THE UNIVERSITY *of* EDINBURGH

THE UNIVERSITY *of* EDINBURGH
Centre for Engineering Biology

FUJIFILM Diosynth biotechnologies

Improving Monoclonal Antibody Production from Chinese Hamster Ovary Cells

James Donaldson

Thesis submitted for the degree of Doctorate of Philosophy

The University of Edinburgh

School of Biological Sciences
Institute of Quantitative Biology, Biochemistry and Biotechnology

2023

Declaration

I declare that this thesis has been composed solely by myself and that it has not been submitted, either in whole or in part, in any previous application for a degree. Except where otherwise acknowledged, the work presented is entirely my own.

James Donaldson

31st October 2023

Abstract

Chinese Hamster Ovary (CHO) cells are used for the production of many therapeutic proteins, including the majority of monoclonal antibodies (mAbs). While significant progress has been made in improving mAb production using CHO cells, challenges remain in producing sufficient quantities of next-generation biologics, such as fusion proteins and bi-specific antibodies. Additionally, production instability (i.e., loss of mAb productivity) during long term culture remains a significant problem. Stability studies, which take several months to complete, are a bottleneck in cell line development.

A landing pad system for expression vector component comparison was integrated into a CHO host cell line. To demonstrate the functionality of the system, the strength of various constitutive promoters were tested by quantifying mCherry expression levels. This system will help facilitate the development of future expression vectors and enable systematic identification and optimisation of components to enhance CHO cell productivity.

A system which can predict the production stability of recombinant CHO cell lines, prior to the completion of a stability study, was envisaged. To pursue this, the production stabilities of 2 monoclonal CHO cell lines (32-124 and 32-121) producing the same mAb were characterised over ~60 generations. Loss of volumetric productivity during long term culture was observed in the 32-121 cell line whereas 32-124 exhibited a minimal decrease. Further analyses of growth characteristics, specific productivity, metabolite profiles, gene copy number, and transgene mRNA expression were subsequently conducted to investigate the underlying cause of productivity loss. The Berkley Lights Beacon® system was subsequently used to compare the two cell lines at generation 15. The results showed differences in both production and growth variability between the two cell populations, as well as an increased frequency of low-producing fast-growing cells in the 32-121 cell line. The method developed during this study will add to existing strategies for identifying early indicators of instability in CHO cell lines.

Lay Summary

This project aimed to improve our ability to manufacture antibody-based medicines. Antibodies play an important role in our immune system, helping us to fight disease. Antibody-based medicines are designed to replicate the function of antibodies in the immune system and are used to treat a broad range of diseases, including cancer, neurological, autoimmune and metabolic diseases. Nowadays, many of the antibody-based medicines are produced using living cells, particularly cells which have been derived from mammals. To produce antibody-based medicines, cells are provided with new genetic instructions that tell the cell how to produce the medicine. These instructions convert the cell into a miniature factory for making the medicine. Chinese Hamster Ovary cells, also known as CHO cells, are a particular type of mammalian cell that are very good at producing medicines. To make high quantities of medicine, CHO cells are grown in large bioreactors which contain all the nutrients that it needs to grow.

Although many improvements have been made in the way we use CHO cells to make antibody-based medicines, there are still some challenges that remain. For example, some new medicines are difficult to make in large quantities using these cells. Additionally, some CHO cells switch off the genetic instructions that tell them how to make the medicine, meaning that they stop producing medicine. This reduces the amount of medicine that a cell factory can produce over time. A lot of time and money may be wasted if cells that have stopped producing the medicine are grown in the large bioreactors. Therefore, it is important that, after the cells have been given the genetic instructions, cells are grown for approximately 2 months and their ability to produce medicine in high quantities is routinely measured. Cells that stop producing the medicine are not taken forward for growth in large bioreactors. However, this long study slows down the process of making new medicines.

The genetic instructions that tell a cell how to make the medicine contains lots of different components. The combination of these individual components can help to determine whether a cell will produce small or large quantities of the medicine. Specific combinations of individual components have previously been shown to help with the production of certain medicines. However, to know which components need to be included in the genetic instructions for each new medicine, we need a way of quickly comparing different component combinations, before the cells are put in the large bioreactors. Therefore, the first chapter in this thesis developed a system that can be used to rapidly compare different compositions of individual components in the genetic instructions.

Next, we look at ways to tackle the problem of cells that stop making medicine over time. To remove the requirement for cells to be grown for 2 months during the development process, we need a way of predicting whether a cell is likely to switch off production, before the long-term study has to be undertaken. To do this, two groups of cells are studied that have been given the same genetic instructions for the same medicine. We grew them for approximately two months and show that one group of cells starts to make less medicine over time. However, the other group of cells continued to produce equal quantities of medicine over the two-month period. We investigated why this happens by looking at how fast they grow, their genetic information and how they perform in mini bioreactors. Next, we compared these two groups of cells again but in a different way. We used a machine that can assess how well each cell in a group of cells is growing and producing medicine. We found that the group of cells which makes less medicine over time has a higher proportion of cells that grow fast but do not make much medicine.

Overall, this research developed methods which will help us make CHO cells better at producing medicines and avoid problems, such as cells stopping producing medicine over time.

Acknowledgements

I want to express my gratitude to everyone who played a part in my PhD journey. The support and encouragement I received from so many individuals has been priceless, and I want to extend my thanks to each and every one of you.

Firstly, thank you to Susan Rosser for giving me the opportunity to pursue my PhD in your lab and for being there with advice and support when I needed it. Also, thank you to my thesis committee, Filippo Menolascina and Elise Cachat, for helping me with your feedback and guidance.

Thank you to the members of the Rosser lab, you have not only been great colleagues but many of you have also become lifelong friends. Your support has meant the world to me. Vivek Senthivel, your patience when explaining new concepts, willingness to engage in discussions and your encouragement during challenging times has been a tremendous help. Also, I have really enjoyed our occasional badminton matches! Matthew Dale, thank you for helping me to get set up in my project and for reviewing my manuscripts! Thank you also to Caroline Wardrope and the Edinburgh Genome Foundry for our collaboration on the Beacon project and to Martin Waterfall for training me in the use of flow cytometers.

I would like to thank FUJIFILM Diosynth Biotechnologies (FDB) and IBioIC for funding my project and for their unwavering support throughout my PhD journey. Special thanks to Alison Young and Devika Kalsi for answering all my questions, helping with new experiment strategies and always being a source of encouragement. I want to express my gratitude to everyone in the mammalian cell culture department at FDB for their warm welcome during my placement and for making me feel like a part of the team. I have been consistently impressed, not only by the exceptional work carried out at FDB, but also by the genuine kindness exhibited by every individual within the company. A special shout-out also has to go to those who subcultured my cells on weekends; you are the true MVPs! I am also thankful to IBioIC for

the support provided during the various training days, and to my cohort for the fun and solidarity we have shared over our PhD journeys. In particular, I want to thank Ian Archer for his immense enthusiasm for industrial biotechnology and for instilling in me the confidence to explore different aspects of the industry.

To my friends at lunch club, your friendship has been a constant source of laughter and unwavering support. I feel so fortunate to have shared so many laughs with you all! Marcus Price, even though your taste in sport and music is questionable, you have been a great source of career, science and life advice, and for that I am very grateful. Alex Arrese, I think we have now dissected everything from “why isn’t my experiment working?” to “who should be Liverpool’s next signing?”! There is truly no one else I would rather do mammalian cell culture/ 5-aside football alongside! Ben and Jess, I'm immensely proud of how far we've come together, and I am so glad that we could do this PhD journey together. You are both so kind and have always been there to listen to my many rants when experiments haven’t worked! I wish you both the absolute best for the future and I know that you will succeed in any avenue that you choose to pursue.

I have had the privilege of developing and growing some amazing friendships during my PhD. To the City on a Hill church community, you have been instrumental in my faith journey over the past 4 years. To Dan Everett and Pascal Sime, thank you for your spiritual guidance and for the many phone calls, during which you patiently listened and never judged. To everyone in my various small groups over the years (particularly Josiah, Rebekah, Adedoyin, Kirsty, Evie, Milo, Alastair, Veronica and Lewis), you made Edinburgh feel like home and were like a family to me. To Grange Cricket Club and everyone in my various 5-aside football groups, you provided not only fun and a healthy outlet for my competitive spirit, but also a great way to unwind outside of the lab. Thank you also to Amy McMillan, Oliva McNeill and Cara Coughlan for welcoming me into London and for providing such a warm and comfortable environment for me to write my thesis in.

Thank you also to my Mum, Dad and Laura. Even though you may have struggled to understand what my lab problems were(!), you have always been there at the end of the phone, providing unwavering support and encouragement. I feel so grateful to have a family like ours and I will never take that for granted.

Lastly, and most importantly, I want to thank Tega, my fiancé. You have been my constant support throughout these four years. You have been there during the toughest of times and you are the first to celebrate with me in the good times. I cannot wait for our next chapter as a married couple. You are my greatest blessing, and I could not have done this without you.

Table of Contents

Declaration 3
Abstract 4
Lay Summary 5
Acknowledgements 7
List of Figures 17
List of Tables 20
Abbreviations 21

Chapter 1 – Introduction 23

1.1 The Biopharmaceutical Industry Continues to Grow 23

1.2 mAbs Continue to Dominate the Biopharmaceutical Industry 24

1.3 mAbs Enable Precise Targeting of Pathogens and Cancer Cells 24

1.4 The History of Monoclonal Antibodies 26

1.5 Recombinant Production of mAbs Relies on Mammalian Cell Expression Systems, with CHO Cells being the Preferred Host 28

1.6 Fed-Batch Cultures (FBCs) Remain the Primary Cultivation Mode for CHO Cell-based mAb Manufacture 31

1.7 Conventional CLD Strategies Tend to be Time Costly and Labour Intensive 33

1.8 The Growing Biopharmaceutical Industry Continues to Drive Innovation in CLD 36

1.8.1 Increasing Product Titre and the Challenge of New Modalities 37
1.8.2 Reducing CLD Timelines and the Stability Study Bottleneck 39

1.9 Areas for Optimisation in CLD – Recent Developments and Future Opportunities 41

1.9.1 Expression Vector Optimisation 41
1.9.1.1 Promoter Engineering to Boost Transcription and Resist Gene Silencing 41
1.9.1.2 The Incorporation of Epigenetic-Regulatory Elements May Prevent Transgene Silencing 43
1.9.1.3 Expression Vector Engineering to Optimise Subunit Dosage 47
1.9.1.4 Signal Peptide Engineering to Improve Recombinant Protein Secretion 50
1.9.1.5 Inducible Gene Expression Strategies Will Add Flexibility to Manufacturing Processes 50

1.9.2 Cell Line Optimisation 56
1.9.3 High Throughput Single Cell Cloning and Screening Methods to Reduce CLD Timelines 59
1.9.4 Targeted Integration 63
1.10 Project Aims 68
1.10.1 A Targeted Integration Strategy for Expression Vector Component Comparison (Chapter 2) 68
1.10.2 Developing a Method for Early Prediction of Production Instability (Chapter 3 and Chapter 4) 69
Chapter 2 - Developing a Targeted Integration System for Expression Vector Component Comparison 71
2.1 Chapter Summary 71
2.2 Introduction 72
2.2.1 Limitations of Transient Transfections 73
2.2.2 Limitations of Stable Clones Generated by RI 75
2.2.3 Limitations in Comparing Stable Pools 76
2.2.4 Considerations for Designing an Expression Vector Component Comparison System: Towards Targeted Integration 77
2.2.5 Proposed Expression Vector Comparison Strategy 80
2.3 Materials 81
2.3.1 Antibiotics for Molecular Biology 81
2.3.2 Reagents and Solutions for Molecular Biology 81
2.3.3 Buffers 81
2.3.4 Molecular Biology Kits and Enzymes 82
2.3.5 Bacterial Culture Media 82
2.3.6 Ordered/ Gifted Constructs 82
2.3.7 Assembled Constructs 83
2.3.8 Tissue Culture Media 86
2.3.9 CRISPR/ Cas9 sgRNA Sequences 87
2.3.10 Primers for Diagnostic PCRs 87
2.4 Methods 90
2.4.1 Preparation of Competent Cells 90
2.4.2 Bacterial Transformation 90
2.4.3 Plasmid DNA Preparation 90
2.4.4 Plasmid and Sequence Verification 91
2.4.5 EMMA Assembly of LPs 91
2.4.6 Gibson Assembly of Promoterless mCherry-hygro/ hygro-mCherry Plasmids 91

2.4.7 Maintaining Mammalian Cell Lines 91
2.4.8 Calculating Generation Number 92
2.4.9 Generation of Cas9 Edited Cell Lines 92
2.4.10 FACS 92
2.4.11 Diagnostic PCRs 92
2.4.12 Flow Cytometry 92
2.4.13 RMCE-Integration of Expression Cassettes 93
2.5 Results 94
2.5.1 Identification of a Suitable Cell Line for LP integration 94
2.5.2 Identification of a Suitable Locus for LP Integration 95
2.5.3 Design and Assembly of the LP 98
2.5.4 Integration of the LP into the Chosen Loci 101
2.5.5 Expansion and Selection of LP Clones 108
2.5.6 Flow Cytometry Analysis of the Selected Clones 112
2.5.7 PCR Analysis Reveals Integration of the Plasmid Backbone in the LemD2 Cell Line 112
2.5.8 Further Analysis of the Fer1l4 c8 Cell Line Indicates Successful LP Integration at the Fer1l4 Locus 115
2.5.9 A Stability Study for the Fer1l4 c8 Cell Line 117
2.5.10 The Mrpl4_c4 cell line Showed Poor Growth Over 60 Generations and PCR Analysis is Inconclusive for LP Integration 119
2.5.11 Further Analysis of the Fer1l4_c8 clone 120
2.5.12 Bxb1 Mediated Cassette Exchange in the Fer1l4_c8 Clone 123
2.5.13 Diagnostic PCRs Confirm That Integration into the LP has been Successful 126
2.5.14 FACS Single Cell Sorting and Subsequent Diagnostic PCRs Suggest there is an Additional, Non-Functioning LP in the Fer1l4_c8 Cell Line 129
2.5.15 Retrospective Analysis of Fer1l4 Clones 133
2.5.16 Optimising Integration of Constructs into the LP 134
2.5.17 An Initial Exploration into Comparing the Strength of Different Promoters Using the Fer1l4_c8 cell line 137
2.6 Discussion 140
2.6.1 Further Analysis of the Fer1l4_c8 Cell Line 141
2.6.2 Future Avenues for LP Integration at Stable Loci in Apollo™ X Host Cells for Expression Vector Component Comparison 142
2.6.3 Expression Vector Component Comparison Using the Fer1l4_c8 Cell Line 146
2.6.4 Future Avenues for Expression Vector Comparison 146
Chapter 3 - Characterising the Production Stability of Two Monoclonal CHO Cell Lines 150

