical procedures.

2.5.3.4.4 Preparation of Sample and 'Check' sample

The cDNA samples obtained (Section 2.5.3.4.3) were quantified using a NanoDrop® ND-1000 UV-Vis Spectrophotometer and diluted in ddH$_2$O to a final concentration of 10 ng/µl. One sample was dedicated as the 'check' sample. This sample was run on all plates to normalise the total cDNA content in each well. The 'check' sample was diluted in ddH$_2$O to final concentrations of 20, 10 and 5ng/µl, aliquoted and stored at -80 ºC until been used.

2.6. Characterisation of activity and stability of SV40 and CMV promoters

SV40 and CMV promoter activities were charactersised by measurement of expression of fluorescent proteins from copies of their genes that had been integrated into the genome of CHO-FRT mini-pools in a site-directed manner through use of pS/FRT and pcDNA5/FRT or pS/FRT/MS2 and pcDNA5/FRT/MS2 vectors (Section 2.3). The stability of these promoters was characterised by evaluation of the expression of these fluorescence proteins at stages during long-term culture.

2.6.1. Generation and characterisation of CHO-SV40-GFP and CHO-CMV-GFP polyclonal cell lines

CHO-SV40-GFP and CHO-CMV-GFP were generated by site-directed integration of pS/FRT-GFP and pcDNA5/FRT-GFP vectors (Section 2.3) into CHO-FRT mini-pools (Section 2.6.1.1). Transfectants were selected in a polyclonal manner (Section 2.6.1.2) and GFP expression was measured by Flow Cytometery (Section 2.6.1.3). The stability of GFP expression in cell lines was characterised by monitoring GFP expression during continuous culture of the cell lines in long-term culture (Section 2.6.1.4).

2.6.1.1. Site-directed integration of pS/FRT-GFP and pcDNA5/FRT-GFP vectors into CHO-FRT mini-pools

Site-directed integration of pS/FRT-GFP and pcDNA5/FRT-GFP vectors into CHO-FRT mini-pools was performed by co-transfecting each of these vectors with pOG44 (Section 2.1.4), using FreeStyle™ MAX Reagent (Life Technologies). Two days prior to transfection, CHO-FRT cells were seeded at 0.2×10^6 cell/ml (Section 2.4.2) in CD CHO medium supplemented with 1X GlutaMax. On the day of the transfection, viable cell density and cell viability were measured (Section 2.4.3) and, if cell viability was greater than 95%, 10.0×10^6 cells were added 125 ml vented capped Erlenmeyer flasks in a total volume of 30 ml of CD CHO medium supplemented with 1X GlutaMax.

15 µg pOG44:(pS/FRT-GFP or pcDNA5/FRT-GFP) vector DNA in a 9:1 (w/w) ratio were added to a 1.5 ml tube containing 1.2 ml CD CHO supplemented with 1X

GultaMax. 18 µl FreeStyle™ MAX Reagent were added to the mixture, the tube was gently inverted twice for mixing and incubated for 10-20 minutes (no longer than 20 minutes) at room temperature. After the incubation period, the mixture was slowly added into the CHO-FRT cell suspension whilst gently swirling the flask and cells were incubated for 48 hours at 37 ºC with 5 % CO_2 and 130 rpm.

2.6.1.2. Selection of CHO-SV40-GFP and CHO-CMV-GFP polyclonal cell lines

After the 48 hour incubation period, 0.5 ml and 1.0 ml of transfected cells were added to a T75 cm^2 flask containing 19.5 ml and 19.0 ml (respectively) of Hygromycin selective medium (CD CHO, 1X GlutaMax, 0.5 mg/ml Hygromycin B). Cells were incubated at 37 ºC with 5 % CO_2 for a period of 4 weeks for selection of site-directed integrated cell lines. Typically, the flasks to which 0.5 ml of transfectants had been added were the ones used for adaptation to suspension culture. The flasks seeded with 1.0ml of transfectants were processed as back-up flasks.

The viable cell density of the selected cells was calculated (Section 2.4.3) after the selection period, and when greater than $1.0x10^6$ cells/ml with a cell viability greater than 80%, the cells were adapted to suspension culture. For the suspension adaption process, 10 ml selected cell stocks were cultured in a 125 ml vented capped Erlenmeyer flask containing 20 ml Hygromycin selective medium at 37 ºC with 5 % CO_2 and 130 rpm. The cells were cultured for 2 passages (Section 2.4.2), after which 10 vials were frozen as stocks (Section 2.4.4). After the second passage, cells were maintained in Hygromycin maintenance medium (CD CHO, 1X GlutaMax, 0.25 mg/ml Hygromycin B) (Section 2.4.2).

2.6.1.3. Measurment of GFP fluorescence by Flow Cytometry

CHO-SV40-GFP and CHO-CMV-GFP cells were harvested and $1.0x10^6$ cells were resuspended in 400 µl PBS. The GFP fluorescence intensity was measured using a CYAN ADP flow cytometer, using a 488nm excitation laser. The voltage applied to the photomultiplier tube was adjusted according to the manufacturers instructions, to ensure the histogram plots were within range. Fluorescence data were acquired using a 530/30nm bandpass filter and analysed using Summit V4.3 software. Live cells were gated using a forward scatter versus side scatter log plot and single cells were gated using a forward scatter versus pulse width plot. GFP fluorescence intensity of each cell line was calculated by measuring the median of level of

fluorescence (530/30 Log) in the flow cytometry histograms plots. Analysis of the flow cytometry histograms was performed using the FlowJo X 10.0.7r2 software. The auto-fluorescence of the CHO-FRT mini-pool was measured in every experiment as a control for non-expressing cells.

2.6.1.4. Long-Term continuous culture of CHO-SV40-GFP and CHO-CMV-GFP polyclonal cell lines

CHO-SV40-GFP and CHO-CMV-GFP cell lines were continuously cultured for 60 days as described in Section 2.4.2. Cells were sub-cultured with a 3/4 day regime at an initially cell density of 0.2×10^6 cells/ml into a final volume of 25 ml of hygromycin maintenance media (CD CHO, 1X Glutamax, 0.25 mg/ml hygromycin) in 125 ml vented capped Erlenmeyer flasks at 37 ºC with 5 % CO_2 at 130 rpm. At each sub-culture, cell density and cell viability a calculated using the Countess® Automated Cell Counter (Section 2.4.3).

2.6.2. Generation and characterisation of CHO-SV40-YFP and CHO-CMV-YFP polyclonal cell lines

CHO-SV40-YFP and CHO-CMV-YFP were generated by site-directed integration of pS/FRT/MS2-YFP and pcDNA5/FRT/MM2-YFP vectors (Section 2.3) into CHO-FRT mini-pools (Section 2.6.2.1). Transfectants were selected on a polyclonal basis (Section 2.6.2.2) and YFP expression was measured by Flow Cytometry (Section 2.6.2.3). A "Top 10%" high YFP expressing population was selected by Fluorescence-activated cell sorting (FACS) (Section 2.6.2.2) to generate the top 10% cell lines. The stability of YFP expression in the polyclonal and top 10% cell lines was characterized by monitoring YFP expression during continuous culture of cell lines in a long-term culture study (Section 2.6.2.4).

2.6.2.1. Site-directed integration of pS/FRT/MS2-YFP and pcDNA5/FRT/MS2-YFP vectors into CHO-FRT mini-pools

Site-directed integration of pS/FRT/MS2-YFP and pcDNA5/FRT/MS2-YFP vectors into CHO-FRT mini-pools was performed by co-transfecting each of these vectors with pOG44 using FreeStyle™ MAX Reagent (Life Technologies) (Section 2.6.1.1)

2.6.2.2. Selection of CHO-SV40-YFP and CHO-CMV-YFP polyclonal and top 10% YFP expressing cell lines

CHO-SV40-YFP and CHO-CMV-YFP polyclonal cell lines were selected using methods described in Section 2.6.1.2. The "top 10% high" expressing population was selected by Fluorescence-activated cell sorting (FACS) using the BD FACSAria™ Fusion.

A 5 ml sample of CHO-SV40-YFP and CHO-CMV-YFP polyclonal cells at day 4 of batch culture was collected and cell density and viability were calculated (Section 2.4.3). Cells were diluted to a viable cell density of $5-10x10^6$ cell/ml in 4 ml hygromycin maintenance medium (CDH CHO, 1X GlutaMax, 0.25 mg/ml Higromycin) supplemented with 10mM of HEPES. Cells were filtered to remove possible clumps using a 50 µm BD™ Medimachine Filcon, Sterile, Cup-Type and transfer to a sterile 5 ml collection tube.

The sort of the cells with the top 10% of YFP fluorescence was measured using a 488nm excitation laser. One sample was used to adjust the voltage applied to the photomultiplier tube, according to the manufacturer's instructions, and live cells were gated using a forward scatter versus side scatter log plot and single cells were gated using a forward scatter versus pulse width plot. The auto-fluorescence of the cells was measured using a sample of parental CHO-FRT mini-pool. YFP fluorescence histograms were acquired and analysed using the BD FACSDiva™ software and the Top 10% high expressing population was gated. $2.5x10^5$ cells were sorted and collected into a 5 ml collecting tube containing 0.5 ml of hygromycin maintenance medium.

The collected cells were diluted in a total volume of 7 ml hygromycin maintenance medium and placed in a T25 cm^2 flask. The cells were incubated for seven days at 37 ºC with 5 % CO_2 at 130 rpm with the T25 cm^2 flask placed vertically to allow a horizontal wave formation during the incubation period. After this incubation period, the cells were diluted in a total volume of 25 ml hygromycin maintenance medium in a 125 ml vented capped Erlenmeyer flask. The cells were cultured at 37 ºC with 5 % CO_2 and 130 rpm, passaged twice (Section 2.4.2), after which 10 vials were frozen for stock (Section 2.4.4).

2.6.2.3. Measurement of YFP fluorescence by Flow cytometry

A 1 ml sample of cell culture was collected, cell density and viability was calculated (Section 2.4.3). YFP fluorescence intensity was measured using the BD Accuri™ C6 flow cytometer according to the manufacturer's instructions. YFP was excited using a 488nm excitation laser and fluorescence data was acquired using a 530/30nm bandpass filter and analysed using the BD Accuri™ C6 software. Live cells were gated using a forward scatter versus side scatter log plot and single cells were gated using a forward scatter versus pulse width plot. YFP fluorescence intensity of each cell line was calculated by measurement of the median of fluorescence of the YFP expression population of the flow cytometry histograms. The analysis of the flow cytometry histograms was performed using the FlowJo X 10.0.7r2 software. The auto-fluorescence of the parental CHO-FRT mini-pool was measured in every experiment as a control for non-expressing cells.

2.6.2.4. Long-Term continuous culture of CHO-SV40-YFP and CHO-CMV-YFP polyclonal cell lines

CHO-SV40-YFP and CHO-CMV-YFP polyclonal and top 10% cell lines were continuously culture for 32 as described in Section 2.6.1.4.2. YFP expression of these cell lines was monitored by measurement of YFP fluorescence (Section 2.6.2.3) at every passage.

2.7. Characterisation of transcriptional activity of SV40 and CMV promoters

The transcriptional activity of SV40 and CMV promoters was measured in CHO-SV40-mStrawberry and CHO-CMV-mStrawberry cell lines by live fluorescent microscopy cell imaging based on the method described by Yunger *et. al.* (2013).

2.7.1. Generation and characterisation of CHO-SV40-mStrawberry and CHO-CMV-mStrawberry polyclonal cell lines

2.7.1.1. Generation and selection of CHO-SV40-mStrawberry and CHO-CMV-mStrawberry polyclonal cell lines

CHO-SV40-mStrawberry and CHO-CMV-mStrawberry polyclonal cell lines were generated by site-directed integration of pS/FRT/MS2-mStrawberry and pcDNA5/FRT/MS2-mStrawberry vectors into CHO-FRT mini-pools as described in Section 2.6.1.1.

2.7.1.2. Measurement of mStrawberry fluorescence by Flow Cytometry

mStrawberry fluorescence was measured and analysed using the BD Accuri™ C6 flow cytometer and the FlowJo X 10.0.7r2 software as described in Section 2.6.2.3. mStrawberry was excited using a 488nm excitation laser and fluorescence data was acquired using a 585/40 nm bandpass filter.

2.7.2. Preparation of CHO-SV40-mStrawberry and CHO-CMV-mStrawberry cells for fluorescent microscopy

Prior to fluorescent microscopy studies, pMS2-GFP vector was transiently transfected into CHO-SV40-mStrawberry and CHO-CMV-mStrawberry cells using FreeStyle™ MAX Reagent (Section 2.7.2.1).

2.7.2.1. Transient transfection of pMS2-GFP vector

CHO-SV40-mStrawberry and CHO-CMV-mStrawberry cell lines were seeded at 0.5×10^6 cell/ml in 30 ml of CD CHO medium supplemented with 1X GlutaMax and incubated for 24 hours at 37 ºC with 5 % CO_2 and 130 rpm. Cell Density and cell viability were calculated (Section 2.4.3) and cells were diluted to 1.0×10^6 cell/ml in 30 ml of CD CHO medium supplemented with 1X GlutaMax in a 125 ml vented capped Erlenmeyer flasks.

18µg of pMS2-GFP DNA vector and 18 µl of FreeStyle™ MAX Reagent were added to two separate 1.5 ml tubes containing 0.6 ml CD CHO medium supplemented with 1X GlutaMax and incubated at room temperature for 5 minutes. The diluted DNA vector was added to the tube containing the FreeStyle™ MAX Reagent mixture and the final mixture was incubated for 30 minutes at room temperature. The mixture was added dropwise whilst gently swirling cell suspensions in the 125 ml vented capped

Erlenmeyer flasks previously prepared. The cells were incubated at 37 ºC with 5 % CO_2 and 130 rpm for 24 hours.

2.7.3. Live Cell Imaging

After the 24 hours incubation, 1ml of the transiently transfected CHO-SV40-mStrawberry and CHO-CMV-mStrawberry cells were diluted in 9 ml of fresh CD CHO medium supplemented with 1X GlutaMax. 2 ml of diluted cells were added to a 35 mm glass bottom cell culture dish (Greiner) pre-coated with poly-D-lysine and incubated for 24 hours at 37 ºC with 5 % CO_2 prior to live cell microscopy studies.

Images were collected on a Leica TCS SP5 AOBS inverted confocal using a (100x/ 1.49 Apo TIRF) objective (and 3x confocal zoom). The confocal settings were as follows, pinhole 1 airy unit, scan speed 1000Hz unidirectional, format 1024 x 1024. Images were collected using the following detection mirror settings; FITC 494-530nm and Texas red 602-665nm using the 488nm (20%) and 594nm (100%) laser lines respectively. When acquiring 3D optical stacks the confocal software was used to determine the optimal number of Z sections. Only the maximum intensity projections of these 3D stacks are shown in the results.

2.8 Statistical analysis

All data presented is represented as a mean ± range when biological replicates (n) n=2 or mean ± standard deviation (SD) when n> 2, unless otherwise.

Range =   Higher value – lower value

Standard Deviation (SD) = $(\sqrt{[\Sigma\{x-m\}2]/\{n-1\}})$

x: observed value

m: mean of n observations

n-1: degrees of freedom

The correlation coefficient (r value) was calculated in the plots using Microsoft Excel.

Chapter 3. Generation and stability characterisation of CHO-FRT mini-pools

The main aim of this project is to describe transcriptional activity and stability of recombinant promoters in CHO cells. To fully achieve this, it is important to study the impact of promoter type on the stability of recombinant gene expression when in the same genomic environment. This first Chapter of results describes the work undertaken to generate the CHO cell lines that have been used to address the aim of the project.

The first aim of this Chapter is to generate a set of CHO cell lines to be used as host cells in the work described in Chapter 4 and 5. It is intended that these cell lines shall enable the integration of recombinant promoters and/or genes towards specific genomic locations in a site-directed manner. This Chapter describes the generation of the cell lines and the validations performed at each step of cell line generation in order to monitor the appropriate properties of each cell line (Section 3.2).

Due to the possible impact that this project could have in the bioprocessing industry, the methodology used on the generation of the cell lines was intended to be as close as possible to the industry standards within the academic resources. For that reason, due to the nature of the industry protocols for the achievement of high expression cell lines, the experimental protocol described in this chapter can introduce caveats that can impact the overall propose of the project. For instance, even though appropriate measures to reduce the high-copy of recombinant genes integrated within the host genome were taken it is still highly possible the formation of multiple copy number cells lines. Furthermore, as will be discussed in the Section 3.2, the single cell selection method used does not fully guarantees that all the cell lines selected is originated from a single cell. For that reason, it is possible that some cells lines generated can have a low degree of initially heterogeneity and so cannot be considered clonally-derived cells lines but instead, mini-pools of cells with low degree of heterogeneity. Nevertheless, due to the low degree of initially heterogeneity of each mini-pool, the mini-pools generated using this selection process have different and distinctive phenotypes. Because of these reason, it is essential to characterise these mini-pools with a special focus on the long-term stability of the mini-pool, before using them as hosts cells in Chapters 4 and 5.

Figure 3.1 illustrates the steps and rational used to generate and characterise CHO-FRT mini-pools. Due to the large number of mini-pools selected, at the end of each step of the generation and characterisation process, the data was analysed and a smaller number of mini-pools was rationally progressed to the following step. This rational selection based on the data of each step, allowed to minimise the number of

mini-pools characterised without considerably affecting the phenotype variation of cells lines originally obtained.

The genomic environment has a strong impact on the promoter status and activity and stability of promoter-driven transcription (Section 1.7). The majority of studies that report on expression stability of recombinant mammalian cell lines engineered to produce therapeutics proteins, utilises cells generated by random integration of recombinant vectors into the genome of the host cell (Bailey et al. 2012; Ho et al. 2015; Kim et al. 2001; Yoshikawa et al. 2000b). Due to the different genomic environments resulting from random integration, it is inappropriate to directly compare the expression and stability of expression of different recombinant promoters, unless the nature and consequences of the specific genomic can be predicted and assigned. For a truly comparison, ideally, a set of CHO mini-pools with different phenotypes but that allow the site-directed integration of recombinant promoters to pre-determined genomic locus should be used. With such panel of cell, the potential for direct comparison of expression and stability of different recombinant promoters becomes possible and would open the potential to interrogate the mechanisms by which different genomic environments characteristics interacts with each promoter. Unfortunately, the mini-pools generate this Chapter do not fully fulfilling these ideal characteristics. However, due the different phenotyoes observed and the fact the site-directed integration can be used, allow to suggest that these mini-pools still represent a step forward for this type of comparative studies, when compared to traditional methods.

To keep track of each mini-pool at different stages of development, throughout this Chapter each CHO-FRT mini-pool is named by the number (n) that was given during the mini-pool selection process (CHO-FRT n).

Figure 3.1. Overview of the strategy and rational used for generation and characterisation of CHO-FRT mini-pools.

This Figure illustrates the sequence of steps used to generate the CHO-FRT mini-pools and the characteristics that were assessed at each step of this process.
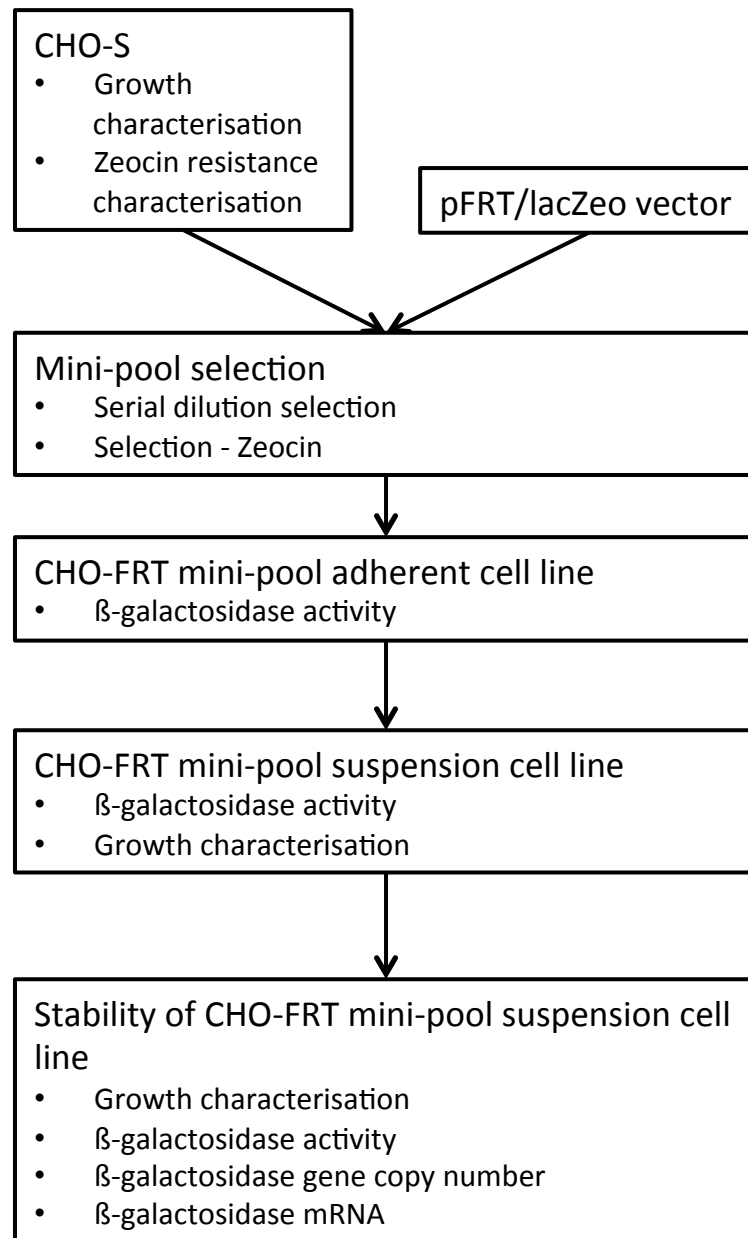
3.1. CHO-S cell line characterisation

This Section characterises cell growth (Section 3.1.1) and Zeocin resistance (Section 3.1.2) of the parental CHO-S cell line (CHO-S), for comparison with CHO-FRT mini-pools. Also, as Zeocin is the antibiotic used for the selection of cells that have stably integrated the pFRT/lacZeo vector, it is important to characterise Zeocin resistance of this cell line to determine the antibiotic concentration for mini-pool selection process and limit the selection of false positive cells.

3.1.1. CHO-S cell growth characterisation

Viable cell density and cell viability were measured during batch culture (Figure 3.2). Batch culture revealed four phases - an initial lag phase (day 0-2), an exponential growth phase (day 2-5), a stationary/slow decrease phase (day 5-7) and a decline phase (day 7-8).

CHO-S had a log growth rate of 0.25 day^{-1} during exponential phase, reaching a maximal viable cell density of 9.0×10^6 cell/ml at day 5. During the stationary phase a slight decrease of viable cell density was observed but cell viability was maintained above 95%. In the decline phase, a dramatic drop of the viable cell density from 7.0×10^6 cell/ml to a 0.3×10^6 cell/ml was observed over a period of a single day, after which the culture was stopped. During the culture period, the percentage of viable cells in culture was maintained above 95% throughout the lag, exponential and stationary phases of culture.

The results obtained were similar to previous data from the Dickson laboratory (Page 2012). Page observed that CHO-S cells in batch culture had a duration of culture of 10 days and achieved a maximum cell density of approximately 9.0×10^6 cell/ml at day 7. The results shown in Figure 3.2 are also very similar to data reported by Benton et al (2002) and Kleman et al (2008) where cells were cultured in similar conditions (Benton et al. 2002, Kleman et al.2008). Benton et al (2002) characterised CHO-S growth in an early stage cell line and reported a maximum viable cell density of $9-11 \times 10^6$ cell/ml during an 8 day batch culture (Benton et al. 2002). Similar data were reported by Kleman et al (2008) who characterised CHO-S cell growth across a 6 week period of prolonged culture (Kleman et al. 2008). Kleman and colleagues only showed viable cell density during the initial lag and exponential phases but in his study reported a maximum cell density of $9-11 \times 10^6$ cells/ml (Kleman et al. 2008).

Compared with results obtained with other CHO host cell lines, under the conditions of culture in the current project, CHO-S had a shorter overall period of culture than CHO-K1 (9-10 days) and CHO-DG44 (18 days) (SIgma-Aldrich). However, during culture CHO-S achieved considerable greater maximum cell densities compared to maximum densities of 6-7x10^6 cell/ml and 2-3x10^6 cell/ml for CHO-K1 and CHO-DG44, respectively (Sigma-Aldrich).

3.1.2. CHO-S resistance to Zeocin

CHO-S cell proliferation in the presence of Zeocin was monitored by light microscopy (Figure 3.3). The results revealed that 500 µg/ml was the minimal concentration of Zeocin in which no cell growth was observed after 10 days of culture (Figure 3.3-D). This result is in accordance to previous reports (400-500µg/ml) of the use of Zeocin for selection in COS, HEK and CHO cell lines (Bennett et al. 1998, Delacote et al. 2007).

Figure 3.2. CHO-S growth characterisation in batch culture.
The growth of CHO-S cell line was characterised by daily monitoring the viable cell density (━■━) and cell viability (━◢ ●) during a batch culture as described in Section 2.4.5. Viable cell density and cell viability were measured daily and calculated as described in Section 2.4.3. Values are means ± Range (n=2).

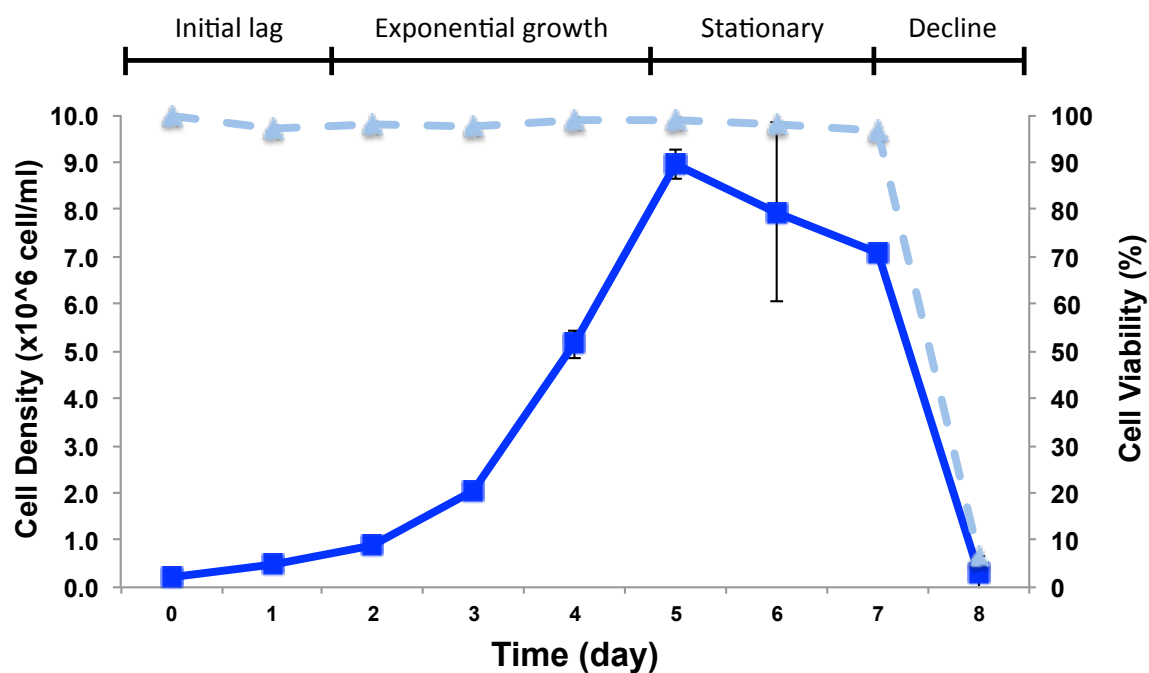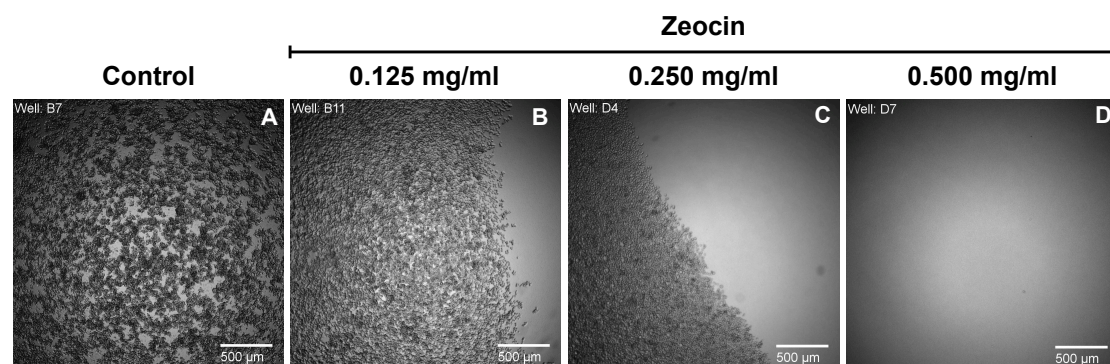**Figure 3.3. CHO-S Zeocin resistance characterisation.**

The Zeocin resistance assay was performed as described in Section 2.4.6. CHO-S cells were seeded at initial cell density of 5.0×10^3 cell/ml in 96 well plates. CHO-S growth was monitored over a maximal period of 10 days, in the presence of a range of concentrations of Zeocin (A – Contro(-); B – 0.125 mg/ml; C – 0.250 mg/ml; D – 0.500 mg/ml).

3.2. Generation and initial characterisation of CHO-FRT mini-pools

This Section describes the generation of the CHO-FRT mini-pools and the characterisation of these mini-pools at each step of cell line development. Firstly, this section 3.2.1 explains the mini-pools selection and mini-pool amplification processes. Then, the ß-galactosidase expression and growth properties of CHO-FRT mini-pools are characterised in Sections 3.2.2 and 3.2.3 and related in Section 3.2.4 to the identification of different phonotypical patterns.

3.2.1. Mini-pool selection and amplification

The CHO-S cells were transfected by electroporation with pFRT/lacZeo (Section 2.5.1.4). Transfectants were subjected to a serial limiting dilution cloning process (0.25, 1.0 and 5.0×10^3 cells/ml) following the methods described in Section 2.5.1.5. A total of 15 positive wells (wells containing small clusters of cells) were identified by light-microscopy after 15 days of selection, with the majority of them, 14, from the plates seeded at 5.0×10^3 cell/ml. In order to obtain a higher number of colonies, a second transfection was performed. Using the results obtained for the first transfection and mini-pool selection, in this second experiment, transfectants were only seeded at an initial cell density of 5.0×10^3 cell/ml. A total of 34 positive wells were observed after 10 days of selection and 73 additional wells were identified after 15 days of selection leading to a total of 107 positive wells identified from this second transfection and selection process.

Although serial limiting dilution method was used, the cells cannot be considered clonally-derived cell lines. The definition of clonal-derived cell line has been a long-time discussed issue. Limiting dilution methods relay on statistical probability to claim achievement of monoclonality and this often is accomplished by the performance of multiple repeat rounds of this process (Evans et al 2015). For monoclonality to be achieved by only one round of limiting dilution, it is necessary visual confirmation by light microscopy imaging to confirm that at the time of plating, only a single cell was presented within the well (Evans et al 2015). Since the limiting dilution method described in Section 2.5.1.5 only shows one round of limiting and no visual confirmation of the presence of a single cell at the time of plating, it is not possible to guarantee that the cell lines generated using this method are monoclonal cell lines. However, due to the nature of the limiting dilution method used that

confers a low heterogeneity of the cells that generate each cell line, we can describe each positive well as mini-pools (Wolfel et al 2013).

The cells from each mini-pool were sequentially passaged from 96 well plates to a 24, 12, 6 well plates, T25 and T75 cm^2 flasks in order to increase cell biomass. A total of 82 mini-pools were successfully expanded to the T75 cm^2 flask stage and used for the suspension adaption process.

The 40 mini-pools that did not successfully expand to the T75cm2 flask stage stopped growing either in 6 well plate or T25 cm^2 flask stage. The majority of these mini-pools that failed to thrive originated from the second transfection and they were only identified after 15 days of selection. The fact that they were only observed at late stages of selection could indicate a slow growing phenotype, and may relate to their inability to show continued selection and growth during subsequent expansion. The slow growing phenotype of these mini-pools could be due to an incomplete adaption of the mini-pools to the presence of Zeocin in the medium. As the lacZ-Zeocin gene is inserted randomly into the cell genome, if was integrated to a locus with low transcriptional activity or in a locus that undergoes silencing, the cells may reveal a low degree of resistance to Zeocin and, consequently, a slow growing phenotype in the presence of this selection marker. Another possible explanation is the loss of resistance to Zeocin due gene silencing or to the loss of the lacZ-Zeocin gene during the amplification process. As discussed in Section 1.5.3, CHO cells are known to exhibit a natural genomic instability that in some cases can led to gene silencing and gene disruption.

All the 82 CHO-FRT mini-pools in T75 cm^2 flask stage were used for adaption to suspension cultures (Section 2.5.1.6). 22 CHO-FRT mini-pools successfully adapted to suspension culture with the remaining 60 mini-pools failing to grow in suspension after three attempts of this adaption process.

3.2.2 ß-galactosidase activity of CHO-FRT mini-pools

The ß-galactosidase activity of CHO-FRT mini-pools was analysed with a ß-galactosidase kinetic assay that uses CPRG as the enzyme substrate (Section 2.5). Other substrates, such as the commonly used ONPG, have been also been used to assay ß-galactosidase activity in mammalian cells. However, CPRG has been reported to be more sensitive than ONPG, for detection of ß-galacatosidase activity due to its higher affinity for this enzyme (Eustice et al. 1991). Due to the fact that most of the reports use ONPG as substrate, the results described in Sections 3.3.2

and 3.3.3 cannot be directly compared to previous reports of the use of ß galactoasidase in CHO cells.

Two endogenous ß-galactosidases have been reported in mammalian cells, a lysosomal ß-galactosidase and a senescence-associated ß-galactosidase (SA-b-galactosidase), with optimal pH's of 4 and 6.0, respectively, and both are reported to be inactive at pH's greater than 7 (Kurz et al. 2000). The recombinant ß-galactosidase to be expressed in CHO-FRT cells is encoded by the gene lacZ from *E. coli* with an optimal pH of 7.5 (Watson et al. 2008). The CPRG assay is performed at pH 7-8, so in this assay, the ß-galactosidase activity measured should arise from the activity of the recombinant ß-galactosidase.

Figure 3.4 shows the increase in absorbance at 550 nm, measured on the ß-galactosidase activity reaction of the CHO-S host cell line (Figure 3.4 – A) and of three different CHO-FRT mini-pools that can exemplify a low, medium and high rate of ß-galactosidase activity (Figure 3.4 – B, C and D). Distinct ß-galactosidase activities are easily identified between mini-pools based on rate at which absorbance at 550nm increased during the reaction period. Also, no (or very little) endogenous ß-galactosidase activity was observed when this assay was performed on the non-transfected CHO-S cell line.

Figure 3.4. ß-galactosidase assay exploratory assessment.

The degradation of CPRG by ß-galactosidase leads to the formation of chlorophenol red that can be monitored by measuring the absorbance of the enzymatic reaction at 550nm. A 1.5 ml cell culture sample was harvested (Section 2.5.1) and the assay was performed as described Section 2.5.3. (A – Host cell line [CHO-S]; B - Low ß-galactosidase activity [CHO-FRT 112]; C – Medium ß-galactosidase activity [CHO-FRT 113]; D – High ß-galactosidase activity [CHO-FRT 46]). The ß-galactosidase activity was calculated by using the slope of each graph (bold number in line equation) to calculate the rate of degradation of substrate (Section 2.5.3). ß-galactosidase is presented in unites (U) (1 U = 1 µg of substrate degraded per minute).
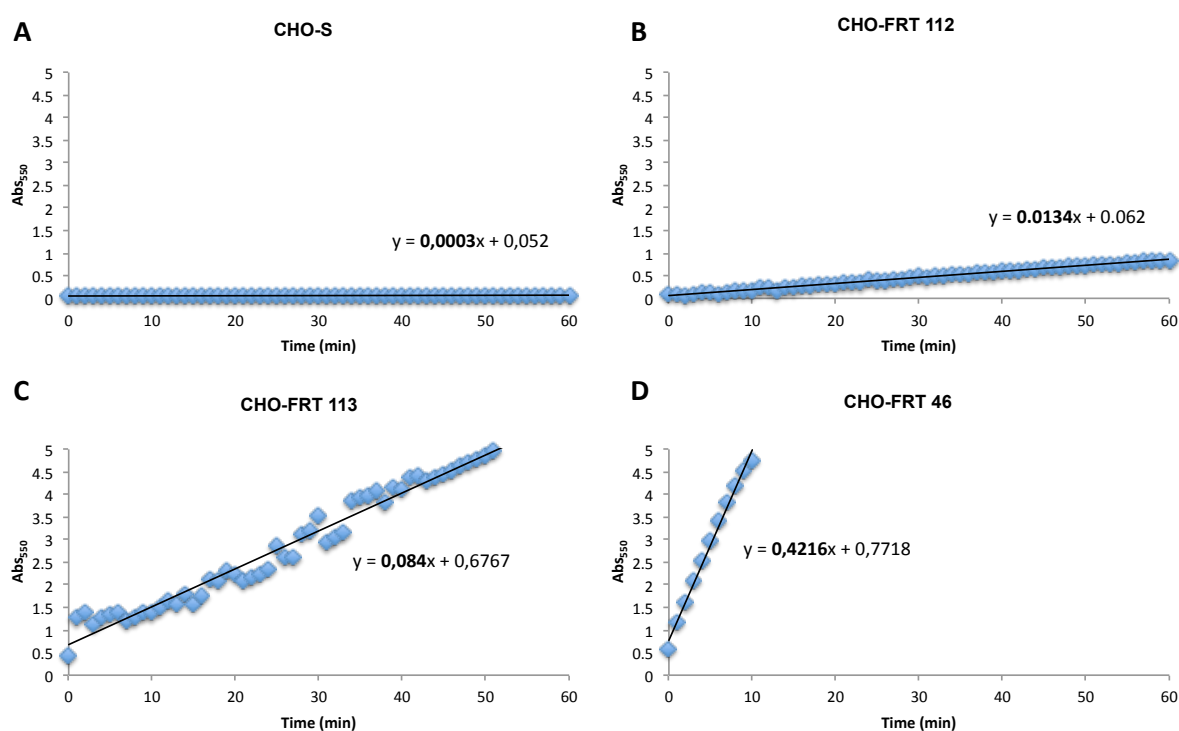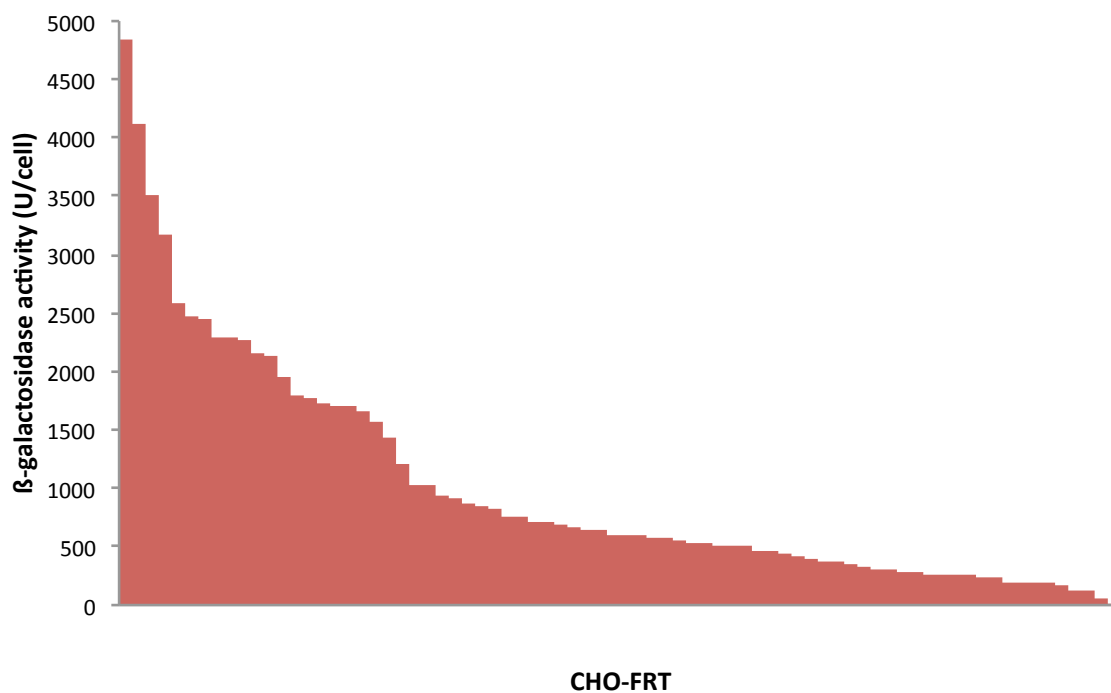
3.2.2.1. ß-galactosidase activity of CHO-FRT mini-pools at T75 cm2 flask stage

Figure 3.5 ranks ß-galactosidase expression per 10^6 cells for the adherent CHO-FRT mini-pools. The normalisation of ß-galactosidase activity per amount of protein in the cell extract may be a more appropriate method to normalise the ß-galactosidase activity analysed in each extract as it allows a more accurate comparison of the ß-galactosidase expression between mini-pools. However, due to the high number of cell extracts handled in this initial screen as an intermediate step of the mini-pool generation process, the values were normalised per cell. Nevertheless, as all cell extractions were performed in a similar manner, these normalisation towards cell number will have limited affect on relative ß-galactosidase expression of mini-pools and it remains a valid method for the direct comparison of ß-galactosidase expression between the mini-pools.

All CHO-FRT mini-pools expressed the ß-galactosidase, yet the amounts produced vary between mini-pools. Most of mini-pools expressed relative low amounts of ß-galactosidase, when compared to the highest expresser mini-pool. Approximately 75% of the mini-pools produced less then half of the amount of ß-galactosidase produced by the highest expresser (mini-pool 46) and approximately 60% of the mini-pools expressed less then a quarter of ß-galactosidase of mini-pool 46. Further discussion and analysis of the ß-galactosidase activity in CHO-FRT mini-pools will be made in Section 3.2.3.2 on the analysis of the ß-galactosidase activity of at the stage of suspension adaption of CHO-FRT.

Figure 3.5. ß-galactosidase analysis of CHO-FRT mini-pools in T75 cm2 flask stage
ß-galactosidase analysis was performed as described in Section 2.5. A T75ml cell culture sample of each mini-pool was harvested as described in Section 2.5.1 and the ß-galactosidase assay was performed as described in Section 2.5.3. The ß-galactosidase activity results for each mini-pool was calculated as unite (U) per cell.