3.1 Chapter Summary 150

3.2 Introduction 151

3.2.1 Defining Instability 151

3.2.2 A Stability Study Characterises the Production Stability of the Cell Line, Prior to Manufacture 152

3.3 Materials 154

3.3.1 Cell Culture Media 154

3.3.2 FBC Supplements 154

3.3.3 Molecular Biology Reagents 154

3.4 Methods 155

3.4.1 Stability Study Design 155

3.4.2 Revival of Cell Lines from Cryopreserved Stocks 157

3.4.3 Routine Subculture of Recombinant Apollo™ X Cell Lines (32-121 and 32-124) 157

3.4.4 Cryopreservation of Cell Banks 157

3.4.5 Stability Study: FBC Shake Flask Screen 158

3.4.6 Product Concentration Analysis 158

3.4.7 Calculations 158

3.4.7.1 Generation Number 158

3.4.7.2 Time Integral of Viable Cell Density (IVC) 159

3.4.7.3 Relative Product Concentration 159

3.4.7.4 Relative Specific Productivity 159

3.4.7.5 Relative Metabolite Concentration 160

3.4.8 Linear Regression Analysis to Assess Cell Line Stability 160

3.4.9 Quantitative real-time PCRs 162

3.4.9.1 Sample Preparation 162

Genomic DNA for Gene Copy Number Analysis 162

cDNA for mRNA Expression Analysis 162

3.4.9.2 Threshold Cycle (CT) Determination 162

3.5 Results 163

3.5.1 A previous FDB study had Produced and Analysed the Stability of 48 Monoclonal Cell Lines Expressing the same mAb 163

3.5.2 The 32-121 and 32-124 Cell Lines were Selected for Further Investigation 167

3.5.3 Designing a Study to Validate the Stability of the 32-121 and 32-124 Cell Lines 168

3.5.4 Changes in Growth Rate during Routine Subculture over 60 Generations 170

3.5.5 The 32-121 Cell Line Exceeded the Alert Limit for Cell Line Instability 172

3.5.6 Linear Regression Models were used to Complement the Analysis of Production Instability 175
3.5.7 The Relationship Between Changing Growth Characteristics and Productivity over 60 Generations 177
3.5.8 The Drop in Q_p-Rate for the 32-121 Cell Line (Minus MTX) exceeded the Alert Limit for Cell Line Instability 179
3.5.9 Comparing Flask FBC Results with the initial Ambr®15 Study 182
3.5.10 Drop in sp-Q_p was not Specific to the Early or Late Phase for the 32-121 Cell Line 185
3.5.11 Analysis of FBC Metabolite Data for the 32-121 and 32-124 Cell Lines 187
3.5.12 Design and Setup of Gene Copy Number Experiments using Real-Time Quantitative PCR 190
3.5.13 Gene Copy Number Analysis Suggests the Loss of Transgene Copies was Unlikely to be the Primary Cause of Productivity Loss in the 32-121 Cell Line 192
3.5.14 Design and Setup of Transgene mRNA Expression Quantification Using Real-Time Quantitative PCR 198
3.5.15 A Significant Relationship Between the CT Of Actb and Generation Number was Observed 200
3.5.16 Screening Reference Genes for Gene Expression Studies 200
3.5.17 Gene Expression Analysis of 32-121 and 32-124 Cell Lines 205

3.6 Discussion 208

3.6.1 Experiment Limitations and Challenges of Experimental Design 209
3.6.2 Understanding the Cause of Loss of Productivity in the 32-121 Cell Line 212
3.6.3 Future Avenues for Understanding the Production Instability of the 32-121 Cell Line 215

Chapter 4 - Comparing two cell lines with different production stabilities at an early generation using the Berkley Lights Beacon® System 217

4.1 Chapter Summary 217

4.2 Introduction 218

4.2.1 Predicting CHO Cell Production Instability using Sequencing and Omics is Inherently Difficult and Time Consuming 219
4.2.1.1 Predicting Production Stability from the Locus/Loci at which the Transgene Integrates 219
4.2.1.2 Distal Factors which Impact Stability 222
4.2.1.3 Alternative Approaches for Predicting Cell Line Stability are Needed 224
4.2.2 Identifying Low-Producing Fast-Growing Subpopulations in Early Generation Cell Lines 224

4.3 Materials 228
4.3.1 Cell Culture Media for Beacon® Runs 228
4.3.2 Cell Lines 228
4.4 Methods 228
4.4.1 Cell Culture 228
4.4.2 Analysis of Cell Lines using the Berkeley Lights Beacon® instrument 228
4.4.3 Calculations 229
4.4.3.1 Doubling Time 229
4.4.3.2 Cell Specific Productivity 229
4.4.4 Modelling Changing Cell Populations over 60 Generations 230
4.4.4.1 Starting Cell Population 230
4.4.4.2 Batch Cultures 231
4.5 Results 232
4.5.1 A Simple Model to Illustrate a Population of Cells Consisting of 2 Subpopulations 232
4.5.2 Generation 15 was Chosen as the Early Timepoint for Comparing the 32-124 and 32-121 Cell Lines (Minus MTX) 236
4.5.3 Optimisation of Cell Growth on the OptoSelect™ chip 238
4.5.4 Filtering Process Prior to Beacon® Data Analysis 241
4.5.5 Comparison of Mean Doubling Time and Cell Specific Productivity Between the 32-121 and 32-124 Cell Lines 241
4.5.6 Beacon® Analysis Reveals an Increase in the Proportion of low-producing fast-growing cells in the 32-121 cell line on day 4 243
4.5.7 The Beacon® Data Should be Normalised Prior to Analysis 245
4.5.8 Normalisation of the Beacon® Data Confirms Previous Conclusions 248
4.5.9 Increased Variation Around the Regression Line in the 32-121 Cell Line 251
4.5.10 Increased Variation in The 32-121 Cell Line Stems From Increased Variation in Both Growth and Specific Productivity 251
4.5.11 Using the day 4 Beacon® Results to Develop an Initial Model of Instability 253
4.6 Discussion 257
4.6.1 Experiment Limitations and Future Avenues for exploration using the Beacon® System 258
4.6.2 Future Avenues for Developing Predictive Methods for Production Instability in Recombinant CHO Cell Lines 262
Chapter 5 - Final Discussion 265
5.1 Thesis Summary and Key Findings 266
5.1.1 A Landing Pad System for Expression Vector Component Comparison 266

5.1.2 Developing a System for Early Prediction of Production Instability _____ 267

5.2 Alternative Applications for the Fer1l4_c8 Cell Line _____ 269

5.2.1 Self-Regulating CHO Cells _____ 269

5.2.2 Solving Product-Specific Production Challenges _____ 270

5.3 Beacon® Analysis of LP Cell Lines _____ 271

5.4 Combining Predictive Measures with Stability Enhancers _____ 272

5.5 Final Conclusions _____ 272

Appendix 1 _____ 274

Appendix 2 _____ 282

Appendix 3 _____ 292

References _____ 296

List of Figures

Figure 1.1 - The Structure of an Antibody 25
Figure 1.2 - Schematic of Culture Processes 32
Figure 1.3 - Traditional CLD Approaches are Labour Intensive and Time Consuming. 35
Figure 1.4 - Components of a mAb Expression Vector 42
Figure 1.5 - Barrier Elements Prevent the Spread of Heterochromatin 46
Figure 1.6 - The Architecture of Inducible Gene Expression Systems 52
Figure 1.7 - Engineering Self-Regulating CHO cells to Improve Biopharmaceutical Production. 57
Figure 1.8 - Single Cell Sorting Systems for CLD 61
Figure 1.9 - CRISPR/Cas9 for Targeted Gene Integration. 64
Figure 1.10 - SSR Mediated by a Serine Integrase 66
Figure 2.1 - Previous work identified 3 candidate loci for LP integration. 97
Figure 2.2 - Design and Assembly of the LP for TI 100
Figure 2.3 - Antibiotic Selection of the Transfected Population. 102
Figure 2.4 - Single Cell Sorting of the Transfected Populations. 106
Figure 2.5 - Selection of Clones at the 24-well plate Stage 110
Figure 2.6 - Analysis of LP Clones using Flow Cytometry 114
Figure 2.7 - Diagnostic PCRs reveal integration of the LP backbone in the LemD2 cell lines. 115
Figure 2.8 - Diagnostic PCRs for LP Integration at the Fer1l4 Locus. 116
Figure 2.9 - The Fer1l4_c8 Cell Line Shows Unstable Production of mNG over 60 Generations. 118
Figure 2.10 - Analysis of the Mrpl4 Cell Line. 120
Figure 2.11 - Successful integration of a promoterless hygro-P2A-mCherry construct. 125
Figure 2.12 - Diagnostic PCRs show successful integration of the hygR-p2A-mCherry construct into the LP 127
Figure 2.13 - Single Cell Sorting of hygR-P2A-mCherry Cells and Subsequent PCR Analysis 130
Figure 2.14 - Potential Silencing of Additional mNG Copies. 134

Figure 2.15 - Retrospective Analysis of Alternative Fer1l4 Clones Suggests that other Clones were the Single Integration Clones. 135
Figure 2.16 - Insertion of an mCherry-p2A-hygro construct into the LP 136
Figure 2.17 - Initial Comparison of Constitutive Promoters in the LP Cell Line 138
Figure 3.1 - The Stability Study. 153
Figure 3.2 - Design of the Stability Study 156
Figure 3.3 – Calculating specific productivity rate using linear regression analysis of the product concentration against IVC 161
Figure 3.4 - Stability analysis of 48 monoclonal cell lines over 60 generations using the Ambr®15 bioreactor system 165
Figure 3.5 - Further analysis of the 32-121 and 32-124 cell lines 169
Figure 3.6 - Growth characteristics of cell lines under routine subculture conditions with and without MTX over 60 generations 171
Figure 3.7 - Difference in product concentration between the early and late FBC for the 32-121 and 32-124 cell lines 174
Figure 3.8 - Change in day 14 product concentration over 60 generations for regression line analysis of production stability 176
Figure 3.9 - Change in growth characteristics over 60 generations 178
Figure 3.10 - Difference in Q_p- Rate between the Early and Late FBC for the 32-121 and 32-124 cell lines 180
Figure 3.11 - Regression line analysis for the change in q_p- rate over 60 generations 182
Figure 3.12 - Comparison of phase-specific sp-Q_p 186
Figure 3.13 - Metabolite analysis of the 32-124 and 32-121 cell lines 188
Figure 3.14 - Selection of primers for gene copy number analysis 193
Figure 3.15 - Gene copy number analysis using Cog1 as a reference gene 194
Figure 3.16 - Gene copy number analysis using ActB as a reference gene 195
Figure 3.17 – Reference gene comparison for gene copy number analysis 197
Figure 3.18 – Amplification efficiencies of primer sets for gene expression analysis 199

Figure 3.19 - CT values for the ActB gene exhibit a significant linear relationship with generation number. 201
Figure 3.20 – Selection of reference genes for gene expression analysis. 204
Figure 3.21 - mRNA expression analysis of the 32-121 and 32-124 cell lines 206
Figure 4.1 - The Spread of Heterochromatin 221
Figure 4.2 - The Beacon® System 226
Figure 4.3 – Modelling growth and productivity changes in a mixed CHO cell population over 60 generations. 234
Figure 4.4 - Analysis of Generation 15 Cell Lines Using the Berkley Lights Beacon® Optofluidic System. 238
Figure 4.5 - Optimisation of Growth in the OptoSelect™ chip using the 32-124 Cell Line. 240
Figure 4.6 - Filtering Process prior to Beacon® Data Analysis 242
Figure 4.7 - Comparison of Mean Doubling Time and Cell Specific Productivity between the 32-121 and 32-124 cell lines 243
Figure 4.8 – Beacon® Analysis Reveals an Increase in the Proportion of low-producing fast-growing cells in the 32-121 cell line. 245
Figure 4.9 - Analysis of the day 3 and day 5 data reveals challenges for the Beacon® stability screening method. 246
Figure 4.10 - Normalisation Increases the Proportion of Cells in Grid Square E 249
Figure 4.11 - Increased variation around the regression line in the 32-121 cell line. 252
Figure 4.12 - Increase in Clonal Variation is due to Increased Variability in Cell Specific Productivity and Doubling Time 252
Figure 4.13 - A Simple 9-Cell Model for Production Instability. 256
Figure 5.1 - Sequencing Results from Diagnostic PCRs of the Fer1l4_c8 Cell Line. 294
Figure 5.2 - Sanger Sequencing of the PCR products show recombination at the AttB sites has occurred. 295

List of Tables

Table 1.1- Epigenetic regulatory elements that have been tested in CHO cells 44
Table 1.2 - Examples of Inducible Gene Expression Systems Which Switch on Gene Expression in Response to Small Molecule Addition 54
Table 2.1 - Ordered/ Gifted Constructs 82
Table 2.2 - Assembled Constructs 83
Table 2.3 - CRISPR/ Cas9 sgRNA Sequences 87
Table 2.4 - Primers for Diagnostic PCRs 87
Table 2.5 - Outgrowth of Cloned LP Cells following Single-Cell Sorting. ... 108
Table 3.1 – Cell culture samples were taken from the routine subculture flasks at specific subculture numbers to seed early, early-mid, late-mid and late FBCs. 172
Table 3.2 – Comparison between the initial Ambr®15 and flask FBC study 183
Table 3.3 – Selection of reference genes from the literature and Apollo™ X transcriptomic data 202
Table 4.1 - Parameters used for the 9-Subpopulation Model 255

Abbreviations

This section provides a list of the abbreviations used in this thesis. To provide clarity, abbreviations are redefined in each subsequent chapter. Abbreviations that have been employed only once, such as for figures or widely recognized methods (e.g., ELISA), have been omitted from this list.

- **Berkley Lights Beacon® System** – Beacon®
- **BSD** – Blasticidin
- **Cas9** – CRISPR Associated System Protein 9
- **CDR** – Complementarity-Determining Region
- **CHO** – Chinese Hamster Ovary
- **cHS4** – DNaseI Hypersensitive Site 4 Of The Chicken Beta-Globin Locus Control Region
- **CLD** – Cell Line Development
- **CMV** – Cytomegalovirus
- **CRISPR** – Clustered Regularly Interspaced Short Palindromic Repeats
- **crRNA** – Clustered Regularly Interspaced Short Palindromic Repeats RNA
- **DHFR** – Dihydrofolate Reductase
- **DSB** – Double-Stranded Break
- **EF-1α** – Elongation Factor 1α
- **EMMA** – Extensible Mammalian Modular Assembly
- **ERE** – Epigenetic Regulatory Element
- **FACS** – Fluorescence-Activated Cell Sorting
- **FBC** – Fed-Batch Culture
- **FDA** – Food and Drug Association
- **FDB –** FUJIFILM Diosynth Biotechnologies
- **GS** – Glutamine Synthetase
- **HA –** Homology Arm
- **HC** – Heavy Chain
- **HDR** – Homology Directed Repair

- **LC** – Light Chain
- **LDC** – Limiting Dilution Cloning
- **LP** – Landing Pad
- **LTC** – Long Term Culture
- **mAb** – Monoclonal Antibody
- **MAR** – Matrix Attachment Region
- **mNeonGreen** – mNG
- **MTX** – Methotrexate
- **NHEJ** – Non-Homologous End Joining
- **PAM** – Protospacer Adjacent Motif
- **PTM** – Post-Translational Modification
- **RI** – Random Integration
- **RMCE** – Recombinase-mediated Cassette Exchange
- **sgRNA** – Single Guide RNA
- **SSR** – Site-Specific Recombination
- **STAR** – Stabilising Antirepressor Element
- **TALEN** – Transcription Activator Like Effector Nuclease
- **TF** – Transcription Factor
- **TFBS** – Transcription Factor Binding Sites
- **TI** – Targeted Integration
- **tracrRNA** – Trans-Activating Clustered Regularly Interspaced Short Palindromic Repeats RNA
- **UCOE** – Ubiquitously Acting Chromatin Opening Elements
- **UTR** – Untranslated Region
- **VCD –** Viable Cell Density
- **ZFN** – Zinc Finger Nuclease

Chapter 1 – Introduction

The objective of this chapter is to provide a broad overview of the biopharmaceutical industry, with a specific focus on monoclonal antibodies (mAbs) and their production using Chinese Hamster Ovary (CHO) cells. Optimisation efforts in cell line development (CLD) are explored, and the ongoing challenges in current methods are identified. Several of these challenges will be the focus of further investigation in this research. This section will thus provide the groundwork for the discussions presented in the following chapters. Each subsequent results chapter will provide a specific literature review concerning its respective topic.

1.1 The Biopharmaceutical Industry Continues to Grow

Biopharmaceuticals, often referred to as biologics, are therapeutic products sourced from biological materials (such as proteins, nucleic acids and cells) that are designed for the treatment of diseases (Kesik-Brodacka, 2018). This category of pharmaceuticals encompasses a broad range of therapeutics, such as vaccines, blood components, allergenics, somatic cells, gene therapies, tissues, recombinant therapeutic proteins, and cell-based therapies (Niazi and Lokesh, 2021). From January 2018 to June 2022, a total of 197 biopharmaceutical products received approval in either the United States or the European Union, contributing significantly to the industry's estimated market value of $350 billion in 2021, according to NextMSC (NextMSC, 2022; Walsh and Walsh, 2022). Further reports have indicated that more than 7,800 biopharmaceutical products are currently in various stages of clinical development worldwide, with over 1,000 having progressed to phase 3 trials (Walsh and Walsh, 2022). With no signs of slowing down, the biopharmaceutical sector is forecasted to reach nearly $975 billion by 2030, with a compound annual growth rate of 11.7% from 2022 to 2030 (NextMSC, 2022). Increasing cases of chronic diseases such as cancer, cardiovascular conditions, and diabetes within an aging middle-class

demographic are likely to contribute significantly to this continued market growth (NextMSC, 2022).

1.2 mAbs Continue to Dominate the Biopharmaceutical Industry

In the biopharmaceutical industry, mAb-based therapies continue to dominate. A significant milestone was reached in 2021 when the U.S. Food and Drug Association (FDA) approved its 100th mAb treatment, marking 35 years since the approval of the first FDA-approved mAb therapy (Mullard, 2021). mAb-based therapeutics are currently used to treat a broad range of indications, including cancer, neurological diseases, autoimmune diseases and metabolic diseases (Mahmuda et al., 2017). The success of mAbs is evident, with these treatments comprising 53.5% of all biopharmaceutical approvals between 2018 and 2022 (Walsh and Walsh, 2022). Furthermore, in 2021, mAbs accounted for 80% of total protein-based global biopharmaceutical sales, underscoring their continued dominance within the industry (Walsh and Walsh, 2022).

1.3 mAbs Enable Precise Targeting of Pathogens and Cancer Cells

mAb therapeutics are designed to replicate the function of antibodies in the immune system, enabling precise targeting of pathogens and cancer cells. The high level of specificity and affinity that antibodies show to their target molecule achieves a high level of efficacy whilst minimising side effects, making them a valuable therapeutic modality (Suzuki et al., 2015). Antibodies, also known as immunoglobulins, play an essential role in the immune system, serving as the principal effector proteins of the humoral adaptive response (Alejandra et al., 2023). To defend against infection, antibodies recognise and bind to antigens, which can neutralise pathogens, activate the complement system and recruit various types of white blood cell (Alberts et al., 2002). Antibodies exhibit a symmetrical Y-shaped structure consisting of four polypeptide chains. The antibody structure is illustrated further in Figure 1.1.

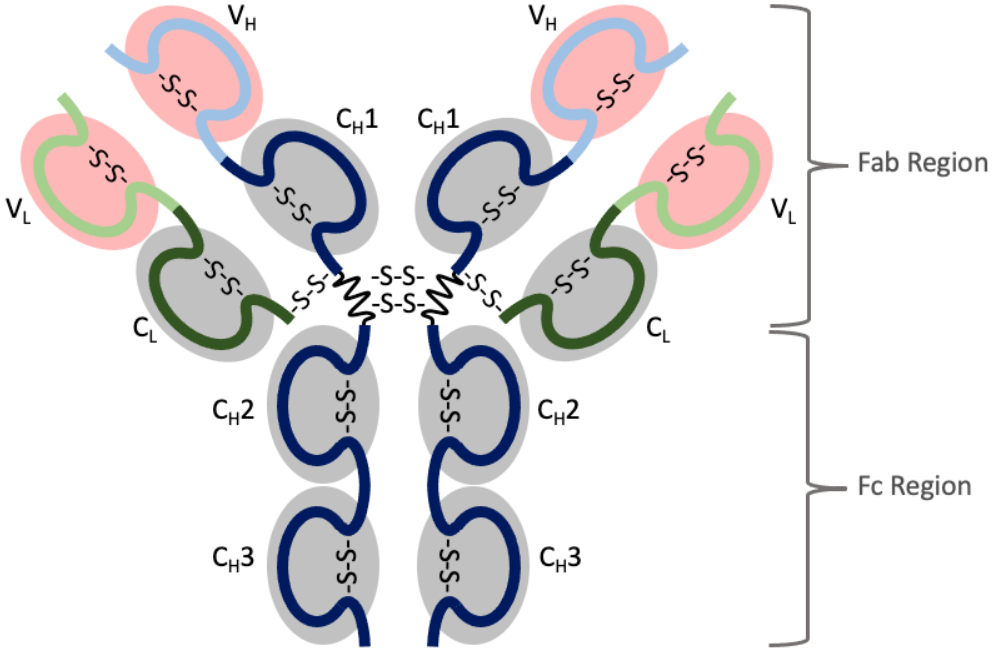


Figure 1.1 - The Structure of an Antibody

An antibody is composed of two identical heavy chains (HC) and two identical light chains (LC). The LC is composed of one variable (V_L) and one constant (C_L) domain, while the HC contains one variable (V_H) and three or four constant (C_H) domains. Pairing between an HC and LC occurs via the formation of a disulfide (S) bond and the association between two HC-LC heterodimers forms the immunoglobulin (Alberts et al., 2002; Alejandra et al., 2023; Chiu et al., 2019). The antibody structure can be functionally divided into two components: the fragment antigen-binding (Fab) region and the fragment crystallisable (Fc) region. The Fc region, located at the antibody's tail, interacts with receptor molecules and proteins in the complement system (Alejandra et al., 2023). The Fab region, comprising one constant and one variable domain from both HC and LC, provides antigen specificity through its two variable domains, which contain 3 hypervariable loops (referred to as complementarity-determining regions, CDRs) forming the antigen-binding site. Constant domains are highlighted in grey. Variable domains are highlighted in red. The figure is adapted from *Alberts et al., 2002*.

There are five classes of antibodies (IgA, IgD, IgE, IgG, and IgM), each with a distinct HC (α, δ, ε, γ, and μ). Additionally, there are two types of LC (κ or λ), and either type can associate with any class of heavy chain, influencing specificity for antigens rather than antibody properties (Alberts et al., 2002).

1.4 The History of Monoclonal Antibodies

Early insights into the structure and function of antibodies in combating infections laid the groundwork for the development of monoclonal antibody-based therapies. Pioneering work by Behring and Kitasato during the late 19th century demonstrated the therapeutic potential of transferring serum from immunized animals to treat infections like diphtheria, an approach termed "serum therapy" (Kaufmann, 2017). Subsequently, Paul Ehrlich coined the term "antibody" during his proposal of the side-chain theory in 1900. This theory suggested that antibody molecules possess branched structures, later termed "receptors," which interact with antigens (Maehle, 2009). Ehrlich's contributions laid a foundational understanding of antibody-antigen interactions, shaping subsequent research in immunology and therapeutics. In 1948, Dr. Astrid Fagraeus established a link between plasma B cells and antibody production (Fagraeus, 1948). This theory was further elaborated upon by Frank Burnet and David Talmage in 1957 through the introduction of the clonal selection theory. According to this theory, each plasma B cell displays a single species of antibody on its cell surface with a unique antigen-binding site (Alberts et al., 2002). When activated by an antigen, a B cell proliferates and transforms into an antibody-secreting effector cell, releasing soluble antibodies with the same antigen-binding site as the cell-surface receptors used earlier as antigen receptors (Alberts et al., 2002). The term "monoclonal" in monoclonal antibodies refers to antibodies derived from homologous B cells originating from a single parent clone, ensuring high specificity for a single antigenic site. This contrasts with polyclonal antibodies, which are a mixed population of immunoglobulins capable of binding to various epitopes of an antigen or target molecule. In 1959, Gerald Edelman and Rodney Porter independently elucidated the

molecular structure of antibodies, and in 1973, the first atomic resolution structure of an antibody fragment was subsequently published.

A revolutionary step in the research and discovery of monoclonal antibody-based therapies came from Köhler and Milstein with the development of the hybridoma method in 1975 (Kaunitz, 2017; Posner et al., 2019). This method involves isolating B cells from an animal, such as mice, that have been exposed to the target antigen. The antibody-producing B cell is then fused with a myeloma cell, which is cancerous and immortal, using either electrofusion or polyethylene glycol (PEG). This fusion process generates a hybridoma cell line capable of continuous growth in culture and consistent production of immunoglobulins with identical amino acid sequences (Köhler and Milstein, 1975). Electrofusion uses a pulsed electrical field to merge the cells, while PEG facilitates the fusion of their plasma membranes.

The first mAb therapeutic approved by the FDA, Orthoclone OKT3 (muromonab-CD3), was developed to prevent kidney transplant rejection. However, its murine origin often led to adverse side effects, with most patients developing strong anti-mouse antibody responses, limiting its application to acute cases (Liu, 2014; Posner et al., 2019). To overcome such challenges, genetic engineering methods and transgenic animals were developed to alter the amino acid composition of the antibody making it closer to that occurring in humans (Morrison et al., 1984; Posner et al., 2019). For example, the CDR grafting technique, which incorporates the CDRs of murine-derived antibodies into the constant regions of a human antibodies, accelerated the development of mAb therapeutics and made it possible to target diseases that require long-term treatment, such as cancer and autoimmune diseases (Alejandra et al., 2023; Lu et al., 2020).

In recent years, there has been a notable shift in monoclonal antibody (mAb) development techniques from traditional animal models to more efficient systems, such as phage display (Liu, 2014). This method utilises genetically engineered bacteriophages to present antibody libraries on their surface (Burioni et al., 1997; Smith, 1985). Initially, a phage library is introduced into

a well or vial containing the target antigen. Phages which present antibodies with the highest affinity bind to the epitopes of the antigen. Washing steps are used to remove non-binding phages. The bound phages are then eluted, often through enzymatic digestion, and helper phages are introduced to enable the amplification of the desired phages. Phage display thus allows for a more rapid generation of antibodies against a target antigen compared with traditional methods such as animal immunisation, which rely on the natural production of antibodies by B cells, a process that can span several months (Davis, 2020). Using phage display, the DNA sequences of the high-affinity antibodies can subsequently be recovered and used for further mAb screening and development (Nixon et al., 2014). For example, these DNA sequences can be utilised to construct expression vectors for the recombinant production of mAbs using cell-based expression systems (Posner et al., 2019). Compared with hybridoma techniques, recombinant methods offer better process repeatability and consistency, shorter manufacturing times, smoother transitions between antibody formats, and an animal-free production process (Baghini et al., 2023). For these reasons, the large-scale manufacturing of therapeutic mAbs now primarily relies on recombinant DNA technology using cell-based expression systems.