3.2.2.2. ß-galactosidase activity of CHO-FRT mini-pools at suspension-adapted stage.

Figure 3.6 shows the activity of ß-galactosidase per µg cell of protein of the CHO-FRT mini-pools that were adapted into suspension culture. Similar to the profile of data observed for the same mini-pools at T75cm2 flask stage, 70% of the mini-pools produced less then half of the amount of ß-galactosidase of the highest expresser (CHO-FRT 54) and approximately 40% of the mini-pools produce less then a quarter of the amount of ß-galactosidase produced by the CHO-FRT 54.

Similar distribution patterns for the distribution of cell productivity were previously reported from external studies (Liu et al. 2014; Porter et al. 2010)) and in previous observations in the Dickson laboratory (Page 2012). These studies did differ in terms of the percentages of cells that expressed half and a quarter of the amount of protein expressed compared the highest producer mini-pool. Although the CHO mini-pool used (and the selection process and product) is different in the current study from those previously reported, all the highlighted studies exhibited only a small percentage of cell lines that revealed a high expressing phenotype. In CHO-S transfected to express a model Ig4 antibody, Page (2012) generated a profile in which approximately 90% of the selected cell lines expressed less then a quarter of the high expressing selected cell line (Page 2012).

Figure 3.7 correlates the ß-galactosidase activity of the CHO-FRT mini-pools in adherent and suspension stages. No direct correlation between the ß-galactoasidase activities is observed between these two stages of these mini-pools. This lack of correlation can be observed with a change of profile of CHO-FRT 46 from the highest CHO-FRT expressing mini-pool in adherent stage (Section 3.2.2.1) to a relative medium expressing mini-pools when cultured in suspension (Figure 3.6). Similar results were observed by Porter et al (2010) in a study that tracked the productivity of the Top 10 high expressing cell lines during GS selection and cell line amplification (Porter et al. 2010). Porter et al (2010) observed that the cell lines that were ranked as high expressing cells at suspension stage were marked as medium and poor expressing cells at the previous stages of cell line development (Porter et al. 2010). The author suggested that the main differences observed are due to differences of environment of which cells are cultured. For instance, Porter et al 2010 observed that the biggest changes in the cell ranking occurred when the environment of culture or culture method were drastically changed from adherent to suspension culture and also when cells were analysed in suspension cultured with and without feed (Porter et al. 2010). It is possible that these shifts of environment

interfere on the cell genome and epigenetic environment leading to possible genomic instability and consequently gene silencing.

Figure 3.6. ß-galactosidase activity in suspension adapted CHO-FRT mini-pools.
The ß-galactosidase analysis was performed as described in Section 2.5. A 1.5 ml mid-exponential culture sample (day 4) of each CHO-FRT mini-pool was harvested as described in Section 2.5.1 and the total protein of each lysate was quantified as described in Section 2.5.2. The ß-galactosidase assay was performed as described in Section 2.5.3 and the ß-galactosidase activity of each CHO-FRT mini-pool was calculated as U per µg of total protein of lysate.
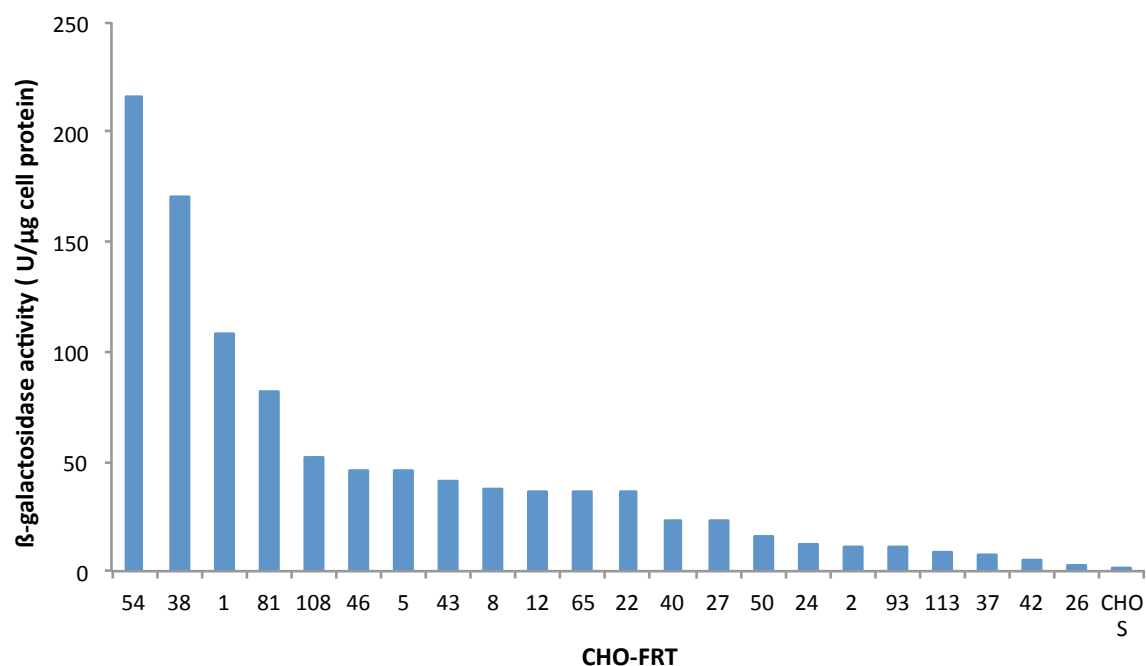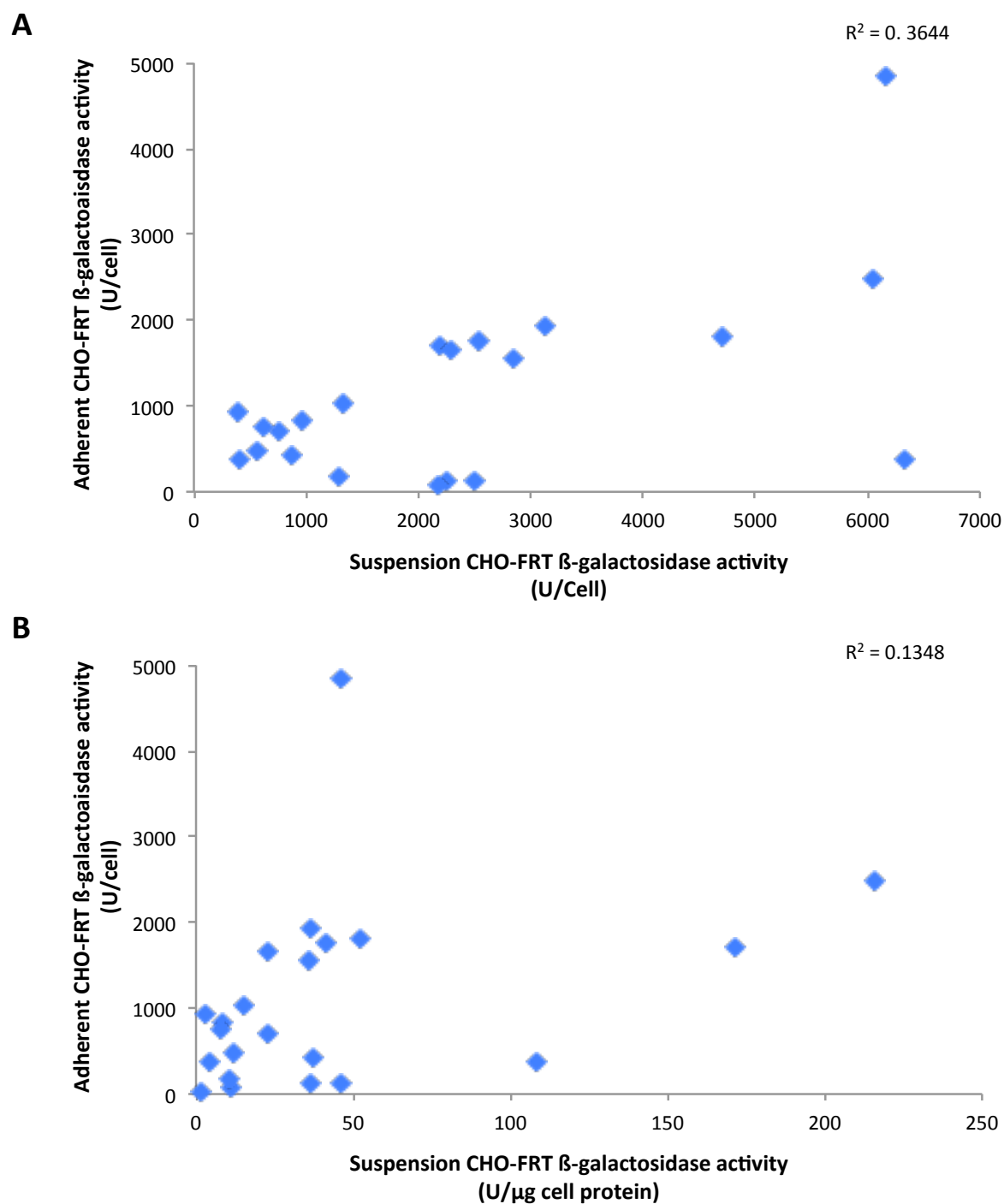
Figure 3.7. Correlation between the ß-galactosidase activity of mini-pools at adherent and suspension stages.

Adherent CHO-FRT ß-galactosidase activity (Section 3.3.1) was plotted against the ß-galactosidase activity of the same mini-pool adapted to suspension culture (Section 3.3.2), **A**- U/cell vs U/cell; **B**- U/cell vs U/cell; protein. R^2 – correlation coefficient.

3.2.3. Characterisation of the growth of CHO-FRT mini-pools

CHO-FRT mini-pool growth patterns in batch culture showed the same 4 stage profile observed for CHO-S cell line. Figure 3.9 shows that, as was observed for the distribution of ß-galactoasidase activity, the growth characteristics of CHO-FRT mini-pools varied from mini-pool to mini-pool. Most of the mini-pools had a similar or higher growth rate than the CHO-S cell line (Table 3.2), and CHO-FRT mini-pools achieved a range of maximum cell density from 5 $x10^6$-13x10^6 cell/ml. For four of the CHO-FRT mini-pools (CHO-FRT 1, 42 108 and 119) the maximum cell density was 50% less than that observed for CHO-S, ranging from 1.6-2.8x10^6 cell/ml (Table 3.2).

Such variability has been previously reported in NS0 myeloma and CHO cells (Barnes et al. 2006; Kim et al. 1998). Both studies analysed clone variability of selected cells after limiting dilution cloning and observed that even after two or three rounds of limiting dilution cloning the variability of the growth properties in the selected cell lines was considerably greater in both expression systems. These authors proposed that although such variability could be due to a metabolic burden upon the cells associated with production of recombinant protein, it might also have been caused by genomic differences between cell lines due to the different gene integration locus or to a natural phenotype drift in culture (Barnes et al. 2006; Kim et al. 1998).

Figure 3.8. Batch culture analysis of CHO-FRTmini-pools.

CHO-FRT mini-pool batch culture analysis was performed as described in Section 2.4.5. Each CHO-FRT mini-pool was cultured with an initial cell density of 0.2x106 cell/ml and cell density was measured daily as described in Section 2.4.3. (▬ CHO-FRT; ▬ CHO-S)
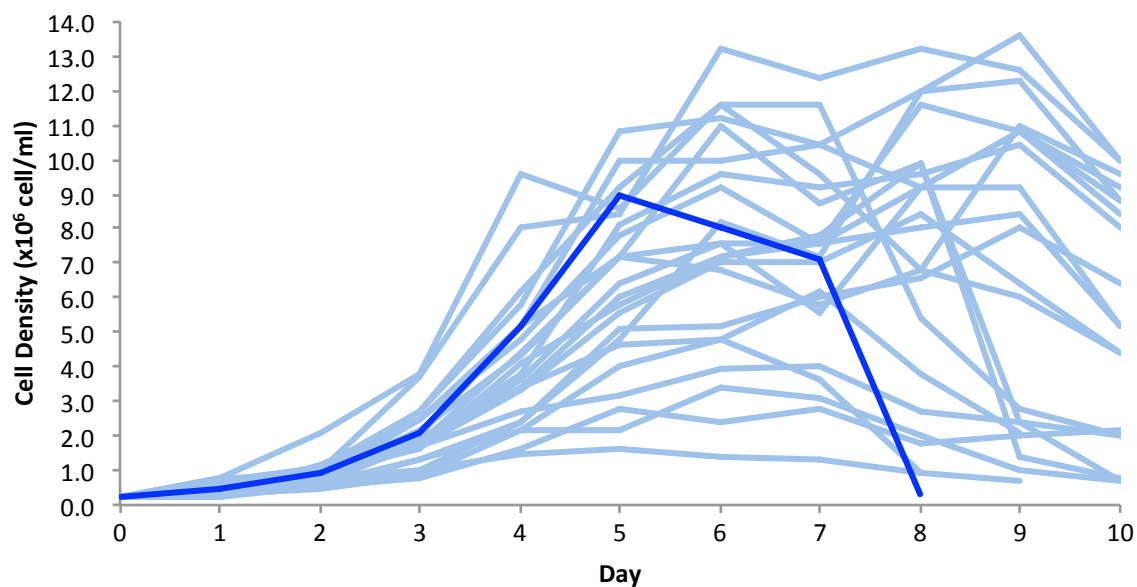
Table 3.1 CHO-FRT cell growth rate and maximum cell density

CHO-FRT mini-pool batch culture analysis was performed as described in Section 2.4.5. The growth rate of each mini-pool was calculated as described in Section 2.4.6. Maximum cell density is the highest viable cell density achieved by each CHO-FRT during batch culture (Section 2.4.4).

| CHO-FRT	1	2	5	8	12	22	24	26	27	37	38	40	42	43	46	50	54	65	81	93	108	113	CHO-S
log cell growth rate	0.23	0.38	0.20	0.31	0.31	0.27	0.31	0.25	0.33	0.27	0.31	0.28	0.17	0.36	0.31	0.31	0.27	0.34	0.28	0.31	0.22	0.15	0.25
Maximum cell density (10^6 cell/ml)	2.8	9.6	11.6	11.6	5.1	4.8	10.0	7.0	9.2	7.2	8.2	7.6	1.9	11.2	7.2	13.2	7.2	7.6	4.8	11.0	3.4	3.9	9.0

3.2.4. Correlation between ß-galactosidase activity and cell growth rate of CHO-FRT mini-pools

This section describes the ß-galactosidase activity and growth properties of CHO-FRT mini-pools analysed in Sections 3.4 and 3.5.2. The ultimate objective of this analysis is to allow a rational decision of which CHO-FRT mini-pools to be used on the study described in Section 3.3.

Figure 3.9 exhibits the correlation between ß-galactosidase activity and growth rate of CHO-FRT mini-pools. No direct relationship was observed between these two characteristics. Most of the mini-pools express relatively low amounts of ß-galactosidase varying slightly on the growth properties. However it is possible to identify six outliers mini-pools that can be grouped in three different pairs with similar phenotypes. CHO-FRT 54 and 38 (Figure 3.9 – green circle) have a high expression and medium cell growth rate phenotypes while oppositely CHO-FRT 42 and 113 (Figure 3.9 – red circle) revealed a distinctive low growth/low expression profile. CHO-FRT 1 and 81 (Figure 3.9 – yellow circle) formed a third group of medium expression/medium growth mini-pool phenotype. These clusters have the only purpose to ease the selection of the cell to progress for the next stage of characterisation (Section 3.3). By clustering these outliers mini-pools in similar groups it is possible to perform a more rational decision on which mini-pools to progress without losing to much phenotypic diversity observed at this stage.

Figure 3.9. ß-galactosidase expression and cell growth rate of CHO-FRT mini-pools.
CHO-FRT ß-galactosidase activity (Section 3.4.2) was plotted against the growth rate of the mini-pool in batch culture (Section 2.5). Most of the cell express relatively low amounts of ß-galactosidase varying only on growth properties. Six mini-pools can be easily identified as outliers and grouped in three groups with different characteristics of ß-galactosidase expression and growth properties; (R^2 – correlation coefficient; ‒ ‒ High ß-galactosidase expressing mini-pools; ‒ ‒ Medium ß-galactosidase expressing mini-pools; ‒ ‒ Low ß-galactosidase expressing and low cell growth mini-pools; ◆CHO-FRT 1; ■ CHO-FRT 2; ▲CHO-FRT 5; ✕ CHO-FRT 8; ✻ CHO-FRT 12; ⬤ CHO-FRT 22; + CHO-FRT 24; ‒ CHO-FRT 26; ‒CHO-FRT 27; ■ CHO-FRT 37; ▲ CHO-FRT 38; ✕ CHO-FRT 40; ✳ CHO-FRT 42; ⬤CHO-FRT43; + CHO-FRT 46; ‒CHO-FRT 50; ‒CHO-FRT 54; ■ CHO-FRT 65; ▲ CHO-FRT 81; ✕ CHO-FRT 93; ✻ CHO-FRT108; ⬤ CHO-FRT113)
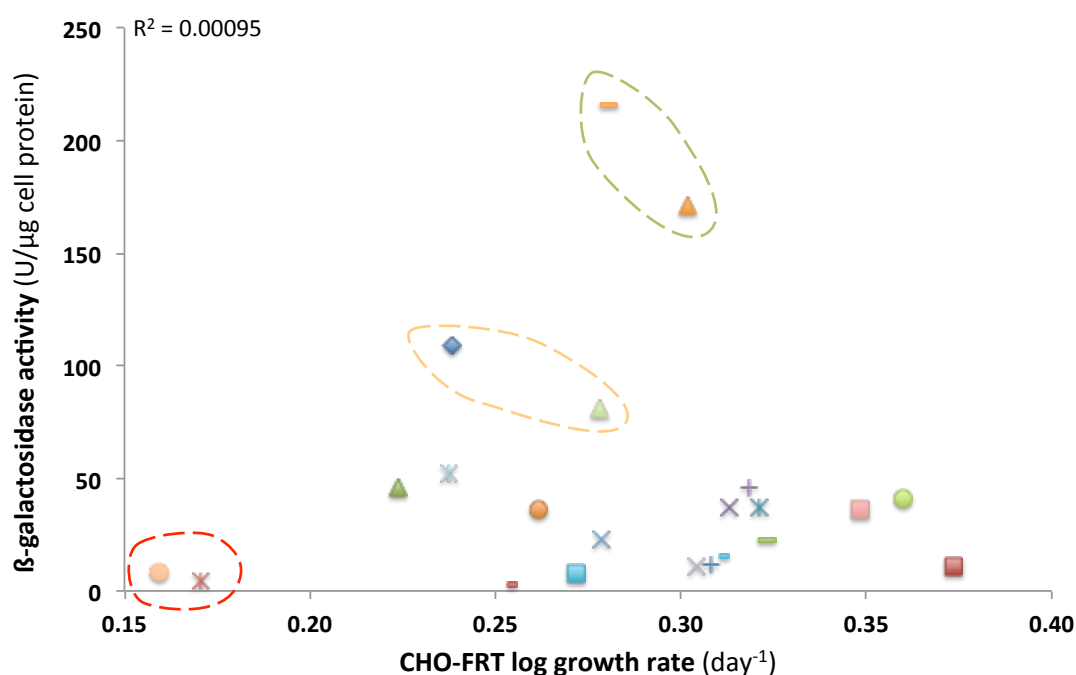
3.3. Long-term stability of selected CHO-FRT mini-pools

This Section describes the stability of CHO-FRT mini-pools during a long-term culture study where the mini-pools were continuously cultured throughout a period of 67 days (Section 2.5.3.1). To identify possible phenotype drifts, several characteristics of CHO-FRT such as, growth properties (Section 3.3.2) and ß-galactosidase activity (Section 3.3.3), gene copy number and mRNA (Section 3.3.4) of CHO-FRT were analysed over the period of culture. During this Section, the terms "early stage" and "late stage" will be used to refer to cells of the same mini-pool but analysed before and after, respectively, the long-term stability study. Also, although in this Section different cell characteristics are evaluated during the stability study, the expression "stable mini-pool" will be used to refer to cells that retain >70% of their initial protein expression during prolonged culture (Bailey et al. 2012).

Due to the technical challenges of performing a long-term culture study with all the CHO-FRT mini-pools, based on the analysis of Section 3.2.4, ten CHO-FRT mini-pools (CHO-FRT 1, 5, 22, 43, 46, 50, 54, 65, 81 and 108) were selected for stability characterisation. This selection of mini-pools of different phenotypes was undertaken in order to maximise the heterogeneity of the mini-pools used and minimise the risk of missing mini-pools representative of particular phenotypes. However, despite representing a very distinctive outlier phenotype (Figure 3.9) due to their considerably poor expression and growth characteristics CHO-FRT 42 and 113 were not selected for this study.

3.3.1. CHO-FRT long-term viable cell density

Figure 3.10 shows the viable cell density of CHO-FRT mini-pools at day 4 of culture during the long-term culture study. The Figure reveals that the number of viable cells in culture was maintained relatively constant in most of the mini-pools examined. Nevertheless, there was a considerable variability in growth of these mini-pools, with maximal achievable cell densities varying between 4.5×10^6 cell/ml for CHO-FRT 1 and 9.0×10^6 cell/ml for CHO-FRT 50. Cell viability was maintained above 95% throughout study, except at day 39 (as commented upon below).

At day 39 of culture it was necessary to perform an unscheduled passage due to a technical problem on the robotic system. During the night of the 38th day of culture, the robot shut down its shaking mechanism leaving the cells deposited on the bottom of the culture flasks for several hours. When the viable cell density was

measured, this had dropped to values between $1.0\text{-}2.0 \times 10^6$ cell /ml and so a decision was made to passage the cultures one day earlier that planned. Passaging the cells gave fresh medium to accelerate recovery from this potentially stressed situation and also to dilute dead cells in the culture, minimising the release of proteases and apoptotic factors that could affect the recovery process of the healthy cells (Goswami et al. 1999).

Most of the mini-pools took at least 8 days to recover from this incident. CHO-FRT 1 took 12 days for cell density and cell viability to return to healthy levels, whereas CHO-FRT 5 took 4 days to fully recover. As previously discussed in Section 3.2.4, CHO cell lines have a variability of cell phenotypes in their cell population with different productivity and cell growth properties (Barnes et al. 2006; Davies et al. 2013; Kim et al. 2001; Kim et al. 1998). It is possible that in a process similar to the "natural selection", the cells that survived this event had a more robust phenotype, associated with more rapid growth and, consequently this event could have changed many features of the overall population of cell lines. Based on these facts, a possible explanation for the significant phenotype change in CHO-FRT 5 is that perhaps the initial heterogeneity of cell population in this mini-pool was considerably greater compared to the others CHO-FRT. A greater heterogeneity gives a greater pool of cells, increasing the possibility of the existence of more rapidly growing cells that were selected by this event.

On day 16 of culture the robot also had a technical failure that did not allow the robot arm to work, so on that day, CHO-FRT cells were passaged manually. The incident was minor and did not have any influence on the cell behaviour and phenotype.

Although these incidents may have introduced stress factors that were not envisaged in the original long-term stability study, the results obtained in this study are still valid as the incidents were consistent to all the mini-pools.

Figure 3.10. CHO-FRT viable cell density variation during long-term culture.

CHO-FRT mini-pools were cultured for a period of 67 days as described in Section 2.6. Each CHO-FRT mini-pool was cultured at an initial cell density of 0.2x106 cell/ml and passaged with a four day period regime using a robotic system; Values are means ± Range (n=2). (Viable cell density; Cell viability)
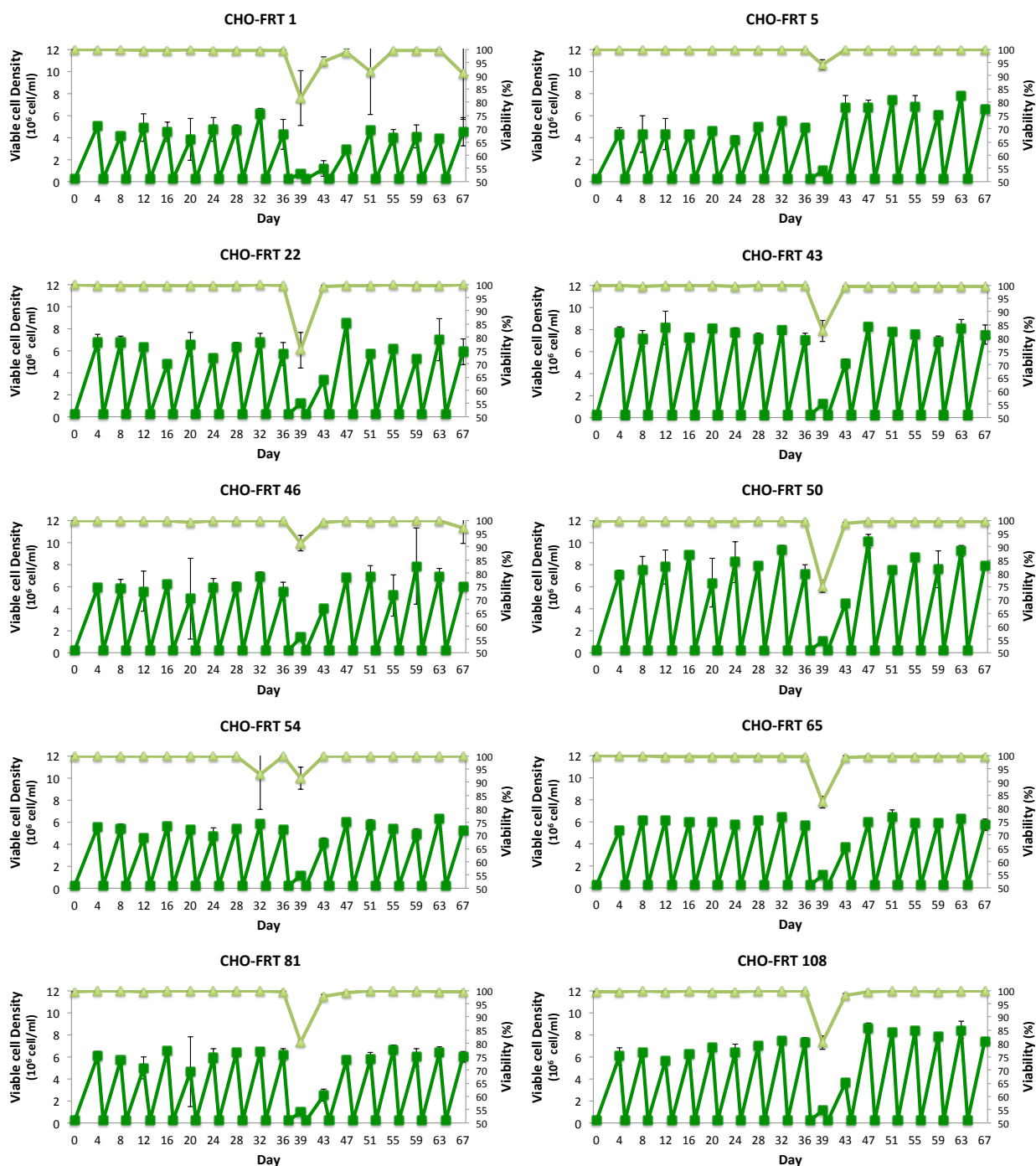
3.3.2. Characterisation of the growth of CHO-FRT mini-pools at early and late stages

Figure 3.11 shows growth profiles obtained in batch culture of CHO-FRT mini-pools assessed early and late in continuous culture. All the CHO-FRT mini-pools revealed an batch culture of 8 days similar to that observed for CHO-S, with the exception of CHO-FRT 54 analysed at early stages that had a batch culture of 9 days. Also, all CHO-FRT mini-pools revealed a similar viable cell density profile in batch culture (Figure 3.11). Due to a technical problem with the robotic system used, it was not possible to count the viable cell density of the 6th day of culture during the early batch culture analysis.

Nevertheless, CHO FRT mini-pools showed a general increase of growth rate, an effect that was more noticeable in CHO-FRT 1, 22 and 46 where an increase greater than 0.1 d^{-1} was observed (Table 3.3). This general increase of growth rate was also observed by Kim et al (1998) in a study that described the stability of CHO-DG44 clonally derived cell lines expressing a chimeric antibody (Kim et al.1998).

Figure 3.11. Batch culture properties of CHO-FRT mini-pools at early and late stages. CHO-FRT mini-pool batch culture analysis was performed as described in Section 2.6. CHO-FRT cells were cultured in the Sonata robotic system and seeded with an initial cell density of 0.2×10^6 cell/ml. Cell density and cell culture viability was measured daily as described in Section 2.6; Values are mean ± Range (n=2) (Viable cell density ■ Early stage; ◆ Late stage. Cell viability ■ Early stage; ◆ Late stage)
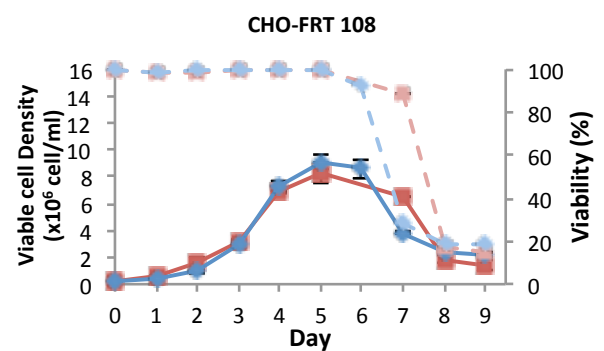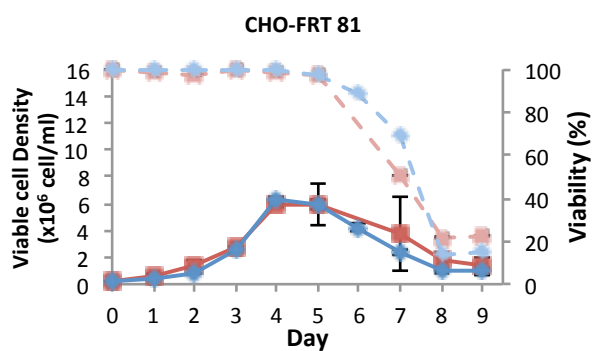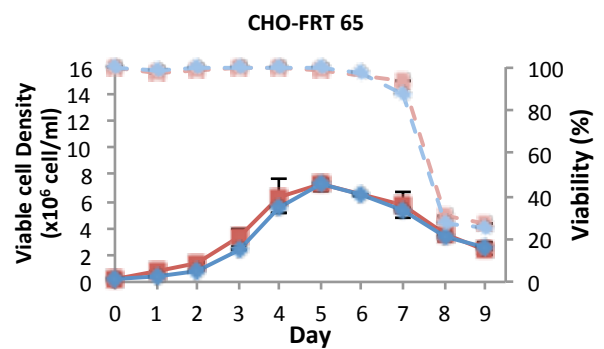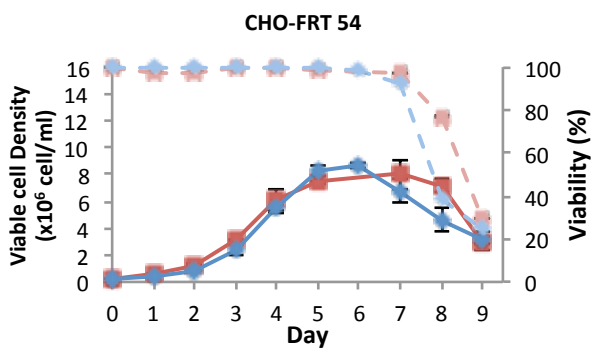
CHO-FRT 1   CHO-FRT 5

CHO-FRT 22   CHO-FRT 43

CHO-FRT 46   CHO-FRT 50

CHO-FRT 54   CHO-FRT 65

CHO-FRT 81   CHO-FRT 108
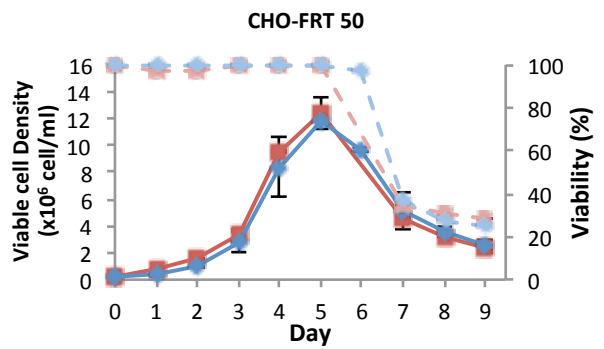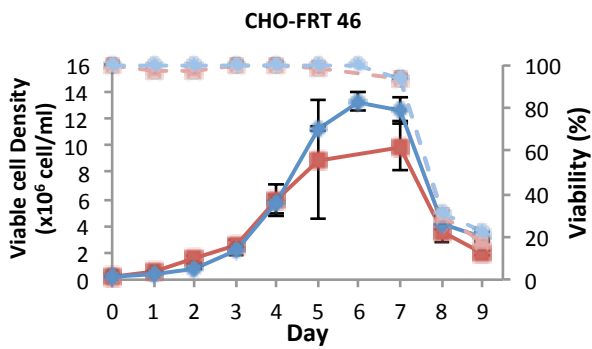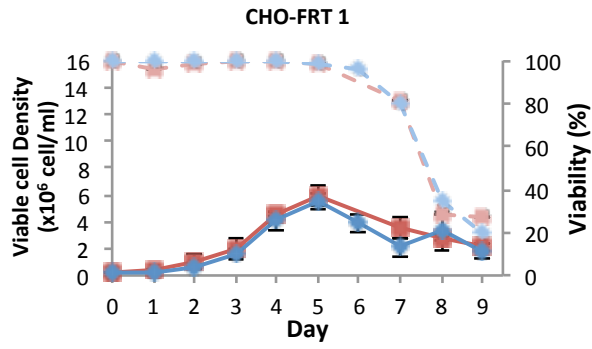
Table 3.2 CHO-FRT cell growth rate and maximum cell viable density at early and late stages.

Early and Late CHO-FRT mini-pools batch culture analyses were performed as described in Section 2.6. The growth rate of each mini-pool was calculated as described in Section 2.4.6. Maximum cell density is the highest viable cell density achieved by each CHO-FRT during batch culture (Section 2.6). Values are average (n=2)

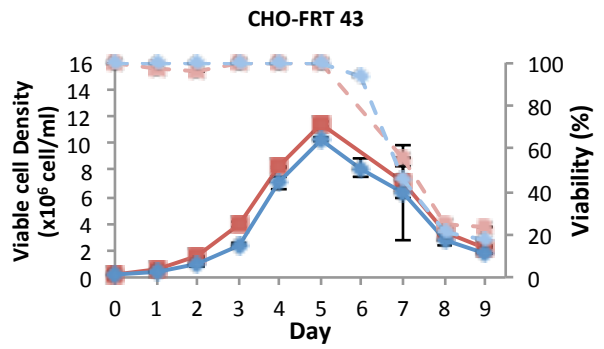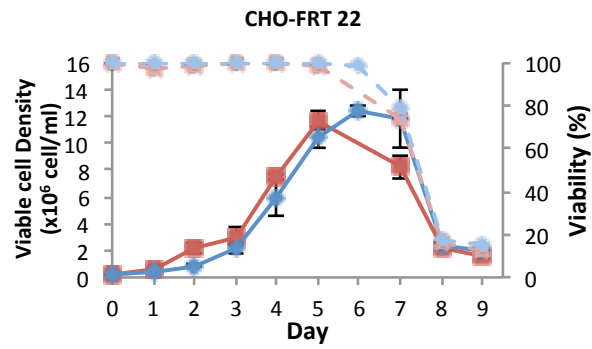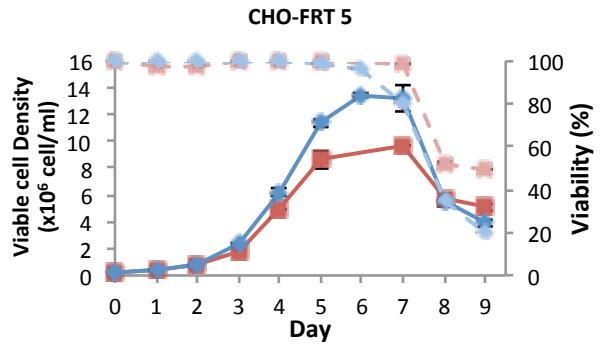|	CHO-FRT	1	5	22	43	46	50	54	65	81	108
Early stage	Log cell growth rate (day 2-5)	0.24	0.33	0.24	0.29	0.25	0.30	0.26	0.23	0.20	0.25
	Maximum cell density (10^6 cell/ml)	5.93	9.71	11.67	11.41	9.94	12.32	7.99	7.27	5.96	8.31
Late stage	Log cell growth rate (day 2-5)	0.31	0.37	0.36	0.35	0.38	0.36	0.32	0.30	0.28	0.33
	Maximum cell density (10^6 cell/ml)	5.55	13.15	12.36	10.31	13.25	11.81	8.69	7.22	5.92	9.16

3.3.3. CHO-FRT ß-galactosidase expression, gene copy number and mRNA stability

Figure 3.12-A, reveals the stability profile of CHO-FRT mini-pools in relation to ß-galactosidase expression. The results obtained for CHO-FRT mini-pools revealed that most of the CHO-FRT mini-pools analysed have a stable profile with the exception of CHO FRT 1, 22 and 81 that had an unstable profile with a drop greater than 40% of ß-galactosidase expression.

The majority of studies that analyse and describe protein expression stability report a greater proportion of instable cell lines between similar cell lines (Chusainow et al. 2009; Fann et al. 2000; Jun et al. 2006; Kim et al. 1998). There are considerable differences in the system used in this study compared to other reports. In this study a non-secreted small protein was used as a gene reporter and its transcription is regulated by the SV40 promoter, while other reports used antibodies or other complex and secreted proteins under the influence of CMV promoter. The use of a weaker promoter in the current study, one that is less sensitive to methylation and transcriptional silencing (Ho et al. 2015) with a small, simple non-secreted protein, is felt to decrease possible complications from translation and secretory processes that could have an influence on the overall cell line stability. In addition the use of the robotic system improved the quality of cell culture handling, and had been verified to produce close reproducibility between replicate flasks.

Based on the results obtained with ß-galactosidase enzyme activity, the number of CHO-FRT mini-pools subsequently analysed in terms of gene copy number and mRNA was limited due to the technical challenges that arise with both techniques when dealing with a high number of samples. To make sure that all the different cell phenotypes were analysed, six CHO-FRT mini-pools were selected covering the three different phenotypes based on their stability of ß-galactosidase activities (Figure 3.12-A). Two unstable mini-pools (CHO-FRT 1 and 22) were selected as well four stable mini-pools, two high producing mini-pools (CHO-FRT 54 and 108) and two low producing mini-pools (CHO FRT 5 and 50).

Figures 3.12-B and 3.12-C show, that contrary of what was observed during protein expression analysis, gene-copy number and mRNA were maintained for most of CHO-FRT mini-pools with the exception of CHO-FRT 54 (defined as high producing and stable for ß-galactosidase activity).

CHO-FRT 1 and 22, had an unstable profile at the level of ß-galactosidase protein expression, but both mini-pools were stable as assessed by gene-copy number and

mRNA. Therefore, since no instability was observed in both genomic and transcription stages and the protein that is used as reporter agent is a non-secreted ß-galactosidase, is possible to propose that the instability observed in this two mini-pools can either occur during the translation process or due an increase of protein degradation in later stages of the mini-pools.

In contrast, CHO-FRT 54 revealed an unstable profile at recombinant gene copy number with a loss of recombinant mRNA present in this mini-pool. Despite this recombinant gene loss, the CHO-FRT 54 mini-pool maintained protein expression during long-term culture. Since the same amount of protein activity was detected despite the decrease of mRNA it is possible to suggest that in early stages of this mini-pool, translation was the limiting factor that limited greater ß-galactosidase expression.

Figure 3.12. Early and Late stage analysis of ß-galactosidase activity, gene copy number and mRNA in CHO-FRT mini-pools.