1.5 Recombinant Production of mAbs Relies on Mammalian Cell Expression Systems, with CHO Cells being the Preferred Host

Significant optimisation efforts in CLD processes, media formulation, and bioreactor conditions over the past two decades have transformed mammalian cells into the preferred expression system for recombinant mAb production (Wurm, 2004). This transformation has occurred despite historical challenges, including low yield, medium complexity, serum dependency, and shear sensitivity (Li et al., 2010; Walsh and Walsh, 2022). In fact, 107 out of 159 approved recombinant products between 2018 and 2022 (67%) were manufactured using mammalian cells (Walsh and Walsh, 2022). A key advantage of mammalian cells is their ability to perform the correct post-translational modifications (PTMs). These PTMs ensure proper protein folding, human-like glycosylation patterns, and prevention of protein

aggregation, all of which are crucial for the efficient secretion, drug efficacy, and stability of mAbs (Jenkins et al., 2008).

In particular, CHO cells have emerged as the dominant mammalian cell line for the production of mAbs. CHO cells were originally chosen for their safety (as they are less prone to propagating human pathogenic viruses), genetic plasticity (allowing for straightforward integration of foreign DNA into their genome), and their robust growth characteristics (Wurm and Hacker, 2011). CHO cells were used to produce the first approved recombinant biopharmaceutical (tissue plasminogen activator) from mammalian cells in 1986 (Kunert and Reinhart, 2016). More recently, CHO cells were used for the manufacture of 95 out of the 107 mAb-based therapeutics approved between 2018 and 2022 that were manufactured using mammalian cells (89%) (Walsh and Walsh, 2022). Over the same time period, NS0 mouse myeloma cells were used for the manufacture of 7 mAb-based products. However, the broad-scale application of NS0 mouse myeloma cells has been limited by their production of two immunogenic glycan epitopes, namely N-glycolylneuraminic acid and galactose-α1,3-galactose (Dumont et al., 2016; Li et al., 2010). Several other mammalian systems were also employed over the same time period, each for the production of a single product, including baby hamster kidney (BHK) cells, human embryonic kidney (HEK) cells, sp2/0 mouse myeloma cells, and PER C6 immortalised primary human embryonic retinal cells. Notably, some products, particularly biosimilars and 'me-too' type products (products exhibiting an incremental improvement on an existing active pharmaceutical ingredient) approved between 2018 and 2022, did not require the glycosylation or other PTMs typically associated with mammalian systems. This enabled cost-effective production in non-mammalian systems, most commonly *Escherichia coli* (Walsh and Walsh, 2022).

CHO cells have gained recognition as a reliable system for mAb production, consistently securing regulatory approval (Li et al., 2010). However, it is important to note that the term “CHO cell” represents a diverse range of parental cell lines, including CHO-S, CHO-Pro^{-}, CHO-K1, CHO-DG44, CHO-

S, CHO-DXB11, each characterised by distinct genetic backgrounds and phenotypic traits (Wurm and Wurm, 2021). The origin of all CHO cells can be traced back to an immortalisation process initiated from an ovarian biopsy of an adult Chinese hamster within Prof. Theodore T. Puck's laboratory (Puck et al., 1958; Wurm and Wurm, 2021). The absence of comprehensive documentation tracking the movements of CHO cell lines between different laboratories and researchers, combined with undefined numbers of subcultivations (using diverse media and culture systems) from the early CHO-*ori* cultures to the present, presents substantial challenges in precisely outlining the lineage of the different CHO cell lines (Wurm and Wurm, 2017). A decade after the original CHO-*ori* cell line was generated, the proline-deficient CHO-K1 cell line was derived from the original CHO-*ori* cell line (Kao and Puck, 1968). The CHO-S cell line was produced by adapting an early-variant CHO cell line for suspension growth by Thompson et al in 1974.

Several mutagenesis experiments have produced auxotrophic CHO cell lines, enabling the development of selection systems for recombinant cell lines in mAb manufacturing. While alternative auxotrophic-based selection systems have been explored, such as the proline (Budge et al., 2021; Sun et al., 2020) and arginase (Roca et al., 2019) systems, the two most widely used CHO selection systems are the dihydrofolate reductase (DHFR)- and the glutamine synthetase (GS)-based systems (Baghini et al., 2023). In 1980, the CHO-DXB11 cell line, with a deletion of one DHFR allele and an inactivating mutation in the second allele, was produced through chemical mutagenesis (Urlaub and Chasin, 1980). Subsequently, the CHO-DG44 cell line was generated by deleting both DHFR gene alleles from one of the first isolated CHO cell lines (not the K1 cell line) (Urlaub et al., 1983). The GS system, initially designed for cell lines lacking sufficient GS expression, was adapted for use in CHO cell lines by employing the GS inhibitor, methionine sulfoximine (MSX) (Cockett et al., 1990). The efficiency of the GS selection system was enhanced by generating GS-knockout CHOK1SV cell lines using zinc finger nuclease (ZFN) technology (Fan et al., 2012). In DHFR-deficient or GS-deficient CHO cell lines, external sources of the appropriate nutrients, hypoxanthine and thymidine for the DHFR system and glutamine for the GS

system, are essential for cell survival. Selection is achieved by introducing a functional DHFR or GS transgene on the expression vector, rescuing cell growth when grown without an external hypoxanthine and thymidine or glutamine nutrient source. To boost mAb productivity, transgene copy number amplification can be achieved by increasing selection stringency over several rounds, via the addition of MSX or methotrexate (MTX, a DHFR inhibitor).

1.6 Fed-Batch Cultures (FBCs) Remain the Primary Cultivation Mode for CHO Cell-based mAb Manufacture

The manufacture of mAb-based therapeutics usually relies on the large-scale suspension cultures of genetically engineered CHO cells in bioreactors (Kelley, 2009). Manufacturing processes can be broadly categorised into two stages: upstream and downstream processing. Upstream processes encompass the cell culture and harvest steps, while downstream processes involve the purification of the mAb product through a series of chromatography and filtration steps (Hong et al., 2018). The primary focus of this PhD research is on upstream processes, and thus, no further discussion on downstream processing will be provided. A comprehensive discussion of the recent advances and future directions in the field of downstream processing can be found in a recent review by *Matte, 2022*.

Two of the most commonly used methods for CHO cell upstream processing are fed-batch cultures (FBCs) and perfusion cultures. FBC, a modified version of a batch culture, involves the sequential addition of nutrients into a fixed culture volume, to prevent nutrient limitation, prolong culture duration, and maximise product titre. Once culture viability drops below a certain threshold, the culture is harvested and the product is purified. Alternatively, perfusion culture is characterised by a constant inflow of fresh media and outflow of spent medium at the same rate (Bielser et al., 2018). For suspension cultures, such as CHO cells, a cell retention device is required to keep the cells inside the bioreactor while media and product are removed. Perfusion cultures were initially designed for the manufacture of unstable

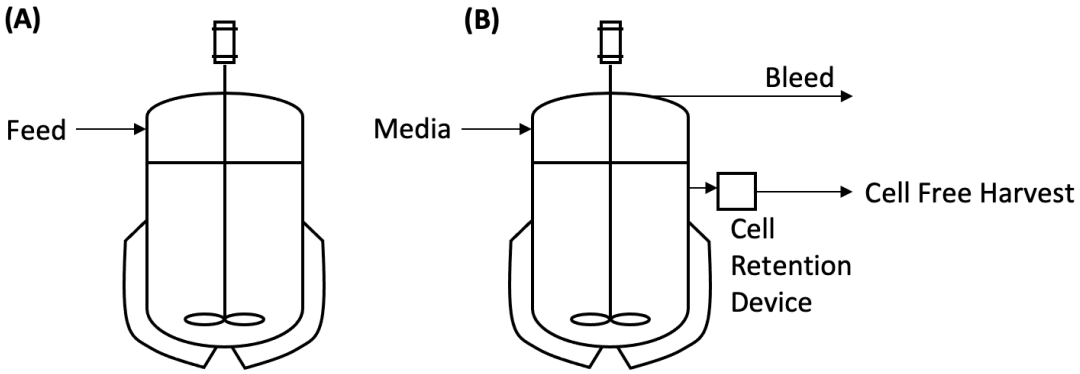


Figure 1.2 - Schematic of Culture Processes

Figure adapted from *Bielser et al., 2018*.

(A) Fed-batch culture. Feed is added during the run and product accumulates in the bioreactor.

(B) Perfusion culture. Media is fed into the bioreactor continuously and product is continuously removed (i.e. does not accumulate). A cell retention device keeps the cells inside the bioreactor whilst product is removed. The bleed stream is used to remove excess cells, maintaining a steady cell concentration.

products, such as blood factors and enzymes, with the aim of reducing the residence time (Wong et al., 2022). However, in comparison with FBC, perfusion cultures can also achieve higher cell densities and can be maintained for extended periods, with repeated product harvests throughout that duration (Wurm, 2004). There is also a growing interest in the concept of completely continuous manufacturing, where cells are continuously grown in perfusion culture, product harvests are continually taken, and continuous downstream processes purify the product. Such systems could potentially reduce the required size of bioreactors and thereby lower capital costs (Bielser et al., 2018).

While continuous manufacturing provides an intriguing avenue for exploration, FBCs remain the predominant cultivation mode for stable biopharmaceuticals such as mAbs, mainly due to their ease of operation (Wong et al., 2022). Over the past two decades, substantial efforts have been directed towards optimising FBCs (Xu et al., 2023). Initially conducted

on murine cell lines, the FBC titres were limited by suboptimal media, feed, and bioprocessing conditions, resulting in lower productivities (Wong et al., 2022). However, extensive optimisation of media, cell lines, and processes, particularly for CHO-cell based culture, has enabled mAb titers of up to 10 g/l to be achieved (Kelley, 2009). Consequently, the majority of platform technologies developed for optimising mAb production predominantly employ scale-down models that help develop cell lines and processes for FBC methods, rather than perfusion (Bielser et al., 2018). Therefore, the primary focus of this PhD is the continued optimisation of current platform technologies to improve mAb manufacture, with a particular emphasis on strategies for developing cell lines that will perform well in FBCs.

1.7 Conventional CLD Strategies Tend to be Time Costly and Labour Intensive

CLD refers to the processes used to generate clonal recombinant CHO cell lines for producing mAbs or other proteins of interest. The primary objectives of CLD are to achieve high mAb productivity, long-term expression stability, and the desired quality attributes, such as glycosylation patterns. To provide an initial overview of a conventional CLD timeline, this section outlines the workflow adapted from the Thermo Fisher Scientific's Freedom™ CHO-S™ kit user guide (Thermo Fisher, 2020) and references the timeline described by *Tejwani et al., 2021* to illustrate the duration of each step (Figure 1.3A). Using these methods, a typical CLD process will take approximately 5-6 months to complete. Additionally, the lack of specialised CLD equipment involved makes the process labour-intensive.

The first step in the CLD workflow is the transfection of the host CHO cell line with a mAb expression vector, which contains both the HC and LC expression cassettes (Figure 1.3B). To construct the expression vector, DNA assembly techniques traditionally rely on restriction digestion and the ligation of DNA fragments using type II restriction enzymes. For example, when cloning the HC and LC transgenes into Thermo Fisher's Freedom® pCHO 1.0 plasmid, a combination of AvrII and Bst717I or EcoRV and Pac1 enzymes

are used (Thermo Fisher, 2020). It is worth noting that the Thermo Fisher protocol offers the choice of either MTX or puromycin-mediated selection for the CLD process. Consequently, a DHFR expression cassette is included in the expression vector. Although the Thermo Fisher CHO-S cell line is not DHFR deficient, like CHO-DG44 cells, the inclusion of the DHFR gene means that cells that have successfully integrated the expression vector will exhibit enhanced resistance to MTX compared with the parental host.

The majority of CLD strategies currently employ transfection methods whereby the expression vector is integrated randomly into the host cell line, yielding heterogeneous pools. These pools are subsequently subjected to selection using MTX. The MTX concentration may be incrementally increased to enhance cell protein productivity through gene amplification, which is characterised by an increase in recombinant gene copy number in cells (Kelley, 2009). To identify the most productive pools and assess their product quality in terms of charge, glycan profile, and aggregation, a productivity assessment is conducted. This assessment may involve batch studies conducted over 5 days, FBC studies spanning 2 weeks, or a combination of both, depending on the specific evaluations required.

Clonality is considered a crucial factor in CLD processes, as clonal lineages are believed to produce mAbs with a stable and consistent product quality profile (Tihanyi and Nyitray, 2020). Traditionally, limiting dilution cloning (LDC) was a common technique for isolating single cells in individual wells (Priola et al., 2016). In this process, a stable pool of transfected cells is diluted to a concentration whereby, on average, each well receives fewer than 1 cell, statistically increasing the likelihood that some wells will end up with only a single cell (Priola et al., 2016). Often, two rounds of LDC are necessary to achieve a sufficiently high statistical probability of clonality (Quiroz and Tsao, 2016).

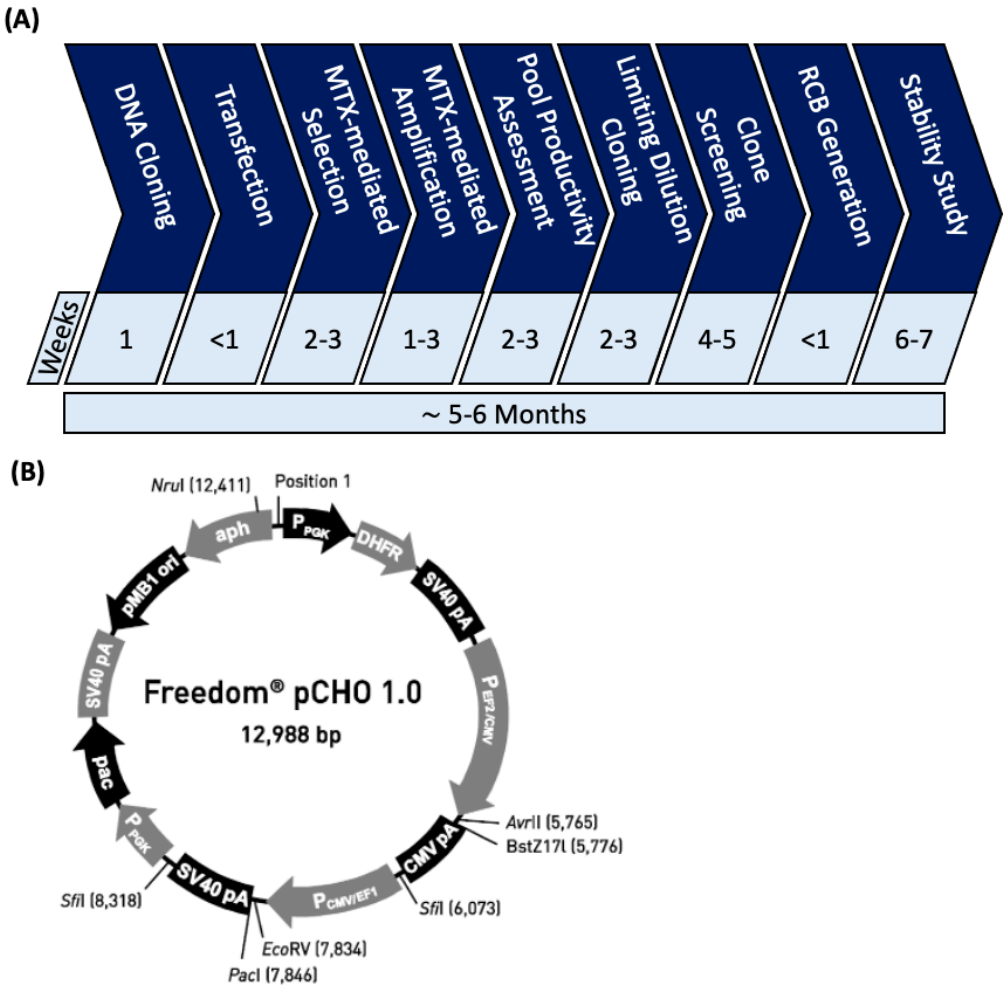


Figure 1.3 - Traditional CLD Approaches are Labour Intensive and Time Consuming.

(A) The timeline of a traditional CLD approach without the use of specialised equipment. Figure is adapted from Tejwani et al., 2021 and the Thermo Fisher Scientific's Freedom™ CHO-S™ kit user guide (Thermo Fisher, 2020). **(B)** Plasmid map of the Thermo Fisher's Freedom® pCHO 1.0 taken from Thermo Fisher's User Guide (Thermo Fisher, 2020).

Following the single-cell cloning step, clones are expanded into individual populations, and subsequent screening steps are conducted to select clones that are producing high quantities of mAb. Productivity assessments are typically conducted using SDS-PAGE or protein-specific ELISA. Thermo Fisher's protocol suggests three rounds of screening. Firstly, productivity is assessed in 96-well plates to identify the top 100 high-producing clones. After clone expansion, each clone is screened in 6-well plate format to identify the top 30. Finally, clones are screened in flasks, FBCs, to identify the top 10 high-producing clones.

One of the ongoing challenges in CLD strategies, which will be explored further in section 1.8.2, is the unpredictable loss of productivity that recombinant CHO cells can experience. Therefore, to prevent unstable cell lines from progressing to the manufacturing stage, a stability study is conducted. The Thermo Fisher protocol suggests that the top 10 clones undergo an evaluation of production stability over at least 60 generations, adhering to regulatory requirements. During such studies, culture samples are taken at various generation timepoints and used to seed FBCs. Cell lines that do not maintain >70% of their volumetric product concentration over the FBCs are usually identified as unstable and are not taken forward to the manufacture stage (Dahodwala and Lee, 2019).

1.8 The Growing Biopharmaceutical Industry Continues to Drive Innovation in CLD

The growing biopharmaceutical market, particularly the increasing number of mAbs under development, has driven the search for innovation in CLD. These innovations aim to generate cell lines capable of producing high quantities of mAbs more quickly and more cost-effectively, while maintaining product quality and long-term production stability (Li et al., 2010). However, several challenges still exist for CHO cell-based biopharmaceutical production.

1.8.1 Increasing Product Titre and the Challenge of New Modalities

The pursuit of higher production titres has been an important driving force behind innovations in CLD. Improvements in host cell lines, expression vectors, media formulations, and process conditions have enabled the routine achievement of titres ranging from 3 to 8 g/l for most mAbs, representing a remarkable (approximately 100-fold) increase in production titres over the past 25 years (Wurm, 2004). In fact, mAb titres have reached such high product concentrations that questions have arisen about whether increases in mAb titres beyond 8 g/l may yield diminishing returns (Kelley et al., 2018). Such concerns are rooted in the likelihood of downstream processing bottlenecks that may prevent the full recovery of mAb products when high product concentrations from production-scale bioreactors are achieved (Kelley et al., 2018). Common bottlenecks include constraints in in-process pool tank volumes, buffer make-up volumes, and chromatographic or ultrafiltration capacities (Kelley et al., 2018).

Despite the overall success of CLD strategies, generating cell lines capable of achieving industrially acceptable titres for certain therapeutic products, including some mAbs, remains a challenge (Alves and Dobrowsky, 2017). For example, certain mAbs are difficult to express at industrial-level titres (Kaneyoshi et al., 2019; Pybus et al., 2014; Reinhart et al., 2014). In particular, the secretion of specific mAbs can be problematic as cause protein aggregation, triggering the unfolded protein response (Kaneyoshi et al., 2019; Reinhart et al., 2019). The commercial and therapeutic success of mAbs has also spurred the development of the next generation of antibody therapies, including engineered products such as Fc-fusion proteins and bispecific/multispecific antibodies. Fc-fusion proteins, for example, have a tendency to aggregate within cells, posing challenges for the secretory capacity of the host cell (Johari et al., 2015). Alternatively, protein engineering strategies have yielded novel proteins, such as the hyperactive salt- and actin-resistant variant of human deoxyribonuclease I, that are difficult to produce due to their toxicity in the cell (Lam et al., 2017). Another

significant challenge lies in the correct assembly of multispecific formats. For example, bispecific IgG antibodies consist of two different HCs and two different LCs. Theoretically, random pairing of these different LCs and HCs can result in up to 10 different chain combinations, with only one combination being the correctly assembled bispecific IgG. Ensuring the accurate pairing of LCs and HCs represents a major challenge in bispecific IgG production (Sawant et al., 2020; Brinkmann and Kontermann, 2017).

The intracellular processes involved in recombinant therapeutic protein production in CHO cells are complex and encompass multiple steps. These steps include transgene transcription, transport of mRNA to the endoplasmic reticulum membrane, translation, translocation, protein folding, various PTMs, and eventual protein secretion (Reinhart et al., 2014). Any of these individual steps could provide a bottleneck in protein expression and may be attributed to limitations in the host cell machinery (such as secretory capacity), suboptimal expression vector composition (such as subunit dosage or codon usage) or inherent properties of the molecule which make it more susceptible to degradation, aggregation, or other unfavourable inter-protein interactions that can lead to cytotoxicity (Alves and Dobrowsky, 2017). A platform-based approach to CLD has proven successful for the manufacture of most mAbs, with standardised steps and expression vector backbones. However, the growing variety of alternative product motifs entering the biologics market, which each present unique production manufacturing challenges, means that the development of specific strategies for CLD may be required. For example, one method to increase the percentage of correctly assembled bi-specific antibody product is to optimise subunit dosage through expression vector optimisation (Blanco et al., 2020; Carver et al., 2020; Kaneyoshi et al., 2019). To circumvent the issue of toxic protein expression, decoupling growth and production has been proposed as an alternative manufacturing strategy to constitutive production (Donaldson et al., 2021; Lam et al., 2017). Some of the methods that are being developed for these purposes are discussed in the section 1.9.1.5.

1.8.2 Reducing CLD Timelines and the Stability Study Bottleneck

Reducing CLD timelines has also been a significant focus for the industry. For example, the current FUJIFILM Diosynth Biotechnologies (FDB) CLD strategy can produce cell lines that yield mAb titers of up to 10 g/l in just 10 weeks (Pybus et al., 2020). Speeding up the development of novel therapies is not only beneficial for patients but also advantageous from a pharmaceutical company's perspective. Reducing the time required to move through development stages can provide a competitive advantage by enabling companies to be the first to market with a new product (Cha and Yu, 2014). Moreover, from a sales perspective, the pace of development is a critical factor for revenue generation. The time spent in development affects the period of patent protection once the drug is on the market. Once the patent expires, generic competition significantly reduces sales revenue (Rang and Hill, 2013, p. 14). Additionally, recent examples, such as those during the COVID-19 pandemic, have demonstrated the potential for accelerated development timelines from lead mAb identification to phase 1 studies (Kelley, 2020).

However, unpredictable loss of mAb productivity (termed “production instability”) during long term culture (LTC) in certain cell lines remains a significant issue for CHO cell bioproduction. To ensure that an unstable cell line does not advance to the manufacture stage, the productivity of lead candidate cell lines, identified during CLD, must be evaluated over a period of ~60-70 generations, covering the period of time it takes to scale-up from research cell bank to production volume. As mentioned in section 1.7, cell lines that do not maintain >70% of their volumetric product concentration over 60-70 generation time period are usually identified as unstable and are not taken forward to the manufacture stage (Dahodwala and Lee, 2019). However, the stability study process can take up to 3 months, thus representing a significant bottleneck in the CLD timeline (Wurm and Wurm, 2017).

Understanding the sources of CHO cell instability during LTC is a complex issue, given the diverse and often cell-specific molecular mechanisms involved (Dahodwala and Lee, 2019). There are two well-established causes of CHO cell instability, namely the loss of transgene copy number through chromosomal rearrangements, resulting in reduced productivity (Chusainow et al., 2009; Kim et al., 2011), and epigenetic modifications, such as methylation or histone modification, leading to transgene silencing and decreased productivity (Osterlehner et al., 2011; Veith et al., 2016). Loss of recombinant gene expression has also been associated with the inherent genomic plasticity of CHO cells (Wurm and Wurm, 2017, 2021). Continuous changes in the genomic landscape of the cells, due to chromosomal rearrangements and chromosome loss, can result in increased DNA damage and accumulation of genomic mutations, influencing the stability of CHO cell phenotype over time (Zhang et al., 2016). These genotypic changes may contribute to alterations in metabolism, global gene expression (Jamnikar et al., 2015; Li et al., 2015; Tzani et al., 2021), sensitivity to cell stress (Bailey et al., 2012; Chusainow et al., 2009), DNA repair mechanisms (Torres et al., 2023) and propensity to undergo apoptosis (Dorai et al., 2012), all of which have been associated with production instability during LTC of recombinant CHO cell lines.

To address the challenge of production instability, substantial efforts have been made to understand its root causes and develop methods to enhance production. These efforts often involve modifying the cells or the expression vector (discussed further in section1.9). However, a significant opportunity to substantially reduce CLD timelines could be in the development of a system which enables early prediction of CHO cell instability. While regulatory requirements may still necessitate a stability study, predictive strategies could provide greater confidence in a cell line's stability at an early stage. This means that cell lines could progress to the next manufacturing stages before the stability study is completed.

1.9 Areas for Optimisation in CLD – Recent Developments and Future Opportunities

1.9.1 Expression Vector Optimisation

To maximise recombinant protein production in CHO cells, the expression vector offers multiple opportunities for optimisation. The mAb expression vector typically includes epigenetic regulatory elements (EREs), strong promoters, 5'-untranslated regions (5'-UTRs), signal peptides, codon-optimised sequences for both the light and heavy chains, and 3'-untranslated regions (3'-UTRs) which contain polyadenylation sequences (Figure 1.4).

The optimisation of individual vector components has led to the creation of platform vector backbones, widely utilised in the manufacturing of most mAb products. However, persisting challenges, such as ensuring expression stability during LTC, as well as the emerging complexities associated with the expression of difficult to express proteins, such as bispecifics and toxic proteins, underscore the need for ongoing expression vector development. Such advancements could be instrumental in shaping the development of product specific CLD strategies necessary to address these unique challenges.

1.9.1.1 Promoter Engineering to Boost Transcription and Resist Gene Silencing

To boost transcription, and reduce transgene silencing, expression vector optimisation strategies have largely focused on promoter engineering. Traditionally, viral (e.g., cytomegalovirus, CMV) or endogenous mammalian (e.g., Elongation Factor 1α, EF-1α) promoters have been used to drive recombinant gene expression in CHO cell lines (Romanova and Noll, 2018). However, epigenetic silencing of CMV (via methylation of the promoter sequence) can lead to loss of gene expression (Osterlehner et al., 2011; Yang et al., 2010). Removal of CpG dinucleotides from CMV reduced promoter methylation but did not significantly improve long term expression stability due to the accumulation of repressive histone modifications (Ho et al., 2016).