The ß-galactosidase activity (A) analysis was performed as described in Sections 2.5. A day 4 batch culture sample was harvested as described in Section 2.5.1 and the total protein of each lysate was quantified as described in Section 2.5.2. The ß-galactosidase assay was performed as described in Section 2.5.3 and the ß-galactosidase activity of each CHO-FRT mini-pool was calculated U per µg of total protein of lysate; The ß-galactosidase gene copy number (B) and mRNA (C) analysis was performed as described in Section 2.6 an 2.7. A culture sample from day 4 of batch culture was harvested as described in Section 2.6.1 and genomic DNA and total mRNA were extracted as described in Sections 2.6.2 and 2.7.1. Gene copy number and mRNA quantification were performed by qPCR as described in Sections 2.6.3 and 2.7.2. Values are mean ± Range (n=2) (■Early stage; ■Late stage)
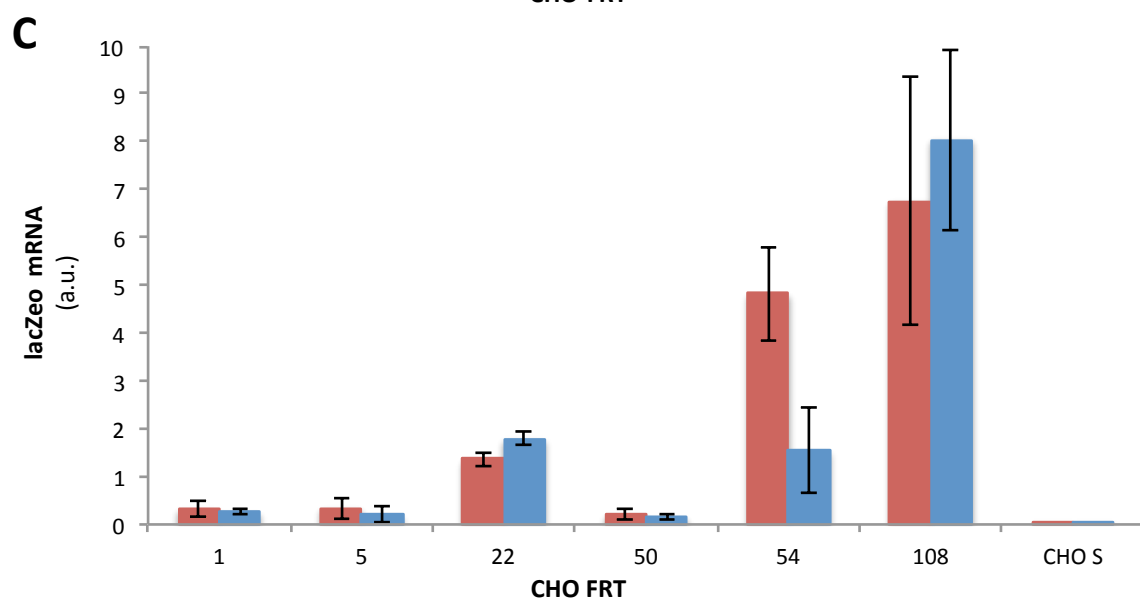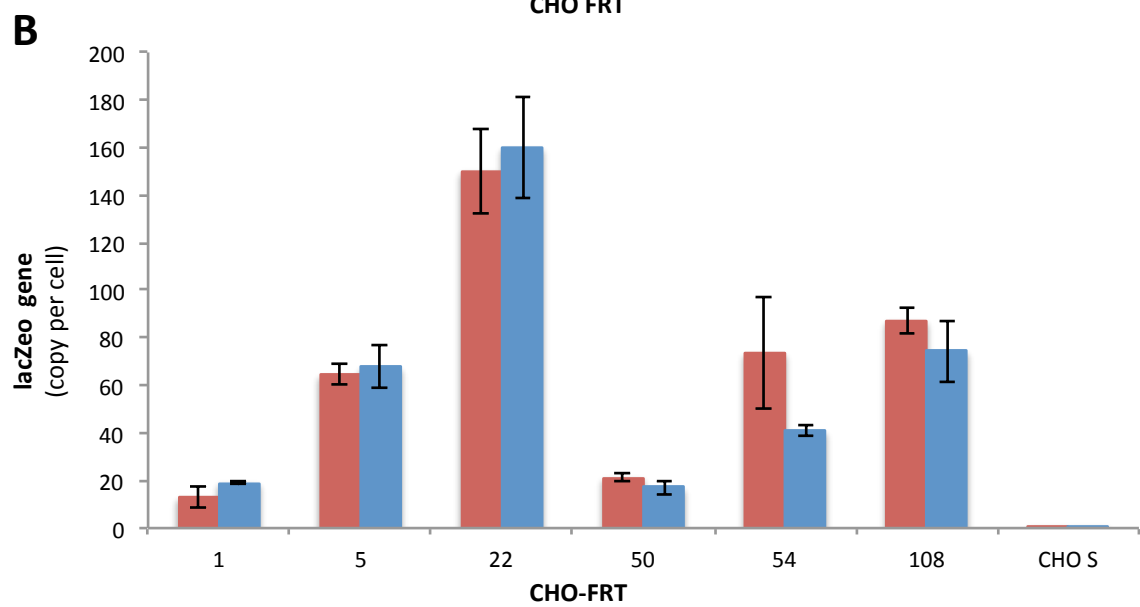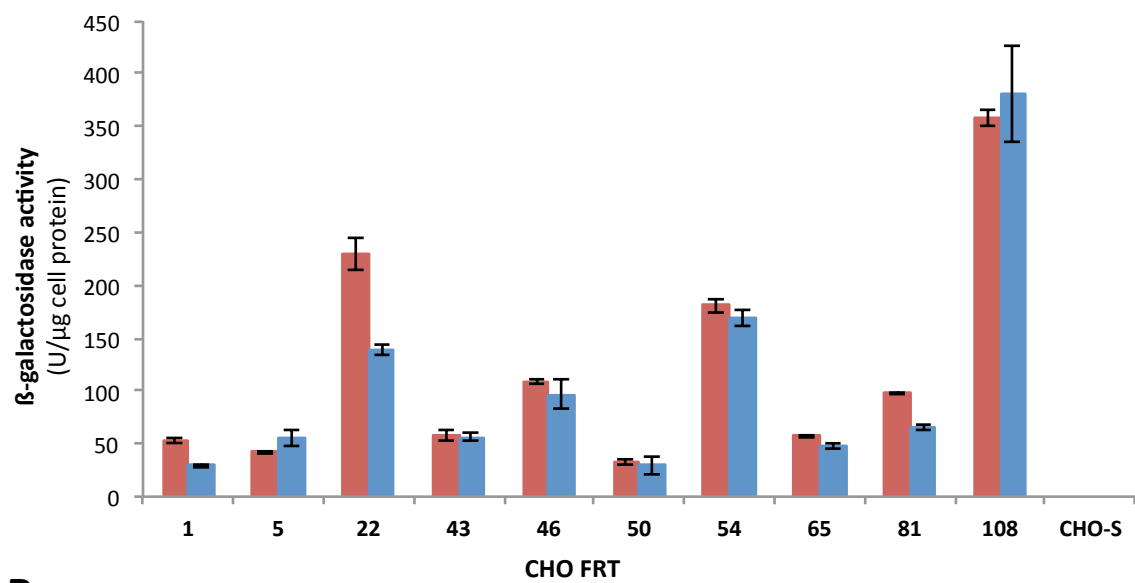
3.3.4. Correlation between ß-galactoaisdase activity, gene copy number and mRNA in early and late CHO-FRT mini-pools

This Section addresses the correlations between ß-galactosidase activity, gene copy number and mRNA at early and late stages of the CHO-FRT mini-pools analysed in Sections 3.3.4. The main objective of these correlations was to identify differences and similarities of stability phenotypes of the CHO-FRT mini-pools.

Figure 3.13-A relates the ß-galactosidase activity with the ß-galactosidase gene copy number in both early and late stages of CHO-FRT mini-pools but showed no correlation between activity and gene copy number. This has been previously observed in the Dickson laboratory (Betts and Dickson 2015). This study observed no direct correlation between the number of recombinant gene copies and the recombinant protein expression when expressing GFP and EPO in CHO-K1, CHO-S and CHO-DG44 cell lines. However, the authors showed that when UCOEs were used, there was a good correlation between gene copy and expression of recombinant protein (Betts and Dickson 2015).

In contrast a good correlation was observed between ß-galactosidase activity and mRNA (Figure 3.13-B) despite a lack of correlation between ß-galactosidase gene-copy number and mRNA (Figure3.13-C). Similar observations were reported by Betts and Dickson (2015) when expressing GFP in CHO-DG44 cells (Betts and Dickson 2015). The authors observed that despite a lack of correlation between protein production and gene copy number, there was a good correlation between mRNA and the amount of protein produced (Betts and Dickson 2015). The combination of the observations made in this study with the results described by Betts and Dickson (2015) suggest that the genomic locus and the influence of its environment on the recombinant gene have a pivotal influence on the overall cell line productivity (Betts and Dickson 2015).

Figure 3.13. Correlation between ß-galactosidase activity, gene copy number and mRNA stability in CHO-FRT mini-pools.

ß-galactosidase expression for each CHO-FRT mini-pool (Section 3.7.3) were plotted against the ß-galactosidase gene copy number (**A**) and mRNA (**B**) stability results (Section 3.7.4). Numbers against symbols indicate a specific mini-pool. ß-galactosidase gene copy number and mRNA stability results were also compared (**C**); R^2 – correlation coefficient. (■ Early stage; ■ Late stage)
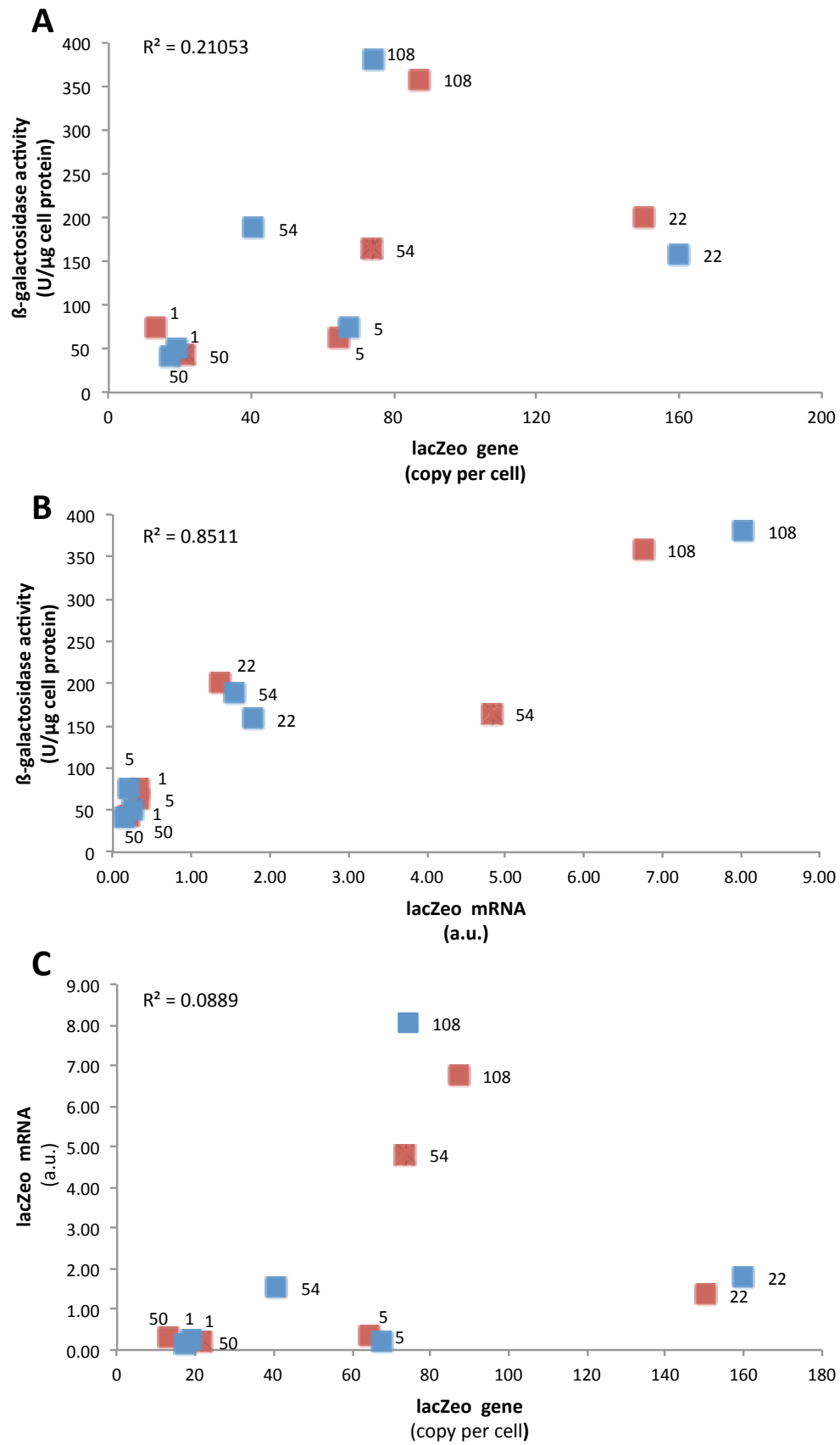
3.4. Discussion

The first objective of this Chapter was to generate a set of CHO cell lines that enable the integration of different recombinant promoters and/or genes into cell line-specific genomic locations. The results described in Section 3.2 revealed that this objective was achieved as twenty-two CHO-FRT mini-pools containing FRT and ß-galactosidase sequences in their genome were generated. These CHO-FRT were generated by stable integration of sequences of the vector pFRT/LacZeo vector (that encodes ß-galactosidase gene, as gene reporter, and a FRT sequence) into the genome of CHO-S cells. The integration of FRT sequences offers the possibility of integration, in a site-directed manner, of different recombinant promoters and/or genes to pre-determined genomic locus.

The CHO-FRT mini-pools generated had considerably different phenotypes. Although the nature of the selection method does not confer a single-clone characteristic to the cell line, the low heterogeneity origin of these mini-pools, awards distinctive phenotypical characteristics to each mini-pool generated (Wolfel et al 2013). The majority of the selected mini-pools had good growth properties when compared to the CHO-S host cell line (Section 3.2.4). Most of the mini-pools expressed ß-galactosidase relatively poorly (Section 3.2.3). Nevertheless, a heterogeneous population of mini-pools with different growth properties and ß-galactosidase expression was obtained. When both characteristics were related (Section 2.3.5), it was possible to observed the formation of a major population of mini-pools with similar poor ß-galactosidase expression but slightly varying on their growth properties. Besides this major group of cells, three pairs of outliers mini-pools were observed with two pairs having medium and high ß-galactosidase expression characteristics and the third pair with very poor expressing and growing phenotype.

In addition to generation of CHO-FRT mini-pools, a second objective of this Chapter was to characterise the stability profile of these mini-pools during a long-term culture. As the main intention of these mini-pools is to be used as host cells for the study of the stability of recombinant promoters, it was imperative to characterise the stability of CHO-FRT mini-pools before inserting any promoter and/or gene.

The results in Section 3.3 show that this objective was achieved as a full characterisation of different parameter of the CHO-FRT mini-pools was performed during long-term culture. In terms of ß-galactosidase activity, most of the CHO-FRT mini-pools revealed a stable expression profile, with the exception of three mini-pools that showed a drop of ß-galactosidase activity greater than 30% (Section 3.3.4). As

discussed in Section 3.3.4, the CHO-FRT that revealed an unstable profile did not show any loss in both recombinant gene-copy number and mRNA. Since a decrease of protein production occurred without a loss of recombinant gene and mRNA, it was proposed that, for these mini-pools, instability occurred either due to a change at a mRNA translation process or due to an increase in rates of protein degradation.

CHO-FRT mini-pools showed a good correlation between the ß-galactosidase activity and ß-galactosidase mRNA present in their cell (Section 3.3.5). On the other hand, in these cells, no correlation between ß-galactosidase activity and gene copy number and between ß-galactosidase gene-copy number and mRNA was observe (Section 3.3.5).

From the results described in Section 3.3 it is possible to conclude that the final goal of this Chapter was also achieved as four CHO-FRT mini-pools with different phenotypes and stability profiles were selected to be used in experiments to be described in Chapters 4 and 5 (CHO-FRT 108 [good expressing/stable mini-pool]; CHO-FRT 1 [poor expressing/unstable mini-pool]; CHO-FRT 22 [medium-good expressing/highly unstable mini-pool]; CHO-FRT 54 [medium-good expressing/stable mini-pool]).

Despite the fact that both objectives for this Chapter were achieved, this mini-pools will introduce a caveat for further studies. None of the CHO-FRT cell generated has just a single FRT site in its genome. Ideally, in Chapter 4 and 5, more than one CHO-FRT mini-pools containing one FRT sequence should be used. The existence of more than one FRT sequence in the CHO-FRT mini-pools, means that when a site-directed is used to target genes to the FRT sites, the population of cells that will be selected will be heterogeneous. This heterogeneity occurs due to the inefficiency of the FRT site-integration method used and also due to the selection process that does guarantee that all the FRT sites of the cell are integrated. This possibility had been expected and, in an attempt to minimise it, a limited amount of DNA was used in transfections when generating the CHO-FRT mini-pools. Yet, due to the nature of the transfection process and the selection methodology used, all the mini-pools generated have more than one gene copy.

Although the presence of a possible heterogeneous population can influence the results described in Chapters 4 and 5, CHO-FRT mini-pools can still be used to study the transcriptional activity and stability of recombinant promoters. For instance, the non-random integration of the second recombinant promoter that will be studied results that although heterogeneous, the heterogeneity of the cells obtained is limited

compared to random integration as the amount of FRT sites is also limited in each of the CHO-FRT sites. Furthermore, Chapter 4 and 5 uses fluorescent proteins as second gene reporters. By using cell sorting selection methods, it is possible to select specific sub-populations of the pool of selected cell and study the expression behaviour of those sub-populations when different promoters are used. This strategy will be further described and discussed in Chapter 4.

Chapter 4. The influence of SV40 and CMV promoters on the long-term recombinant protein expression in CHO cells

The main goal of this Chapter is to describe a methodology that can be used to study the influence and the behaviour of recombinant promoters on the stability of recombinant protein expression in CHO cell lines. Ideally, to understand the impact of the promoter on the stability of recombinant gene expression, the promoters used should be in the same genomic environment. In an attempt to achieve similar genomic environment, fluorescent proteins (GFP and YFP) under the influence of two promoters (SV40 and CMV), within a common expression cassette, have been integrated (in a site-directed manner to the FRT sites) into the genome of CHO-FRT mini-pools generated in this study (Chapter 3).

SV40 and CMV promoters are commonly used for the expression of recombinant proteins in mammalian systems and have been extensively characterised (Foecking and Hofstetter 1986; Ho et al. 2015; Zarrin et al. 1999). Also, fluorescent proteins have been widely used to screen and evaluate relative promoter strength in different cellular systems (Alper et al. 2005; Chalfie et al. 1994; Lissemore et al. 2000). By using two of the major well-characterised recombinant promoters (in combination with fluorescent proteins which offer convenient expression screening) this Chapter also aims to validate the methodology used as a general protocol that could be extended to study and characterise the activity and stability of novel synthetic promoters for use in CHO cells.

The main process used to screen and characterise novel and synthetic recombinant promoters is via transient transfection (Brown et al. 2014) or stable integrative transfections (Mariati et al. 2014a). However, both processes cannot accurately characterise recombinant promoters. Transient expression is a methodology that gives a quick output of relative promoter strength, but does not consider the genomic environment that can influence the activity and stability of the promoter. On the other hand, stable integrative expression can characterise both the activity and stability of recombinant promoter expression. Nevertheless, due to the fact that recombinant promoters are generally randomly integrated, the genomic environment that influences each promoter varies, limiting true comparisons of different promoters.

The work describes in this chapter attempts to show an alternative characterisation method that uses the CHO-FRT mini-pools developed in the work described in Chapter 3 and the FRT site-direct integration methods. As discussed in Section 3.4, the CHO-FRT mini-pools have some caveats that make them not the ideal cells to be used for these types of comparative studies. The presence of multiple gene FRT sites, the inefficiency of the FRT-site directed integration process and the mini-pool origin of this cells, possibly introduces heterogeneity of genomic environments that

influences the activity recombinant promoters and the heterogeneity of phenotypes in the overall cell culture population.

However, because site direct-integration methods are used for the integration of the recombinant promoters, the heterogeneity of genomic environments of the polyclonal populations observed is limited. This limited number of FRT sequence, specially in mini-pools with low FRT sites such as CHO-FRT 1, possibly diminished the influence of a totally random integration. Although not ideal, this possible limited heterogeneity can represent a small improvement of the currents methods used to characterise and compare the activity of stably integrated recombinant promoters.

Figure 4.1 shows a flow chart of the structure of this Chapter. Two linked but independent studies where undertaken to measure the expression of GFP (Section 4.1) and YFP (Section 4.2) under the control of SV40 and CMV promoters during a long-term culture period. The study described in Section 4.1 was initially performed and solely looked to the expression of GFP under the influence of the two recombinant promoters on polyclonal populations. Because the results of this analysis of the polyclonal population were not conclusive, based on the results obtained, the study described in Section 4.2 was performed to analyse and compare the expression of recombinant proteins between polyclonal population and in a high expressing sub-population of the polyclonal cell line. For this second study new polyclonal cell lines were generated expressing YFP, which gene construct also contained an MS2 loop mRNA tag. This MS2 mRNA tag was introduced in this cell lines to allow similar fluorescent microscopic studies to the ones described in Chapter 5 if a full methodology had been developed. Finally, based on the two previous sections, Section 4.3 characterises and discusses the influence of SV40 and CMV promoters on the recombinant expression profile during log-term culture.

Throughout this Chapter the terminology "SV40-GFP", "CMV-GFP", "SV40-YFP" and "CMV-YFP" will be used as abbreviations to define the expression of the fluorescent proteins (GFP or YFP) under the influence of the defined recombinant promoter (SV40 or CMV). These abbreviations can also be preceded by the abbreviation "CHO" and/or followed by the numbers 1, 22, 54 and 108 (e.g. CHO-SV40-YFP, SV40-YFP 108 or CHO-SV40-YFP 108) which refer to the analyses of the recombinant promoter/fluorescence protein combination integrated in a specific CHO-FRT mini-pool (Section 3.3).

Figure 4.1. Overview of the strategy used to characterise SV40 and CMV promoter activity.

This Figure illustrates the sequence of steps used in this study to characterise SV40 and CMV promoters using CHO-FRT mini-pools.
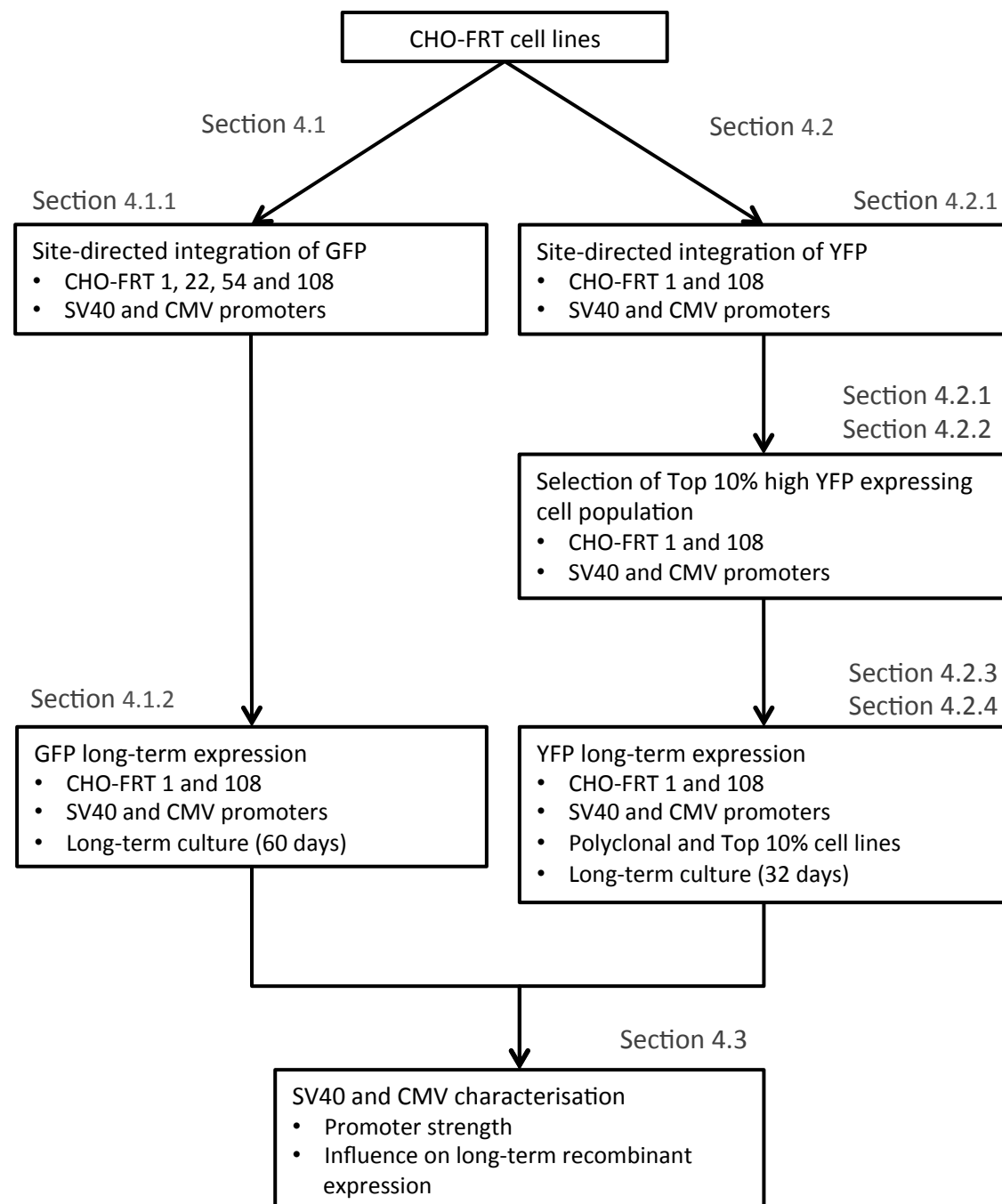
4.1 Expression of GFP in CHO-SV40-GFP and CHO-CMV-GFP polyclonal cell lines

This Section describes the initially work that was undertaken to study the expression of recombinant proteins under the influence of SV40 and CMV promoters using GFP as a reporter protein. CHO-SV40-GFP and CHO-CMV-GFP polyclonal cell lines were generated by site-directed integration of pS/FRT-GFP and pcDNA5/FRT-GFP vectors (Section 2.3), into the FRT sites existent in CHO-FRT mini-pools. Firstly, this section analyses the SV40-GFP and CMV-GFP expression in CHO-FRT 1, 22, 54 and 108 mini-pools (Section 4.1.1). After this early GFP expression analysis, the GFP expression in CHO-SV40-GFP 1, CHO-CMV-GFP 1, CHO-SV40-GFP 108 and CHO-CMV-GFP 108 was analysed during a long-term culture (Section 4.1.2) in an attempt to characterise the stability of SV40 and CMV promoters in these cell lines. Because the results of this section were not conclusive due to the fact that only relate to one transfection event, a second similar experiment was designed in Section 4.2. Nevertheless, the results obtained on the Section 4.1 were essential for the design of the experiments of Section 4.2, where recombinant expression was not only analysed in polyclonal populations but also in high expressing sub-populations of the polyclonal cell lines (Section 4.2).

4.1.1. Analysis of GFP expression in CHO-SV40-GFP and CHO-CMV-GFP polyclonal cell lines

Figure 4.2 and 4.3 shows the flow cytometry histograms of CHO-SV40-GFP (Figure 4.2) and CHO-CMV-GFP (Figure 4.3) polyclonal cell lines at the first and second culture passage after cell line selection (Section 2.6.1). Figure 4.4 shows GFP expression for CHO-SV40-GFP and CHO-CMV-GFP polyclonal cell lines (Figure 4.4-A) and the quantification of GFP expression population in both CHO-SV40-GFP and CHO-CMV-GFP polyclonal cell lines (Figure 4.4-B) analysed in Figures 4.2 and 4.3.

The use of flow cytometry in this study brings an advantage to understanding the dynamics of recombinant expression in a cell line. This technique not only allows the quantification of the overall cell line protein expression but can also show the heterogeneity of the cell population. Other quantitative methods, such as western blot, ELISA or enzymatic assays, only measure the overall protein expression by the

cell population, offering limited assessment of the possible existence of non-expressing cells or cells with differing specific productivities.

In a flow cytometry histogram, the analysis of the distribution of the cell population along the GFP fluorescence axis (x axis) reflects the heterogeneity of the cell population. In addition, the height of these histograms (y axis) illustrates the relative number of cells that have a particular GFP expression phenotype. Combining these two characteristics enables evaluate the GFP intensity of the GFP expressing population by calculating the median of the distribution of GFP expressing cells and the and an assessment of the heterogeneity of the cell population, by identification of the abundance of particular GFP expressing cell phenotypes within the heterogeneous cell population.

CHO-SV40-GFP polyclonal cell lines had a very similar profile of GFP expression for all cell lines analysed. The CHO-SV40-GFP histograms showed the presence of a single GFP-expressing cell population with relative low heterogeneity (Figure 4.2). This low heterogeneity can be observed by the relatively narrow distribution of cells in this population. A minor non-expressing cell population was also observed in all cell lines, being more pronounced in CHO-SV40-GFP 22 cell line (Figure 4.2.B). Despite a similar pattern of cell line histograms, the CHO-SV40-GFP 108 cell line showed the highest GFP intensity whilst the lowest GFP intensity was observed in the CHO-SV40-GFP 22 cell line (Figure 4.4-A). For CHO-SV40-GFP 1 and 54 the quantification of GFP intensity revealed that both cell lines expressed similar amounts of GFP (Figure 4.4-A).

After the first passage of CHO-SV40-GFP cell lines, the non-expressing GFP population disappeared in all the cell lines (Figure 4.2) that consequently led to an increase of GFP expressing population (Figure 4.4-B). Coupled with these observations, the GFP expressing population profile revealed a small shift of the main population towards higher GFP intensity (Figures 4.2) that reflected on a small increase of GFP expression in all CHO-SV40-GFP cell lines (Figure 4.4-A)

In contrast, the flow cytometry histograms for CHO-CMV-GFP cell lines differed between cell lines with distinct profiles observed for each cell line (Figure 4.3). Compared to CHO-SV40-GFP polyclonal cell lines (Figure 4.2), CHO-CMV-GFP revealed a general increase of heterogeneity in the cell line, shown by an increase in cell distribution along the x axis in the histograms for these polyclonal cell lines (Figure 4.3).

The CHO-CMV-GFP 108 cell line had a relatively high GFP expression profile with little presence of non-expressing and low GFP-expressing cells (Figure 4.3-D). Consequently, CHO-CMV-GFP 108 revealed the highest GFP intensity (Figure 4.4-A) and expressing population (90% of cells expressed GFP) (Figure 4.4-B). In contrast, CHO-CMV-GFP 22 cell line revealed a low-GFP expressing profile with only 20% of the cells expressing GFP (Figure 4.4-B). This fact underpins the overall low GFP intensity of this cell line (Figure 4.4-A).

CHO-CMV-GFP 1 (Figure 4.3-A) and CHO-CMV-GFP 54 (Figure 4.3-C) showed the presence of a mixture of GFP-expressing and non-GFP-expressing cells. In both cell lines 70% of the cells were expressed GFP (Figure 4.4-B). Despite these data, the GFP intensity in CHO-CMV-GFP 1 was greater than CHO-CMV-GFP 54 (Figure 4.4-A) showing that the GFP-expressing cells in CHO-CMV-GFP 1 cell line expressed greater amounts of GFP than the CHO-CMV-GFP 54 cell line.

After the first passage, CHO-CMV-GFP 1 and CHO-CMV-GFP 108 revealed an increase of relatively high-expressing cells (Figure 4.3-A and D) that led to and increase of GFP intensity in both cell lines (Figure 4.4-A). In contrast, CHO-CMV-GFP 22 and CHO-CMV-YGFP 54 (Figure 4.3- B and C) cell line maintained their GFP expressing population profile (Figure 4.4-B) and GFP intensity (Figure 4.4-A).

Some of the observations presented for the cell lines were not expected. For instance, because the expression of GFP is regulated by the SV40 promoters in the CHO-SV40-GFP polyclonal cell lines it was expected that these cell lines ranked similarly CHO-FRT mini-pools in Figure 3.13. Because CHO-FRT 1 was the mini-pool in Section 3.3.4 that had the lower ß-galactosidase expressing profile it was expected that the GFP fluorescence in CHO-SV40-GFP 1 would be considerably lower than other CHO-SV40-GFP.

Another unpredicted result was that CMV-GFP fluorescence was considerable lower than SV40-GFP fluorescence in CHO-FRT 1, 22 and 54 mini-pools. CMV promoter is described to be a considerably stronger promoter than SV40 (Foecking and Hofstetter 1986; Ho et al. 2015) and so it was expected that CMV-GFP fluorescence would be considerably greater than SV40-GFP fluorescence in all the CHO-FRT mini-pools.

As the results observed in this Section are the outcome of one transfection, is possible that these unpredicted observations could be limited by a possible variation of transfection the site-directed integration process efficiency of each cell line. Due to this fact, further experiments were undertaken on the generation of the CHO-SV40-

YFP and CHO-CMV-YFP polyclonal cell lines in the same manner that will be described in Section 4.2.1.

Nevertheless, because at the time of experiment there was no predictability of the behaviour of the GFP expressing population in this cell lines, a full characterisation of the long-term expression of both SV40-GFP and CMV-GFP in CHO-FRT 1 and 108 was performed. It was necessary to reduce the number of mini-pools used in this long-term characterisation due to the costs involved in this type of studies.

CHO-FRT 1 and CHO-FRT 108 were chosen because both mini-pools revealed the greatest SV40-GFP and CMV-GFP expression, with expression being greater for both constructs in CHO-FRT 108. These initial observations were in accordance with the characterisation of CHO-FRT 1 and 108 (Chapter 3), where both mini-pools revealed different stability and gene copy number profiles. In Section 3.3, CHO-FRT 1 was characterised as a low expressing, slightly unstable mini-pool with a low copy-number, while CHO-FRT 108 was characterised as a high expressing stable mini-pool with a considerable number of gene copies.
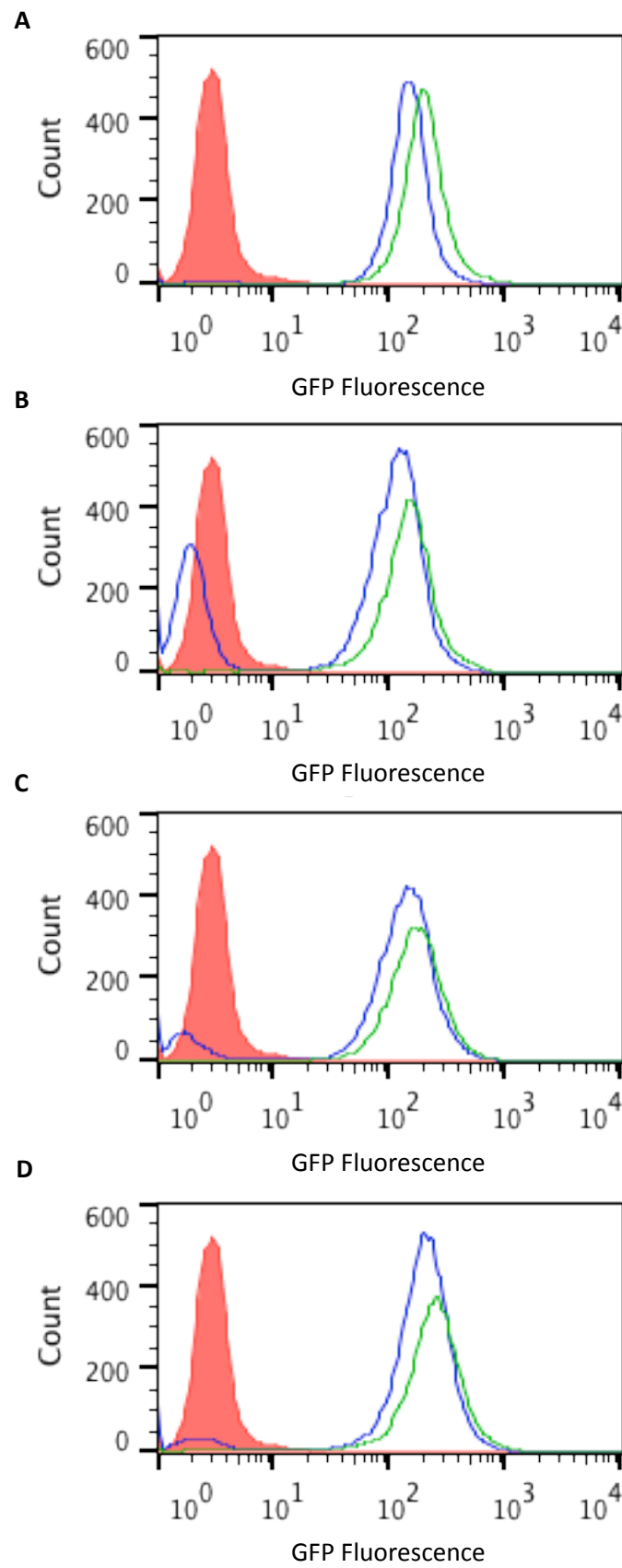
Figure 4.2. Flow cytometry histograms of CHO-SV40-GFP polyclonal cell lines.
CHO-FRT mini-pools were transfected with pS/FRT-GFP vector and selected as described in Section 2.6.1. After selection, at the first and second passages of these cell lines, 1.0×10^6 cells were harvested and re-suspended in 400 µl PBS and analysed in CYAN ADP flow cytometer, using a 488nm excitation laser (Section 2.6.1.3). Fluorescence data was acquired using a 530/30nm band-pass filter and cell line histograms were analysed using the FlowJo X 10.0.7r2 (Section 2.6.1.3). The auto-fluorescence of the CHO-FRT mini-pool was measured in every experiment as a control for non-expressing cells and used to determine the positive GFP expressing population. **A**- CHO-SV40-GFP 1; **B**- CHO-SV40-GFP 22; **C**- CHO-SV40-GFP 54; **D**- CHO-SV40-GFP 108. (▬ CHO-FRT; ▬ 1st passage; ▬ 2nd passage)

A

GFP Fluorescence

B

GFP Fluorescence

C

GFP Fluorescence

D

GFP Fluorescence

Figure 4.3. Flow cytometry histograms of CHO-CMV-GFP polyclonal cell lines.
CHO-FRT mini-pools were transfected with pcDNA5/FRT-GFP vector and selected as described in Section 2.6.1. The resulting cells were treated and analysed similarly to figure 4.2. (**A**- CHO-CMV-GFP 1; **B**- CHO-CMV-GFP 22; **C**- CHO-CMV-GFP 54; **D**- CHO-CMV-GFP 108).(━ CHO-FRT; ━ 1st passage; ━ 2nd passage)
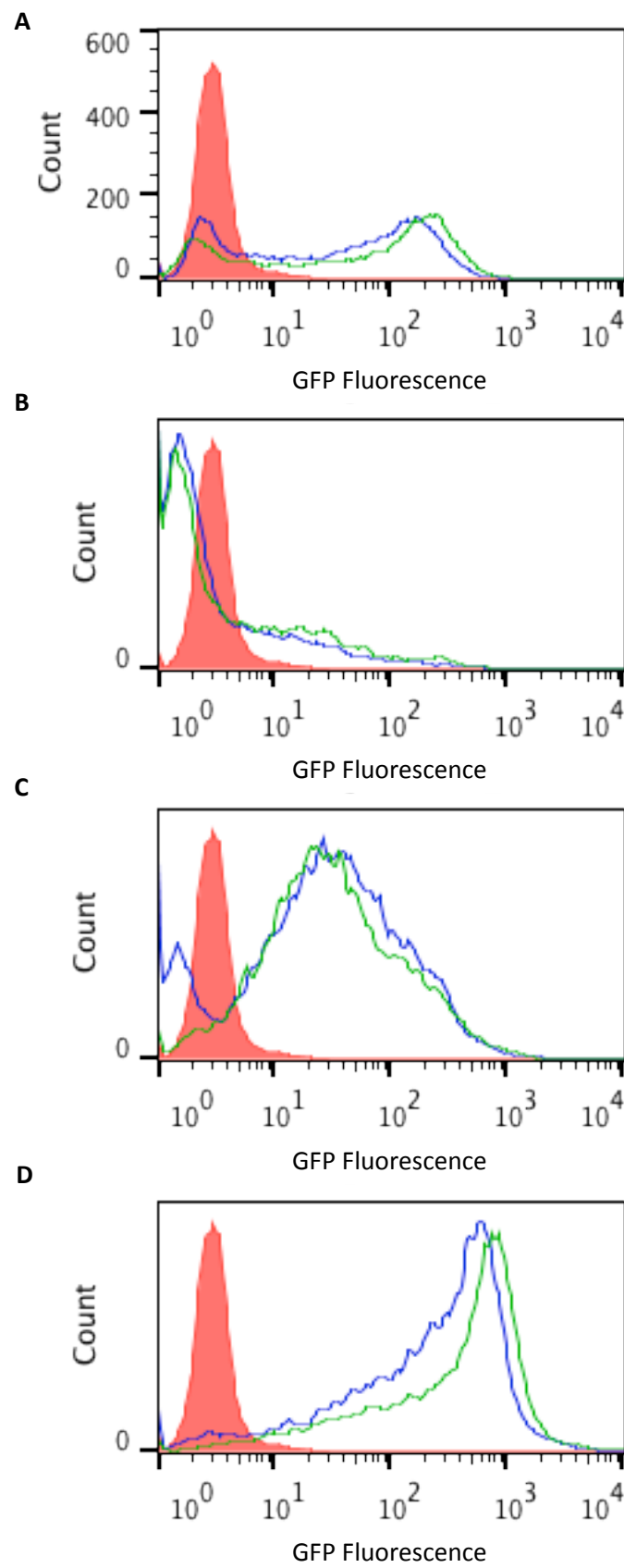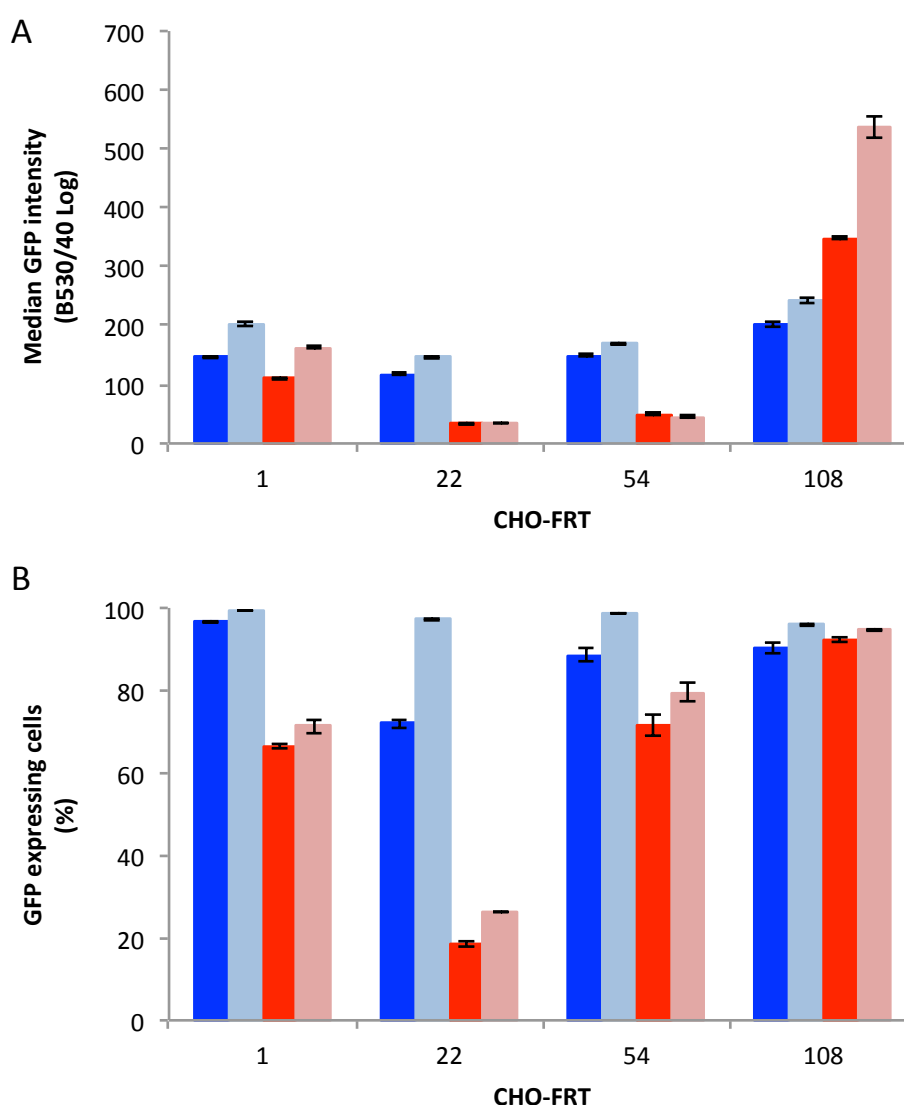
Figure 4.4. Analysis of GFP expression and GFP expressing population in CHO-SV40-GFP and CHO-CMV-GFP polyclonal cell lines.

A- The GFP expression in each CHO-SV40-GFP and CHO-CMV-GFP polyclonal cell line at the first and second passage of culture after selection. The GFP expression was calculated by measuring the median level of GFP fluorescence intensity (530/30 Log) of the GFP expressing population of the histograms in Figure 4.2. **B**- Quantification of the GFP expressing cells in each CHO-SV40-GFP and CHO-CMV-GFP polyclonal cell line at the first and second passage of culture after selection. The percentage of the GFP expressing cells in each cell line was calculated using the Summit V4.3 software (Section 2.6.1.3) Data are mean ± SD (three technical replicates) (■ SV40-GFP 1st Passage; ■ SV40-GFP 2nd Passage; ■ CMV-GFP 1st Passage; ■ CMV-GFP 2nd Passage)

4.1.2. SV40-GFP and CMV-GFP long-term expression in CHO-FRT mini-pools

Figure 4.5 shows the flow cytometry histograms obtained during long-term culture of polyclonal cell lines CHO-SV40-GFP 1 (Figure 4.5-A), CHO-SV40-GFP 108 (Figure 4.5-B), CHO-CMV-GFP 1 (Figure 4.5-C) and CHO-SV40-GFP 108 (Figure 4.5-D). Similar to the analyses of the long-term culture of CHO-FRT mini-pools (Section 3.3), in this analysis the terms "early stage" and "late stage" are used to define the cell lines at the start (day 0) and at the end (day 60), respectively, of long-term culture.

In both CHO-SV40-GFP and CHO-CMV-GFP polyclonal cell lines, the cell population profile was maintained during the period of culture with the presence of a stable GFP expressing population (Figure 4.5). For the CHO-CMV-GFP 1 polyclonal cell line (Figure 4.5-C) the small non-expressing population observed at day 0 and day 18 of culture was no longer detectable by day 32.

The preservation of the cell population profiles of CHO-SV40-GFP and CHO-CMV-GFP was reflected by the maintenance of GFP intensity during the period of culture (Figure 4.6-A). Based on the comparison of GFP expression at early and late stages, both CHO-SV40-GFP and CHO-CMV-GFP cell lines appeared to be stable as no decrease of GFP expression was observed (Figure 4.6-B).

Several reports have shown a considerable loss of CMV-driven recombinant protein expression in CHO cells (Du et al. 2013; Kim et al. 2011; Mariati et al. 2014b; Osterlehner et al. 2011; Spencer et al. 2015; Williams et al. 2005). Furthermore, a direct comparison of expression of GFP under the control of SV40 and CMV promoters during long term culture has described SV40 as a stable promoter and CMV as an unstable promoter (Ho et al. 2015). Using a definition of stability that a stable cell line was one that maintained 70% of its initial expression (Bailey et al. 2012), the authors described SV40 promoter as having a positive influence towards stable recombinant production (Ho et al. 2015). In contrast, the authors described CMV promoter as an unstable promoter as they observed a loss greater than 50% of initial expression. (Ho et al. 2015).

From Figures 4.4 and 4.6 it is clear that CHO-SV40-GFP and CHO-CMV-GFP polyclonal cell lines do not reflect the expected relative differences of SV40 and CMV promoter strengths and stability (Ho et al. 2015). As the results are the outcome of one transfection, it is not possible to make absolute conclusions from the observations. This approach was repeated in Section 4.2 using YFP and

consequently, the reinforcement obtained form both studies allows a more detailed discussion will be further discussed in Sections 4.2.2 and 4.2.4 .
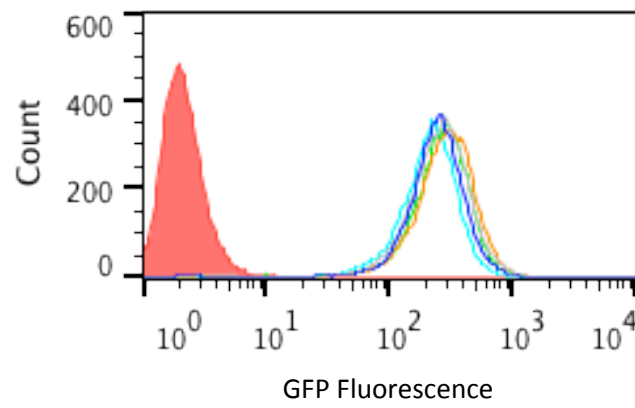
Figure 4.5. Flow cytometry histograms of CHO-SV40-GFP and CHO-CMV-GFP polyclonal cell lines over long-term culture.

CHO-SV40-GFP 1 (A) and CHO-SV40-GFP 108 (B) polyclonal cell lines and CHO-CMV-GFP 1 (C) and CHO-CMV-GFP 108 (D) polyclonal cell lines were continuously cultured for 60 days as described in Section 2.6.1.4. At day 0, 18, 32, 46 and 60 of the long-term culture, a cell culture sample was harvested and 1.0×10^6 cells of cells were re-suspended in 400 µl of PBS. The re-suspended cells were analysed in CYAN ADP flow cytometer, using a 488nm excitation laser (Section 2.6.1.3) and fluorescence data was acquired using a 530/30nm bandpass filter. Cell line histograms were analysed using the FlowJo X 10.0.7r2 software (Section 2.6.1.3).(▬CHO-FRT; ▬ Day 0; ▬ Day 18; ▬ Day 32;▬ Day 46; ▬ Day 60)
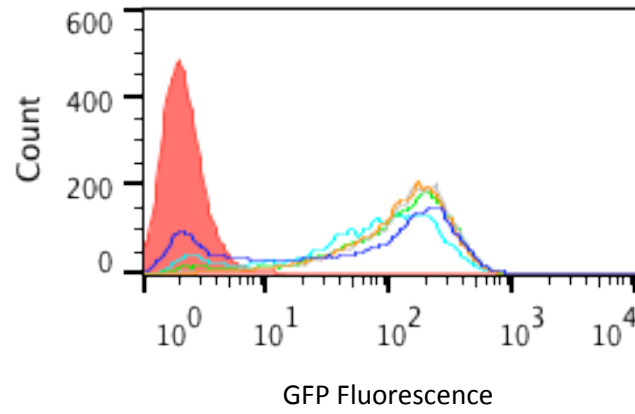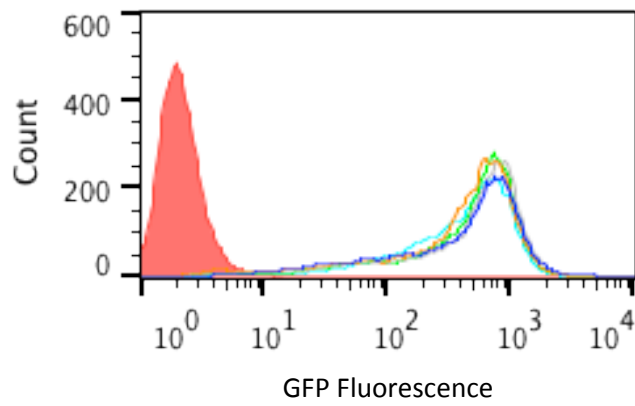
A
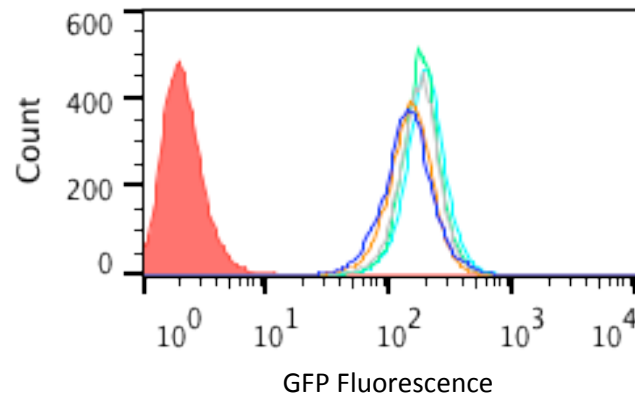
B

C

D

Figure 4.6. GFP expression in CHO-SV40-GFP and CHO-CMV-GFP polyclonal cell lines during long-term culture.

CHO-SV40-GFP 1 (◆), CHO-SV40-GFP 108 (■), CHO-CMV-GFP 1 (■) and CHO-CMV-GFP 108 (◆) cell lines were continuously cultured for 60 days as described in Section 2.6.1.4. At day 0, 18, 32, 46 and 60 of the long-term culture, GFP fluorescence intensity of each cell line was analysed (**A**) in CYAN ADP flow cytometer (Section 2.6.1.3) and calculated by measuring the median of level of fluorescence (530/30 Log) using the Summit V4.3 software. To characterise the stability of each cell lines, the GFP fluorescence intensity at early and late stages (**B**) of the each cell line were compared. Data are mean ± Range (n=2) (■ CHO-SV40-GFP early stage; ■ CHO-SV40-GFP late stage; ■ CHO-CMV-GFP early stage; ■ CHO-CMV-GFP late stage)

151
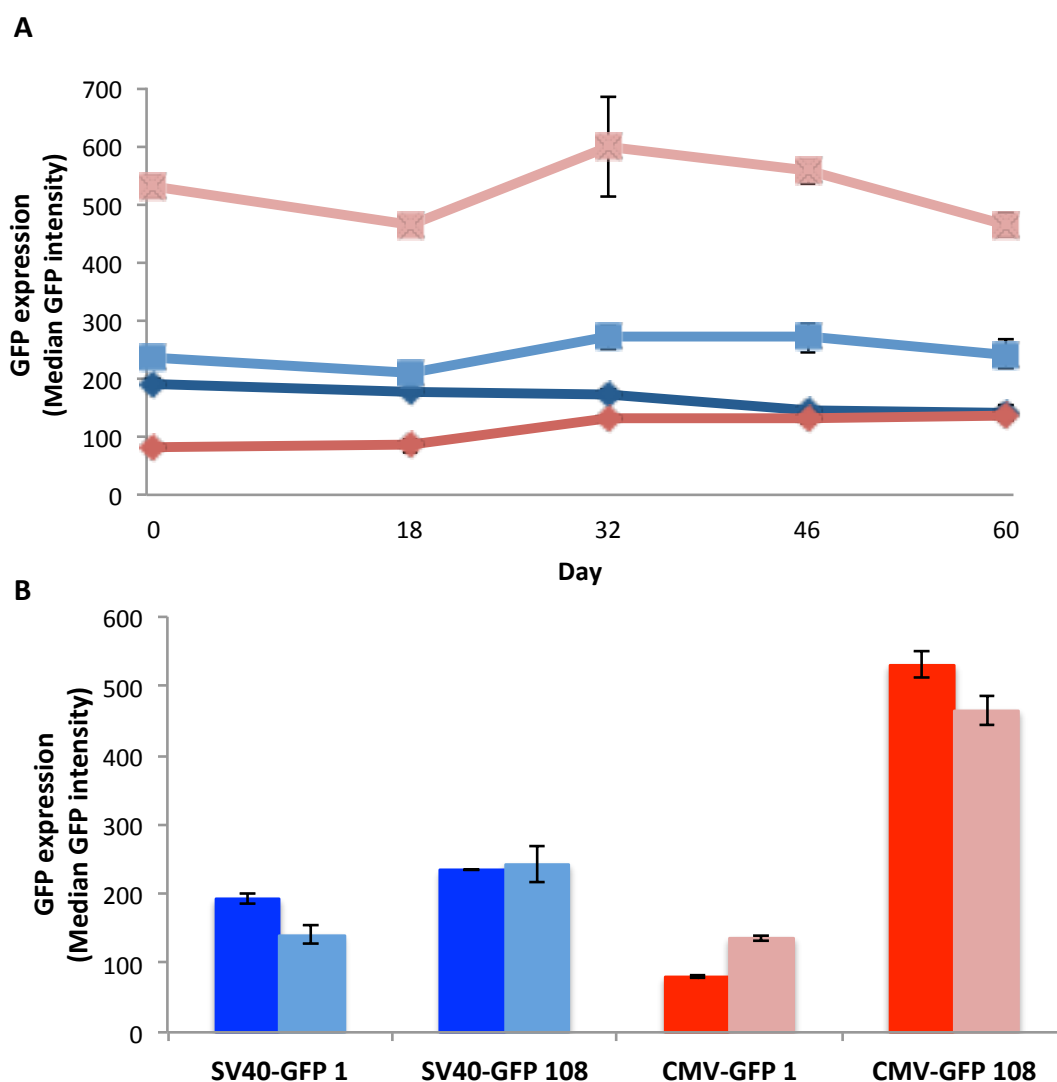
4.2. Characterisation of the influence of SV40 and CMV promoters on the stability of YFP expression in CHO-SV40-YFP and CHO-CMV-YFP polyclonal cell lines

This Section characterises the influence of SV40 and CMV promoters on the stability of YFP expression in CHO-SV40-YFP and CHO-CMV-YFP cell lines. These cell lines were generated in the same manner as the CHO-SV40-GFP and CHO-CMV-GFP cell lines (Section 4.1). The only difference was that pS/FRT/MS2-YFP and pcDNA5/FRT/MS2-YFP vectors (Section 2.3) were used for the site-directed integration process.

Firstly, this Section describes the selection by FACS of the top 10% high expressing population of CHO-SV40-YFP and CHO-CMV-YFP cell lines and compares the YFP expression in both polyclonal and top 10% cell lines (Section 4.2.2). Similar to the studies described in Section 4.1, the influence of the SV40 and CMV promoters in the stability of recombinant expression was characterised by measuring the expression SV40-YFP and CMV-YFP during long-term culture (Section 4.2.3).

The top 10% cell lines were selected as an attempt to diminish the impact of possible site-directed integration inefficiency that can consequently lead to multiple cassette insertions due to the presence of multiple FRT site in CHO-FRT mini-pool. This second round of selection for high-expressing cells was based on three assumptions:

(1) Since the Flp-In system was used YFP integration would not be a random process (Gronostajski and Sadowski 1985)

(2) Due to inefficiency of the Flp-recombinase, site-directed integration would not be 100% efficient for all the selected cells, but there would be cells in the polyclonal cell population with YFP integrated in all FRT sites (Buchholz et al. 1996);

(3) Due to a limited number of FRT targets, the cells with higher YFP expression would have greater numbers of YFP integrations for each site.

Based on these assumptions, by selecting the top 10% high expressing population in both polyclonal cell lines, the variability of cell genotypes should be diminished, as cells with similar high-expressing phenotype will probably have similar genotype composition. Furthermore, the use of FACS as a second selection method for high-expressing cells has been successfully used to study CHO productivity and long-term recombinant stability (Du et al. 2013; Spencer et al. 2015; Tornøe et al. 2002).

Although this method to select for high expressing populations has been successfully used in previous studies, this method has some caveats. It is possible that a stringent selection threshold for high expressing cells not only improves the likelihood of isolating high producing cells but also for cells that do not have an intrinsic high expressing profile but are perturbed above steady state at the time of selection (Pilbrough et al 2009).

To diminish the influence of the efficiency of the site-directed integration and to give statistical significance to the characterisation of both SV40 and CMV promoters, in this Section, two independent transfections were made to specifically integrate SV40-YFP and CMV-YFP constructs in CHO-FRT 1 and CHO-FRT 108 mini-pools, generating two independent cell lines for each construct/cell line combination (Section 4.2.2). From each independent cell line, two biological replicates were generated leading to a total of four biological samples used for the characterisation of these promoters (Section 4.2.4). To help the comparison and analysis of the results, the two independent cell lines are named with a ".1" or ".2" after the name of the cell line that generated (e.g. CHO-CMV-YFP 108.1 and CHO-CMV-YFP 108.2). In this Section, the histograms shown for each independent cell line are from just one of the biological replicate as no notable differences of cell line histograms were observed between biological replicates. The biological replicates for each independent cell line can be found in Appendix 3.

4.2.1. Characterisation of YFP expressing cell population in CHO-SV40-YFP and CHO-CMV-YFP polyclonal and Top 10% cell lines

Figure 4.7 shows the flow cytometry 530/30 Log vs Side scatter plots of CHO-SV40-YFP and CHO-CMV-YFP cell lines from which the top 10% cell population was selected by FACS (Section 2.6.2.2). Figure 4.8 shows the resulting cytometry histograms after the polyclonal and top 10% selection of the same cell lines.

Figures 4.7 and 4.8 appear to show different YFP expressing values for the same cell line. The fluorescent of the cells of the polyclonal cell lines and the top 10% sub-population measures during cell sorting using BD FACS Aria does not correspond to the same values of the fluorescence of the same population in the BD Accuri due to differences in voltage used in both instruments. The voltage on the BD FACS Aria machine used was tuned in each sort shown in Figure 4.7 to ease the selection of the top 10% sub-population. On the other hand, BD Accuri analyser voltage is set-up by the manufacture, not allowing any cell line specific voltage tuning. Because the

voltage used during sorting was difference of the standard voltage of the BD Accuri analyser, the fluorescence values measured of the same cell lines are different in Figure 4.7 and 4.8.

Similar to what was observed for the expression of SV40-GFP and CMV-GFP, the histograms in Figure 4.8 show that the SV40-YFP and CMV-YFP have different profiles of cell population from different CHO-FRT host cells.

CHO-SV40-YFP 1.1 and CHO-SV40-YFP 1.2 revealed the biggest difference in YFP-expressing population (Figure 4.8-A). Both CHO-SV40-YFP cell lines revealed well-defined populations but with opposite characteristics. In CHO-SV40-YFP 1.1 the cells expressed medium amounts of YFP compared with CHO-SV40-YFP 1.2 that expressed high amounts of YFP. On the other hand, both CHO-SV40-YFP 108 cell lines revealed very similar profiles with a high heterogeneous population and an even distribution of cells expressing different amounts of YFP (Figure 4.8-B).

CHO-CMV-YFP 1.1 and CHO-CMV-YFP 1.2 revealed different flow cytometry histogram profiles with CHO-CMV-YFP 1.1 showing a low/medium YFP-expressing phenotype population while CHO-CMV-YFP 1.2 displayed a more heterogeneous population with most of the cells displaying a non-YFP-expressing profile (Figure 4-8-C). In CHO-CMV-YFP 108 polyclonal cell lines both CHO-CMV-YFP 108.1 and CHO-CMV-YFP 108.2 revealed similar, low-expressing profiles (Figure 4.9-C). In this cell lines, with the main cell population concentrated in a very distinctive peak of relatively low or non-YFP-expressing cells.

The variations of cell line distribution for these cell lines were not expected, especially as the same promoter/YFP cassette had been integrated in a site-directed manner. It is possible these variations could result from differences in efficiency of site-directed integration between different independent transfections. The Flp-recombinase encoded in the vector pOG44 supplied in the Flp-In system has a mutation in the codon for amino acid 70 (changing from Phe to Leu) (Buchholz et al. 1996). This mutation destabilises Flp-recombinase so that it is more sensitive to temperature, a feature that decreases Flp-recombinase activity at 37 ºC (Buchholz et al. 1996). Based on low Flp-recombinase activity combined with the fact that FreeStyle™ MAX efficiency in CHO-S cells is 60% (FreeStyle™ manual) is possible that overall site-directed integration efficiency was considerably low. A low efficiency in CHO-FRT mini-pools, that have more than one FRT site, will likely lead to a high variability of cells with different amounts of YFP integrated in different locations in the polyclonal cell population.

Furthermore, variability of cell line histogram profiles from the same combination of promoter/fluorescence protein have also been observed for HEK293 and CHO cells by Eyquem et al (2013) and Tornøe et al (2002), respectively (Eyquem et al. 2013; Tornøe et al. 2002). Eyquem and colleagues studied GFP expression in HEK-293 cells under the influence of two recombinant promoters (spleen focus forming virus – SFFV; eukaryotic elongation factor 1 alpha with the first intron - EEF1A1+intron) in specific genomic locations (SH6, FUT8, or DMD21 locus) that had been targeted using meganuclease (Eyquem et al 2013). In that study, the authors observed that the cell histogram profiles were dependent on the genomic locus/promoter combination, with the same promoter leading to different expression of GFP in different loci and also different promoters. The authors also observed that the insertion of more than one GFP gene copy in the same locus could influence the histogram profile (Eyquem et al. 2013). As CHO-FRT 1 and CHO-FRT 108 mini-pools have more than one FRT site and possibly more than one gene copy per locus, it can be suggested that the high variability observed of cell phenotypes in these cells was a result of similar factors coupled with the multitude of different gene copy numbers/genomic locus combinations in the cells of the CHO-SV40-YFP polyclonal cell lines.

Another study, using site-directed integration in a CHO Flp-In cell line to evaluate the activity of synthetic promoters, showed that differences in cell line histograms can be observed between independent transfections (Tornøe et al. 2002). The differences reported by the authors were minor compared to the ones reported in this Section. However, Tornøe et al (2002) observed minor differences after two rounds of FACS selection for the top 10% high expressers (Tornøe et al. 2002), highlighting the strong influence of the variability of the site-directed integration efficiency in these observations.

The flow cytometry histograms obtained for the CHO-SV40-YFP and CHO-CMV-YFP Top 10% cell lines exhibited very similar profiles. All these cell lines showed a well-defined population of relatively high expressing cells. However, in both CHO-SV40-YFP 108 (Figure 4.8-B) and CHO-CMV-YFP 108 (Figure 4.8-D) top 10% cell lines a small number of cells expressing lower amount were observed increasing the heterogeneity of these cell lines. Nevertheless, contrary to observations made for polyclonal cell lines, no non-expressing cells were present in the top 10% cell lines. A well-defined high expressing population was also observed by Tornøe et al (2002). However, the authors also showed the presence of non-expressing cells in cell lines generated by FACS selection (Tornøe et al. 2002).
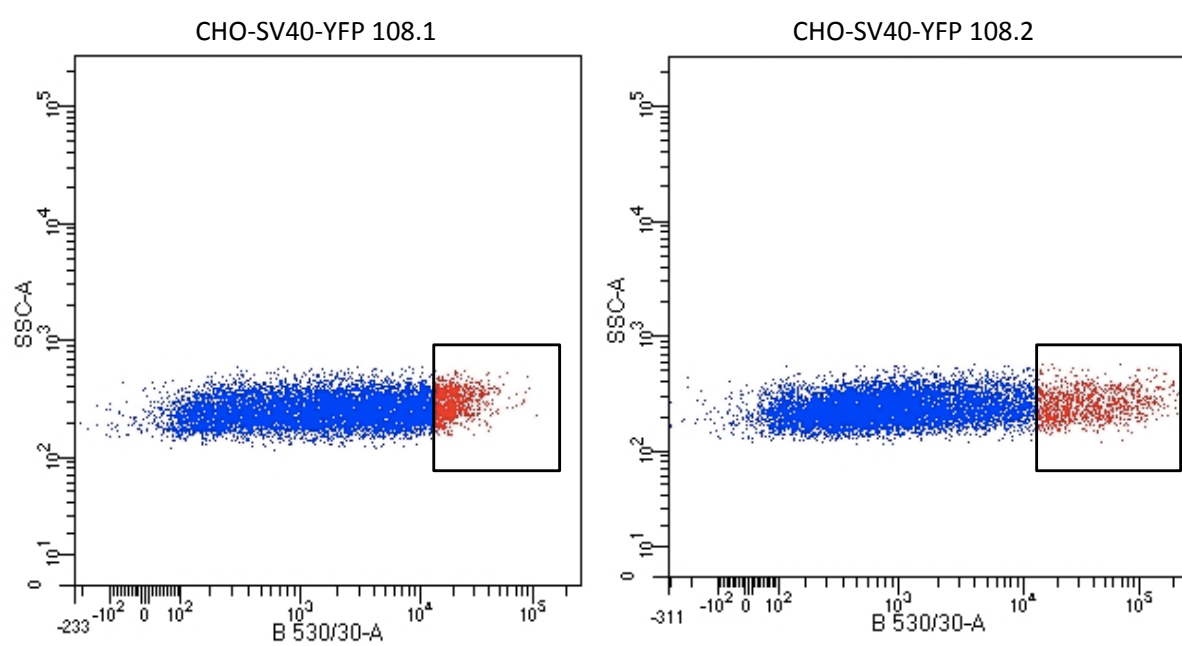
Figure 4.7. CHO-SV40-YFP and CHO-CMV-YFP Top 10% sub-population selection by FACS.

CHO-FRT mini-pools were transfected as described in Section 2.6.2.1 and polyclonal pools selected as described in Section 2.6.2.2. A 5 ml day 4 culture sample of CHO-SV40-YFP and CHO-CMV-YFP polyclonal cells at early stages of the cell line was harvested and prepared as described in Section 2.6.2.2. Prepared cell samples were analysed using the BD FACSAria™ Fusion and the cell line GFP fluorescence vs. side scatter plots were acquired using the BD FACSDiva™ software. The Top 10% high expressing cell population were gated (black square) and 2.5×10^5 cells were sorted and collected into a 5 ml collecting tube containing 0.5 ml of maintenance medium (Section 2.6.2.2). (**A**- CHO-SV40-YFP 1; **B**- CHO-SV40-YFP 108; **C**- CHO-CMV-YFP 1; **D**- CHO-CMV-YFP 108).
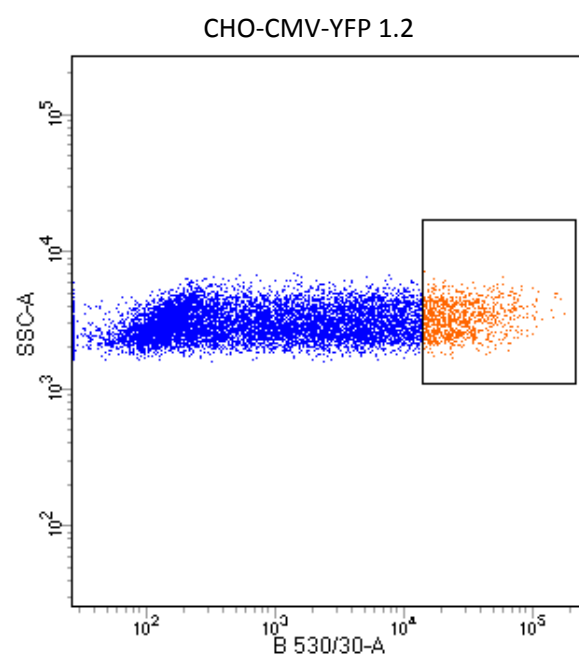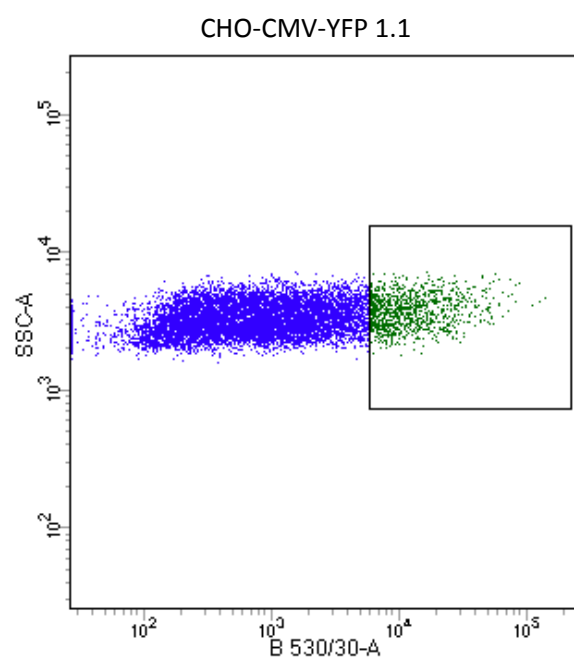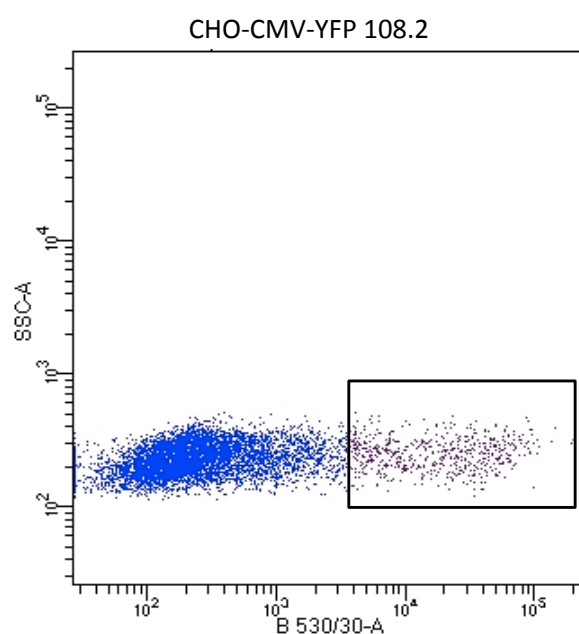
A

CHO-SV40-YFP 1.1

CHO-SV40-YFP 1.2

B

CHO-SV40-YFP 108.1

CHO-SV40-YFP 108.2

C

CHO-CMV-YFP 1.1

CHO-CMV-YFP 1.2

D

CHO-CMV-YFP 108.1
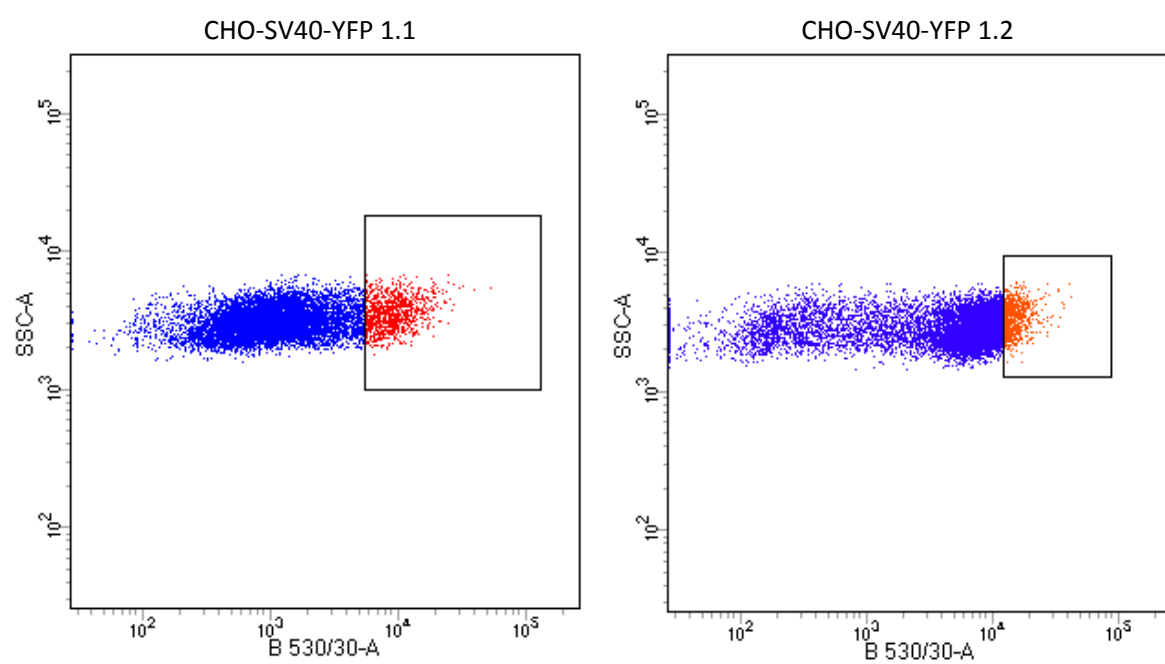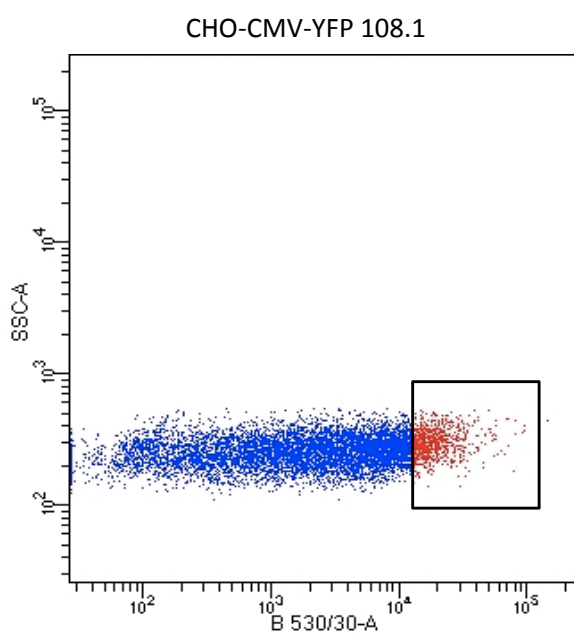
CHO-CMV-YFP 108.2

Figure 4.8. YFP expression in CHO-SV40-YFP and CHO-CMV-YFP polyclonal and top 10% cell lines.

A 1 ml sample of culture was collected and YFP fluorescence intensity was measured using the BD Accuri™ C6 flow cytometer, as described in Section 2.6.2.3. The YFP was excited using a 488nm excitation laser and fluorescence data was acquired using a 530/30nm bandpass filter and analysed using the BD Accuri™ C6 software. The auto-fluorescence of the CHO-FRT mini-pool was measured in every experiment as a control for non-expressing cells. **(A** – CHO-SV40-YFP polyclonal; **B**– CHO-SV40-YFP Top 10%; **C** – CHO-CMV-YFP polyclonal**; D**– CHO-CMV-YFP Top 10%); (☐ CHO-FRT; ☐ Polyclonal; ☐ Top 10%)
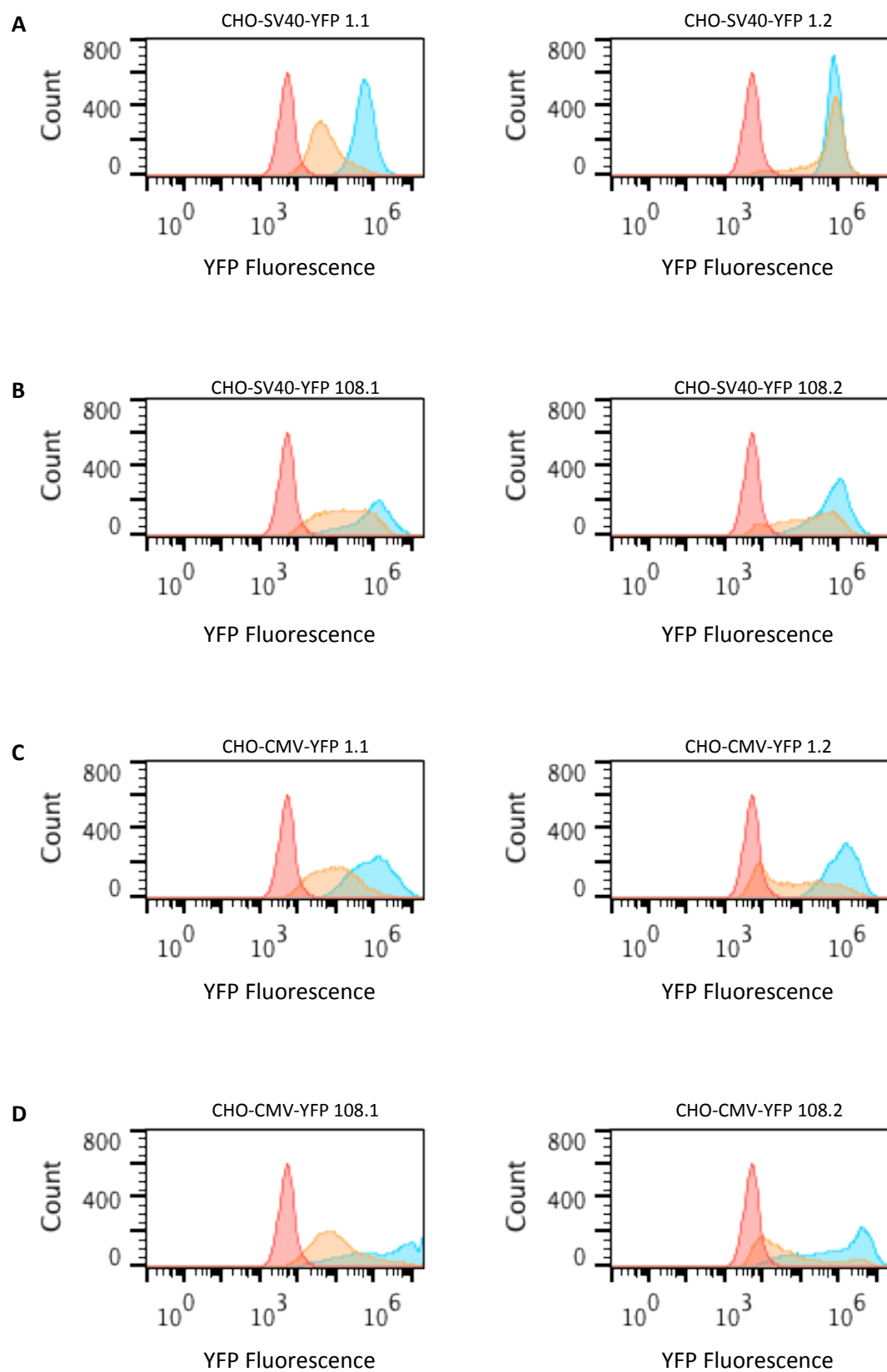
4.2.2. YFP expression in CHO-SV40-YFP and CHO-CMV-YFP polyclonal and Top 10% cell lines

Figure 4.9 shows the quantification of SV40-YFP and CMV-YFP expression in both polyclonal and top 10% cell lines. Similar to what was observed for SV40-GFP and CMV-GFP expression in Section 4.1.2, the expression of SV40-YFP and CMV-YFP was similar in both CHO-FRT 1 and CHO-FRT 108 mini-pools after polyclonal selection. As discussed in Section 4.1.2, such results were not expected as SV40 and CMV promoters to have different strengths and site-directed integration was used to incorporate the promoter/fluorescent protein cassette.

The variation in efficiency of the Flp-In system for CHO-FRT mini-pools, discussed in Section 4.2.1, could in part explain these observations, however it would not explain the similarity in YFP fluorescence for SV40-YFP and CMV-YFP in both CHO-FRT mini-pools. One possible explanation for the similarity in expression after polyclonal selection could be the result of early promoter silencing during the selection process. The period of selection of the polyclonal cell lines was 4 weeks and the initial cell density of transfectants was low (1.6-0.8x10^5 cell/ml) (Section 2.6.1.2). A low initial cell density means that for the cells to reach the optimum cell density to be cultured in suspension (approximately 1.0x10^6 cell/ml), they needed to grow for a relatively high number of generations during which the promoter could be silenced.

Early gene silencing was described by Spencer et al (2015) when GFP was expressed under the control of a CMV promoter. The authors observed that gene silencing occurred as early as 17 generations after clonal selection (Spencer et al. 2015). The authors integrated a single-copy of the CMV-GFP construct using a lentiviral system and selected 55 high-GFP-producing clones by FACS but after the clonal expansion process the majority of the clones had significantly less GFP expression, with 3 of the clones showing no GFP expression (Spencer et al. 2015). The authors concluded that the loss of GFP expression was a result of histone modifications rather than DNA methylation. The degree of promoter methylation was quantified in two of the clones that completely lost GFP expression and no promoter methylation was observed (Spencer et al. 2015).

Besides possible gene silencing occurred during the selection methods, is also possible that the similar recombinant expression of the polyclonal cell lines is due to the fact that these cell lines achieved a steady state of expression during selction process (Pilbrough et al 2009). This basal steady state was describe by Pilbrough

and colleagues in polyclonal cell lines and was mainly attributed for non-heritable genotypic fluctuations (Pilbrough et al 2009).

Unsurprisingly, the top 10% selection generated cell lines that exhibited a significant increase of YFP expression. This increment appears to vary depending on the nature of promoter used (Figure 4.9). For the SV40-YFP expression, the increment observed was 2-fold for CHO-SV40-YFP 1 and 3.5-fold for CHO-SV40-YFP 108 while for the CMV-YFP expression the increment observed was 6-fold in CHO-CMV-YFP 1 and 15-fold in CHO-CMV-YFP 108. In contrast to the finding with the polyclonal cell lines, with the top 10% cell lines the CMV-YFP expression was significantly greater than SV40-YFP (Figure 4.9).

Although the results described in Figure 4.9 appear to reflect the expected differences in promoter strength previously described for these two promoters (Boshart et al. 1985; Foecking and Hofstetter 1986; Ho et al. 2015; Zarrin et al. 1999), is also possible that these observations could not only be due to differences in strength but also an due to state of each cell selected by FACS sorting selection process. When using FACS as a method for selecting high expressing cells, Pilbrough and colleagues observed that this method not only selected high expressing cells but also selected cells that were above steady state (Pilbrough et al 2009). The authors described these cells as cells that have a stochastic gene expression that can lead to mountainous fluctuations of gene expression and consequently masking their natural expression profile by appearing to have higher levels of gene expression (Pilbrough et al 2009). Based on this observations, it is possible that the instead of selecting cells that are naturally high expressing cells with possible similar number of YFP gene copies (as previously assumed in Section 4.2), this methods also selected lower expressing cells that were perturbed above steady state at the time of selection masking the really difference between the strength of the promoters.

Figure 4.9. Expression of YFP under the influence of both SV40 and CMV promoters in polyclonal and Top 10% cell lines.