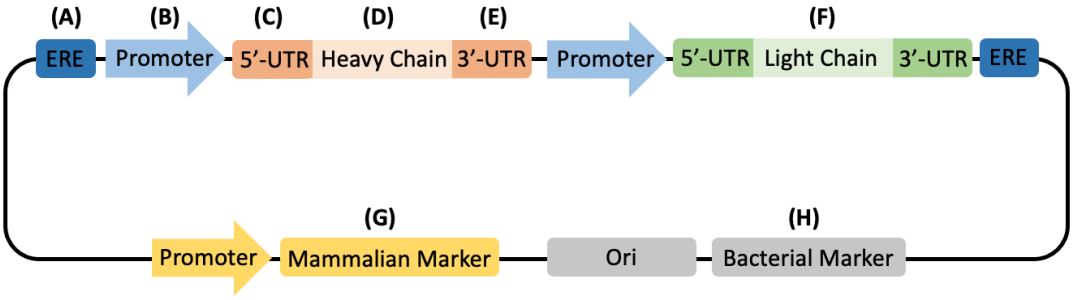


Figure 1.4 - Components of a mAb Expression Vector

(A) EREs are often positioned upstream and downstream of the expression cassette to protect the transgene from silencing. **(B)** Strong constitutive or inducible promoters are used to regulate transcription. **(C)** The 5'-untranslated region (5'-UTR) regulate translations and is positioned upstream of the initiation codon. The 5'-UTR also contains a Kozak sequence which provides a strong translation initiation site. **(D)** A signal peptide is often included to improve product secretion of both the heavy and LC. **(E)** A codon optimised gene sequence encoding the HC of the monoclonal antibody. **(F)** The 3'-untranslated region (3'-UTR) contains a variety of regulatory sequences (such as microRNA response elements, AU-rich elements, and the polyadenylation signal) and is positioned downstream of the translation termination codon. **(G)** A codon optimised gene sequence encoding the LC of the monoclonal antibody. **(H)** A selectable marker is used to enrich cells which have successfully integrated the expression vector. **(I)** The expression vector backbone often contains a bacterial antibiotic resistance marker for DNA assembly purposes.

CHO cell transcriptomic studies are becoming increasingly used for synthetic promoter design. For example, transcriptomic data was used to identify strong endogenous promoters, which were truncated and combined with endogenous enhancers to generate promoters with increased expression stability (Nguyen et al., 2019). Combining transcriptomic data with bioinformatic identification of transcription factor binding sites (TFBSs) offers the opportunity for promoters to be designed with pre-defined characteristics,

such as the ability to drive high levels of transcription and resist epigenetic silencing (Brown et al., 2017, 2014; Cheng and Alper, 2016; Johari et al., 2019). For example, a suite of constitutive promoters were designed to drive high-level recombinant gene expression by combining TFBSs derived from the promoters of highly expressed genes (Johari et al., 2019). The suite included a promoter (1/09) which produced a 2.5-fold increase in SEAP mRNA abundance compared with a standard CMV promoter. Additionally, TFBSs can be arranged to minimise sequence characteristics (CpG islands, repetitive sequences) which make a promoter prone to silencing (Brown et al., 2017).

1.9.1.2 The Incorporation of Epigenetic-Regulatory Elements May Prevent Transgene Silencing

The strength and stability of mAb expression is impacted by the genomic position at which the transgene integrates (termed ‘position effects’). EREs are expression vector components which influence the chromatin environment surrounding the transgene at the site of integration, to promote gene expression and prevent gene silencing, thus minimising the negative impact of a gene integrating into a region of heterochromatin (Harraghy et al., 2015). Therefore, EREs are a particularly useful expression vector component for random integration strategies, which do not control the site of transgene integration. There are a range of DNA sequences considered to be EREs and the mechanisms by which such sequences function are diverse (Table 1.1).

Table 1.1- Epigenetic regulatory elements that have been tested in CHO cells

Element	Origin	Proposed Mechanism of Action	Example(s)
Barrier Insulators	Boundary between heterochromatin and euchromatin	Binding site for DNA-binding proteins which prevent the spread of heterochromatin	cHS4 (Chung et al., 1993; Grosveld et al., 1987)
Stabilising Antirepressor Elements (STARs)	Protects transgene from silencing in region of heterochromatin	Block the spread of histone deacetylation and methylation	STAR40 (Kwaks et al., 2003)
Ubiquitously acting chromatin opening elements (UCOEs)	Promoter regions of ubiquitously expressed housekeeping genes	Methylation-free CpG island helps to maintain open chromatin environment favourable for gene expression	A2UCOE (Allen and Antoniou, 2007; Antoniou et al., 2003; Williams et al., 2005)
Matrix attachment regions (MARs)	DNA elements which bind to the nuclear matrix	Regulate high-order chromatin structure and position the transgene in topologically independent loops	MAR X_S29 (Girod et al., 2005; Saunders et al., 2015)

For example, barrier elements (usually located between regions of heterochromatin and euchromatin in the genome) recruit DNA-binding proteins which protect gene expression by preventing the spread of heterochromatin (Figure 1.5). DNaseI hypersensitive site 4 of the chicken beta-globin locus control region (cHS4) is one of the more commonly used barrier elements. In chicken erythroid cells, cHS4 maintains β-globin genes in an open, active region of chromatin despite the influence of a 16kb heterochromatin domain situated immediately upstream (Prioleau et al., 1999; Ulianov et al., 2012). However, the results of several studies suggest that the impact of cHS4 on transgene expression stability is cell line dependent and only exhibits moderate blocking activity in CHO cells (Izumi and Gilbert, 2000; Otte et al., 2007; Saunders et al., 2015).

Stabilising Anti-Repressor elements, such as STAR40, were identified in a screen for DNA fragments which protect silencing of gene expression, when a transgene is inserted into a region of heterochromatin (Kwaks et al., 2003). Stable transfection of CHO-K1 with a SEAP expression cassette, flanked by STAR40 elements, generated more clones that produced high SEAP (compared with the unshielded control construct) (Kwaks et al., 2003). Additionally, the presence of STAR40 elements in the expression vector prevented the loss of SEAP expression over 60 generations which was observed in clones that had been transfected with a SEAP expression vector lacking STAR40 elements (Kwaks et al., 2003). STAR40 is located upstream of the IL17R gene on chromosome 22, and is thought to exhibit barrier activity by blocking the spread of histone deacetylation and methylation (Kwaks et al., 2003). However, when STAR40 elements were incorporated into a mAb expression vector at different positions relative to the HC and LC antibody genes, the average expression levels of the clones produced from each construct were similar to that of the control vector (which did not contain a STAR40 element), regardless of where the element was positioned (Saunders et al., 2015).

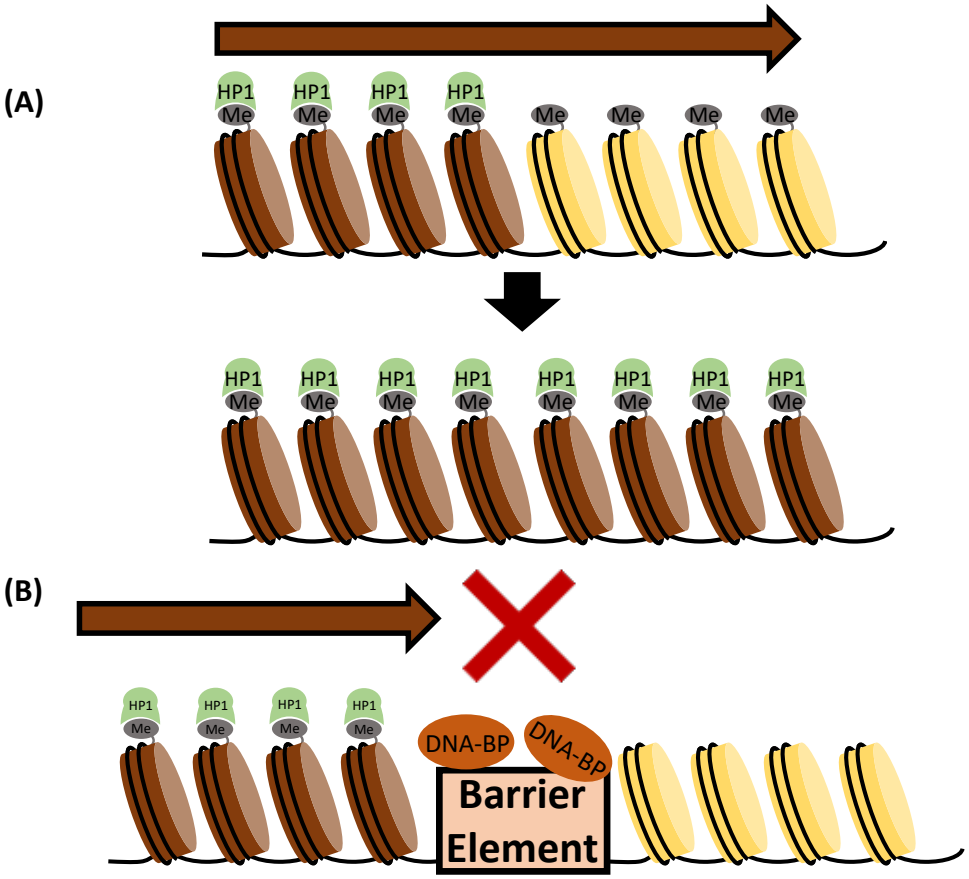


Figure 1.5 - Barrier Elements Prevent the Spread of Heterochromatin

(A) In heterochromatic regions of DNA, c methylates histone 3 lysine 9 (H3K9me2/3). H3K9me2/3 is recognised by binding proteins, such as HP1. The spread of heterochromatin is caused by the recruitment of additional HMTs (via recognition of H3K9me2/3 or binding to HP1), which methylate adjacent nucleosomes (Wang et al., 2016). **(B)** Barrier elements recruit DNA binding proteins (DNA-BP) and histone modifiers which antagonise the HMT-HP1 cycle, thus preventing the spread of heterochromatin.

Another type of ERE are ubiquitously acting chromatin opening elements (UCOEs), which originate from promoter regions of ubiquitously expressed housekeeping genes and consist of methylation-free CpG island fragments (Neville et al., 2017). The ability of UCOEs to regulate the chromatin environment and improve CHO cell bioproduction has recently been reviewed (Sizer and White, 2023). UCOEs are usually positioned upstream of the promoter and have been shown to significantly increase expression and the proportion of high producers (Saunders et al., 2015). However, the impact of UCOEs on expression stability during long-term culture has been inconsistent (Betts et al., 2015; Saunders et al., 2015; Williams et al., 2005). For example, following MTX-mediated amplification, the no-UCOE clones showed higher expression levels than the UCOE-clones.

MARs bind to the DNA matrix, regulate high-order chromatin structure and protect gene expression by positioning the transgene in topologically independent loops (Kwaks and Otte, 2006). Although MARs have been shown to significantly increase long term expression stability (Zhao et al., 2017), MAR X_S29 increased mAb expression significantly less than A2UCOE when the two EREs were compared (Saunders et al., 2015). Therefore, the biomanufacturing industry would benefit from the development of novel EREs.

1.9.1.3 Expression Vector Engineering to Optimise Subunit Dosage

As previously discussed, correct assembly of the 4 different HCs and LCs is a significant challenge for bi-specific IgG production (see section 1.8.1). One method to increase the percentage of correctly assembled product is to optimise subunit dosage (Blanco et al., 2020; Carver et al., 2020; Kaneyoshi et al., 2019). For example, the suite of constitutive promoters designed by combining *Johari et al., 2019*, consisted of promoters with varying transcriptional strengths. A combination of promoters with varying transcriptional strengths could be used to tailor the expression of the light

chain and heavy chain, thus optimising subunit dosage and improving the assembly of bispecific antibodies (Patel et al., 2021).

Alternatively, engineering of the 5'-UTR sequence to alter its secondary and tertiary structure can also be used to control mRNA stability and translational efficiency. The 5'- and 3'- untranslated regions (UTRs) are mRNA domains, located upstream (5'-) and downstream (3'-) of the gene of interest, that regulate many post-transcriptional gene regulation processes. These domains, which consist of a range of regulatory elements involved in pre-mRNA processing, mRNA stability, and translation initiation, are transcribed but rarely translated (Schuster and Hsieh, 2019). GC-rich 5'-UTR sequences can impair translation by forming a stable hairpin structure (Grens and Scheffler, 1990). In fact, the minimum free energy of hairpin structures in the 5'-UTR are a good indicator of translation efficiency (Eisenhut et al., 2020). By generating a suite of 5' UTRs, containing RNA hairpins with varying thermodynamic stabilities, GC-content and position, translation efficiency was regulated (Eisenhut et al., 2020). The strategy was used to reduce translation of HC mRNA in a mAb expression cassette, improving mAb assembly and increasing product titres by 3.5-fold. Such strategies could be adapted to improve the assembly of more complex molecules, such as multispecific formats, by tailoring the translation efficiencies of each component (Blanco et al., 2020). Machine learning methods are being increasingly used to predict 5′ UTR translation efficiency in eukaryotic cells based on sequence data (Cao et al., 2021; Ding et al., 2018; Dvir et al., 2013; Sample et al., 2019). Recently, a predictive model for the impact of 5'-UTR sequences on GFP expression in HEK293T cells was generated using polysome profiling of a library of 280,000 randomised 50 bp synthetic 5′ UTRs. Using this data, a convolutional neural network (CNN) was trained to predict ribosome loading from 5'-UTR sequences (Sample et al., 2019). The model has since been extended to make predictions from longer 5'-UTRs (Karollus et al., 2021). Applying such predictive models in CHO cells will be useful for selecting endogenous 5'-UTR sequences which can be used to tailor translational strength. Alternatively, for the purpose of enhancing protein expression in non-viral DNA therapy, a series of synthetic 5'-UTRs

were generated by evolving endogenous sequences *in silico* (Cao et al., 2021). Using this method, three synthetic 5′-UTRs were generated which yielded moderate increases in GFP expression, when compared with commonly used introns and 5'-UTRs (such as the first intron of the human CMV immediate early gene). Translation is a highly conserved process so strategies to increase production via RNA engineering are likely to function consistently across different CHO production cell lines (Eisenhut et al., 2020; Hershey et al., 2012). In agreement with this, the 5'-UTRs generated by *Cao et al.* and *Eisenhut et al.* were shown to function similarly across a range of mammalian cell types. This contrasts with synthetic promoter strategies, which are often cell line specific (Brown and James, 2017).