The median YFP fluorescence was calculated using the BD Accuri™ C6 software based on the cell line histograms of Figure 4.8. Data are mean ± SD (n=4). The significance of YFP expression between polyclonal and top 10% cell lines was determined using an independent two-tailored t-test (* - $p < 0.05$, ** - $p < 0.01$). The significance of the YFP expression regulated by SV40 and CMV promoter in the same CHO-FRT mini-pools was determined using an independent two-tailed t-test (# - $p < 0.01$, ## - $p < 0.01$).(■ SV40-YFP polyclonal; ■ SV40-YFP top 10%; ■ CMV-YFP polyclonal; ■ CMV-YFP top 10%)
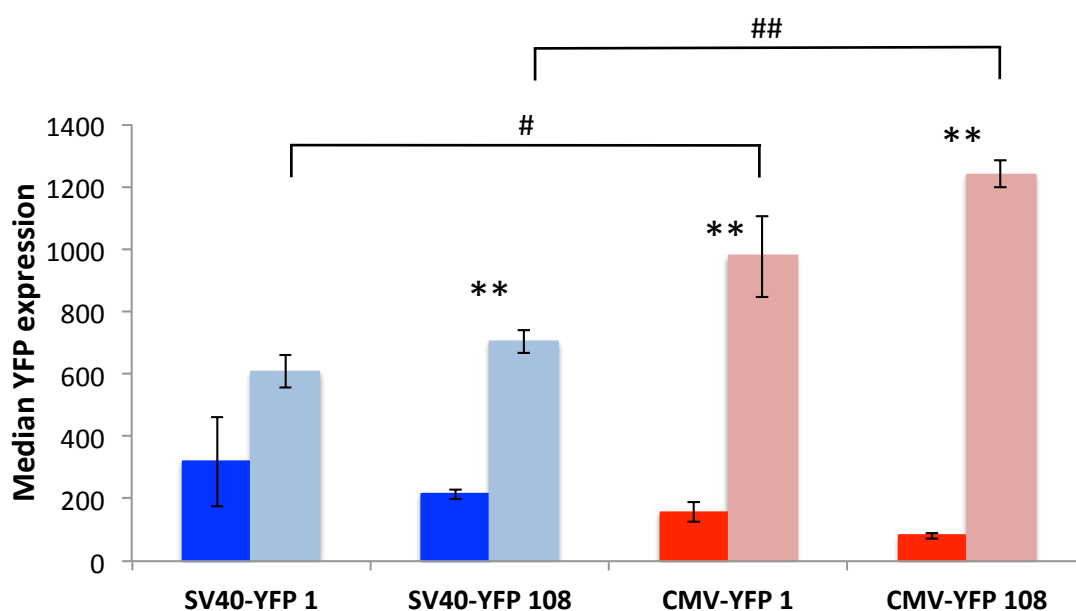
4.2.3. Characterisation of YFP expressing cell populations in CHO-SV40-YFP and CHO-CMV-YFP polyclonal and Top 10% cell lines during long-term culture

This Section analyses expression profiles of early and late stages of both polyclonal (Figure 4.10) and top 10% (Figure 4.11) CHO-SV40-YFP and CHO-CMV-YFP cell lines. Similar to the analyses of the long-term culture of CHO-FRT and GFP expressing cell lines, in this analysis the terms "early stage" and "late stage" will be used to define the cell lines at the start of long term culture (day 0) and after 32 days of culture, respectively.

Only CHO-SV40-YFP 1.2 (Figure 4.10-A), CHO-SV40 108.1 (Figure 4.10-B) and CHO-CMV-YFP 1.1 (Figure 4.10-C) polyclonal cell lines revealed changes in cell population during long-term culture with the appearance of non-expressing cells. At late stages, the non-expressing cells in CHO-SV40 108.1 and CHO-CMV-YFP 1.1 polyclonal cell lines formed the dominant cell population of the cell lines, while in CHO-SV40-YFP 1.2 polyclonal cell line, the dominant cell population was from the high YFP-expressing cells. This was evidenced by the presence of the same distinctive peak of high expressing cells in both early and late stage cells.

In contrast to observations for polyclonal cell lines, in the top 10% cell lines, the change in histogram profiles during the long-term culture were promoter-dependent. For instance, for the expression of SV40-YFP, all the top 10% maintained their histogram profile with the exception of CHO-SV40-YFP 108.2 top 10% which revealed a decrease of the main high YFP-expressing population, leading to an even distribution of medium-expressing and high- expressing cells (Figure 4.11-B).

On the other hand, all CHO-CMV-YFP top 10 % cell lines revealed an unstable cell population with almost all the cell lines showing a shift of main cell population from high-expressing cells to low YFP-expressing and non-expressing cells at late stages (Figure 4.11- C, D). Only CHO-CMV-YFP 1.2 top 10% cell line revealed a lower degree of cell population instability, as despite an appearance of low-expressing cells at later stages, the main cell population maintained a relatively high expressing profile.

No studies have been published describing the variation of cell populations during long-term culture using the SV40 promoter. However, for the CMV promoter the variations in cell population have been described previously in CHO and HEK293 cell lines (Du et al. 2013; Eyquem et al. 2013). When analysing the heterogeneity and instability of mAb-expressing CHO cells, Du et al (2013) observed shifts of

populations from high-expressing to non-expressing cells in histogram profiles of monoclonal cell lines (Du et al. 2013). The authors observed that high-expressing populations could be partially maintained after two rounds of FACS selection for the top 10% cell line. Their observations made for the top 10% cell population differ from the ones observed in Figure 4.11. The fact that Du et al (2013) did two rounds of FACS selection for the top 10% cell population from monoclonal cell lines could explain the differences observed. Furthermore, the observation of stable population could be masked by the short period of which the cells are maintained in culture as the author's evaluation was restricted only 20 days after FACS selection (Du et al. 2013).

Similar changes cell line population during long-term culture were also observed for site-integrated genes in HEK293 cells (Eyquem et al. 2013). Using meganucleases to integrate a single-copy of GFP to three specific genomic locations, Eyquem and colleagues observed that when just one copy of the GFP was inserted, the histogram profile of the cells was maintained during a long-term culture (Eyquem et al. 2013). However, insertion of more than one gene copy into a single locus or multiple loci could either lead to the selection of cell lines with a stable cell population profiles, or shift the cell population towards a non-expressing profile (Eyquem et al. 2013). The authors did not have an explanation for these observations but suggested similar studies should be performed in different cell lines in order to understand which factors might account for, and predict, stable and unstable cell line profiles (Eyquem et al. 2013).

Based on the observations made in this Section for the top 10% cell lines, is possible to suggest that the variation of cell line populations during a long-term culture appears to be promoter- dependent. However due to the mix population nature of the top 10% sub-population (discussed in Section 4.2.2), the variation of the cell line population observed during long-term culture is influenced by the these mix population nature that could mask this apparent promoter dependency.

Figure 4.10. Flow cytometry histograms of CHO-SV40-YFP and CHO-CMV-YFP polyclonal lines at early and late cell line stage.
CHO-SV40-YFP 1 and CHO-SV40-YFP 108 and CHO-CMV-YFP 1 and CHO-CMV-YFP 108 polyclonal cell lines were continuously cultured for 32 days as described in Section 2.6.2.4. A 1 ml culture sample at early and late stage of each cell line was collected and YFP fluorescence intensity was measured using the BD Accuri™ C6 flow cytometer, as described in Section 2.6.2.3. The YFP was excited using a 488nm excitation laser and fluorescence data was acquired using a 530/30nm bandpass filter and analysed using the BD Accuri™ C6 software. The auto-fluorescence of the CHO-FRT mini-pool was measured in every experiment as a control for non-expressing cells. (**A** – CHO-SV40-YFP 1 polyclonal**; B**– CHO-SV40-YFP 108 polyclonal; **C** – CHO-CMV-YFP 1 Polyclonal; **D**– CHO-CMV-YFP 108 polyclonal). (☐ CHO-FRT; ☐Early stage; ☐ Late stage)
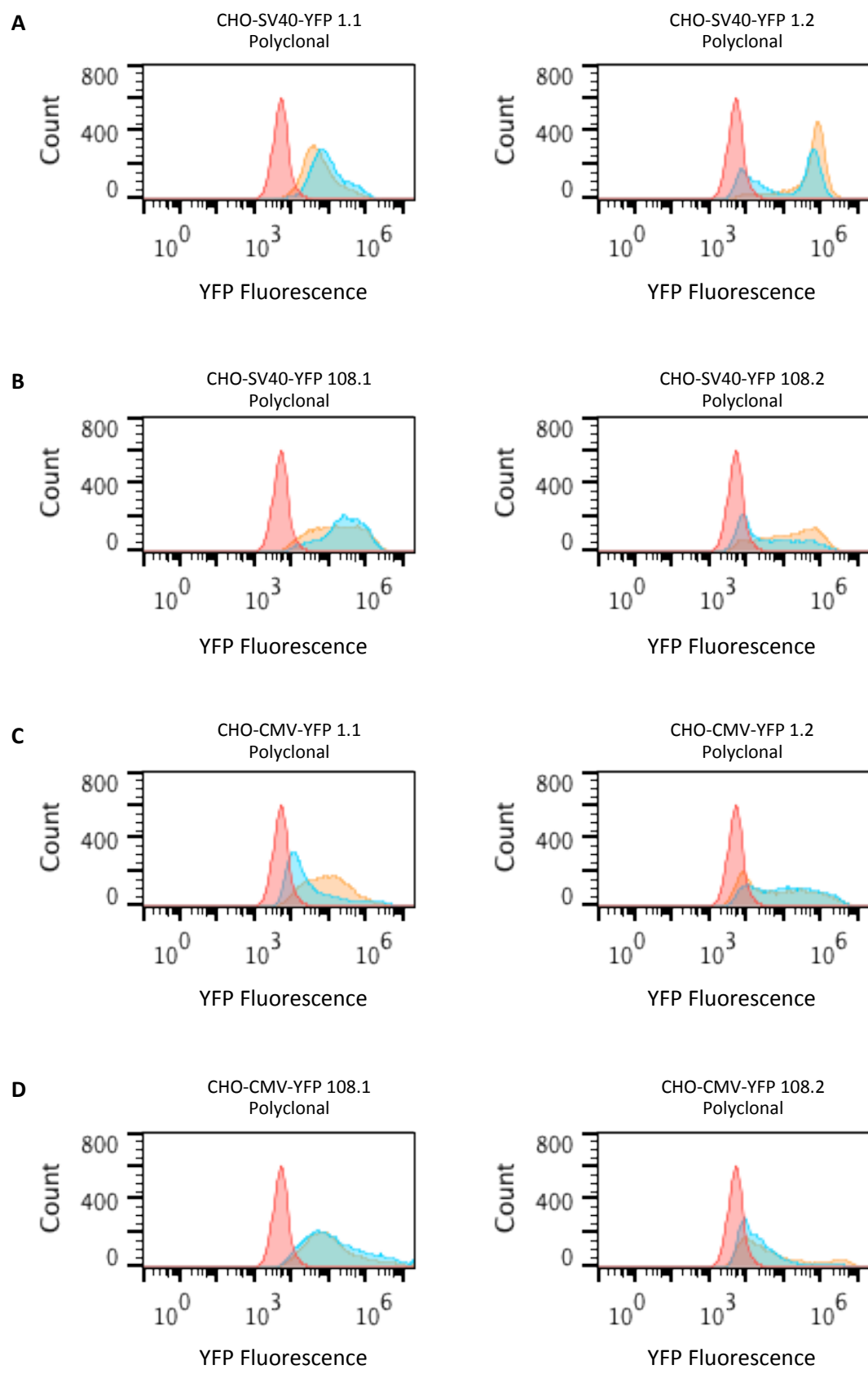
Figure 4.11. Flow cytometry histograms of CHO-SV40-YFP and CHO-CMV-YFP top 10% cell lines at early and late cell line stages.

CHO-SV40-YFP 1 and 108 and CHO-CMV-YFP 1 and 108 top 10% cell lines were continuously cultured for 32 days as described in Section 2.6.2.4. A 1 ml culture sample at early and late stages of each cell lines was collected and YFP fluorescence intensity was measured using the BD Accuri™ C6 flow cytometer, as described in Section 2.6.2.3. The YFP was excited using a 488nm excitation laser and fluorescence data was acquired using a 530/30nm bandpass filter and analysed using the BD Accuri™ C6 software. The auto-fluorescence of the CHO-FRT mini-pool was measured in every experiment as a control for non-expressing cells. (**A**– CHO-SV40-YFP 1 Top 10%**; B**– CHO-SV40-YFP 108 Top 10%; **C**– CHO-CMV-YFP 1 Top 10%; **D**– CHO-CMV-YFP 108 Top 10%).(■ CHO-FRT; ■ Early stage; ■ Late stage)

A. CHO-SV40-YFP 1.1 Top 10%, CHO-SV40-YFP 1.2 Top 10%

B. CHO-SV40-YFP 108.1 Top 10%, CHO-SV40-YFP 108.2 Top 10%

C. CHO-CMV-YFP 1.1 Top 10%, CHO-CMV-YFP 1.2 Top 10%

D. CHO-CMV-YFP 108.1 Top 10%, CHO-CMV-YFP 108.2 Top 10%

169
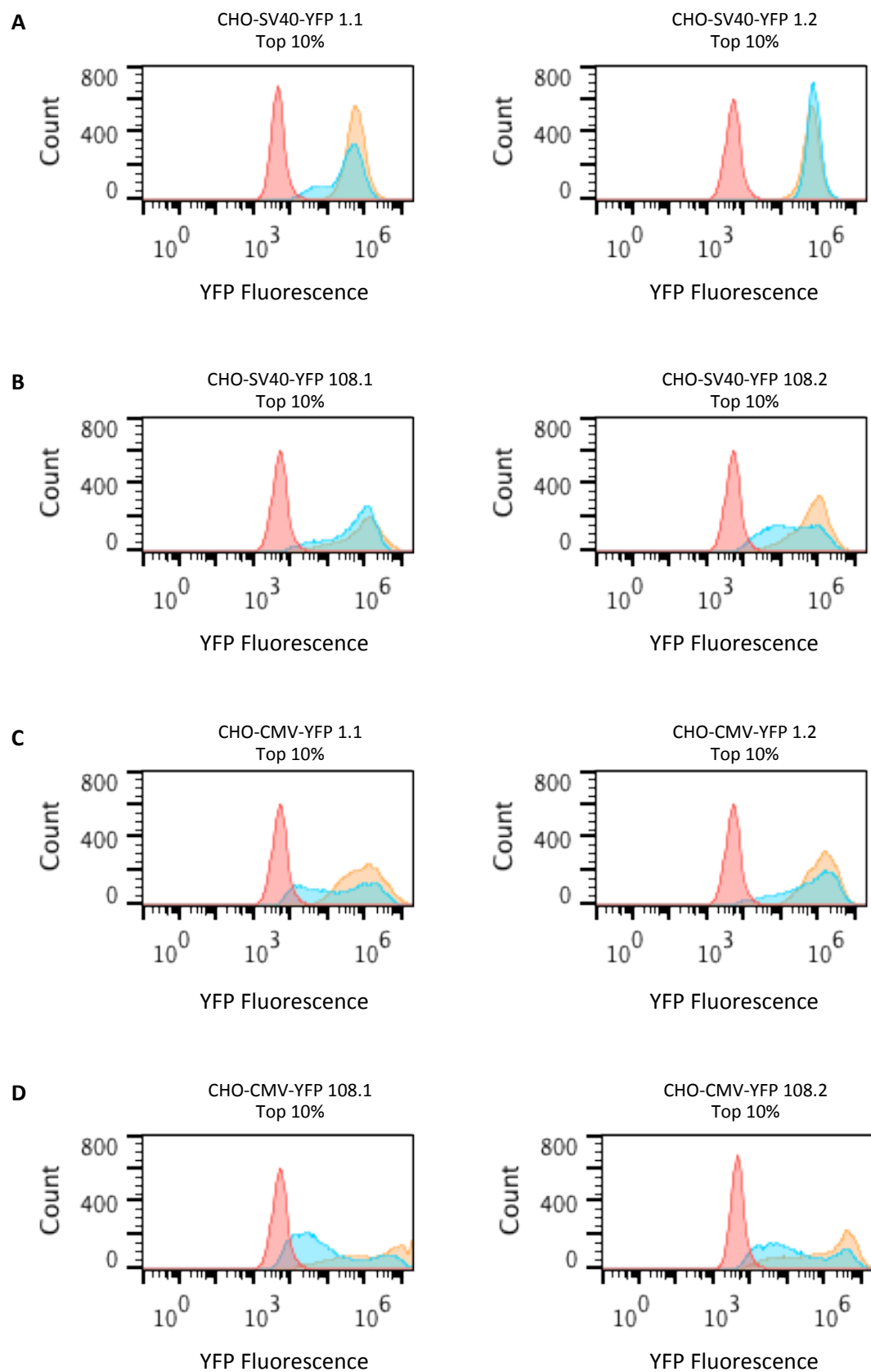
4.2.4. Stability of YFP expression in CHO-SV40-YFP and CHO-CMV-YFP polyclonal and Top 10% cell lines

Figure 4.12 shows YFP expression in CHO-SV40-YFP and CHO-CMV-YFP polyclonal (Figure 4.12-A) and top 10% (Figure 4.12-C) cell lines during long-term culture and compares SV40-YFP and CMV-YFP expression at early and late stages of culture for polyclonal (Figure 4.12-B) and top 10% (Figure 4.12-D) cell lines.

Similar to what was described for SV40-GFP and CMV-GFP in (Section 4.1.2), the expression of SV40-YFP and CMV-YFP in the polyclonal cell lines was maintained during long-term culture (Figure 4.12-A,C). Based on the maintenance of expression in polyclonal cell lines alone, both promoters could be characterised as stable in CHO cells (Figure4.12-B, D). As discussed in Section 4.1.2, these observations contradict what has been previously described for both promoters, especially for CMV for which loss of CMV-driven recombinant expression in CHO cells has been widely reported (Du et al. 2013; Kim et al. 2011; Mariati et al. 2014a; Osterlehner et al. 2011; Spencer et al. 2015; Williams et al. 2005).

The non-expected results of GFP and YFP expression in polyclonal cell lines (Figures 4.6, 4.9, and 4.12) permit to determine that these polyclonal cell lines were not appropriate to study the activity and stability of these promoters. On the other hand, as observed in Figure 4.9 for the comparison between the activity of SV40 and CMV, Figure 4.12 appear to show that the top 10% expressing population also reflects the expected differences in promoter stability.

Based on Bailey's definition of stability (Bailey et al. 2012), in the top 10% sub-population, SV40-YFP expression appears to be slightly unstable in CHO-FRT cells while CMV-YFP appears to have a considerable unstable profile. SV40-YFP expression was considered slightly unstable because although only approximately 40-45% of SV40-YFP expression was loss in both CHO-FRT mini-pools used (Figure 4.12-B), these value relatively closed to threshold of Bailey's definition of stability (Bailey et al. 2012). In contrast, CMV-YFP expression was considered unstable in both CHO-CMV-YFP 1 and 108 top 10% cell lines as approximately 50% and 80%, respectively, of the initial YFP expression was lost during the culturing period (Figure 4.12-D).

These observations in relation of the stability of these promoters are slightly different to those observed by Ho et al (2015) (Ho et al. 2015). In their study, Ho and colleagues described SV40 promoter as stable with an average loss of expression of 22% while CMV promoter as unstable with an average loss of expression of 80% (Ho

et al. 2015). Although the retention of expression observed in this study is similar to the observations of Ho et al (2015) for the CMV promoter, they differ for the SV40 promoter. The values presented by Ho and colleagues for each promoter were calculated by averaging the retention rates of 18 clonal cell lines while in Figure 4.12-B they are cell line specific. Despite described as stable promoter, Ho et al (2015) also observed instability when SV40 promoter was used, with 7 of the 18 cell lines using SV40 promoter revealing unstable characteristics (Ho et al. 2015).

As YFP expression was monitored at several time points during this study it was possible to evaluate the change in SV40-YFP and CMV-YFP expression during long-term culture (Figure 4.12-A,C). Figure 4.12-A and Figure 4.12-C reveal that SV40 and CMV appear to have different patterns of loss of expression during long-term culture. When SV40 promoter was used, the drop of YFP expression during long-term culture appears to happen constantly during the culturing period as (Figure 4.12-A). On the other hand in both CMV-YFP top 10% cell lines the drop of CMV-YFP expression observed was logarithmic with the majority of the loss of CMV-YFP expression occurring during the first 4 days of the culture period (Figure 4.12-C).

A similar pattern for the loss of recombinant expression for CMV in CHO cells was also observed by Pilbrough et al (2009). The authors co-expressed mAbs with fluorescent protein to monitor the cell line productivity and for clonal selection using FACS (Pilbrough et al 2009). Similar to the observation in Figure 4.12-C, Pilbrough and colleagues observed the loss of productivity of each cell line had a logarithmic behaviour with the majority of the loss occurring in the first 20 days after selection and that cells need between 30 to 50 days after selection to reach a steady expression state (Pilbrough et al. 2009). The authors related this logarithmic loss of expression not to the gene silencing but to the stochastic gene expression of these cells. Pilbrough and colleagues suggested that when cells are selected for high expressing cells, is possible that some of the cells selected as high expressing cells could be actually just average expressing cells that at the time of selection were at above steady state masking their natural expressing nature (Pilbrough et al. 2009).

Although the expression of SV40-YFP and CMV-YFP appear to be unstable in CHO-FRT 1 and 108 mini-pools, is not possible to conclude that these unstable profiles are purely due to the nature of the promoter. Due to the FACS selection method used to select these high expressing sub-populations and the similarity of the results observed by Pilbrough et al (2009), is likely that the differences in instability described in Figure 4.12 are mainly due to the stochastic variation of gene expression, as reported by Pilbrough et al (2009). Nevertheless, based on these

observations is possible to suggest that the fact CMV is stronger promoter than SV40 (Boshart et al. 1985; Foecking and Hofstetter 1986; Ho et al 2015; Xu et al. 2001; Zarrin et al. 1999), possibly increases the number of events that influence the stochastic gene expression. These increment of number of events could exacerbated the variation of gene expression when CMV is used, possibly explaining the differences of the degree of instability the behaviour of the YFP expression observed for both promoters during the long-term.

Figure 4.12. YFP expression of CHO-SV40-YFP and CHO-CMV-YFP polyclonal and top10% cell lines during long-term culture.

CHO-SV40-YFP and CHO-CMV-YFP polyclonal and top 10% cell lines were continuously cultured for 32 days as described in Section 2.6.2.4. YFP fluorescence intensity of each SV40-YFP (**A**) and CMV-YFP (**C**) cell lines was measured at days 4, 11, 18, 25 and 32 of culture using a BD Accuri™ C6 flow cytometer (Section 2.6.2.3) and the median of YFP fluorescence was calculated using the BD Accuri™ C6 software. To characterise the stability YFP expression of each SV40-YFP (**B**) and CMV-YFP (**D**) cell line, the median YFP expression at early and late stages was compared. Data are mean ± SD (n=4). The difference in YFP expression between early and late stages was determined using an independent two-tailored t-test (* - $p < 0.05$, ** - $p < 0.01$), R2/R2 – correlation coefficient (linear), R2/R2 – correlation coefficient (logarithmic) (━■━ SV40-YFP 1 polyclonal ; ━▲ SV40-YFP 1 top 10%; ━◆━SV40-YFP 108 polyclonal; ━■━ SV40-YFP 108 top 10%; ━◆━ CMV-YFP 1 polyclonal; ━■━CMV-YFP 1 top 10%; ━■━ CMV-YFP 108 polyclonal; ━▲━CMV-YFP 108 top 10%; ■ SV40-YFP early stage; ■SV40-YFP late stage; ■ CMV-YFP early stage; ■CMV-YFP late stage)

174
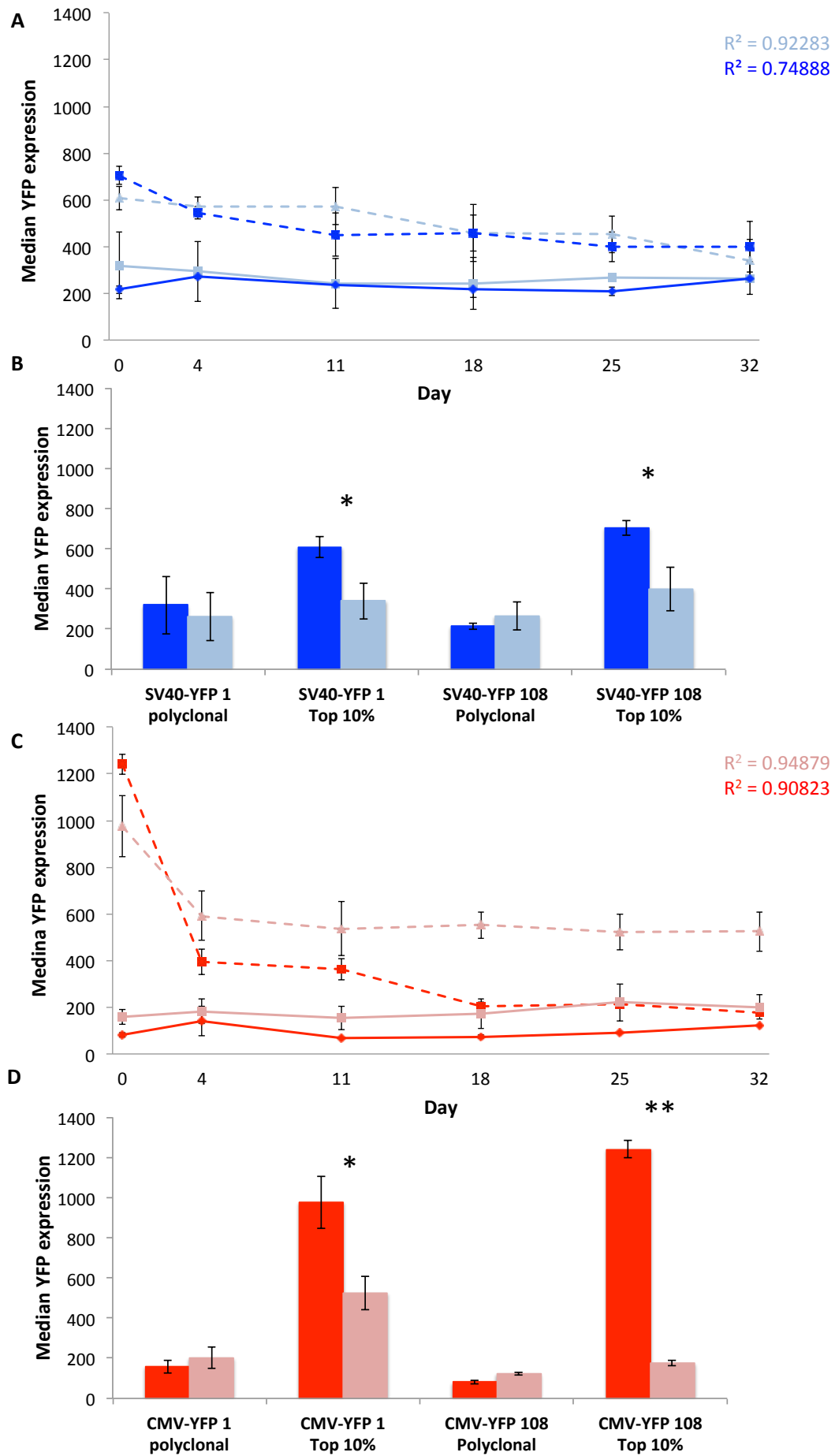
4.3 Discussion

The main goal of the work reported in this Chapter was to develop a methodology that combined the use of site-directed integration methods and fluorescent proteins to study the influence of recombinant promoters on the stability of recombinant protein expression in CHO cell lines. Furthermore, this Chapter also aimed to endorse this methodology as a possible alternative for this type of studies by using two of the most well-characterised recombinant promoters, SV40 and CMV.

SV40 and CMV promoters are two of the most well characterised recombinant promoters with strengths that have been previously characterised (Boshart et al. 1985; Foecking and Hofstetter 1986; Ho et al. 2015; Xu et al. 2001; Zarrin et al. 1999). As these previous studies used either transient or stable transfections, the results observed neglect the influence of the different genomic environment in which the both promoters are inserted. The methodology described in this chapter attempted to overcome the influence of the genomic by using FRT-site directed integration methods to target the insertion of the recombinant gene under the influence of SV40 or CMV promoters to the same genomic location.

The results presented in this chapter can not demonstrate with 100% certainty that this methodology can be used to characterised different promoters due to the nature of the CHO-FRT mini-pools used. Despite FRT-site directed integration is used, the CHO-FRT mini-pools generated in Chapter 3 contain multiple FRT sequences that potentially offer targets for integration of recombinant promoter/fluorescent protein expression cassette (Chapter 3.8). Furthermore the mini-pool nature of CHO-FRT mini-pools (Chapter 3.8) increasing the factors that could contribute for the existence of heterogeneous populations. These two non-ideal characteristics introduce some caveats that can undermine the certainty of the similarity of the genomic location where both promoters were inserted, reducing the advantages of the use of site-directed integration.

These non-ideal characteristics can be demonstrate by the analysis of the expression of the GFP and YFP in polyclonal cell lines. In both cases, differently of what was expected, no considerably differences were observed between both promoters. The inefficiency of the Flp-FRT site directed integration method used coupled with the multiple FRT sites does not guarantees that the after selection all the FRT sites were successfully integrated with recombinant promoter-gene combo. This critical factor could help to explain the lack of consistency and the heterogeneity of cell population obtained in polyclonal cell lines.

Although the existence of multiple FRT sequences in CHO-FRT mini-pools was identified as a potential caveat for the correct characterisation of recombinant promoter (Section 3.8), it can bring an advantage, in comparison to the use of a single cell line with a single FRT site (Tornøe et al 2002), if full integration of the recombinant promoter/fluorescent protein cassette is guaranteed. The site-directed integration in multiple cell lines to target the recombinant promoter towards the same multiple genomic locations will allow a comparison of the results of the same genomic environment but will dilute the influence of the environment of a specific locus.

Assuming that due to a limited number of FRT targets, the cells with higher YFP expression would have greater numbers of YFP integrations for each site, (Section 4.2) the selection of top 10% expressing population was performed in an attempt to select fully or high YFP integrated cell lines. As Figure 4.8 reveals, this second round of selection was successful on diminishing the cell line heterogeneity and increasing the overall cell line productivity. As no YFP gene copy number assessment was performed after site-directed integration, it is not possible to evaluate if the second round of selection was successful in guaranteeing that all the FRT sites present in CHO-FRT sites were integrated with promoter/YFP cassettes. Despite the expected differences in YFP expression observed between cell lines and promoters it still possible to that due to the multiple number of FRT sites, still exists heterogeneity of number of YFP copies in cell lines studied. Because of this possible herogeinity, although Figure 4.9 appear to reveals that YFP expression in this top 10% cell lines reflect the expected differences in promoter strength (Ho et al. 2015), is not possible to guarantee that this method can prove with 100% certainty that CMV is stronger than SV40 promoters in the same genomic environment.

Despite still non-ideal, the results of the top 10% cell lines lead to interesting observations that show that this methodology allows to study the behaviour of both recombinant promoters during long-term culture. The data presented in this study shows that both promoters have distinct influence on the maintenance of the long-term recombinant expression (Section 4.2.4). The results presented in this Chapter, suggest that SV40 promoter is a slightly unstable (Figure 4.12-A, B) while in similar conditions, CMV appears to be highly unstable (Figure 4.12-C, D). The loss of expression for SV40 promoter happens gradually during the long-term culture with a linear correlation (Figure 4.12-A). On the other hand, for CMV promoter, this loss of

expression has a logarithmic behaviour with most of the expression been lost in the first third of the long-term culture period (Figure 4.12-C).

The instability of CMV promoter in recombinant expression in CHO cells has been extensively reported either for gene reporters, such as GFP, or expression of complex proteins, such as antibodies, (Du et al. 2013; Kim et al. 2011; Mariati et al. 2014b; Osterlehner et al. 2011; Spencer et al. 2015; Williams et al. 2005; Yang et al. 2010). Most of these studies identified DNA methylation as the main cause for recombinant gene silencing under the control of a CMV promoter (Kim et al. 2011; Mariati et al. 2014b; Osterlehner et al. 2011; Williams et al. 2005; Yang et al. 2010. However histone modifications (Spencer et al. 2015) and increase of non-expressing cells in the culture population (Du et al 2013) have also been reported as causes of loss of recombinant expression.

Strangely, although the SV40 promoter has been extensively characterised and has been described as a stable promoter (Ho and Yang 2014), there are limited studies that report on the influence of this promoter on long-term recombinant expression. Just recently, the influence of the SV40 promoter on the long-term expression in CHO was reported for the first time in 2015 (Ho et al. 2015). Ho and colleagues observed that recombinant expression regulated by SV40 was maintained during long-term culture in the majority of the cell lines studied (Ho et al. 2015). Despite characterise it as stable promoter, Ho et al (2015) also observed instability for SV40 promoter with 39% of the cell lines used in his study revealing unstable characteristics.

Due to the similarity between the results described in Figure 4.12 for CMV promoter and the observation of the instability reported by Pilbrough et al (2009), it is possible that the main cause of instability is the stochastic gene expression fluctuation discussed in Section 4.2.4. If the stochastic gene fluctuations is considered the main cause for the instability observed in study, is possible then, to suggest that because CMV is stronger promoter than SV40 (Boshart et al. 1985; Foecking and Hofstetter 1986; Ho et al 2015; Xu et al. 2001; Zarrin et al. 1999) may lead to and increase of the number of events that influence the stochastic gene expression in CMV promoter (Pilbrough et al 2009). A higher number of events could not only explaining the differences of the degree of instability for both promoters as also differences of long-term YFP expression behaviour observed for both promoters (Pilbrough et al. 2009).

Nonetheless, it is also possible that different molecular characteristics of both promoters would influence the differences in instability observed in this project. For

instance, the content of CpG dinucleotide in each promoter (as potential targets for methylation) is considerably different with 9 GpG dinucleotides been reported for the SV40 promoter (Bryans et al. 1992) and 33 for the CMV promoter (Osterlehner et al. 2011). This could be evaluated by calculating the methylation pattern of each promoter by using techniques such as methyl-assisted qRT-PCR (Yang et al. 2010) or bisulfite sequencing (Mariati et al. 2014b; Osterlehner et al. 2011). Nevertheless, it is also possible that histone modifications and gene silencing could also have a strong impact in these observations (Kim et al. 2011; Spencer et al. 2015) .

During this project variability and heterogeneity was a recurring finding for different cell populations (either with GFP or YFP) after the polyclonal selected cells that have undergone site-directed integration (Figure 4.2, 4.3 and 4.8). Besides the inefficiency of the site-directed integration and possible early gene silencing during the polyclonal selection process that were suggested as causes for this observations (Sections 4.1.2, 4.2.2 and 4.2.4) is not possible to neglect the natural capacity of CHO cells for undergo phenotypic drift and generate heterogeneity even in clonally-derived cell lines.

When studying the functional heterogeneity in CHO cell populations, Davies et al (2013) observed that within a cell line, there was a considerably high heterogeneous population of cells with different growth properties and productivity (Davies et al 2013). The author isolated 100 single clones from a commonly-used parental CHO cell host, CHOK1SV, and observed that all the resulting clonally-derived cell lines had different growth properties (Davies et al. 2013). Furthermore, the transient and stable productivity of this clonally-derived cell lines was different when compared to the parental cell line (Davies et al. 2013).

Based on the results this Chapter for the SV40 and CMV promoters and the discussion above, its is possible to propose that, although not ideal, the method developed in this Chapter can be used as an alternative protocol to study the expression of recombinant promoters during long-term culture.

Chapter 5. The use of MS2-RNA tag to measure the transcriptional activity of recombinant promoters in CHO-FRT cells

The main goal of this Chapter was the implementation of a methodology to characterise the transcriptional activity of recombinant promoters in CHO-FRT mini-pools, by quantification of the transcriptional output of recombinant reporter genes in living single cells.

As Chapter 4 described, SV40 and CMV promoters exhibited different strengths and stability characteristics. These characterisations were made on related cell lines that expressed a small, non-secreted and non-glycosylated protein and it is possible that the differences observed at the protein level would be due to differences of the transcriptional activity of each promoters. Quantification of the transcriptional output from a recombinant promoter directly would confirm interpretations about the strength and stability of recombinant promoters whilst increasing the robustness of the methodology for studying genomic environment.

The methodological approach in this Chapter to quantify the transcriptional output is based on combination of Flp-recombinanse site-directed and MS2-tag/MCP-GFP methods described in HEK 293 cells by Yunger *et a*l (2013) (Section 1.3.4.3). In Yunger et al (2013) a sequence of tandem bacteriophage sequence repeats (MS2) were cloned downstream of the recombinant gene and then underwent site-directed integration to a pre-selected genomic location (Yunger et al 2013). The use of this strategy to quantify transcription in living CHO cells is a novel approach for the study of the transcriptional mechanism of recombinant genes in these cells.

In this Chapter the terminology STRB will be used to refer to mStrawberry fluorescent protein. As in Chapter 4, the terminology "SV40-STRB" or "CMV-STRB" will be used to distinguish between the promoter/fluorescent protein constructs used. Firstly, this Chapter describes the generation of CHO-SV40-STRB and CHO-CMV-STRB cell lines to be used and then describes the results obtained when the methodology developed by Yunger et al (2013) is used in CHO-SV40-STRB and CHO-SV40-STRB cells. Figure 5.1, summarises the approach used in this Chapter to characterise the transcriptional activity of SV40 and CMV promoters using CHO-FRT mini-pools.

Figure 5.1. Overview of the strategy to characterise the transcriptional activity of SV40 and CMV promoters.

This Figure illustrates the steps used in characterisation of the transcriptional activity of SV40 and CMV promoters using the CHO-FRT mini-pools.
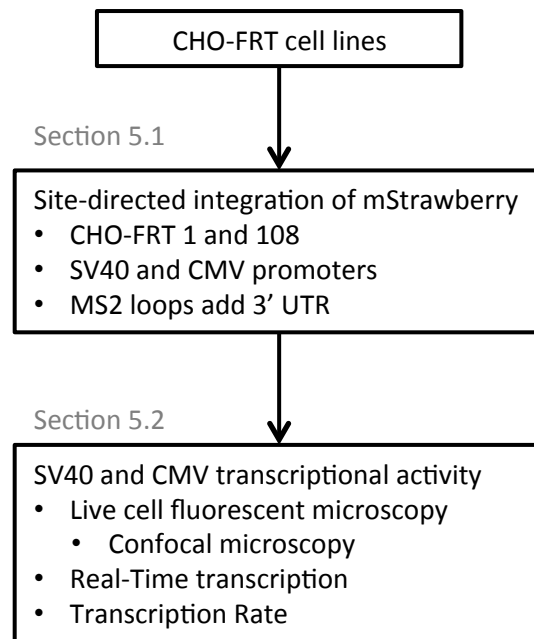
5.1 Analysis of mStrawberry expression in CHO-SV40-STRB and CHO-CMV-STRB polyclonal cell lines

As described for CHO-SV40-YFP and CHO-CMV-YFP polyclonal cell lines analysed in Chapter 4, CHO-SV40-STRB and CHO-CMV-STRB polyclonal cell lines were generated by site-directed integration of the pS/FRT/MS2-mStrawberry and pcDNA5/FRT/MS2-mStrawberry vectors, respectively, into the FRT sites within the genome of CHO-FRT 1 and CHO-FRT 108 mini-pools. The pS/FRT/MS2 and pcDNA5/FRT/MS2 vectors (Section 2.3), encode a 22x MS2-loop sequence between the 3' end of the recombinant gene and the polyadenylation sequence encoded by these vectors.

Figure 5.2 shows flow cytometry histograms for CHO-SV40-STRB and CHO-CMV-STRB polyclonal cell lines after cell line selection (Section 2.6.1). The mStrawberry fluorescence was measured in a BD Accuri™ C6 because the comparison of the histograms obtained for the CHO-SV40-STRB and CHO-CMV-STRB with the histogram of a non-mStrawberry expressing cell line (CHO-FRT 108 mini-pool) presented a rapid and efficient method to measure a relative mStrawberry expression in CHO-SV40-STRB and CHO-CMV-STRB and conclude that these cell lines successfully express mStrawberry. However, despite allowing identifying positive to negative mStrawberry expressing cell lines, due to a lack of sensitivity of the BD Accuri™ to accurately quantify the fluorescence of mStrawberry, the histograms presented in Figure 5.2 do not correctly represent the mStrawberry expressing cell distribution in CHO-SV40-STRB and CHO-CMV-STRB polyclonal cell lines

The BD Accuri™ C6 flow cytometer has a limited set of excitation lasers and filter sets. Because of that, the cell lines histograms in Figure 5.2 were obtained using a 488 nm laser as excitation source and a 584/40 nm filter. mStrawberry excitation and emission spectrum reveal that this fluorescent protein has a maximum excitation at 574 nm and a maximum emission at 596 nm (Shaner et al. 2004). Based on these spectrums, Piatkevich and Verkhusha (2011) calculate that a 488 nm excitation laser has a 12% excitation efficiency and an effective brightness of 3 (Piatkevich and Verkhusha 2011). Comparing with the 568 nm laser, suggested by the authors as ideal for mStrawberry, that has a 94% excitation efficiency and an effective brightness of 24 (Piatkevich and Verkhusha 2011), is possible to conclude that the 488 nm excitation laser does not accurately measures the mStawberry fluorescence.

Figure 5.2. Analysis of the SV40-STRB and CMV-STRB expression site-directed integrated into CHO-FRT mini-pools.

CHO-FRT 1 and CHO-FRT 108 mini-pools were transfected with pS/FRT/MS2-Strawberry and pcDNA5/FRT/MS2-mStrawberry vectors and selected as described in Section 2.6.1. After being adapted to suspension culture, a 1 ml culture sample from day 4 of culture was collected and analysed in BD Accuri™ C6 flow cytometer, using a 488nm excitation laser (Section 2.6.2.3). Fluorescence data was acquired using a 5850/40nm bandpass filter and cell line histograms were analysed using the BD Accuri™ C6 Software (Section 2.6.2.3). The auto-fluorescence of a CHO-FRT mini-pool was measured as a control for non-expressing cells. (CHO-FRT; CHO-SV40-STRB 1; CHO-SV40-STRB 108; CHO-CMV-STRB 1; CHO-CMV-STRB 108)
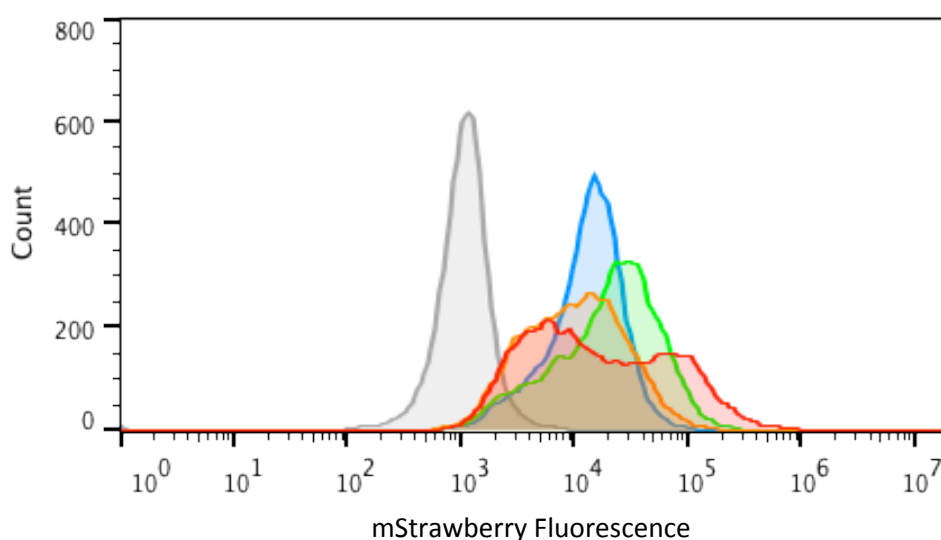
5.2. Live-cell imaging of CHO-SV40-STRB and CHO-CMV-STRB polyclonal cell lines

Figures 5.3 and 5.4 show fluorescent microscopy images obtained by live-cell imaging fluorescent microscopy of CHO-SV40-STRB 108, CHO-CMV-STRB 108 cell lines, as well as CHO-FRT 108 mini-pool used as negative control, transiently transfected with pMS2-GFP vector (Section 2.7.2). mStrawberry fluorescence was distributed in the cytoplasm of CHO-SV40-STRB 108, CHO-CMV-STRB 108 cell lines, whereas GFP fluorescence (MCP-GFP protein) was more highly concentrated in the nucleus of cells, presumably due to the presence of the NLS in the MCP-GFP encoded by the pMS2-GFP vector (Fusco et al. 2003).