The 3'-UTR (also known as the terminator region) is located immediately downstream of the termination codon of the GOI, impacting polyadenylation, translation efficiency, localization and mRNA stability (Plass et al., 2017; Tanguay and Gallie, 1996). A generic structure for the 3'-UTR has been proposed which consists of an upstream sequence element (USE), a highly conserved hexameric polyadenylation signal (PAS), a cleavage/polyadenylation site, and a downstream sequence element (DSE). By modulating each element, a suite of rationally-designed synthetic 3'-UTRs were produced (Cheng et al., 2019). The 3'-UTR sequences were subsequently used to optimise the expression of both GFP and SEAP transgenes in HT1080 WT and HEK293F cells. Increases in recombinant protein production, as a result of optimising the 3'-UTR sequence, were shown to be due to an increase in mRNA stability. However, it must be noted that the synthetic 3'-UTRs produced in this study failed to significantly increase production above commonly used viral 3'-UTRs, such SV40 (Cheng et al., 2019).

1.9.1.4 Signal Peptide Engineering to Improve Recombinant Protein Secretion

Difficulties with post-translational processes, such as mis-folding and protein aggregation, have often been identified as significant bottlenecks in CHO cell protein production (reviewed by Torres et al., 2022). The first step in the classical secretory pathway occurs whilst translation is still in progress, during which the unfolded polypeptide chain is translocated from the cytosol into the lumen of the endoplasmic reticulum. This first step is reliant on the signal peptide, a 20 amino acid long sequence at the N-terminus of the polypeptide chain, which is recognised and bound by a signal recognition particle (Cooper, 2000). Association of the signal peptide with the signal recognition particle inhibits further translation and targets the complex to the signal recognition particle receptor on the endoplasmic reticulum membrane. Various studies have shown that signal peptide optimisation can significantly improve product secretion in CHO cells (Haryadi et al., 2015; Kober et al., 2013; Park et al., 2022; Srila et al., 2022). It is worth noting that the optimal signal peptide can be different for different products (Haryadi et al., 2015; Srila et al., 2022). For example, the human serum albumin preproprotein (SpB) and synthetic and codon-optimised signal peptide (SC) were the optimal SPs for both Trastuzumab and Adalimumab (Srila et al., 2022). However, the best combination for Trastuzumab was SpB_HC + SpSC_LC, whereas the best combination for Adalimumab was SpSC_HC + SpB_LC. It is thought that this is due to the impact of the variable region on protein secretion (Haryadi et al., 2015; Srila et al., 2022). Additionally, these results point to the need for product-specific expression vector optimisation strategies.

1.9.1.5 Inducible Gene Expression Strategies Will Add Flexibility to Manufacturing Processes

Decoupling growth and production has been proposed as a product-specific strategy for certain proteins, such as cytotoxic proteins (e.g. Human DNase I (Lam et al., 2017) and certain mAbs (Misaghi et al., 2014)), that are difficult to produce using standard methods. To implement this strategy, the culture

process is divided into two phases; the growth phase and the production phase. During the growth phase, cells are grown to a high cell density in the absence of transgene expression. Cell proliferation is then halted and transgene expression is induced (the production phase) These ideas, and how they could be implemented, are discussed further in a review publication written during this PhD, found in Appendix 1 (Donaldson et al., 2021).

To decouple growth and production, an important part of the expression vector toolkit will be components that enable inducible transgene expression. Interactions between bacterial response regulator proteins and specific operator DNA sequences have been exploited to create repression- and activation-based inducible gene expression systems which switch on gene expression in response to the addition of a small molecule to the media (Weber and Fussenegger, 2007) (Figure 1.6). In repression-based systems, a regulator protein is often used to sterically block transcription by binding to an operator sequence, which is positioned between a constitutive promoter and the transgene (Figure 1.6A). The addition of a small molecule inhibits the regulator-operator interaction, allowing transcription to occur. The dynamic range of the regulatory system can be improved by fusing mammalian repressor proteins (such as dCas9 (dead Cas9) or KRAB (Krüppel-associated box) to the bacterial regulator protein. Activation-based systems have been created by fusing an activation domain (such as VP16) to a DNA-binding domain to produce a transactivator (Figure 1.6B). In the presence of a small molecule inducer, binding of the transactivator to the operator sequence, which is positioned upstream of a minimal promoter, induces transcription.

Initially, the only small molecule-inducible gene expression systems available responded to antibiotics (such as tetracycline, doxycycline (Gossen et al., 1995) and erythromycin (Weber et al., 2002)) and steroid hormones (such as oestrogen (Braselmann et al., 1993) and mifepristone (Wang et al., 1994)).

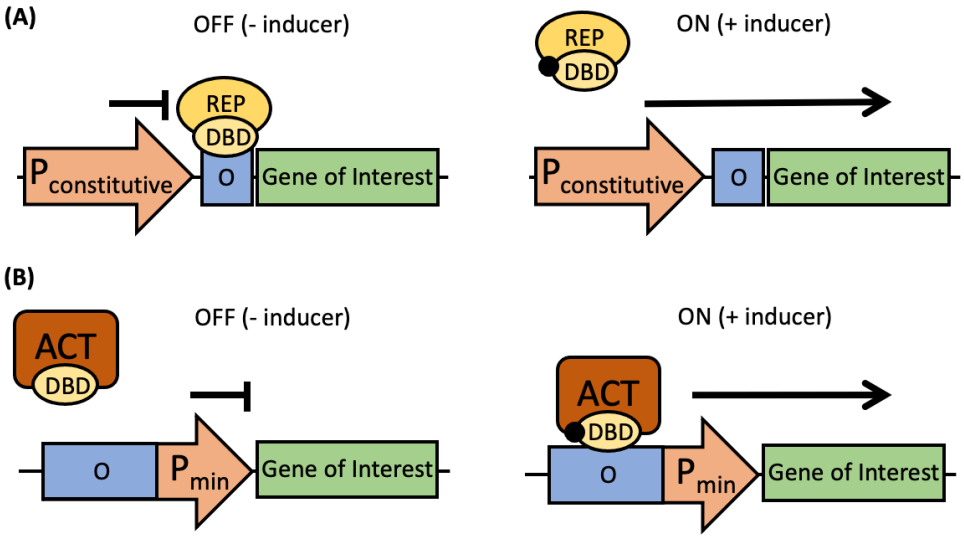


Figure 1.6 - The Architecture of Inducible Gene Expression Systems

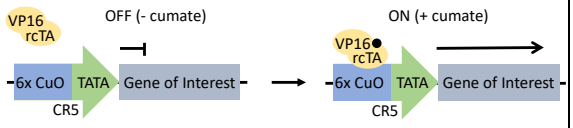
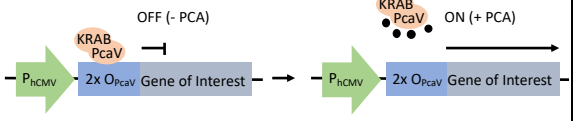
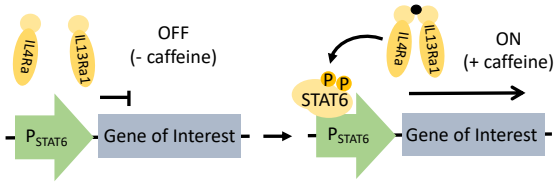
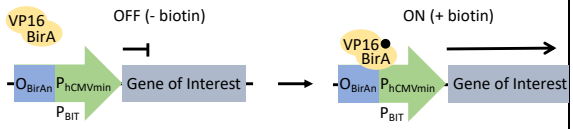
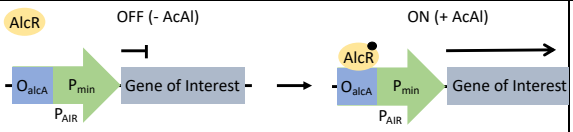
(A) Repressor-based inducible gene expression systems are often constructed by placing an operator sequence (O), downstream of a constitutive promoter. When the small molecule inducer is absent, a DNA-binding domain (DBD), fused with a repressor protein (REP), binds to the operator sequence, sterically inhibiting transcription. Upon addition of the small molecule inducer, the interaction between the regulatory fusion protein and the operator sequence is inhibited, enabling transcription to occur.

(B) Alternatively, activation-based inducible gene expression systems can be generated by placing an operator sequence (O) upstream of a minimal promoter (P_{min}). Upon addition of a small molecule to the media, a transactivator protein (consisting of a DBD fused with an activator (ACT) domain) binds to the operator site and triggers transcription.

However, to be used for bioproduction, a small molecule inducer must be non-toxic, physiologically inactive (excluding antibiotics and steroid hormones), inexpensive and easily removed during downstream processing (Weber and Fussenegger, 2007). Additionally, to avoid media changes during culture, gene expression should be switched on upon addition of the inducer (rather than removal of a repressor). Such criteria have resulted in the development of inducible gene expression systems that respond to physiologically inert substances (such as cumate (Mullick et al., 2006), protocatechuic acid (Yin et al., 2019) and caffeine (Bojar et al., 2018)) or a component of the production medium (such as biotin (Weber et al., 2009) and acetaldehyde (Weber et al., 2004)) (Table 1.2).

Of particular interest is the cumate inducible gene expression system, which improved production of hCD200Fc (by 4-fold) and Rituximab (3-fold) when compared with a standard CMV5 promoter (Poulain et al., 2017). However, inducible gene expression systems are often limited by the need to express additional proteins in the cell line, such as the cumate reverse transactivator in the cumate inducible system (Poulain et al., 2017). Recombinant protein production exerts a considerable metabolic burden on CHO cells, impacting growth rate (Gu et al., 1995, 1994), production (Kallehauge et al., 2017) and cell stress (Poulain et al., 2019).

Table 1.2 - Examples of Inducible Gene Expression Systems Which Switch on Gene Expression in Response to Small Molecule Addition

Small Molecule Inducer	Mechanism	Tested in CHO cells?	Reference
Cumate		Yes	(Gaillet et al., 2010)
Protocatechuic acid		No	(Yin et al., 2019)
Caffeine		No	(Bojar et al., 2018)
Biotin		Yes	(Weber et al., 2009)
Acetaldehyde (AcAl)		Yes	(Weber et al., 2004)

The table presents 5 small molecule inducible gene expression systems. Each system switches ON production when a small molecule (which is physiologically inert or a component of the production medium) is added to the media.

Alternatively, transcriptomic data can inform the design of inducible promoters which do not rely on the addition of small molecules to the media or the expression of additional proteins in the cell line. The first iteration of such promoters were endogenous promoters, isolated from genes with desired expression dynamics (Le et al., 2013; Nguyen et al., 2020). For example, shifting the culture temperature from 37°C to 30-35°C slows cell growth in late phase culture and can be used to prolong culture duration. To switch on production following a temperature shift, cold inducible promoters have been identified which show up to 11-fold increases in luciferase expression at lowered temperatures (Nguyen et al., 2020). However, endogenous promoters frequently contain TFBSs for a variety of TFs, not all of which exhibit specific activity in the desired cell culture phase, thus resulting in unwanted activity in the off state (Wu et al., 2019).

To improve the dynamic range of such promoters, TFBSs found in the promoters of genes, that do not show the desired expression dynamics, can be removed from the promoter sequence (Johari et al., 2019). Alternatively, synthetic promoters can be built *in silico* by assembling modular repeats of TFBSs (corresponding to TFs that are active in the desired context) upstream of a minimal promoter (Johari et al., 2021, 2019; Martinelli and De Simone, 2005; Saxena et al., 2017). To circumvent the need for pre-existing gene expression studies in synthetic promoter design strategies, 6107 different synthetic promoters (each consisting of tandem repeats of a single TFBS and an adenovirus minimal promoter) were screened for specific activity in cancerous cells (Wu et al., 2019). Downstream sequencing and machine-learning based prediction was subsequently used to identify promoters with the desired expression dynamics. This process could be repeated in CHO cells to generate synthetic promoters which function in late phase culture.

1.9.2 Cell Line Optimisation

Over the past 20 years, cell line engineering has yielded incremental improvements to CHO cell characteristics, such as growth (Doolan et al., 2010; Dreesen and Fussenegger, 2011; Jaluria et al., 2007; Park et al., 2000), productivity (Borth et al., 2005; Chung et al., 2004; Kwon et al., 2006; Ohya et al., 2008; Pybus et al., 2014) and product quality (Yamane-Ohnuki et al., 2004; Yang et al., 2015; Zhang et al., 2012). These efforts have relied on permanent changes to the cell by knocking out/ knocking down genes which negatively impact production, or by constitutively overexpressing genes with positive impact (Figure 1a) (Fischer et al., 2015). The emergence of CRISPR/Cas9 technology has made the knockout process quicker, cheaper and more efficient (Grav et al., 2017). For example, several genes have been knocked out simultaneously by co-transfecting multiple guideRNAs alongside the Cas9 nuclease (Grav et al., 2015). Up to 14 genes, encoding host cell proteins (HCPs; contaminating endogenous proteins that co-purify with the desired product), were knocked out using 4 cycles of CRISPR/Cas9-mediated multi-gene disruption, leading to a decrease in HCP content of up to 70% (Kol et al., 2020). Conversely, constitutive overexpression of beneficial transgenes has also been used to improve CHO cell production. Usually, a strong promoter (such as CMV or EF-1α) drives high transgene expression throughout the duration of the FBC. For example, constitutive overexpression of four genes that encode enzymes in the phenylalanine-tyrosine pathway reduced production of the inhibitory by-products, 3-phenyllactate and 4-hydroxyphenyllactate (Mulukutla et al., 2019).

However, certain traits (such as growth rate) do not benefit from permanent gene knockout or constitutive overexpression approaches. During the early phase of a FBC, cell lines should exhibit high growth rates to ensure production titres are not limited by a lack of biomass formation (Figure 1.7A). Once peak cell density is reached however, cell growth should slow down to prolong culture duration and maximise production (Figure 1.7B).