Based on the method described by Yunger *et al*, (2013) the transient expression of the MCP-GFP protein (encoded by the pMS2-GFP vector) would be expected to lead to fluorescent labelling of the active transcription site from which the recombinant gene is being transcribed and had been shown to exhibit as bright green dot in the nucleus of the cell (Yunger et al. 2013). Figure 5.3 revealed that, in CHO cells, the distinctive green-dots described by Yunger et al (2013) were observed in CHOSV40-STRB 108 and CHOP-CMV-STRB 108 cells expressing MS2-lablled mRNA (Figure 5.3-A, B) but, crucially, were also observed in CHO-FRT 108 cells that did not contain MS2-tagged mRNA (Figure 5.3-C). It is important to stress that the cells used as the negative control reproducibly presented strong "positive signals". The same pattern of results were observed when the amount of pMS2-GFP DNA used for transient transfection was decreased by 50%, to diminish potential aggregation arising in the nucleus due to excess expression (Figure 5.4).

The results described above were unexpected, as the pMS2-GFP vector that encoded the MCP-GFP protein, was the same vector used in the studies of Yunger et al (2013) and was obtained from Addgene (plasmid # 27121) as suggested by the authors (Yunger et al. 2013). Moreover, because the MCP-GFP presents a critical component of this strategy, this vector was not generated in-house.

The observation of GFP fluorescence foci in the nucleus of cells without MS2-labbled mRNA, using the reagents promoted by Yunger et al (2013) presented a major block to the method development work undertaken in this aspect of the project. Consequently, although the approach appeared to offer great potential insight to transcriptional activity from specific promoters, it was not possible to this approach to correctly characterise the transcriptional activity of recombinant promoters.

In addition to the detection of distinctive green dots, MCP-GFP localisation also appeared to be unevenly distributed within the nucleoplasm (Figure 5.5) in both cell lines CHO-SV40-STRB 108 and CHO-CMV-STRB 108 (Figure 5.5-A, B). These observations were also applicable to the CHO-FRT 108 mini-pool (Figure.5.5-C). In a study of mRNA movement in Drosophila embryos, van Gemert et al observed that nuclear GFP fluorescence was preferentially localised to nucleoli when MCP-GFP containing a NLS was expressed in wild-type cells with absence of MS2-labelled mRNA (van Gemert et al. 2009). These observations are similar to those observed in Figure 5.5 as the brighter region detected in the nucleus resembled the nucleoli in the nucleus of the cell. The authors suggested that although MCP had been defined to bind specifically and strongly to MS2 stem-loops it also exhibited a non-specific but weaker affinity for other stem-loop structures such as the stem-loop structures of ribosomal RNA (rRNA) highly present in the nucleoli of the cell nucleus (van Gemert et al. 2009).

Moreover, it has been shown that MCP-GFP protein can form aggregates (that would be fluorescent) in cells that lacked MS2-labelled mRNA (Weil et al. 2010). Based on the comments of van Gemert et al (2009) and Weil et al (2010) and the observations exemplified by the images in Figures 5.3 and 5.4, it is also possible to suggest that generation of interactions with nucleoli, in CHO cells, MCP may also interact with other nuclear structures such as Nuclear Bodies, Cajal Bodies or nucleolar caps as these structures have been shown to contain different types of RNAs, with tertiary structures such as stem-loops (Brasch and Ochs 1992; Shav-Tal et al. 2005; Zimber et al. 2004). When studying the dynamics of the nuclear compartments of Hela cells, Shav-Tal *et al* observed that, when cells were transcriptional arrested with actinomycin D, nuclear elements segregated into nucleolar caps and that also led to the relocalisaton of nucleoplasmic proteins, many of which were RNA binding proteins (Shav-Tal et al. 2005). The author tagged RNA binding proteins with GFP and observed the formation of several distinct GFP foci within nuclei. Although only to limited extent, the authors also observe the formation of distinct GFP foci in non-actinomycin D-treated cells. As MCP is a RNA binding protein, it is possible that the observations illustrated in Figure 5.3 reflect those made by Shav-Tal et al. (2005)

Figure 5.3. Formation of MCP-GFP clusters in CHO nucleoplasm.
CHO-SV40-STRB 108 (**A**) and CHO-CMV-STRB 108 (**B**) were transiently transfected with pMS2-GFP vector as described in Section 2.7.2 and fluorescence of mStrawberry and MCP-GFP proteins were detected and analysed on a Leica TCS SP5 AOBS inverted confocal microcopy as described in Section 2.7.3. CHO-FRT 108 cells (**C**) subject to similar treatment were also analysed as a control. White arrows highlight the location of GFP fluorescent foci.
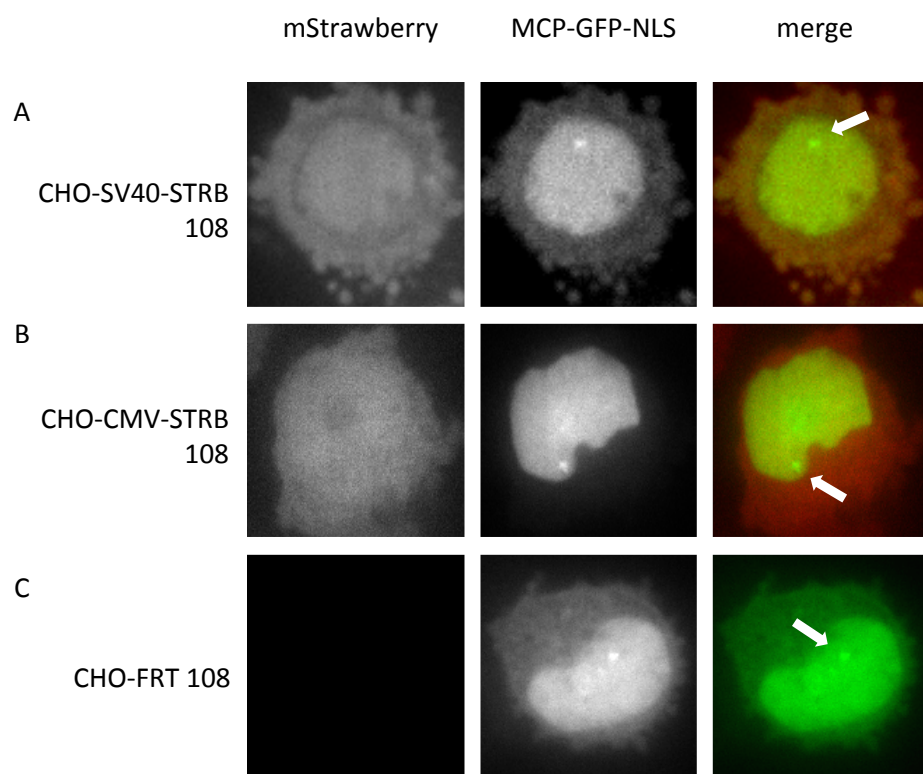
Figure 5.4. Formation of MCP-GFP clusters in CHO nucleoplasm.

CHO-SV40-STRB 108 (**A**) and CHO-CMV-STRB 108 (**B**) were transiently transfected with 50% of the pMS2-GFP vector DNA described in Section 2.7.2 and fluorescence of mStrawberry and MCP-GFP proteins were detected and analysed on a Leica TCS SP5 AOBS inverted confocal microcopy as described in Section 2.7.3. CHO-FRT 108 cells (**C**) subject to similar treatment were also analysed as a control. White arrows highlight the location of GFP fluorescent foci.
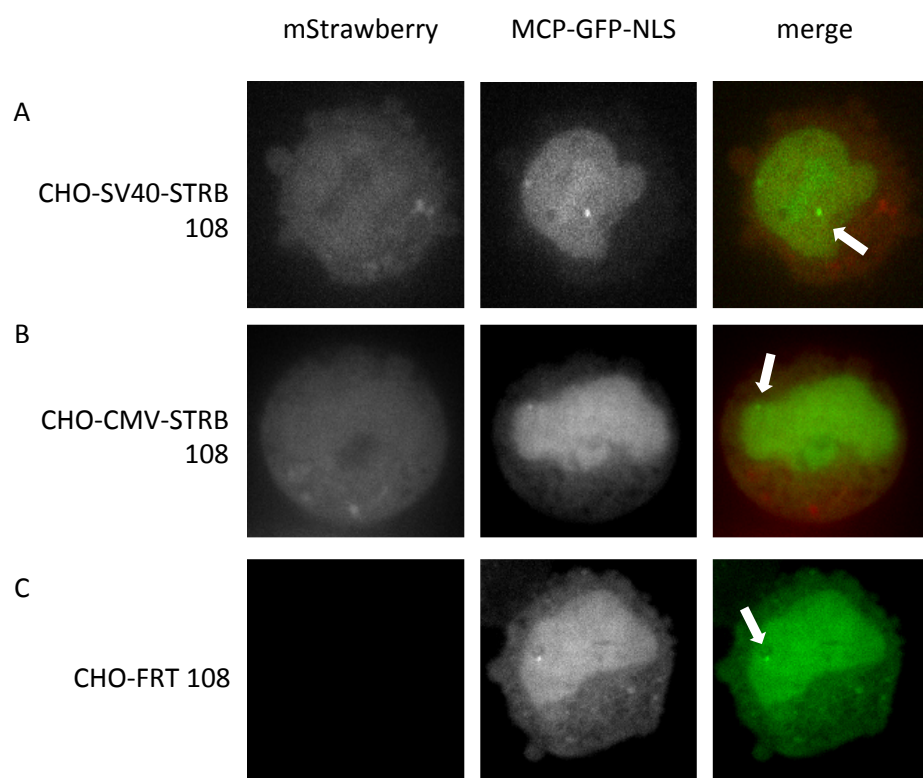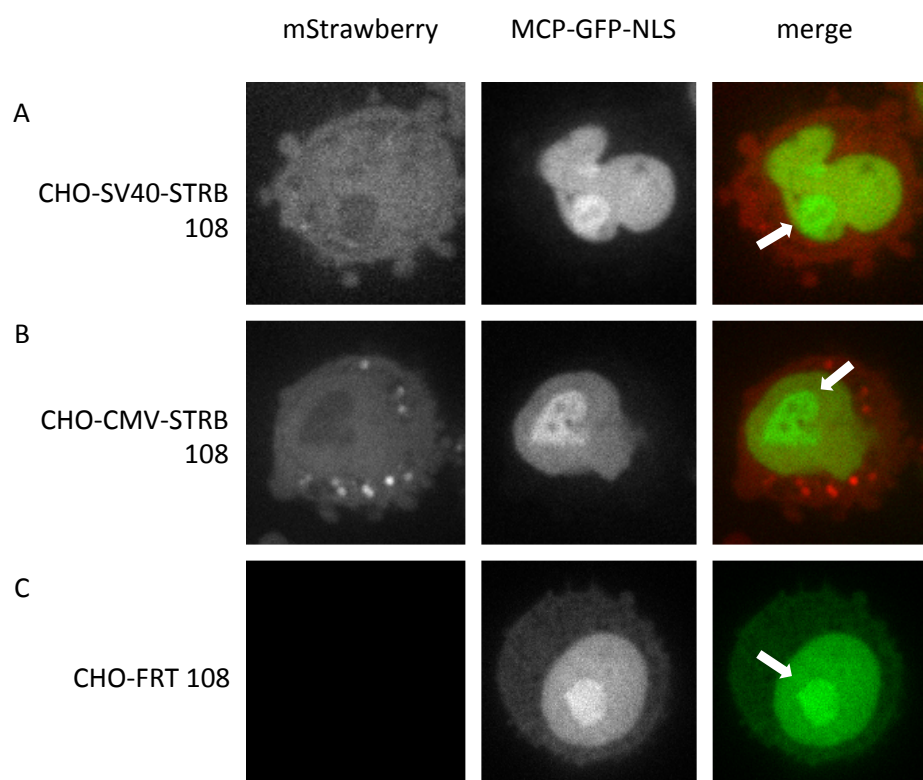
Figure 5.5. Unevenly distribution of MCP-GFP in CHO nucleoplasm.
CHO-SV40-STRB 108 (**A**), CHO-CMV-STRB 108 (**B**) and CHO FRT 108 mini-pools were treated and analysed as described in Figure 5.3. White arrows highlight the regions of higher concentration of MCP-GFP in the nucleoplasma.

5.3 Discussion

The main goal for the studies described in this Chapter was to develop a methodology for measurement of the transcriptional activity and the transcriptional rate of recombinant promoters in CHO-FRT mini-pools. The data presented indicated that all the vectors, transfection and detection methods to implement the methodology described by Yunger et al (2013) (Yunger et al. 2013) in CHO-FRT cell lines were successfully developed. However, based on the results of this Chapter, it can be concluded that the goal was not achieved. The central core of this methodology is based on the application of an MS2-tag that can be used to identify the location of transcription and the quantification of transcription in real-time, by photobleaching the GFP foci coupled to subsequent measurement the rate of the formation a new GFP foci (Yunger et al. 2013).

As discussed in the previous Section, the observation of MCP-GFP foci in both CHO-SV40-STRB 108 and CHO-CMV-STRB 108 as well as in CHO-FRT 108 mini-pool provided a major setback for the potential application of the methodology (at least for this CHO cell system). It was suggested that MCP-GFP might identify and bind MS2-labbled mRNA. However the presence of similar results in CHO-FRT mini-pools raises questions about the robustness the use of this strategy in CHO cells and ultimately about the accuracy of the results obtained for different promoters.

Besides attempting to optimise the transfection methods to decrease the expression of MCP-GFP, no further studies were made to investigate the cause of the formation of the GFP foci in CHO-FRT cells. However based on these results it is possible to suggest three experiments to investigate the cause and dynamics of the MCP-GFP foci, experiments that could form the basis for possible optimisation strategies of this method.

The first experiment would be a truncation of MCP to investigate if the formation of the foci is MCP- or GFP-dependent. If GFP foci were still visible under these conditions, this would prove GFP-dependency and so MCP could still be used in this method if tagged with a different fluorescent protein or other fluorescent marker. The second experiment would involve removal of the NLS from the MCP-GFP construct to diminish the concentration of MCP-GFP in the nucleus of the cell. Because MCP-GFP concentration in the nucleus would be speculated to be less under such conditions, considerably less, non-specific MCP-GFP binding would be expected. Due to the strength of the binding between MCP and MS2 stem-loop (Bertrand et al. 2000) the lesser amounts of free MCP-GFP protein in the nucleus would be expected

to bind the nascent recombinant mRNA. A major problem of this strategy, and the main reason why not it was not applied, is that without the NLS, the concentration of free MCP-GFP in the nucleus would be likely to be too low to perform FRAP studies. A lack of sufficient free MCP-GFP would lead to non-accurate measurements of the rate of transcription as the re-appearance rate MCP-GFP foci would not only be dependent on the rate of transcription of the recombinant mRNA but also on the availability of free MCP-GFP to bind the nascent transcript.

If no GFP foci were observed either, when MCP was truncated or NLS was removed, the third experiment that could be suggested would be to photobleach MCP-GFP foci in all cell lines and compare the rate of re-appearance of the MCP-GFP foci. If the presence of MS2-labbled mRNA altered the rate of which these foci re-appeared, the dynamics of the formation of the MCP-GFP foci could be further studied to determine the influence of the transcription of nascent mRNA in this event and ultimately generate a mathematical model could be use to indirectly and accurately measured the transcription rate.

The successful development of this novel approach to visualise and quantify the transcriptional activity of recombinant genes in CHO cells in real-time would be a major achievement to understand the transcriptional mechanism of recombinant genes in these cells. Moreover, the outcome of such methodology would potentially be a major step forward on the molecular development of novel promoters and regulatory elements that influence the recombinant expression in CHO cells.
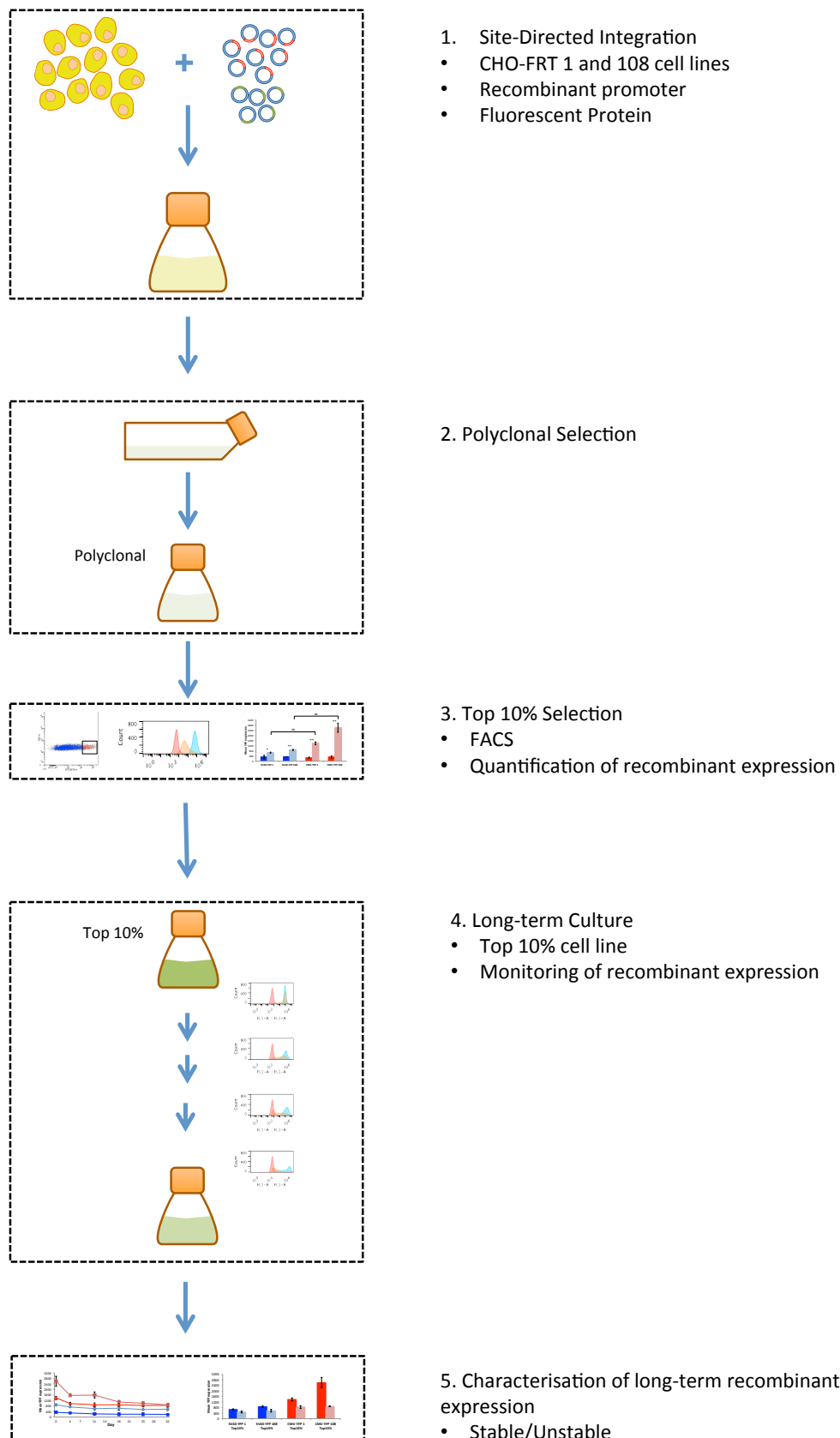
Chapter 6. Discussion

The overall aim of this project was to describe transcriptional activity and stability of recombinant promoters in CHO cells. Therefore, in order to achieve this, the specific aims of proposed for this project were:

a) Generate a cell/vector system to enable the study of recombinant promoters in CHO cells;

b) Generate a methodology to evaluate recombinant promoter activity in CHO cells when promoters were localised in the same genomic environment;

c) Characterise the activity of two widely-used recombinant promoters in CHO cells, SV40 and CMV promoters, in the same genomic environment;

d) Study the long-term expression of both promoters by evaluating the influence of each promoter in long-term expression of recombinant proteins in CHO cells;

e) Determine the transcriptional activity of each promoter in CHO cells *in vivo.*

The results in this thesis were presented in three distinct Chapters with the data discussed in some detail at the end of each Chapter. To achieve the aims initially proposed for this project it was necessary to generate a set of reagents. The first set of reagents was the vector system used and the generation strategy for this was described in Section 2.2. Final vectors maps are presented in Appendix 2. The second set of reagents was a set of CHO-FRT mini-pools and the generation and characterisation of these was described in Chapter 3. Using these two sets of reagents, Chapter 4 and Chapter 5 showed the application of generated reagents towards achievement of the aims of the project.

Figure 6.1 demonstrates a schematic representation of the alternative method developed in this project to evaluate and study the expression of recombinant promoters in long-term culture. The combination of site-directed integration with CHO-FRT mini-pools, in combination with fluorescent reporter proteins, represents a more efficient approach to characterise recombinant promoters in CHO cells.

Figure 6.1. Schematic representation of the method developed to characterise the strength and stability of recombinant promoters in CHO cells.

1. Site-Directed Integration
   - CHO-FRT 1 and 108 cell lines
   - Recombinant promoter
   - Fluorescent Protein

2. Polyclonal Selection

3. Top 10% Selection
   - FACS
   - Quantification of recombinant expression

4. Long-term Culture
   - Top 10% cell line
   - Monitoring of recombinant expression

5. Characterisation of long-term recombinant expression
   - Stable/Unstable

Based on the results obtain on Chapter 4 it is possible to suggest that aim (a) and (d) were, at least, partially achieved. The method developed in Chapter 4, and outlined in Figure 6.1, seems to show, to a certain degree, that as previously described, CMV promoter leads to high protein expression then the SV40 promoter in CHO cells (Section 4.2.2) (Ho et al. 2015). Furthermore, although not totally agreeing with previous descriptions for the stability of both promoters, the same method was capable to show that these two promoters appear to have different expressing behaviours during long term culture. The results in Section 4.2.4 reveal that CMV promoter led to considerably unstable long-term expression, as observed by Ho et al. (2015), and that the loss of protein expression appears to occur exponentially, with the majority of the loss of productivity occurring at early stages of culture (Section 4.2.4). On the other hand, contrary to what was observed by Ho et al. (2005), SV40 promoter appear to also lead to an unstable long-term protein expression. However, despite these observations, the degree of instability observed for SV40 is lower than observed to CMV (Section 4.2.4). Moreover, using this methodology, it was also possible to observed that with SV40 promoter, the loss of productive appears occurs differently to CMV promoter, occurring gradually during the period of culture (Section 4.2.4)

As discussed in Sections 4.3, the method developed in this thesis cannot guarantee with 100% certainty that the observations made for different promoters are made in the same genomic environment. This factor strongly influences the fully achievement of aims b) and c) as the same genomic environments for the comparison of both promoters used was not possible as initially aimed. However, because site-directed integration is still used and these mini-pools have limited amount of FRT sequences, it is possible to suggest that the analysis performed using these mini-pools allows to comparison both promoters in similar genomic environments (Section 4.3).

The final aim of this project was to determine the transcriptional activity of SV40 and CMV promoters in CHO cells *in vivo.* The methodological approach to quantify the transcriptional activity of these promoters *in vivo* was based on the combination of site-directed integration in CHO-FRT mini-pools with MS2-tag/MCP-GFP technology to allow live cell visualisation of tagged recombinant mRNA. This approach was enabled by generation of vectors that contained a MS2 RNA tag sequence contiguously cloned 3'- of the recombinant gene. The presence in this tag of specific stem-loops, in the recombinant mRNA, can be used to monitor recombinant mRNA and quantify the transcriptional output of the recombinant gene in single living cells using fluorescent microscopy.

As discussed in Section 5.3, the data presented in Section 5.1 and Section 5.2 indicated that all the vectors, transfection and detection methods necessary to implement this methodology in CHO cell lines were successfully developed. However, based on the results observed in Figure 5.3 and Figure 5.4, it can be concluded that the aim (e) of this project was not achieved.

The central core of this methodology is based on the application of the MS2-tag that can be identified by MCP-GFP protein, a MS2 coat protein tagged with GFP that has a strong affinity to bind to the MS2 stem-loops (Valegård et al 1997). Based on this strong affinity, it is reported to be possible to identify the location of transcription by the formation of foci of GFP on the emergent MS2-tagged mRNA (Yunger et al 2010; Yenger et al 2013). The results presented in Figure 5.3 and 5.4 revealed that, in CHO cells, this technology gave a strong background artefact with a formation of strong GFP foci in control CHO cells which contained no MS2 tagged mRNA. This discovery of this artefact strongly compromised any further observations as this technology had no further positive controls that could be used to discard possible background observations.

From the discussion above, regarding the achievement of the initial aims of the project there are several questions that were established:

6.1. Can the presence of multiple FRT sites affect the characterisation of recombinant promoters when using a site-directed approach?

Other studies have described promoter strengths to recombinant sequences in CHO cells but frequently this is based on random integration and true comparison of the strength and stability of promoters requires that promoters have limited influence from different genomic environments (Kaufman et al 2008; Mariati et al 2014). Ideally, a true comparison would occur with insertion to the same genomic environment. Thus, the use of site-directed methods offers a powerful potential approach to achieve these ideal conditions.

Site-directed integration has been previously be used to characterise recombinant promoters in CHO cells. Tornøe et al (2002) used a site-directed approach to integrate GFP under the influence of three synthetic promoters into CHO Flp-In cell line (Life Technologies) and compared GFP expression driven by each promoter (Tornøe et al 2002). As the CHO Flp-In cell line only has one FRT site, this approach

was able to characterise these synthetic promoters and compare their relative strengths. However, because the genomic environment has a strong influence in the activity of the recombinant promoter, the results presented using this strategy are highly influenced by the unique genomic environment in which the FRT site in is located, not representing the actual promoter strength of each of those promoters. It is possible that the FRT site in CHO Flp-In cell line has a more adequate genomic environment for one promoter favouring its influence on gene expression. Ideally, to dilute the influence of a specific genomic environment, this characterisation should be performed in a set of different CHO cell lines contained a single FRT site.

In attempt to fulfil this requirement, in the method proposed in this project, a library of CHO-FRT cell lines was generated. However, there remains a caveat to absolute certainty on the proposed method due to the fact that all the CHO-FRT mini-pools generated in Chapter 3 had multiple copies of the inserted recombinant gene.

The possibility that gene copies are distributed over multiple FRT sites could result in selective insertion of specific promoters to different environments. Consequently, whilst the developed method may allow for the characterisation of recombinant promoters in the similar genomic environment, there remains the possibility that, without direct confirmation of recombination insertion site, there may be heterogeneity in assessed mini-pools. This provides reasons to regard aims (b) and (c) to be unsuccessful as to be used as described in this project, further verification of FRT arrangements and re-insertion of reporter genes are required prior to drawing absolute conclusions.

The existence of multiple FRT sites in the genome of CHO-FRT may ultimately lead to the formation of heterogeneous cell populations after the promoter/reporter cassette is integrated. Due to the inherent inefficiency of the site-directed integration approach used (Buchholz 1996), it is not possible to guarantee that all the FRT sites in all the selected mini-pools underwent re-integration. To minimise the impact of potential heterogeneity, it was proposed that a second round of FACS selection would be used select for high expressing cells present in the selected population (Section 4.2). The rational behind this second round of selection was, although not guarantying that all the FRT sites in the selected cells had undergone full re-integration, it would diminishes the heterogeneity of the initial polyclonal cell line and possibly selected for fully or high YFP integrated cells. As it was discussed in Section 4.3, based on the results observed in Section 4.2.3, it is possible that the instead of only selecting cells that are naturally high expressing cells with possible high number of YFP gene copies, this second round of selection also selected lower expressing

cells that were perturbed above steady state at the time of selection. This possible selection for high expressing based on the stochastic gene expression fluctuation ultimately masks the really difference between the strength of the promoters, meaning that this method alone cannot be used to accurately evaluate differences in promoter strength.

Nevertheless, the presence of multiple FRT sites in multiple genomic locations could be an advantage to accurately characterise recombinant promoters if the method of selection used guarantees full integration in all the FRT sites. In cells with all the FRT sites integrated, the strength and stability of the different promoters would be evaluated under the influence of a more broaden genomic environment (not representing any particular locus) but yet the same for the different promoters,

When integrated in a FRT site, the promoter/YFP cassette will disrupt the expression the ß-galactosidase gene. Therefore, the cell lines that have full YFP integration will have no have no ß-galactosidase activity. Based on this, the full integration could be achieved by incubating the selected cells with C12FDG (5-dodecanoylaminofluorescein di-beta-D-galactopyranoside), a ß-galactosidase substrate that when degraded in the cell, generates fluorescein (a fluorescent product). Using this approach it would be possible to sort the cells that were fully integrated (YFP positive/fluorescein negative) by FACS.

From the discussion above, the answer to the original question is; Yes, multiple FRT sites can affect the characterisation of recombinant promoters when using site-directed integration approach, but if full integration is achieved it would give an advantage to the general characterisation of the promoter.

6.2 Judging by the number of FRT sites in CHO-FRT cells, is the developed method still suitable for study the long-term activity of recombinant promoters in CHO cells?

The screening and characterisation of recombinant promoters is traditionally made by transient or stable integrative expression, however both methods do not accurately characterise recombinant promoters (Brown 2014; Mariati 2014). For instance, transient expression is a methodology that gives a quick output of relative promoter strength, but does not consider the influence of genomic the genomic environment. On the other hand, stable integrative expression takes into account the influence of the genomic environment in the promoter strength, however due to its

random nature it does not allow an accurate comparison of strength between two promoters.

Despite the method described does not allow to accurate evaluate the differences in promoter strength, it allows to observed considerably differences in the behaviour of the expression driven by the SV40 and CMV promoter during long-term culture. As observed in Section 4.2.4, both promoters were described by this method as unstable. However the degree and pattern of instability differ between of both promoters SV40 was observed to be slightly unstable with a gradual loss of expression, while CMV was described as considerably unstable with logarithmic loss of expression.

Although not yet ideal, because the methodology proposed in this project combines the use of site-directed integration in CHO-FRT mini-pools combined with a second round of selection that allow selection of a specific cell line-subpopulation represents an improvement over traditional methodologies used for these type of studies.

Therefore, the answer to the initial question is; Yes, although not ideal, this easy and relatively quick approach can be suitable to study the long-term expression driven by different promoters and represents an alternative approach for this type of studies.

6.3. Can *in vivo* transcription be measured in CHO cells?

The successful use of this novel approach to directly quantify the transcriptional activity of recombinant promoters in CHO cells *in vivo* would be a major achievement and open many possibilities to further elucidate transcription of recombinant genes in production CHO cell lines. The application of such know-how using the method represented in Figure 6.1 would be a major technological advance for the molecular development of novel recombinant promoters and regulatory elements to be used in CHO cells. In addition, it would open new approaches to decipher linkages between genomic environment, recombinant gene status and transcription from industrially-relevant systems.

The developed approach to *in vivo* quantify transcriptional output of recombinant promoters in CHO cells was based on a method applied to HEK293 cells (Yunger et al 2013). These authors had previously used a similar strategy to quantify the transcriptional output of CMV promoter and human cyclin D1 gene (CCND1) promoter in HEK 293 cells (Yunger et al 2010). In both papers, the authors failed to provide any negative controls (i.e. cells that did not contain MS2 tagged mRNA) of

the live-cell images of the MCP-GFP protein in HEK 293 cells. Apart from these two examples published by Yunger and colleagues, to date, no further reports of quantification of the transcriptional outputs in mammalian cells has been published using this methodology.

In Section 5.3 it three possible experiments were suggested that could be used to address or diminish the influence of MCP-GFP background on the quantification of the transcriptional activity in CHO cell. Such experiments would be essential to optimise the use of MS2/MCP technology in CHO cells. However, because the background artefact observed was strong, this could be laborious and the desired outcome might not be achievable. For that reason, alternative approaches may be considered to achieve assessment of transcription.

One possible method that can be suggested is to use a PP7-RNA tag in a similar manner as the MS2-RNA tag described in Chapter 5. PP7-stem loops are very well characterised and, in a manner analogous to the MS2 stem, can be identified by a specific PP7 Coat protein (PCP) (Lim and Peabody 2002). Although PP7-stem loops have been used to track the movement of RNA is living cells, its use for quantification of transcription *in vivo* has yet to be reported, so appropriate controls should be put in place during the development of this approach.

Based on the results obtained in Chapter 5 and the discussion above the answer for the initial question is; As it stands the available technology appears not to be able to measure transcriptional activity in vivo in CHO cells. However, it is possibly that the technology presented in this project could accurately preform such sensitive measurement if further studies and methods are developed to generate robust and solid controls.

6.4. Future Work

Beside the lines of research that were proposed above in this Chapter and in the Discussion Sections of each Chapter of results, there are other avenues of research that may be built form the foundations of knowledge and reagents (CHO-FRT mini-pools and vectors system) generated in this project. As this project purposed to study transcriptional stability, the proposed future directions will be focused on this ultimate quest of achieving high and stable productivities in CHO cells.

The vector system developed in this project is unique and versatile as it combines multiple features from different technologies. For instance, the vectors generated

allow the use of different recombinant promoters in combination with any recombinant gene to be targeted to any specific gene location that had been previously inserted with the same FRT sequence. Furthermore, the presence of the MS2 loops not only gives the potential to quantify transcriptional activity (if optimised) as also gives the potential to track mRNA intracellular movement and develop new strategies to study recombinant mRNA processing, stability, degradation and trafficking within the CHO cell.

As different promoters have different strengths, they will lead to the generation of different amounts of mRNA within the cell. High concentration of recombinant mRNA can be advantageous for the achievement for high productivity (Barnes et al 2004), however it can also create bottlenecks that ultimately lead to translational silencing. (Schoenberg and Maquat 2012). For that reason, by cloning tunable promoters or different promoters with different strengths in this vector system, by tracking the mRNA using single-cell live-microscopy technics it would be possible to study the influence of the promoter on the translational silencing, the optimal mRNA concentrations within the cell and mRNA degradation mechanisms.

The CHO-FRT mini-pools can also be a foundation to different studies focused on the transcriptional bottlenecks in CHO cells. In this study, a considerable library of CHO-FRT mini-pools were generated with multiple FRT sites, different productivities, different stability profiles and different growth rates. This variety of phenotypes from mini-pools that were generated in the same manner ultimately represents a toolbox that could be used as a basis for several studies to understand the mechanisms in CHO cells that regulate different cellular events.

However, in a era of genomic revolution, the sequencing of the genome of CHO-FRT 1 and CHO-FRT 108 mini-pools would bring invaluable information about the genome of these unique mini-pools and open the possibility for further developments. For instance due to random nature of the process used for the generation of CHO-FRT mini-pools, is likely that the multiple FRT sites in these mini-pools are located in more than one genomic locus. As the genomic locations demonstrate differences in strengths and stabilities of different promoters, identification of the location of these FRT sites could be crucial to understand their genomic environment. This information may allow the identification of regulatory mechanisms that interact with each promoter and ultimately generate better understating of which mechanisms have more influence on generating high productive and/or stable cell lines.

Furthermore, by using genome editing tools, such as CRISPR-Cas9, ZFN or TALENs, it would be possible to study the influence of multiple genomic locations on the overall protein expression and stability. By knocking out several individual genomic locus it would be possible to evaluate their impact in the overall cell line productivity. Such study would ultimately provide an understanding on how the overall cell productive reflects the influence of the individual genomic environments, a question of whether they produce summation of the influence of individual environments or synergistic or antagonistic effects?

As was observed in Chapter 3, the FRT sites in the generated mini-pools appeared to be in stable genomic locations as no loss of gene copy number was observed during the long-term study. For that reason they could be in optimal genomic locations for integration of genomic landing pads for the insertion of different recombinant genes. By using genome-editing tools, it would be possible to insert a reporter gene, flanked by two site-directed target sequences in the same genomic locus but in a new CHO host. Although genomic locations in CHO-FRT appear to provide, in some cases, a relatively stable expression with SV40 promoters, they totally failed when CMV promoter was used. For that reason, if genomic pads were inserted in these genomic locations, in an attempt to optimise the stability of what it already seem to be relatively stable genomic loci, regulatory elements such as S/Mars or UCOE's could be inserted upstream and downstream of the genomic landing pads.

Finally, the majority of biopharmaceuticals produced in CHO cells are mAbs. These therapeutic proteins are encoded by two recombinant genes, heavy chain and light and its productivity in CHO cells is dependent on the optimal ratio of expression of these two genes (Schlatter et al 2005). By creating cell lines with two different landing pads it would be possible to insert these two genes into two well characterised and stable genomic locations and tune their expression with different promoters in order to achieve the optimal ratio and consequently optimise productivity.

References

Alper H, Fischer C, Nevoigt E. 2005. Tuning genetic control through promoter engineering. P Natl Acad Sci Usa 102(36):12678-12683

Arents G, Burlingame RW, Wang BC, Love WL, Moudrianakis EN. 1991 The nucleosomal core histone octamer at 3.1 A resolution: a tripartite protein assembly and a left-handed superhelix. P Natl Acad Sci Usa 88:10148-10152

Axel R, Melchoir W, Sollner-Webb B, Felsenfeld G 1974, Specific sites of interaction between histones and DNA in chromatin. P Natl Acad Sci Usa 71(10): 4101-4105.

Baik JY, Lee KH. 2014. Toward product attribute control: developments from genome sequencing. Curr Opin Biotech 30:40-4.

Bailey LA, Hatton D, Field R, Dickson AJ. 2012. Determination of Chinese hamster ovary cell line stability and recombinant antibody expression during long‐term culture. Biotechnol and Bioeng 109(8):2093-2103.

Bandaranayake AD, Almo SC. 2014. Recent advances in mammalian protein production. FEBS Lett 588(2):253-260.

Baneyx F. 1999. Recombinant protein expression in Escherichia coli. Curr Opin Biotech 10:411-421.

Barnes LM, Bentley CM, Dickson AJ. 2003a. Stability of protein production from recombinant mammalian cells. Biotechnol and Bioeng 81(6):631-639.

Barnes LM, Bentley CM, Dickson AJ. 2003b. Stability of recombinant protein production in the GS‐NS0 expression system is unaffected by cryopreservation. Biotechnol Progress 19(1):233-237.

Barnes LM, Dickson AJ. 2006. Mammalian cell factories for efficient and stable protein expression. Curr Opin Biotech 17(4):381386.

Barnes LM, Moy N, Dickson AJ. 2006. Phenotypic variation during cloning procedures: Analysis of the growth behavior of clonal cell lines. Biotechnol and Bioeng 94(3):530-537.

Bauer AP, Leikam D, Krinner S, Notka F, Ludwig C, Längst G, Wagner R. 2010. The impact of intragenic CpG content on gene expression. Nucleic Acids Res 38(12):3891-3908.

Bennett RP, Cox CA, Hoeffler JP. 1998. Fusion of green fluorescent protein with the Zeocin-resistance marker allows visual screening and drug selection of transfected eukaryotic cells. BioTechniques 24(3):478-482.

Benton T, Chen T, McEntee M, Fox B, King D, Crombie R, Thomas TC, Bebbington C. 2002. The use of UCOE vectors in combination with a preadapted serum free, suspension cell line allows for rapid production of large quantities of protein. Cytotechnology 38(1-3):43-46.

Berger SL. 2007. The complex language of chromatin regulation during transcription. Nature 447:407-412.

Berlowitz L. 1974. Chromosomal Inactivation And Reactivation In Mealy Bugs. Genetics 78:311-322.

Bertrand E, Chartrand P, Schaefer M, Shenoy SM, Singer RH, Long RM. 2000. Localization of ASH1 mRNA particles in living yeast. Mol Cell 2(4):437-445.

Betts Z, Dickson A. 2015. Ubiquitous chromatin opening elements (UCOEs) effect on transgene position and expression stability in CHO cells following methotrexate (MTX) amplification. Biotechnol J. In press DOI: 10.1002/biot.201500159.

Boshart M, Weber F, Jahn G, Dorsch-Häsler K, Fleckenstein B, Schaffner W. 1985. A very strong enhancer is located upstream of an immediate early gene of human cytomegalovirus. Cell 41(2):521-530.

Boyle S, Gilchrist S, Bridger JM, Mahy NL, Ellis JA, Bickmore WA. 2001. The spatial organization of human chromosomes within the nuclei of normal and emrein-mutant cells. Hum Mol Genet 10(3):211 - 219.

Braell WA, Balch WE, Dobbertin DC, Rothman JE. 1984. The glycoprotein that is transported between successive compartments of the golgi in a cell-free system resides in stacks of cisternae. Cell 39(3):511524.

Branco MR, Pombo A. 2006. Intermingling of chromosome territories in interphase suggests role in translocations and transcription-dependent associations. PLoS Biology 4(5):e138.

Brasch K, Ochs RL. 1992. Nuclear bodies (NBs): a newly "rediscovered" organelle. Exp Cell Res 282:211-223

Brinkrolf K, Rupp O, Laux H, Kollin F, Ernst W, Linke B, Kofler R, Romand S, Hesse F, Budach WE and others. 2013. Chinese hamster genome sequenced from sorted chromosomes. Nat Biotechnol 31(8):694-695.

Brown AJ, Sweeney B, Mainwaring DO. 2014. Synthetic promoters for CHO cell engineering. Biotechnol and Bioeng 111(8):1638-1647

Bryans M, Kass S, Scivwright C, Adams RLP. 1992. Vector methylation inhibits transcription from the SV40 early promoter. FEBS Lett 309(1):97-102.

Buchholz F, Ringrose L, Angrand PO, Rossi F, Stewart AF. 1996. Different thermostabilities of FLP and Cre recombinases: implications for applied site-specific recombination. Nucleic Acids Res 24(21):4256-4262.

Buckley MS, Lis JT. 2014. Imaging RNA Polymerase II transcription sites in living cells. Curr Opin Genet Dev 25:126-30.

Butler M, Spearman M. 2014. The choice of mammalian cell host and possibilities for glycosylation engineering. Curr Opin in Biotechnol 30:107-112

Byrne B, Donohoe GG, O'Kennedy R. 2007. Sialic acids: carbohydrate moieties that influence the biological and physical properties of biopharmaceutical proteins and living cells. Drug Discov Today 12(7/8):319-326

Carlberg C, Molnár F. 2013. Chapter 12. Chromatin architecture. Mechanisms of Gene Regulation 183-193 DOI:10.1007/978-94-007-7905-1_12

Carter D, Chakalova L, Osborne CS, Dai Y-f, Fraser P. 2002. Long-range chromatin regulatory interactions in vivo. Nat Genet 32:623-626.

Carvalho DDD, You JS, Jones PA. 2010. DNA methylation and cellular reprogramming. Trends Cell Biol 20(10):609-617.

Chalfie M, Tu Y, Euskirchen G, Ward WW, Prasher DC. 1994. Green fluorescent protein as a marker for gene expression. Science 263:802-805

Chang G-D, Chen C-J, Lin C-Y, Chen H-C, Chena H. 2003. Improvement of glycosylation in insect cells with mammalian glycosyltransferases. J Biotechnol 102:61-71.

Chen J, Haverty J, Deng L, Li G, Qiu P, Liu Z, Shi S. 2013. Identification of a novel endogenous regulatory element in Chinese hamster ovary cells by promoter trap. J Biotechnol 167(3):255-261.

Cheng JK, Alper HS. 2014. The genome editing toolbox: a spectrum of approaches for targeted modification. Curr Opin Biotechnol 30:87-94.

Chubb JR, Trcek T, Shenoy SM, Singer RH. 2006. Transcriptional pulsing of a developmental gene. Curr Biol 16(10):1018-25.

Chusainow J, Yang YS, Yeo JH, Toh PC, Asvadi P, Wong NS, Yap MG. 2009. A study of monoclonal antibody-producing CHO cell lines: what makes a stable high producer? Biotechnol Bioeng 102(4):1182-1196.

Cisse II, Izeddin I, Causse SZ, Boudarene L, Senecal A. 2013. Real-time dynamics of RNA polymerase II clustering in live human cells. Science 341:664-667

Cook PR. 2002. Predicting three-dimensional genome structure from transcriptional activity. Nat Genet 32(3):347-352.

Corden JL. 1990. Tails of RNA polymerase II. Trends Biochem Sci 15:383-387

Cremer M, Hase Jv, Volm T, Brero A, Kreth G, Walter J, Fischer C, Solovei I, Cremer C, Cremer T. 2001. Non-random radial higher-order chromatin arrangements in nuclei of diploid human cells. Chromosome Res 9:541-567.

Cremer T, Cremer M. 2010. Chromosome territories. Cold Spring Harb Perspect Biol 2010;2:a003889

Cutter AR, Hayes JJ. A brief review of nucleosome structure. 2015 FEBS Lett 589(20A):2914-2922

Dahmus ME. 1995. Phosphorylation of the C-terminal domain of RNA polymerase II. Biochim Biophys Acta 1261:171-182

Dahmus ME. 1996. Reversible phosphorylation of the C-terminal domain of RNA polymerase II. J Biol Chem 271(32):19009–19012.

Darzacq X, Shav-Tal Y, de Turris V, Brody Y, Shenoy SM, Phair RD, Singer RH. 2007. In vivo dynamics of RNA polymerase II transcription. Nat Struct Biol 14(9):796-806.

Davies SL, Lovelady CS, Grainger RK, Racher AJ, Young RJ, James DC 2013. Functional heterogeneity and heritability in CHO cell populations. Biotechnol Bioeng 110(1):260-274

de Poorter JJ, Lipinski KS, Nelissen RG, Huizinga TW, Hoeben RC. 2007. Optimization of short‐term transgene expression by sodium butyrate and ubiquitous chromatin opening elements (UCOEs). J Gene Med 9(8):639-648.

Deaven LL, Petersen DF. 1973. The chromosomes of CHO, an aneuploid Chinese hamster cell line: G-band, C-band, and autoradiographic Analyses. Chromosoma 41:129-144.

Delacote F, Deriano L, Lambert S, Bertrand P, Saintigny Y, Lopez BS. 2007. Chronic exposure to sublethal doses of radiation mimetic Zeocin™ selects for clones deficient in homologous recombination. Mutat Res-Fund Mol M 615(1-2):125133.

Derouazi M, Martinet D, Besuchet Schmutz N, Flaction R, Wicht M, Bertschinger M, Hacker DL, Beckmann JS, Wurm FM. 2006. Genetic characterization of CHO production host DG44 and derivative recombinant cell lines. Biochem Bioph Res Co 340(4):10691077.

Deschênes I, Finkle CD, Winocour PD. 1998. Effective use of BCH-2763, a new potent injectable direct thrombin inhibitor, in combination with tissue plasminogen activator (tPA) in a rat arterial thrombolysis model. Thromb Haemost 80:186–191.

Dietmair S, Hodson MP, Quek L-EE, Timmins NE, Gray P, Nielsen LK. 2012. A multi-omics analysis of recombinant protein production in Hek293 cells. PloS One 7(8):e43394

Dorai H, Corisdeo S, Ellis D, Kinney C, Chomo M, Hawley-Nelson P, Moore G, Betenbaugh MJ, Ganguly1 S. 2012. Early prediction of instability of Chinese hamster ovary cell lines expressing recombinant antibodies and antibody-fusion proteins. Biotechnol Bioeng 109(4):1016-1030

Dostie J, Bickmore WA. 2011. Chromosome organization in the nucleus – charting new territory across the Hi-Cs. Curr Opin Genet Dev 22:1-7.

Du ZM, Mujacic M, Le K, Caspary G, Nunn H, Heath C, Reddy P. 2013. Analysis of heterogeneity and instability of stable mAb-expressing CHO cells. Biotechnol Bioprocess Eng 18(2):419-429.

Durocher Y, Butler M. 2009. Expression systems for therapeutic glycoprotein production. Curr Opin in Biotech 20(6):700-707.

Edelman LB, Fraser P. 2012. Transcription factories: genetic programming in three dimensions. Curr Opin Genet Dev 22(2):110-114.

Eickbush TH; Moudrianakis EM 1978. The histone core complex: an octamer assembled by two sets of protein-protein interactions. Biochemistry 17(23): 4955-4964

Elmore S. 2007. Apoptosis: A review of programmed cell death. Toxicol Pathol 35:495-516.

Eustice DC, Feldman PA, Colberg-Poley AM, Buckery RM, Neubauer RH. 1991. A sensitive method for the detection of beta-galactosidase in transfected mammalian cells. BioTechniques 11(6):739-743.

Evans K, Albanetti T, Venkat R, Schoner R, Savery J, Miro-Quesada G, Rajan B, Groves C 2015 Assurance of monoclonality in one round of cloning through cell sorting for single cell deposition coupled with high resolution cell imaging. Biotechnol Progr 31(5):1172-1178

Eyquem J, Poirot L, Galetto R, Scharenberg AM, Smith J. 2013. Characterization of three loci for homologous gene targeting and transgene expression. Biotechnol Bioeng 110(8):2225-2235.

Fallaux FJ, Bout A, van der Velde I. 1998. New helper cells and matched early region 1-deleted adenovirus vectors prevent generation of replication-competent adenoviruses. Hum Gene Ther 9(13):1909-1917

Fann CH, Guirgis F, Chen G, Lao MS, Piret JM. 2000. Limitations to the amplification and stability of human tissue-type plasminogen activator expression by Chinese hamster ovary cells. Biotechnol Bioeng 69(2):204-212.

Feng HT, Sim LC, Wan C, Wong NSC, Yang Y. 2011. Rapid characterization of protein productivity and production stability of CHO cells by matrix-assisted

laser desorption/ionization time-of-flight mass spectrometry. Rapid Commun Mass Sp 25(10):1407-1412.

Finch JT, Lutter LC, Rhodes D, Brown RS, Rushton B, Levitt M, Klug A. 1977. Structure of nucleosome core particles of chromatin. Nature 29:29-36.

Florin L, Lipske C, Becker E, Kaufmann H. 2011. Supplementation of serum free media with HT is not sufficient to restore growth properties of DHFR−/− cells in fed-batch processes – Implications for designing novel CHO-based expression platforms. J Biotechnol 152(4):189-193.

Foecking MK, Hofstetter H. 1986. Powerful and versatile enhancer-promoter unit for mammalian expression vectors. Gene 45:101-105

Fraser MJ, Clszczon T, Elick T. 1996. Precise excision of TTAA-specific lepidopteran transposons piggyBac (IFP2) and tagalong (TFP3) from the baculovirus genome in cell lines from two species of Lepidoptera. Insect Mol Biol 5(2):141-151

Fraser P, Bickmore W. 2007. Nuclear organization of the genome and the potential for gene regulation. Nature 447(7143):413-417.

Fusco D, Accornero N, Lavoie B, Shenoy SM, Blanchard JM, Singer RH, Bertrand E. 2003. Single mRNA molecules demonstrate probabilistic movement in living mammalian cells. Curr Biol 13(2):161-167.

Gemmill TR, Trimble RB. 1999. Overview of N- and O-linked oligosaccharide structures found in various yeast species. Biochim Biophys Acta 1426:227-237.

Ghamari A, van de Corput MP, Thongjuea S, van Cappellen WA, van Ijcken W, van Haren J, Soler E, Eick D, Lenhard B, Grosveld FG. 2013. In vivo live imaging of RNA polymerase II transcription factories in primary cells. Genes Dev 27(7):767-777.

Golding I, Paulsson J, Zawilski SM, Cox EC. 2005. Real-time kinetics of gene activity in individual bacteria. Cell 123(6):1025-1036.

Goldstein DA, Thomas JA. 2004. Biopharmaceuticals derived from genetically modified plants. QJM 97(11):705-716.

Gonzalez F, Duboule D, Spitz Fo. 2007. Transgenic analysis of Hoxd gene regulation during digit development. Dev Biol 306:847-859.

Goswami J, Sinskey AJ, Steller H, Stephanopoulos GN, Wang DI. 1999. Apoptosis in batch cultures of Chinese hamster ovary cells. Biotechnol Bioeng 62(6):632-640.

Graham FL. 1987. Growth of 293 cells in suspension culture. J Gen Vir 68(Pt 3):937-940.

Graham FL, Smiley J, Russell WC, Nairn R. 1977. Characteristics of a human cell line transformed by DNA from human adenovirus type 5. J Gen Vir 36(1):59-74.

Grewal SIS, Moazed D. 2003. Heterochromatin and epigenetic control of gene expression. Science 301:798-802.

Gronostajski RM, Sadowski PD. 1985. Determination of DNA sequences essential for FLP-mediated recombination by a novel method. J of Biol Chem 260(22):12320-12327,

Hacker DL, Jesus MD, Wurm FM. 2009. 25 years of recombinant proteins from reactor-grown cells - Where do we go from here? Biotechnol Adv 27:1023-1027.

Hamilton SR, Davidson RC, Sethuraman N, Nett JH, Jiang Y, Rios S, Bobrowicz P, Stadheim TA, Li H, Choi B-K and others. 2006. Humanization of Yeast to Produce Complex Terminally Sialylated Glycoproteins. Science 313(1441-1443).

Han Y, Liu XM, Liu H, Li SC, Wu BC, Ye LL, Wang QW, Chen ZL. 2006. Cultivation of Recombinant Chinese hamster ovary cells grown as suspended aggregates in stirred vessels. J Biosci and Bioeng 102(5):430-435.

Heidemann M, Hintermair C, Voß K, Eick D. 2013. Dynamic phosphorylation patterns of RNA polymerase II CTD during transcription. Biochim Biophys Acta 1829:55-62

Ho SC, Mariati, Yeo JH, Fang SG, 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(2):138-144.

Ho SC, Yang Y. 2014. Identifying and engineering promoters for high level and sustainable therapeutic recombinant protein production in cultured mammalian cells. Biotechnol Lett 36(8):1569-1579.

Holliday R, Ho T. 1998. Evidence for gene silencing by endogenous DNA methylation. P Natl Acad Sci Usa 95(15):8727-8732.

Hong L, Schroth GP, Matthews HR, Yau P, Bradbury EM. 1993. Studies of the DNA binding properties of histone H4 amino terminus. J Biol Chemi 268:305-314.

Hopkins D, Gomathinayagam S, Rittenhour AM, Du M, Hoyt E, Karaveg K, Mitchell T, Nett JH, Sharkey NJ, Stadheim TA and others. 2011. Elimination of ß-mannose glycan structures in Pichia pastoris. Glycobiology 21(12):1616-1626.

Horike S-i, Cai S, Miyano M, Cheng J-F, Kohwi-Shigematsu T. 2004. Loss of silent-chromatin looping and impaired imprinting of DLX5 in Rett syndrome. Nat Gene 37:31-40.

Hou JJC, Hughes BS, Smede M, Leung KM, Levine K. 2014. High-throughput ClonePix FL analysis of mAb-expressing clones using the UCOE expression system. New Biotechnol 31(3):214-220

Huang Y, Li Y, Wang YG, Gu X, Wang Y, Shen BF. 2007. An efficient and targeted gene integration system for high-level antibody expression. J Immunol Methods 322(1-2):2839.

Hussain H, Maldonado-Agurto R, Dickson AJ. 2014. The endoplasmic reticulum and unfolded protein response in the control of mammalian recombinant protein production. Biotechnol Lett 36(8):1581-1593.

Iborra FJ, Pombo A, Jackson DA, Cook* PR. 1996. Active RNA polymerases are localized within discrete transcription 'factories' in human nuclei. J Cell Sci 109:1427-1436.

Ishaque A, Thrift J, Murphy JE, Konstantinov K. 2007. Over-expression of Hsp70 in BHK-21 cells engineered to produce recombinant factor VIII promotes resistance to apoptosis and enhances secretion. Biotechnol Bioeng 97(1):144-155.

Jackson DA, Hassan AB, Errington RJ, Cook PR. 1993. Visualization of focal sites of transcription within human nuclei. EMBO J 12(3):1059-1065.

Jackson DA, Iborra FJ, Manders EMM, Cook PR. 1998. Numbers and organization of RNA polymerases, nascent transcripts and transcription units in HeLa nuclei. Mol Biol Cell 9:1523-1536.

Jayapal KP, Wlaschin KF, Hu W-S, Yap MGS. 2007. Recombinant protein therapeutics from CHO cells - 20 years and counting. Chem Eng Progr 103(10):40-47.

Jensen EB, Carken S. 1990. Production of recombinant human growth hormone in Escherichia coli; expression of different precursors and physiological effects of glucose, acetate, and salts. Biotechnol Bioeng 36:1-11.

Jenuwein T, Allis CD. 2001. Translating the histone code. Science 293:1074-1080.

Johnson IS. 1983. Human insulinfrom recombinant DNA technology. Science 219:632-637.

Joly JC, Leung WS, Swartz JR. 1998. Overexpression of Escherichia coli oxidoreductases increases recombinant insulin-like growth factor-I accumulation. P Natl Acad Sci Usa 95:2773-2777.

Jones D, Kroos N, Anema R, van Montfort B, Vooys A, van der Kraats S, van der Helm E, Smits S, Schouten J, Brouwer K and others. 2003. High-level expression of recombinant IgG in the human cell line per.c6. Biotechnol Progr 19(1):163-168.

Jonkers I, Lis JT. 2015. Getting up to speed with transcription elongation by RNA polymerase II. Mol Cell Biol 16:167-177

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

Kao F-T, Puck TT. 1967. Genetics of somatic mammalian cells IV: Properties of Chinese hamster cell mutants with respect to the requirement for proline. Genetics 55:513-524.

Kao F-T, Puck TT. 1968. Gentics of somatic mammalian cells, VII. Induction and isolation of nutrional mutants in Chinese hamster cells. P Natl Acad Sci Usa 60:1275-1281.

Kao F-T, Puck TT. 1969. Genetics of somatic mammalian cells IX: Quantification of mutagenesis by physical and chemical agents. J Cell Physio 74:245-258.

Kaufman RJ, Sharp PA, Latt SA. 1983. Evolution of chromosomal regions containing transfected and amplified dihydrofolate reductase sequences. Mol Cell Biol 3(4):699-711.

Kennard ML, Goosney DL, Diane Monteith, Roe S, Fischer D, Mott J. 2009. Auditioning of CHO host cell lines using the artificial chromosome expression (ACE) technology. Biotechnol Bioeng 104(3):526-539.

Kim J-M, Kim J-S, Park D-H, Kang HS, Yoon J, Baek K, Yoon Y. 2004. Improved recombinant gene expression in CHO cells using matrix attachment regions. J Biotechnol 107:95-105.

Kim JY, Kim Y-G, Lee GM. 2012. CHO cells in biotechnology for production of recombinant proteins: current state and further potential. Appl Microbio Biot 93:917-930.

Kim M, O'Callaghan PM, Droms KA, James DC. 2011. A mechanistic understanding of production instability in CHO cell lines expressing recombinant monoclonal antibodies. Biotechnol Bioeng 108(10):2434-2446.

Kim NS, Byun TH, Lee GM. 2001. Key determinants in the occurrence of clonal variation in humanized antibody expression of CHO cells during dihydrofolate reductase mediated gene amplification. Biotechnol Progr 17(1):69-75.

Kim NS, Kim SJ, Lee GM. 1998. Clonal variability within dihydrofolate reductase‐mediated gene amplified Chinese hamster ovary cells: Stability in the absence of selective pressure. Biotechnol Bioeng 60(6):679-688.

Kim SJ, Lee GM. 1999. Cytogenetic analysis of chimeric antibody‐producing CHO cells in the course of dihydrofolate reductase‐mediated gene amplification and their stability in the absence of selective pressure. Biotechnol Bioeng 64(6):741-749.

Kim SY, Lee JH, Shin HS, Kang HJ, Kim YS. 2002. The human elongation factor 1 alpha (EF-1α) first intron highly enhances expression of foreign genes from the murine cytomegalovirus promoter. J Biotechnol 93:183-187

Kim TK, Ryu JS, Chung JY, Kim MS, Lee GM. 2000. Osmoprotective effect of glycine betaine on thrombopoietin production in hyperosmotic Chinese hamster ovary cell culture: Clonal variations. Biotechnol Progr 16(5):775-781.

Kimura H, Tao Y, Roeder RG, Cook PR. 1999. Quantitation of RNA polymerase II and its transcription factors in an hela cell: Little soluble holoenzyme but significant amounts of polymerases attached to the nuclear substructure. Mol Cell Biol 19(9):5383-5392.

Kito M, Itami S, Fukano Y, Yamana K, Shibui T. 2002. Construction of engineered CHO strains for high-level production of recombinant proteins. Appl Microbiol Biot 60(4):442-448.

Kleman MI, Oellers K, Lullau E. 2008. Optimal conditions for freezing CHO-S and HEK293‐EBNA cell lines: Influence of Me2SO, freeze density, and PEI-mediated transfection on revitalization and growth of cells, and expression of recombinant protein. Biotechnol Bioeng100(5):911-922.

Kling J. 2009. First US approval for a transgenic animal drug. Nat Biotechnol 27(4):302-304.

Kornberg RD. 1974. Chromatin Structure: A repeating unit of histones and DNA. Science 184:868-871.

Kouzarides T. 2007. Chromatin modifications and their function. Cell 128:693-705.

Kurz DJ, Decary S, Hong Y, Erusalimsky JD. 2000. Senescence-associated (beta)-galactosidase reflects an increase in lysosomal mass during replicative ageing of human endothelial cells. J Cell Sci 113 (Pt 20):3613-3622.

Kuystermans D, Al-Rubeai M. 2015. Chapter 11. Mammalian cell line selection strategies for high-producers. Animal Cell Culture DOI 10.1007/978-3-319-10320-4_11

Lattenmayer C, Loeschel M, Steinfellner W, Trummer E, Mueller D, Schriebl K, Vorauer-Uhl K, Katinger H, Kunert R. 2006. Identification of transgene integration loci of different highly expressing recombinant CHO cell lines by FISH. Cytotechnology 51(3):171-182.

Lee DY, Hayes JJ, Pruss D, Wolffe AP. 1993. A positive role for histone acetylation in transcription factor access to nucleosomal DNA. Cell 72(1):7384.

Lewis NE, Liu X, Li Y, Nagarajan H, Yerganian G, O'Brien E, Bordbar A, Roth AM, Rosenbloom J, Bian C and others. 2013. Genomic landscapes of Chinese

hamster ovary cell lines as revealed by the Cricetulus griseus draft genome. Nat Biotechnol 31(8):759-765.

Li G, Ruan X, Auerbach RK, Sandhu KS, Zheng M, Wang P, Poh HM, Goh Y, Lim J, Zhang J and others. 2012. Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. Cell 148:84-98.

Li H, Sethuraman N, Stadheim TA, Zha D, Prinz B, Ballew N, Bobrowicz P, Choi B-K, Cook WJ, Cukan M and others. 2006. Optimization of humanized IgGs in glycoengineered Pichia pastoris. Nat Biotechnol 24:210-215.

Lianchun F, Ibrahim K, Lara EK, Jeffery LL, Daniel MB, Christopher CF. 2013. Development of a highly-efficient CHO cell line generation system with engineered SV40E promoter. J Biotechnol 168(4):652-658.

Ling JQ, Li T, Hu JF, Vu TH, Chen HL, Qiu XW, Cherry AM, Hoffman AR. 2006. CTCF Mediates interchromosomal colocalization between Igf2/H19 and Wsb1/Nf1. Science 312(5771):269-272.

Lissemore J, Jankowski J, Thomas C, Mascotti D, deHaseth P. 2000. Protein as a quantitative reporter of relative promoter activity in E. coli. BioTechniques 28:82-89

Liu Y, Yi X, Zhuang Y, Zhang S. 2015. Limitations in the process of transcription and translation inhibit recombinant human chorionic gonadotropin expression in CHO cells. J Biotechnol 204:63-69.

Luger K, Mäder AW, Richmond RK, Sargent DF, Richmond TJ. 1997. Crystal structure of the nucleosome core particle at 2.8A° resolution. Nature 389:251-260.

Luo W, Johnson AW, Bentley DL. 2006. The role of Rat1 in coupling mRNA 3 -end processing to transcription termination: Implications for a unified allosteric torpedo model. Gene Develop 20(8):954-965.

Maeshima K, Hihara S, Eltsov M. 2010. Chromatin structure: Does the 30-nm fibre exist in vivo? Curr Opin Cell Biol 22:291-297

Mariati, Koh EY, Yeo JH, Ho SC, Yang Y. 2014a. Toward stable gene expression in CHO cells. Bioengineered 5(5):340-345.

Mariati, Yeo JH, Koh EY, Ho SC, Yang Y. 2014b. Insertion of core CpG island element into human CMV promoter for enhancing recombinant protein expression stability in CHO cells. Biotechnol Progr 30(3):523-534.

Martin S, Pombo A. 2003. Transcription factories: quantitative studies of nanostructures in the mammalian nucleus. Chromosome Res 11:461-470.

Matasci M, Baldi L, Hacker DL, Wurm FM. 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(9):2141-2150.

Mellor J, Dudek P, Clynes D. 2008. A glimpse into the epigenetic landscape of gene regulation. Curr Opin Genet Dev 18:116-122.

Merrihew RV, Sargent RG, Wilson JH. 1995. Efficient modification of the APRT gene by FLP/FRT site-specific targeting. Somat Cell Molec Gen 21(5):299-307.

Mirkovitch J, Mirault M-E, Laemmli UK. 1984. Organization of the higher-order chromatin loop: Specific DNA attachment sites on nuclear scaffold. Cell 39:223-232.

Montavon T, Soshnikova N, Mascrez B, Joye E, Thevenet L, Splinter E, Laat Wd, Spitz F, Duboule D. 2011. A regulatory archipelago controls hox genes transcription in digits. Cell 147:1132-1145.

Mutskov V, Felsenfeld G. 2004. Silencing of transgene transcription precedes methylation of promoter DNA and histone H3 lysine 9. EMBO J 23(1):138-149.

Nair AR, Xie Jinger X, Hermiston TW. 2011. Effect of different UCOE-promoter combinations in creation of engineered cell lines for the production of Factor VIII. BMC Res Notes 4(1):178.

Nehlsen K, Schucht R, Gama-Norton Ld, Krömer W, Baer A, Cayli A, Hauser Hr, Wirth D. 2009. Recombinant protein expression by targeting pre-selected chromosomal loci. BMC Biotechnol 9(100).

Nikolov DB, Burley SK. 1997. RNA polymerase II transcription initiation: a structural view. P Natl Acad Sci Usa 94:15-22

Nivitchanyong T, Martinez A, Ishaque A, Murphy JE, Konstantinov K, Betenbaugh MJ, Thrift J. 2007. Anti-apoptotic genes Aven and E1B-19K enhance performance of BHK cells engineered to express recombinant factor VIII in batch and low perfusion cell culture. Biotechnol Bioeng 98(4):825-841.

O'Gorman S, Fox DT, Wahl GM. 1991. Recombinase-mediated gene activation and site-specific integration in mammalian cells. Science 251(4999):1351-1355

Omasa T, Cao Y, Park JY, Takagi Y, Kimura S, Yano H, Honda K, Asakawa S, Shimizu N, Ohtake H. 2009. Bacterial artificial chromosome library for genome-wide analysis of Chinese hamster ovary cells. Biotechnol Bioeng 104(5):986-994.

Ono T, Sonta S. 2001. Chromosome map of cosmid clones constructed with Chinese hamster genomic DNA. Cytogenet Cell Genet 95:97-102.

Osborne CS, Chakalova L, Brown KE, Carter D, Horton A, Debrand E, Goyenechea B, Mitchell JA, Lopes S, Reik W and others. 2004. Active genes dynamically colocalize to shared sites of ongoing transcription. Nat Genet 36(10):1065-1071.

Osborne CS, Chakalova L, Mitchell JA, Horton A, Wood AL, Bolland DJ, Corcoran AE, Fraser P. 2007. Myc dynamically and preferentially relocates to a transcription factory occupied by Igh. PLoS Biol 5(8):1763-1772.

Osterlehner A, Simmeth S, Gopfert U. 2011. Promoter methylation and transgene copy numbers predict unstable protein production in recombinant Chinese hamster ovary cell lines. Biotechnol Bioeng 108(11):2670-2681.

Page C. 2012. Investigating the consequences of exogenous expression of unfolded protein response components in recombinant Chinese hamster ovary cells. PhD Thesis, Faculty of Life Sciences, The University of Manchester.

Pallavicini MG, DeTeresa PS, Rosette C, Gray JW, Wurm FM. 1990. Effects of methotrexate on transfected DNA stability in mammalian cells. Mol Cell Biol 10(1):401-404.

Pandit S, Wang D, Fu XD. 2008. Functional integration of transcriptional and RNA processing machineries. Curr Opin Cell Biol 20:260–265

Parada LA, McQueen PG, Misteli T. 2004. Tissue-specific spatial organization of genomes. Genome Biol 2004, 5(7):R44.1-R44.9.

Park SK, Xiang Y, Feng X, Garrard WT. 2014. Pronounced cohabitation of active immunoglobulin genes from three different chromosomes in transcription factories during maximal antibody synthesis. Gene Dev 28(11):1159-1164.

Pau MG, Ophorst C, Koldijk MH, Schouten G, Mehtali M, Uytdehaag F. 2001. The human cell line PER.C6 provides a new manufacturing system for the production of influenza vaccines. Vaccine 19(17-19):2716-2721.

Paulin RP, Ho T, Balzer HJ, Holliday R. 1998. Gene silencing by DNA methylation and dual inheritance in Chinese hamster ovary cells. Genetics 149(2):1081-1088.

Peserico A and Simone C. 2011. Physical and functional HAT/HDAC interplay regulates protein acetylation balance. J Biomed Biotechnol 371832: 1-10

Pham PL, Kamen A, Durocher Y. 2006. Large-scale transfection of mammalian cells for the fast production of recombinant protein. Mol Biotechnol 34(2):225-237.

Phatnani HP, Greenleaf AL. 2006. Phosphorylation and functions of the RNA polymerase II CTD. Gene Dev 20:2922-2936

Piatkevich KD, Verkhusha VV. 2011. Guide to red fluorescent proteins and biosensors for flow cytometry. Method Cell Biol 102:431-461

Pilbrough W, Munro TP, Gray P. 2009. Intraclonal protein expression heterogeneity in recombinant CHO cells. PloS One 4(12):e8432

Pontiller J, Gross S, Thaisuchat H, Hesse F, Ernst W. 2008. Identification of CHO endogenous promoter elements based on a genomic library approach. Mol Biotechnol 39(2):135-139.

Pontiller J, Maccani A, Baumann M, Klancnik I, Ernst W. 2010. Identification of CHO endogenous gene regulatory elements. Mol Biotechnol 45(3):235-240.

Porter AJ, Racher AJ, Preziosi R, Dickson AJ. 2010. Strategies for selecting recombinant CHO cell lines for cGMP manufacturing: Improving the efficiency of cell line generation. Biotechnol Progr 26(5):1455-1464.

Pradhan S, Bacolla A, Wells RD, Roberts RJ. 1999. Recombinant human DNA (cytosine-5) methyltransferase. J Biol Chem 274(46):33002-33010.

Price DH. 2000. P-TEFb, a cyclin-dependent kinase controlling elongation by RNA polymerase II. Mol Cell Biol 20(8):2629-2634

Pybus LP, Dadehbeigi N, Saunders FL, Porter AJ. 2014. Delivering 'fit for purpose' biomanufacturing CHO cell lines. Fujifilm Dyosinth Biotchnologies www.fujifilmdiosynth.com/pdfs/Delivering%20fit%20for%20purpose%20CHO%20cell%20lines%20(1).pdf (Accessed on 25th March 2015).

Qiu J, Swartz JR, Georgiou G. 1998. Expression of active human tissue-type plasminogen activator in Escherichia coli. Appl Environ Microb 64(12):4891-4896.

Rader RA. 2008. (Re)defining biopharmaceutical. Nat Biotechnol 26(7):743-751.

Ramirez OT, Mutharasan R. 1990. Cell cycle‐and growth phase‐dependent variations in size distribution, antibody productivity, and oxygen demand in hybridoma cultures. Biotechnol Bioeng 36:839-848

Ratner M. 2009. Pharma swept up in biogenerics gold rush. Nat Biotechnol 27(4):299-301

Razin A. 1998. CpG methylation, chromatin structure and gene silencing - a three-way connection. EMBO J 17(17):4905-4908.

Razin SV, Iarovaia OV, Sjakste N, Sjakste T, Bagdoniene L, Rynditch AV, Eivazova ER, Lipinski M, Vassetzky YS. 2007. Chromatin domains and regulation of transcription. J Mol Biol 369:597-607.

Regha K, Sloane MA, Huang R, Pauler FM, Warczok KE, Melikant Bz, Radolf M, Martens JHA, Schotta G, Jenuwein T, Barlow DP. 2007. Active and repressive chromatin are interspersed without spreading in an imprinted gene cluster in the mammalian genome. Mol Cell 27:353-366.

Robinson PJJ, Fairall L, Huynh VAT, Rhodes D. 2006. EM measurements define the dimensions of the ''30-nm'' chromatin fiber: Evidence for a compact, interdigitated structure. P Natl Acad Sci Usa 103(17):6506-6511.

Ruiz JC, Wahl GM. 1990. Chromosomal destabilization during gene amplification. Mol Cell Biol 10(6):3056-3066.

Rundlett SE, Carmen AA, Suka N, Turner BM, Grunstein M. 1998. Transcriptional repression by UME6 involves deacetylation of lysine 5 of histone H4 by RPD3. Nature 392:831-835.

Sainsbury S, Bernecky C, Cramer P. 2015. Structural basis of transcription initiation by RNA polymerase II. Mol Cell Bio 16:129-143

Sanyal A, Baù D, Martí-Renom MA, Dekker J. 2011. Chromatin globules: a common motif of higher order chromosome structure? Curr Opin Cell Bio 23:325-333.

Sarah S, Carrie B, Patrick C. 2015. Structural basis of transcription initiation by RNA polymerase II. Nat Mol Cell Biol 16(3):129-143.

Sauer B. 1994. Site-specific recombination: developments and applications. Current Opinion in Biotechnology 5(5):521-527.

Schalch T, Duda S, Sargent DF, Richmond TJ. 2005. X-ray structure of a tetranucleosome and its implications for the chromatin fibre. Nature 436(7):138-141.

Schlake T, Bode J. 1994. Use of mutated FLP recognition target (FRT) sites for the exchange of expression cassettes at defined chromosomal loci. Biochemistry 33(43):12746-12751.

Schoenfelder S, Sexton T, Chakalova L, Cope NF, Horton A, Andrews S, Kurukuti S, Mitchell JA, Umlauf D, Dimitrova DS and others. 2010. Preferential associations between co-regulated genes reveal a transcriptional interactome in erythroid cells. Nat Genet 42(1):53-61.

Shaner NC, Campbell RE, Steinbach PA, Giepmans BNG, Palmer AE1, Tsien RY 2004 Improved monomeric red, orange and yellow fluorescent proteins derived from Discosoma sp. red fluorescent protein. Nat Biotechnol 22(12):1567-1572

Shav-Tal Y, Blechman J, Darzacq X. 2005. Dynamic sorting of nuclear components into distinct nucleolar caps during transcriptional inhibition. Mol Biol Cell 16:2395–2413

Shopland LS, Johnson CV, Byron M, McNeil J, Lawrence JB. 2003. Clustering of multiple specific genes and gene-rich R-bands around SC-35 domains: evidence for local euchromatic neighborhoods. J Cell Biol 162(6):981-990.

Sieger M, Garweg G, Scwarzacher HG. 1971. Constitutive heterochromatin in microtus agrestis: Binding of actinomycin-d and transcriptional inactivity. Chromosoma 35(1):84-98.

Sigma-Aldrich. CHOZN CHO K1 DHFR Cell Line Testing - Techniccal Information. www.sigmaaldrich.com/content/dam/sigma-aldrich/docs/SAFC/General_Information/1/chozn_dhfr_testing.pdf (Accessed on 25th March 2015).

Smallwood A, Estève P-O, Pradhan S, Carey M. 2007. Functional cooperation between HP1 and DNMT1 mediates gene silencing. Gene Dev 21:1169-1178.

Song F, Chen P, Sun D, Wang M, Dong W, Liang D, Xu RM, Zhu P, Li G. 2014. Cryo-EM study of the chromatin fiber reveals a double helix twisted by tetranucleosomal units. Science 344:376-380

Soo KM, Lee GM. 2008. Use of Flp-mediated cassette exchange in the development of a CHO cell line stably producing erythropoietin. J Microbiol Biotechn 18(7):1342-1351.

Soukharev S, Hammond D, Ananyeva NM, Anderson JA, Hauser CA, Pipe S, Saenko EL. 2002. Expression of factor VIII in recombinant and transgenic systems. Blood Cell Mol Dis 28(2):234-248.

Spencer S, Gugliotta A, Koenitzer J, Hauser H, Wirth D. 2015. Stability of single copy transgene expression in CHOK1 cells is affected by histone modifications but not by DNA methylation. J Biotechnol 195:15-29.

Spilianakis CG, Flavell RA. 2004. Long-range intrachromosomal interactions in the T helper type 2 cytokine locus. Nat Immuno 5(10):1017-1027.

Spilianakis CG, Lalioti MD, Town T, Lee GR, Flavell RA. 2005. Interchromosomal associations between alternatively expressed loci. Nature 435:637-645.

Tarchini B, Duboule D. 2006. Control of hoxd genes' collinearity during early limb development. Dev Cell 10:93-103.

Tessarz P, Kouzarides T. 2014. Histone core modifications regulating nucleosome structure and dynamics. Nat Mol Cell Biol 15:703-708

Tiwari VK, McGarvey KM, Licchesi JDF, Ohm JE, Herman JG, Schübeler D, Baylin SB. 2008. PcG proteins, dna methylation, and gene repression by chromatin looping. PLoS Biol 6(12):e306.

Tjio JH, Puck TT. 1958. Genetics of somatic mammalian cells II: Chromosomal constitution of cells in tissue culture. J Exp Med 108(2):259-268.

Tolhuis B, Palstra R-J, Splinter E, Grosveld F, Laat Wd. 2002. Looping and interaction between hypersensitive sites in the active ß-globin locus. Mol Cell 10:1453-1465.

Tornøe J, Kusk P, Johansen TE, Jensen PR. 2002. Generation of a synthetic mammalian promoter library by modification of sequences spacing transcription factor binding sites. Gene 297:21-32

Tremethick DJ. 2007. Higher-order structures of chromatin: The elusive 30 nm fiber. Cell 128:651-654.

Trimble RB, Lubowski C, Hauer CR, Stack R, McNaughton L, Gemmill TR, Kumar3 SA. 2004. Characterization of N- and O-linked glycosylation of recombinant human bile salt±stimulated lipase secreted by Pichia pastoris. Glycobiology 14(3):265-274.

Vaissière T, Sawan C, Herceg Z. 2008. Epigenetic interplay between histone modifications and DNA methylation in gene silencing. Mutat Res 659:40-48

van Gemert AMC, Jost CR, van der Laan AMA, Dirks RW, Reiber JHC, Tanke HJ, Lelieveldt BPF. 2009. Identification of cellular dynamic patterns resulting from repetitive photobleaching using independent component analysis. IEEE International Symposium on Biomedical Imaging 1382-1385

Venkatesh S and Workman JL. 2015. Histone exchange, chromatin structure and the regulation of transcription. Nature Rev Mol Cell Biol 16: 178-189

Walsh G. 2002. Biopharmaceuticals and biotechnology medicines: an issue of nomenclature. Eur J Pharm Sci 15(2):135-138.

Walsh G. 2014. Biopharmaceutical benchmarks 2014. Nat Biotechnol 32(10):992-1000.

Walsh G, Jefferis R. 2006. Post-translational modifications in the context of therapeutic proteins. Nat Biotechnol 24(10):1241-1252.

Watson CM, Trainor PA, Radziewic T, Pelka GJ, Zhou SX, Parameswaran M, Quinlan GA, Gordon M, Sturm K, Tam PP. 2008. Application of lacZ transgenic mice to cell lineage studies. Meth Mol Biol (Clifton, N.J.) 461:149-164.

Weil TT, Parton RM, Davis I. 2010. Making the message clear: Visualizing mRNA localization. Trends Cell Biol 20(7):380-390.

West S, Proudfoot NJ, Dye MJ. 2008. Molecular dissection of mammalian RNA polymerase II transcriptional termination. Mol Cell 29:600-610

Wiberg FC, Rasmussen SK, Frandsen TP, Rasmussen LK, Tengbjerg K, Coljee VW, Sharon J, Yang C-Y, Bregenholt S, Nielsen LS and others. 2006. Production of target-specific recombinant human polyclonal antibodies in mammalian cells. Biotechnol Bioeng 94(2):396-405.

Williams S, Mustoe T, Mulcahy T, Griffiths M, Simpson D, Antoniou M, Irvine A, Mountain A, 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(1):17.

Wilson C, Bellen HJ, Gehering WJ. 1990. Position effects on eukaryotic gene expression. Annu Rev Cell Biol 6:679-714.

Wolfel J, Wissing S, Hertel S, Kraus S, Schiedner G, Faust N. 2013. Identifying & isolating high-producer clones: Rapid method for use in human production systems for biotherapeutics. Genet Eng Biotechnol News 33(18)

Wolffe AP, Matzke MA. 1999. Epigenetics: Regulation through repression. Science 286:481-486.

Wurm FM. 2004. Production of recombinant protein therapeutics in cultivated mammalian cells. Nat Biotechnol 22(11):1393-1398

Xu M, Cook PR. 2008. Similar active genes cluster in specialized transcription factories. J Cell Biol 181(4):615-623.

Xu X, Nagarajan H, Lewis NE, Pan S, Cai Z, Liu X, Chen W, Xie M, Wang W, Hammond S and others. 2011. The genomic sequence of the Chinese hamster ovary (CHO)-K1 cell line. Nat Biotechnol 29:735-741.

Xu ZL, Mizuguchi H, Ishii-Watabe A, Uchida E. 2001. Optimization of transcriptional regulatory elements for constructing plasmid vectors. Gene 272:149-156

Yang F, O'Brien PCM, Ferguson-Smith MA. 2000. Comparative chromosome map of the laboratory mouse and Chinese hamster defined by reciprocal chromosome painting. Chromosome Res 8:219-227.

Yang Y, Mariati, Chusainow J, Yap MG. 2010. DNA methylation contributes to loss in productivity of monoclonal antibody-producing CHO cell lines. J Biotechnol 147(3-4):180-185.

Yao J, Munson KM, Webb WW, Lis JT. 2006. Dynamics of heat shock factor association with native gene loci in living cells. Nature 442(7106):1050-1053.

Ye J, Alvin K, Latif H, Hsu A, Parikh V, Whitmer T, Tellers M, de la Cruz Edmonds MC, Ly J, Salmon P and others. 2010. Rapid protein production using CHO stable transfection pools. Biotechnol Progr 26(5):1431-1437.

Yoo EM, Chintalacharuvu KR, Penichet ML. 2002. Myeloma expression systems. J Immunol Methods 261:1-20

Yoshikawa T, Nakanishi F, Itami S, Kameoka D, Omasa T, Katakura Y, Kishimoto M, Suga K. 2000a. Evaluation of stable and highly productive gene amplified CHO cell line based on the location of amplified genes. Cytotechnology 33(1-3):37-46.

Yoshikawa T, Nakanishi F, Ogura Y, Oi D, Omasa T, Katakura Y, Kishimoto M, Suga K. 2000b. Amplified gene location in chromosomal dna affected recombinant protein production and stability of amplified genes. Biotechnol Progr 16(5):710-715.

Yunger S, Rosenfeld L, Garini Y, Shav-Tal Y. 2010. Single-allele analysis of transcription kinetics in living mammalian cells. Nat Methods 7(8):631-633.

Yunger S, Rosenfeld L, Garini Y, Shav-Tal Y. 2013. Quantifying the transcriptional output of single alleles in single living mammalian cells. Nat Protoc 8(2):393-408.

Zarrin AA, Malkin L, Fong I, Luk KD, Ghose A. 1999. Comparison of CMV, RSV, SV40 viral and Vλ1 cellular promoters in B and T lymphoid and non-lymphoid cell lines. Biochim Biophys Acta 1446:135-139

Zhang F, Frost AR, Blundell MP, Bales O, Antoniou MN, Thrasher AJ. 2010. A ubiquitous chromatin opening element (UCOE) confers resistance to DNA methylation-mediated silencing of lentiviral vectors. Mol Ther 18(9):1640-1649.

Zhou H, Liu Z-g, Sun Z-w, Huang Y, Yu W-y. 2010. Generation of stable cell lines by site-specific integration of transgenes into engineered Chinese hamster ovary strains using an FLP-FRT system. J Biotechnol 147:122-129.

Zimber A, Nguyen QD, Gespach C. 2004. Nuclear bodies and compartments: functional roles and cellular signalling in health and disease. Cell Signal 16:1085-1104

Appendix 1

REAGENTS AND CHEMICALS

BDH Chemicals Ltd., UK

      4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES)

Bio-Rad Laboratories Inc., USA

      Bradford reagent

Bioline, UK

      Agarose powder

      BIOTAQ™ DNA polymerase

      dNTP Set

      HyperLadder™ I DNA ladder to 10Kbp

      HyperLadder™ V DNA ladder to 500bp

      SensiFast™ SYBR® probe Hi-ROX kit

      Tetro cDNA synthesis kit

Eurofins Genomics, Germany

      Custom made DNA oligonucleotides

      Sequencing reactions

Life Technologies™, USA

      CD CHO medium

      Competent *E.coli* cells (DH5α strain)

      Glutamax

      Hygromycin B

      Trizol® reagent

      Zeocin

NBS Biologicals, UK

      SafeView™ nucleic acid stain

New England Biolabs®, UK

      1kB DNA ladder

      Restriction enzymes

DNase1 enzyme (1x)

DNase1 reaction buffer (10x)

QIAGEN Ltd., UK

QIAprep Spin Miniprep Kit

Plasmid Midi Kit

HiSpeed® Plasmid Maxi kit

Roche, Switzerland

Chlorophenol red-ß-D-galactopyranoside (CPRG)

Restriction enzymes

T4 DNA ligase

Sigma-Aldrich Company Ltd., UK

Ampicilin

β-mercaptoethanol

Dimethyl sulfoxide (DMSO)

Phosphate buffered saline (PBS)

Sodium dodecyl sulfate (SDS)

Tris Base

Tris-HCL

Trypan blue

Tween-20

Thermo Fisher Scientific, USA

Ethanol

Ethylenediaminetetraacetic acid (EDTA)

Glacial acetic acid

Glycerol

Glycine

Isopropanol

Methanol

Sodium Chloride

Appendix 2

Figure A2.1. Site-directed integration expression vectors maps.
A – pcDNA5/FRT **B** – pS/FRT; **C** – pcDNA5/FRT/MS2; **D**– pS/FRT/MS2; **E** – pcDNA5/FRT/MS2-YFP**; F** – pS/FRT/MS2-YFP; **G** – pcDNA5/FRT/MS2-mStrawberry; **H**– pS/FRT/MS2-mStrawberry; **I** – pcDNA5/FRT-GFP; **J** – pS/FRT-GFP (pCMV – CMV promoter sequence; pSV40 – SV40 promoter sequence; FRT – FRT sequence; GFP – GFP gene sequence; YFP – YFP gene sequence; mStawberry – mStrawberry gene sequence; MS2 – 22 MS2 loop gene sequence; Hygromycin – Hygromycin resistance gene sequence; T7 – T7 forward prime site; BGH pA – BGH polyadenilation sequence; SV40 pA – SV40 polyadenilation sequence; pUC ori – pUC origin of replication; Ampicillin – Ampicillin (*bla*) resistance gene sequence)

A

Nhe I
Pme I
Afl II
Hind III
Asp718 I
Kpn I
BamH I
BstX I
EcoR V
BstX I
Not I
Xho I
Apa I
Pme I

T7

P CMV

BGH pA

FRT

Hygromycin

SV40 pA

pUC ori

Ampicillin

pcDNA5/FRT
5070 bp

B

Pme I
Afl II
Hind III
Asp718 I
Kpn I
BamH I
BstX I
EcoR V
BstX I
Not I
Xho I
Apa I
Pme I

P SV40

BGH pA

FRT

Hygromycin

SV40 pA

pUC ori

Ampicillin

pS/FRT
4733bp

C

pcDNA5/FRT/MS2
6520bp

D

pS/FRT/MS2
6183bp

E

pcDNA5/FRT/MS2 YFP 7316bp

YFP

P CMV · MS2 · BGH pA · FRT · Hygromycin · SV40 pA · pUC ori · Ampicillin

F

pS/FRT/MS2 YFP 6979bp

YFP

P SV40 · MS2 · BGH pA · FRT · Hygromycin · SV40 pA · pUC ori · Ampicillin

G

mStrawberry

P CMV — MS2 — BGH pA — FRT — Hygromycin — SV40 pA — pUC ori — Ampicillin

pcDNA5/FRT/MS2
mStrawberry
7313bp

H

mStrawberry

P SV40 — MS2 — BGH pA — FRT — Hygromycin — SV40 pA — pUC ori — Ampicillin

pS/FRT/MS2
mStrawberry
6973bp

I

GFP

P CMV
BGH pA
FRT
Hygromycin
Ampicillin
pcDNA5/FRT
GFP
5793bp
SV40 pA
pUC ori

J

GFP

P SV40
BGH pA
FRT
Hygromycin
Ampicillin
pS/FRT
GFP
5456bp
SV40 pA
pUC ori

Figure A2.2. Flp-In cloning and Flp-recombinase vector maps.
A – pFRT/lacZeo; **B** – pOG44 (pSV40 – SV40 promoter sequence; pCMV – CMV promoter sequence; FRT – FRT sequence; ATG – Initiation codon sequence; lacZ-Zeocin – LacZ-Zeocin fusion gene sequence [LacZ ORF gene (no ATG) sequence + Zeocin resistance gene (no ATG) sequence]; SV40 pA – SV40 polyadenilation sequence; pUC ori – pUC origin of replication; Ampicillin – Ampicillin (*bla*) resistance gene sequence; Intron – Synthetic intron sequence; FLP – Flp-recombinase gene sequence)

A
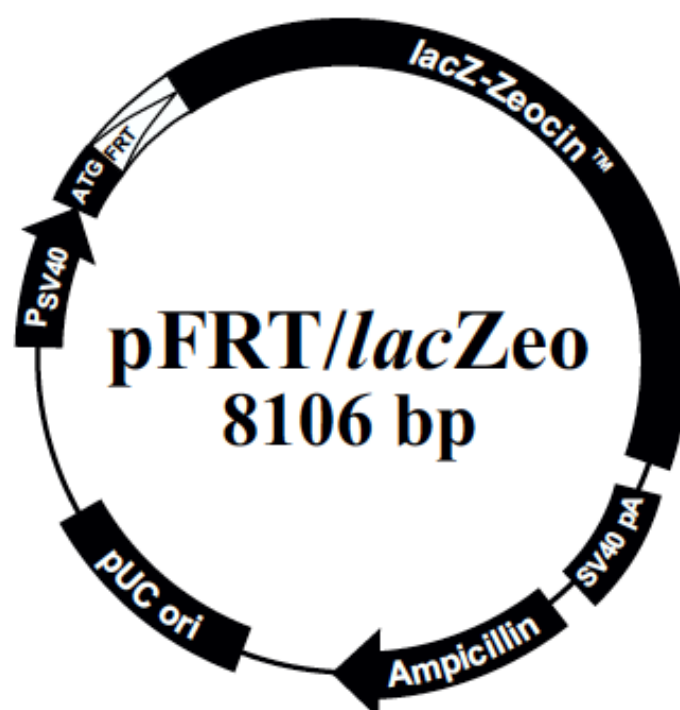
B
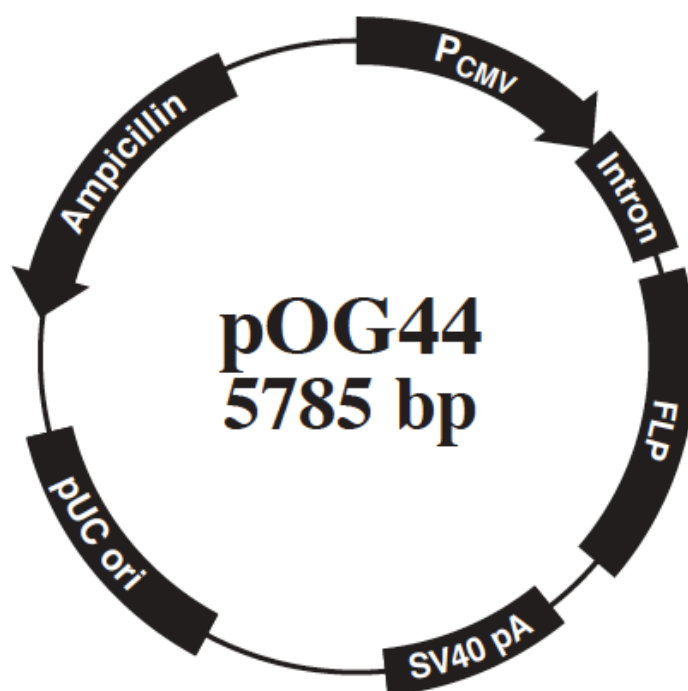
Figure A2.3. pMS2-GFP and p901-GFP vector maps.
A- pMS2-GFP (legend highlighted in figure); **B**- p901-GFP (hCMV – Human CMV promoter sequence; GFP – GFP gene sequence; IRES – Internal ribosome entry site; DHFR – Dihydrofolate reducatse gene sequence; pA – polyadenilation sequence; pMB1 ori – pMB1 origin of replication; bla – Ampicillin (*bla*) resistance gene sequence).

A
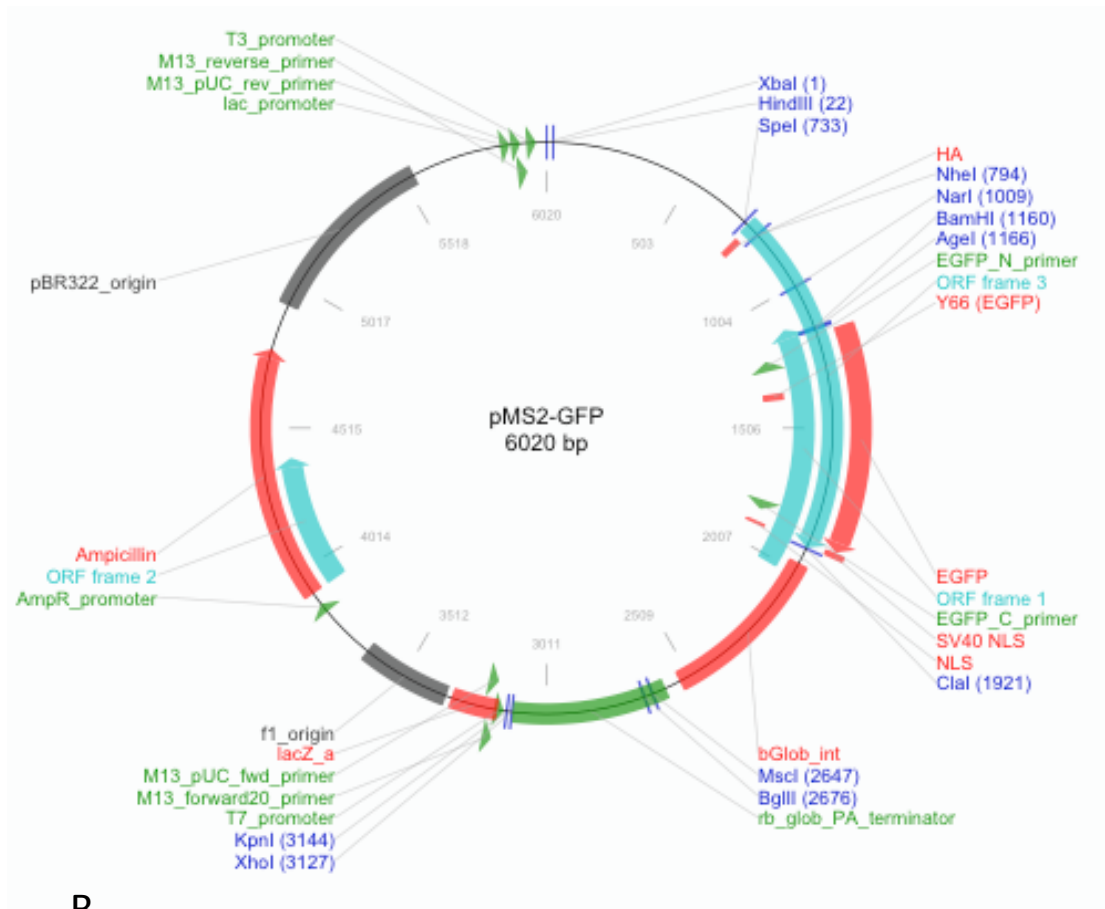
B
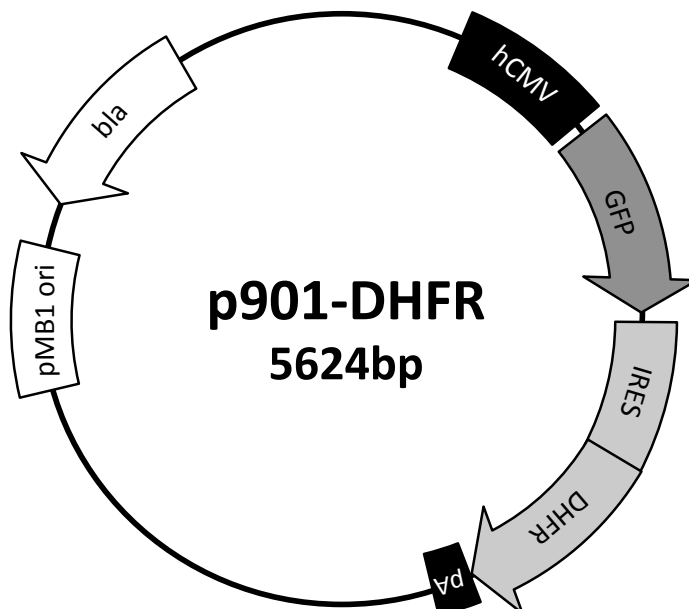
Appendix 3

Figure A3.1. Flow cytometry histograms of CHO-SV40-YFP and CHO-CMV-YFP independent transfections (biological replicates).

A 1 ml culture sample was collected and YFP fluorescence intensity was measured using the BD Accuri™ C6 flow cytometer, as described in Section 2.6.2.3. The YFP was excited using a 488nm excitation laser and fluorescence data was acquired using a 530/30nm bandpass filter and analysed using the BD Accuri™ C6 software. The auto-fluorescence of the CHO-FRT mini-pool was measured in every experiment as a control for non-expressing cells. (**A** – CHO-SV40-YFP 1 polyclonal and CHO-CMV-YFP 1 Polyclonal (Day 0) **;** **B**– CHO-SV40-YFP 1 polyclonal and CHO-CMV-YFP 1 Polyclonal (Day 32) l; **C** – CHO-SV40-YFP 1 Top 10% and CHO-CMV-YFP 1 Top 10% (Day 0); **D**– CHO-SV40-YFP 1 Top 10% and CHO-CMV-YFP 1 Top 10% (Day 32); **E** – CHO-SV40-YFP 1 polyclonal and CHO-CMV-YFP 1 Polyclonal (Day 0) **;** **F**– CHO-SV40-YFP 108 polyclonal and CHO-CMV-YFP 108 Polyclonal (Day 32); **G** – CHO-SV40-YFP 108 Top 10% and CHO-CMV-YFP 108 Top 10% (Day 0); **H**– CHO-SV40-YFP 108 Top 10% and CHO-CMV-YFP 108 Top 10% (Day 32), (▬CHO-FRT mini-pool; ▬Biological replicate 1; ▬ Biological replicate 2
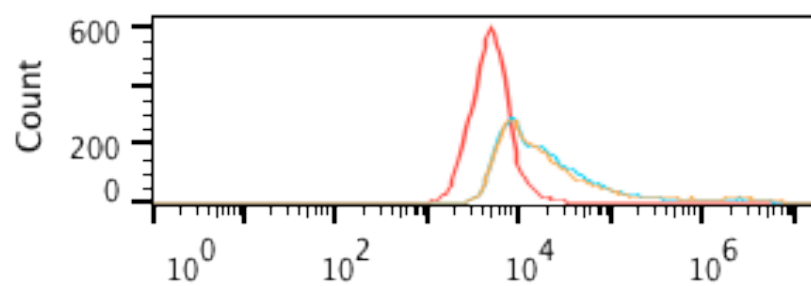
A

SV40-YFP 1.1 Polyclonal
Day 0

CMV-YFP 1.2 Polyclonal
Day 0
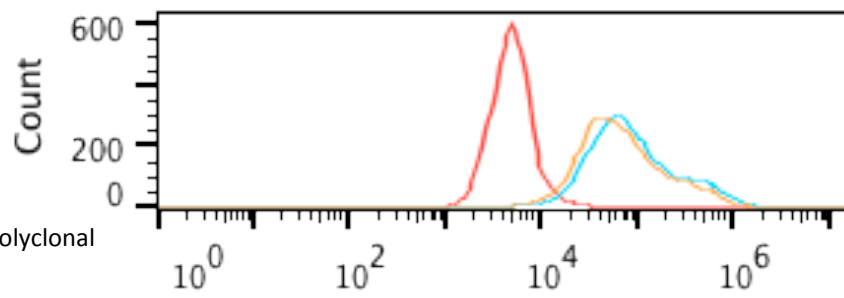
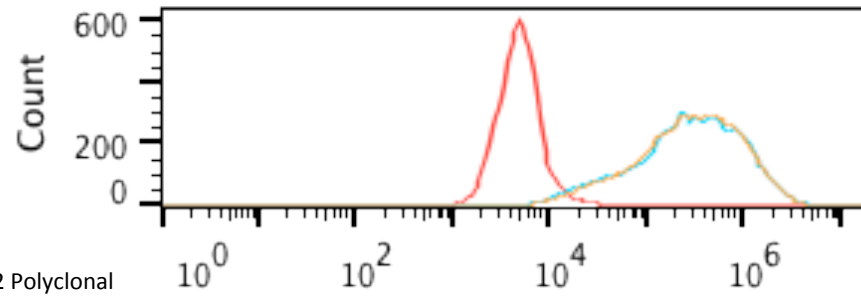SV40-YFP 1.2 Polyclonal
Day 0
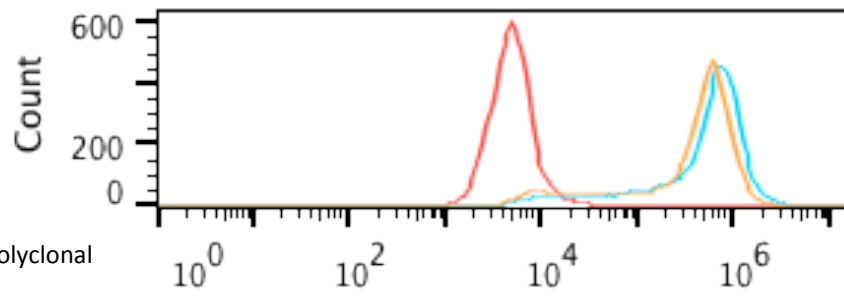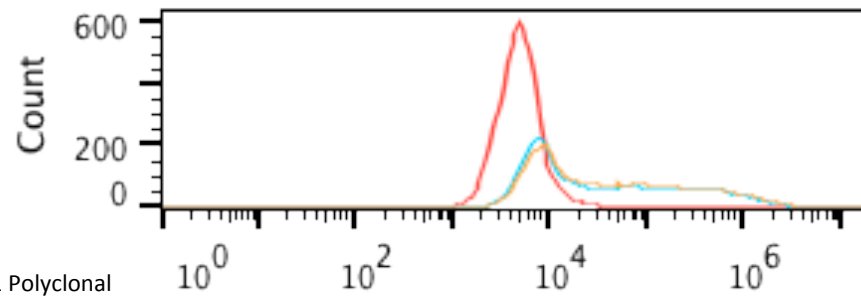
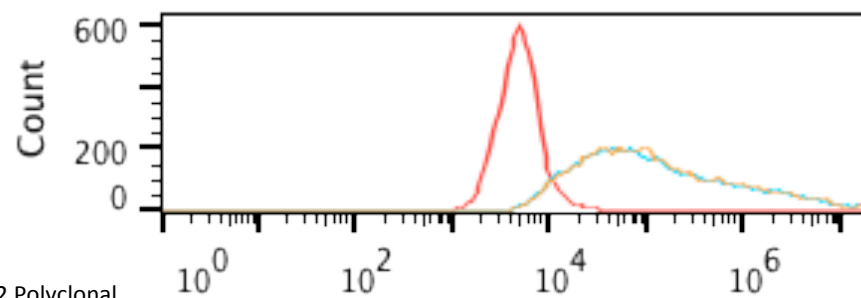CMV-YFP 1.1 Polyclonal
Day 0
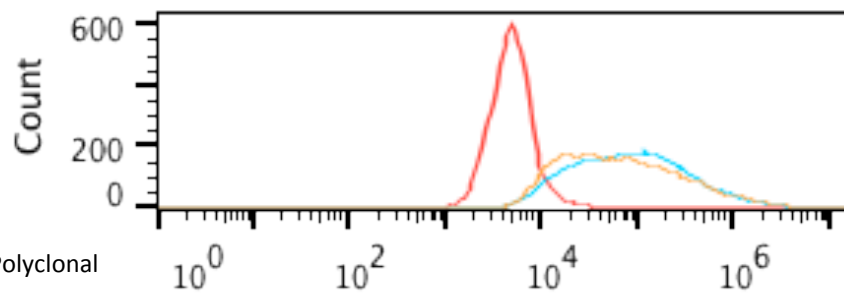
CMV-YFP 1.2 Polyclonal
Day 0
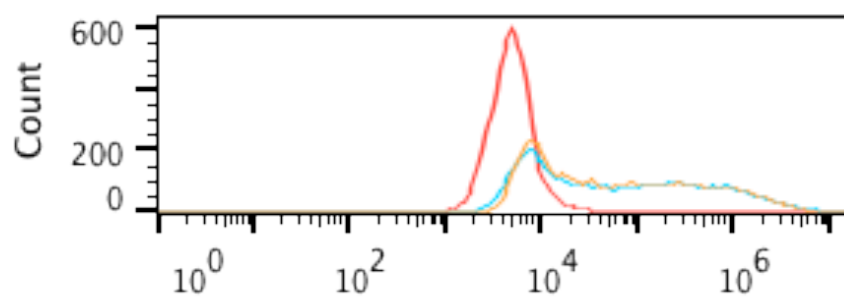
B

SV40-YFP 1.1 Polyclonal
Day 32
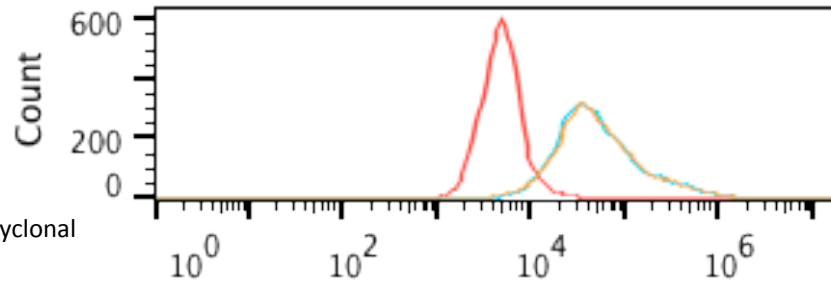
SV40-YFP 1.2 Polyclonal
Day 32
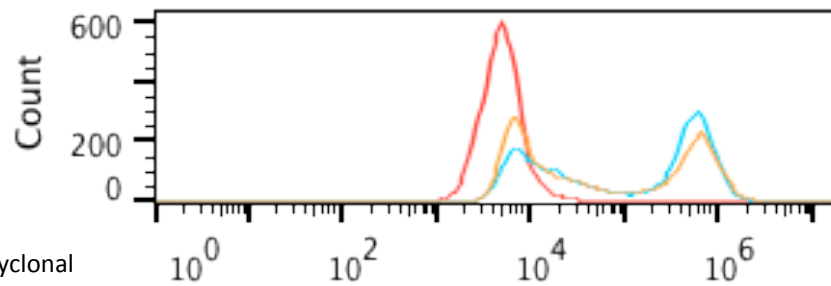
CMV-YFP 1.1 Polyclonal
Day 32
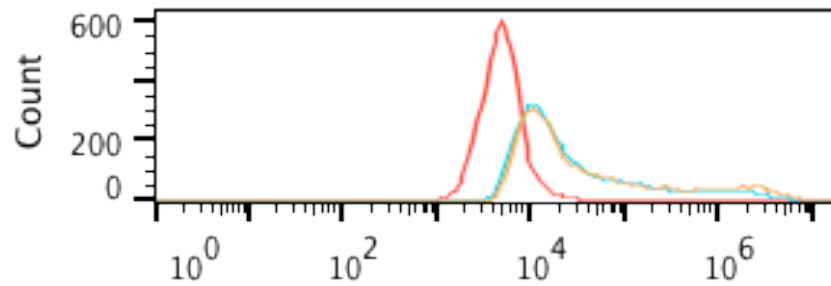

CMV-YFP 1.2 Polyclonal
Day 32
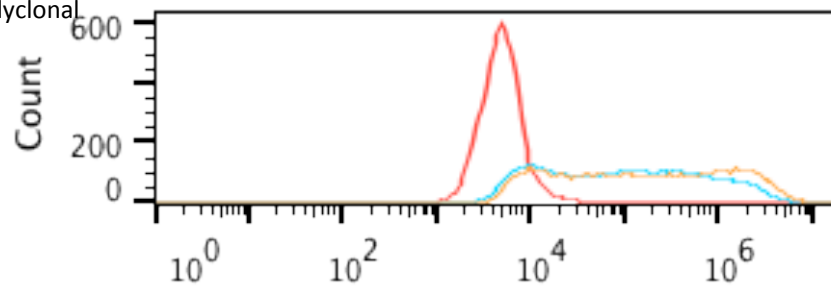
C

SV40-YFP 1.1 Top 10%
Day 0
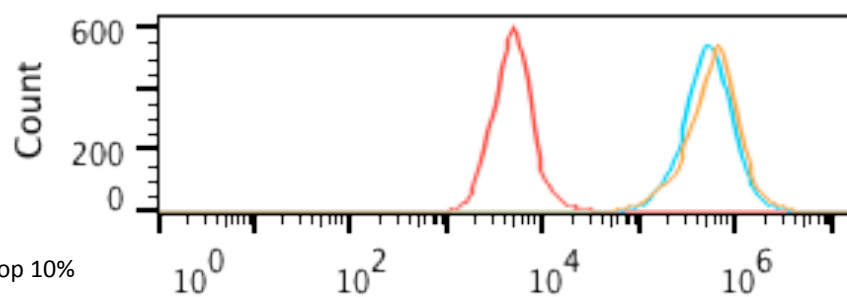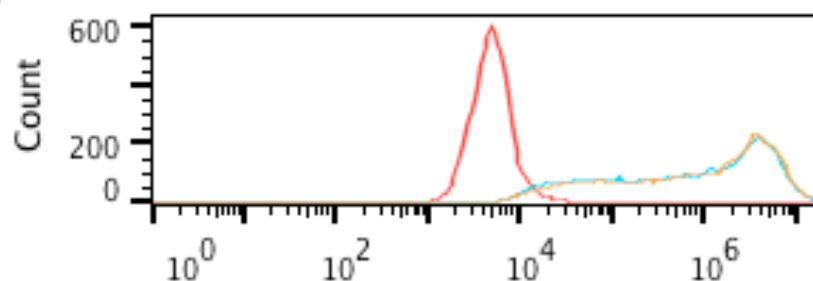
SV40-YFP 1.2 Top 10%
Day 0
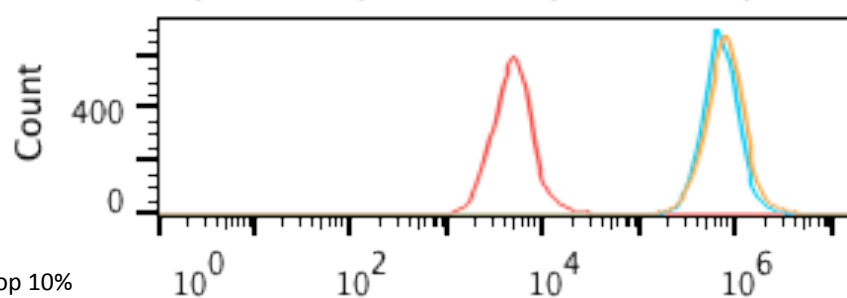
CMV-YFP 1.1 Top 10%
Day 0
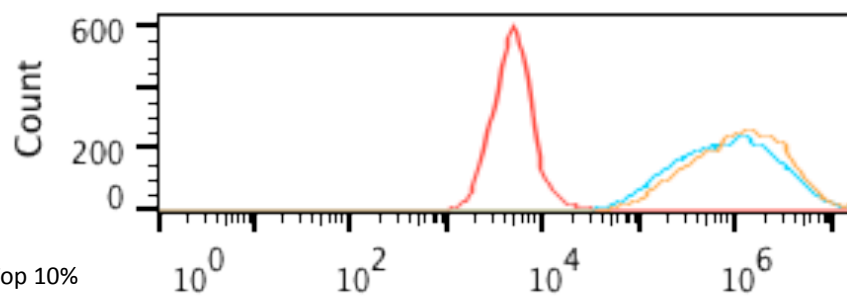
CMV-YFP 1.2 Top 10%
Day 0
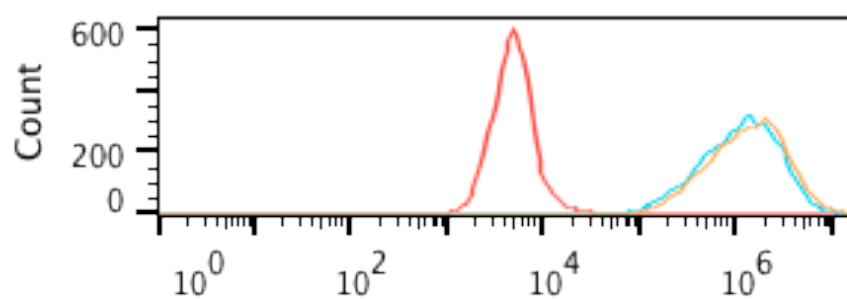
D

SV40-YFP 1.1 Top 10%
Day 32
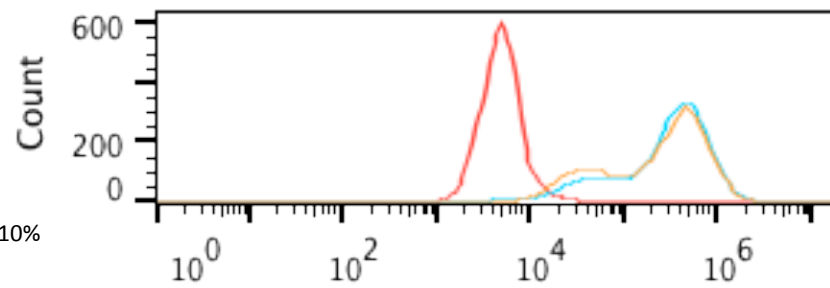
SV40-YFP 1.2 Top 10%
Day 32
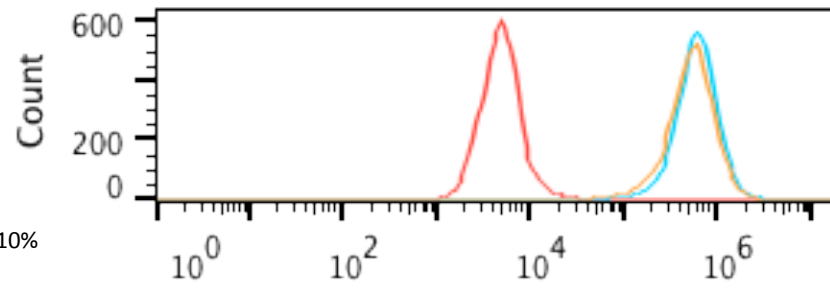
CMV-YFP 1.1 Top 10%
Day 32
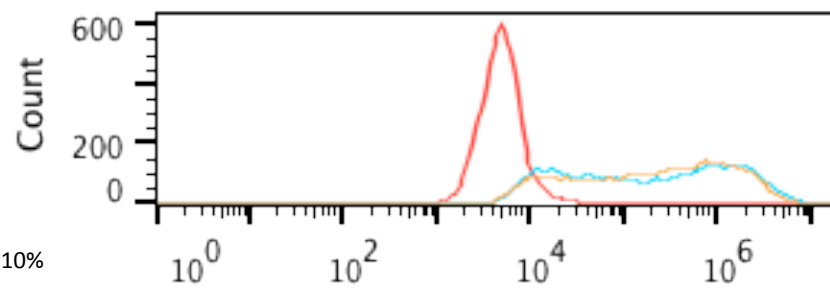
CMV-YFP 1.2 Top 10%
Day 32

233
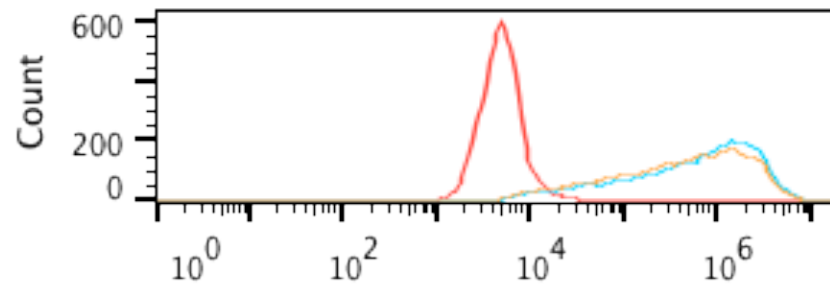
E

SV40-YFP 108.1 Polyclonal
Day 0
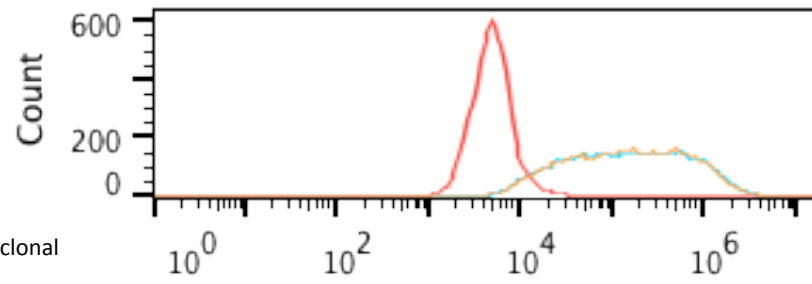
SV40-YFP 108.2 Polyclonal
Day 0
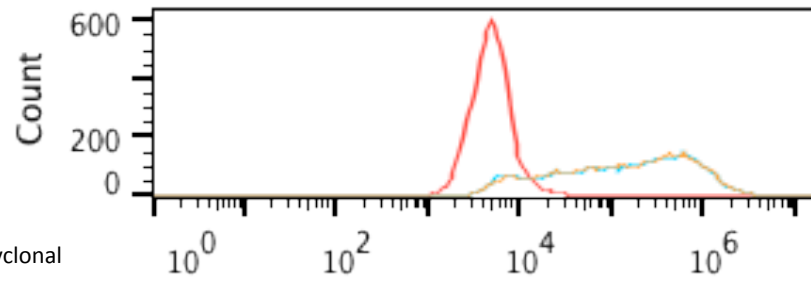
CMV-YFP 108.1 Polyclonal
Day 0
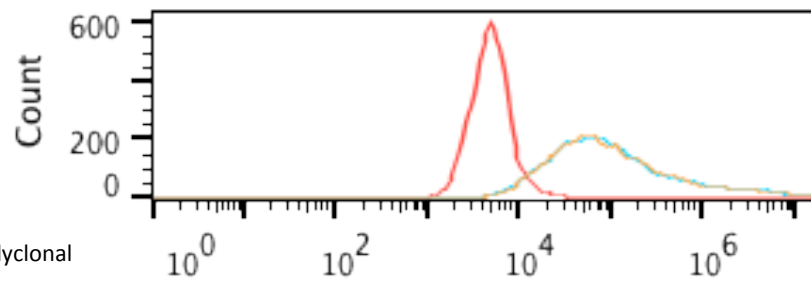
CMV-YFP 108.2 Polyclonal
Day 0

F

SV40-YFP 108.1 Polyclonal
Day 32

SV40-YFP 108.2 Polyclonal
Day 32

CMV-YFP 108.1 Polyclonal
Day 32

CMV-YFP 108.2 Polyclonal
Day 32
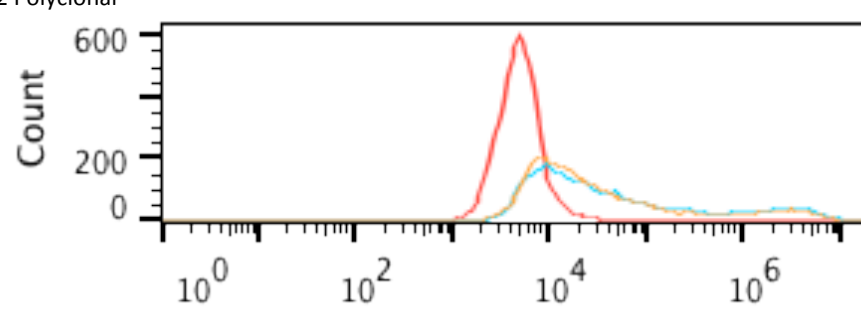
G

SV40-YFP 108.1 Top 10%
Day 0
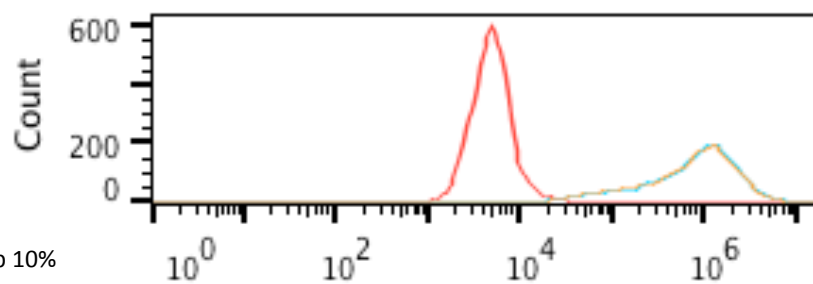
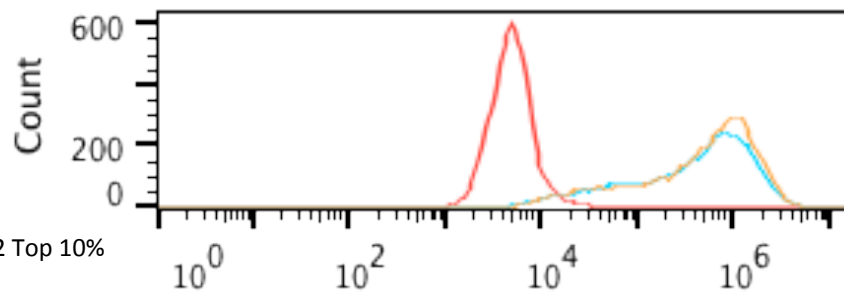SV40-YFP 108.2 Top 10%
Day 0
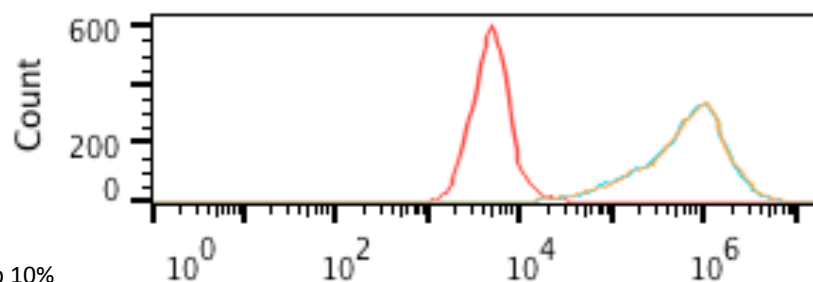
CMV-YFP 108.1 Top 10%
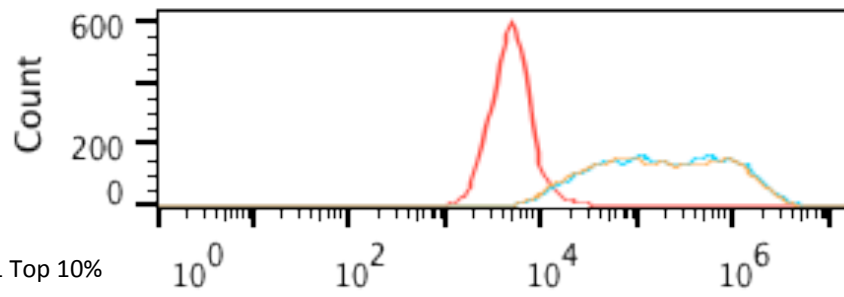Day 0
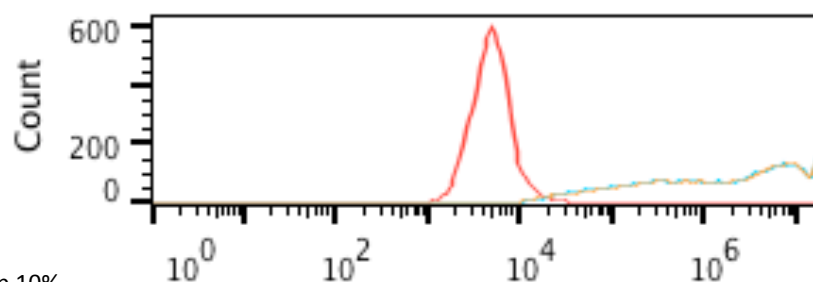
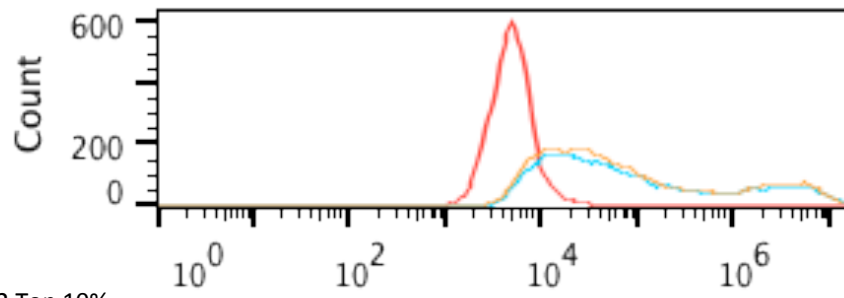CMV-YFP 108.2 Top 10%
Day 0

H

SV40-YFP 108.1 Top 10%
Day 32

SV40-YFP 108.2 Top 10%
Day 32

CMV-YFP 108.1 Top 10%
Day 32

CMV-YFP 108.2 Top 10%
Day 32