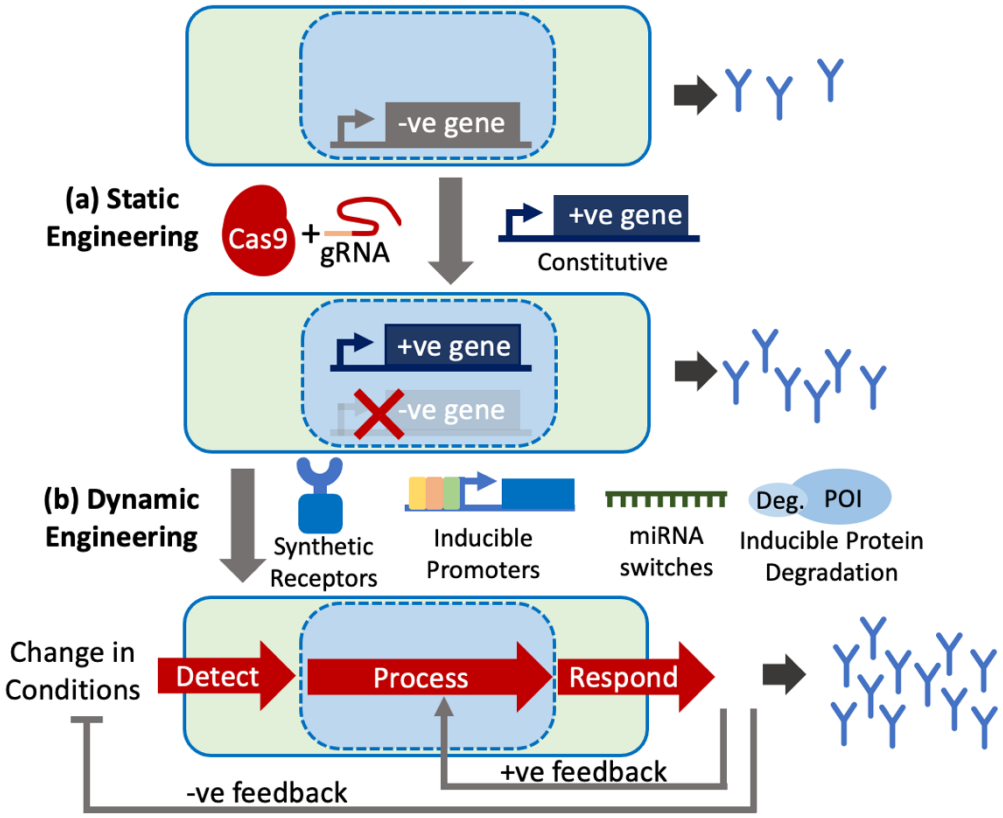


Figure 1.7 - Engineering Self-Regulating CHO cells to Improve Biopharmaceutical Production.

(A) Static engineering approaches have yielded incremental increases in monoclonal antibody production. Mostly, these have used genome engineering tools (such as CRISPR/Cas9) to permanently knock out genes with negative impact on production (-ve gene) or the constitutive overexpression of transgenes with positive impact (+ve gene). **(B)** Dynamic engineering approaches could be used to create next-generation, self-regulating cells which further improve production. Synthetic sense-and-respond programs would enable cells to detect intra- or extracellular changes and generate a response, further augmented by inclusion of negative/ positive feedback circuits. To generate self-regulating CHO cells, a number of recently developed synthetic biology tools such as synthetic receptors, inducible promoters, RNA-level switches and inducible protein degradation systems could be used. Figure and figure legend is taken from (Donaldson et al., 2022).

Constitutive overexpression of certain anti-apoptotic genes (such as Bcl-2 (Lee et al., 2013) or E1B-19K and Aven (Figueroa Jr. et al., 2007)) improves culture duration, but reduces growth rate during exponential phase. Additionally, blocking cell cycle progression, by overexpressing cyclin-dependent kinase inhibitors, has been shown to improve specific productivity and culture duration (Bi et al., 2004; Carvalhal et al., 2003; Fussenegger et al., 1997; Mazur et al., 1998). Alternatively, knockdown/ degradation of cyclins has recently been proposed as an alternative method for proliferation control (Donaldson et al., 2021). Crucially, proliferation control strategies should only be activated in late phase culture, when peak cell density has been reached. Dynamic engineering strategies would enable growth control during the FBC via synthetic cell-to-cell communication (Ma et al., 2022; Weber et al., 2007).

Dynamic engineering strategies may also be advantageous in the regulation of lactate metabolism (Torres et al., 2018). Initially most CHO cells in a FBC are glycolytic (Young, 2013). Glucose is rapidly consumed and lactate is produced to replenish the NAD+ required for glycolysis to continue (Hartley et al., 2018). In high-density CHO cell FBCs, peak lactate concentrations frequently pass 40mM, which severely impacts growth and productivity (Fu et al., 2016; Gagnon et al., 2011). However, as lactate accumulates, certain cultures transition from lactate production to lactate consumption (the lactate switch). Due to the negative impact of lactate accumulation, cultures which undergo a lactate switch tend to produce greater quantities of product (Le et al., 2012) However, not all cultures transition to lactate consumption and the timing of the switch is currently unpredictable. Dynamic engineering strategies could be designed to gain control over the switch by implementing synthetic sense-and-respond programs, whereby genes which shift a cell towards lactate consumption would be upregulated when a threshold lactate concentration is reached. These ideas, and how they could be implemented , are discussed further in a review publication written during this PhD, found in Appendix 2 (Donaldson et al., 2022).

1.9.3 High Throughput Single Cell Cloning and Screening Methods to Reduce CLD Timelines

As previously discussed, CLD relies on single cell sorting methods to generate clonal cell lines (see section 1.7). Ideally, these methods should demonstrate a high cloning efficiency, a term which encompasses several important factors (Scherzinger et al., 2022). Firstly, a high single-cell dispensing efficiency should be achieved, ensuring that single cells are deposited into a high percentage of microplate wells, minimising empty wells or those containing more than one cell. Secondly, these methods should establish a high probability of clonality, increasing the likelihood that each clone originates from a single cell. Additionally, they should ensure a high recovery efficiency by maintaining high cell viability throughout the process, resulting in wells with healthy, growing colonies after single cell dispensing. An ideal single cell cloning method should also offer early insights into clone productivity and growth characteristics.

LDC has traditionally provided a simple method for single-cell cloning that does not require specialised equipment (Priola et al., 2016). However, the method is labour intensive, generates microplates with a high percentage of wells which are empty or contain multiple cells, and often requires multiple subcloning rounds to ensure monoclonality. Furthermore, manual product concentration measurements, using techniques such as the enzyme-linked immunosorbent assay (ELISA), and cell count analyses are typically required to assess clone productivity and growth characteristics (Priola et al., 2016).

To address single-cell sorting limitations, high-throughput methods such as fluorescence-activated cell sorting (FACS) have emerged, enabling rapid analysis of millions of cells per minute. To assess clone productivity, mAb secretion must be converted into a fluorescent signal (Priola et al., 2016). This is typically achieved by linking mAb expression with reporters such as fluorescent proteins or cell surface markers (e.g., CD20), often via reporter-mAb fusions or linker sequences (Gallagher and Kelly, 2017). However, the expression of reporter proteins may increase the metabolic burden on cells,

and the fusion of reporters with mAbs may adversely affect proper protein folding. An alternative approach uses fluorescently labelled antibodies to bind directly to secreted mAb on the cell surface or in proximity to secreting cells. Since secreted proteins are transported to the plasma membrane before release into the extracellular environment, cell surface protein levels should correlate with secreted protein levels, although the accuracy of this correlation remains uncertain (Borth et al., 2000; Dorai et al., 2012; Du et al., 2013; Pichler et al., 2009). To overcome secretion-related mAb staining challenges, immobilisation methods, using encapsulation (Gray et al., 1995; Powell and Weaver, 1990) or a cellular affinity matrix (Holmes and Al-Rubeai, 1999), have been explored. However, these immobilisation techniques tend to be complex, time-consuming, and less high throughput than methods for assessing the expression of cell surface proteins (Priola et al., 2016). Another simpler method, known as cold capture, delays protein release at lower temperatures, allowing for staining of secreted proteins on the cell surface using fluorescent antibodies (Brezinsky et al., 2003; Pichler et al., 2009). Although the proportion of wells plated with only one cell can be much higher with FACS than with the LDC method, high pressures in the FACS system may risk damaging delicate cell lines, potentially impacting cell viability post-sorting (Nawaz et al., 2015).

To address the limitations of FACS, a range of alternative approaches have been developed. Fluorescence based technologies, such as the ClonePix™, combine cell growth in semisolid media with automated fluorescent detection and cell picking. For example, in the ClonePix™ system cells are plated into 6 well plates at low densities, is followed by clone screening which can incorporate both productivity and growth screens (Figure 1.8A). The system is then able to pick the selected colonies and export them into a 96 WP for clonal expansion. Although it has been claimed that the ClonePix™ picker is 10x faster than labour-intensive LDC and FACS (Molecular Devices, 2023), additional rounds of subcloning may be required to achieve the desired probability of clonality (Young et al., 2016).

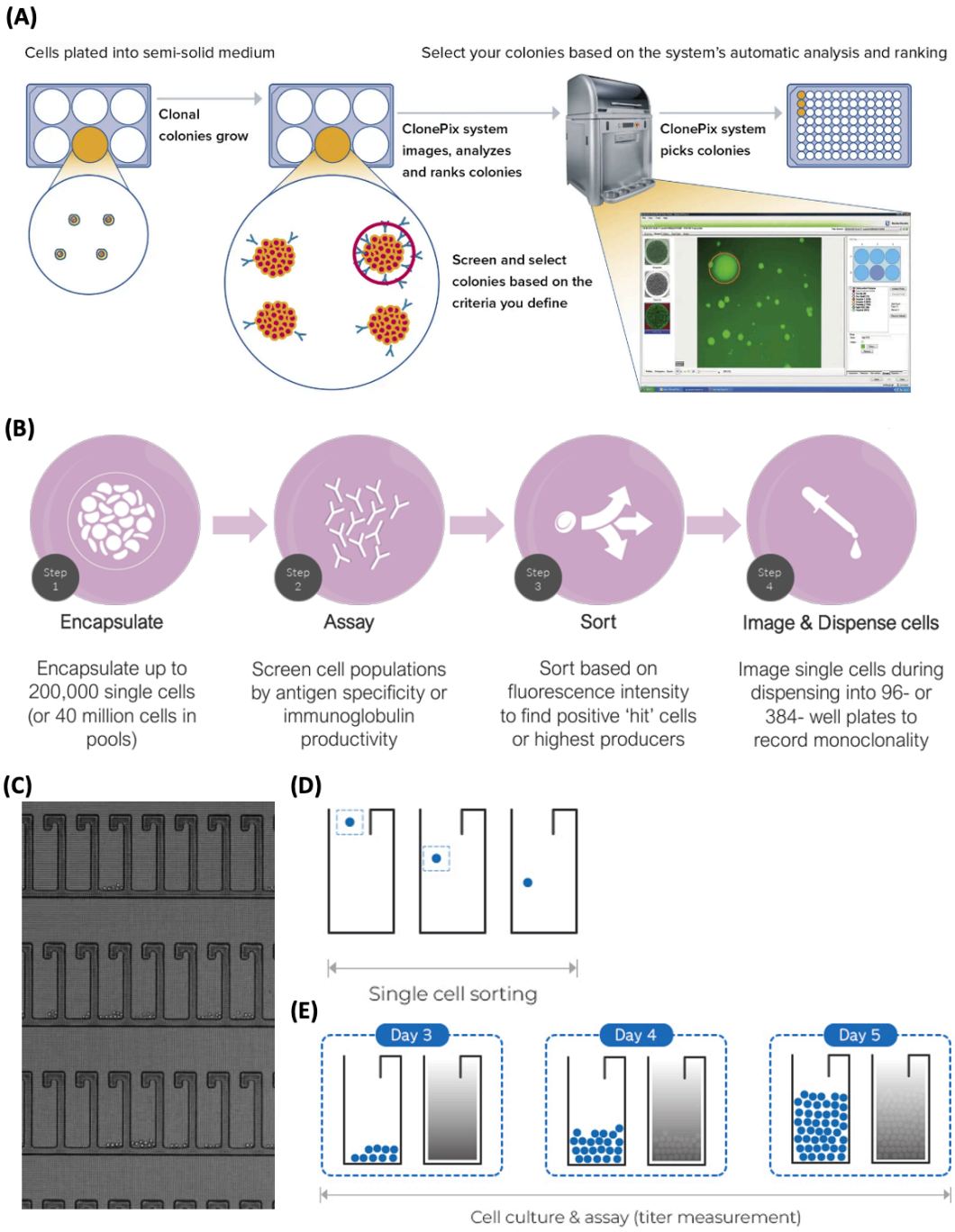


Figure 1.8 - Single Cell Sorting Systems for CLD

(A) Workflow for the ClonePix™ System. Figure was taken from the Molecular Devices website (Molecular Devices, 2023) **(B)** Workflow for the Sphere Fluidics' Cyto-mine® technology. Figure taken from Sphere Fluidics (Sphere Fluidics, 2019). **(C)** In the Beacon system, cells are grown in individual nanopens on a chip. Image from the Berkley Lights Website (Berkley Lights, 2023).

(Continued on next page)

(Figure 1.8 continued)
(D) The Beacon® system uses a light-controlled deposition step to single cell sort individual cells into one of 1758 nanopens. **(E)** Cells are grown in the nanopens for 5 days, with cell count and titre measurements taken on days 3, 4 and 5 using the spotlight assay. **(D)** and **(E)** were taken from the Samsung Biologics Website (Samsung Biologics, 2023).

Of particular significance has been the development of technologies which enable cells to be suspended in picolitre droplets of cell suspension. For example, Sphere Fluidics' Cyto-mine® technology encapsulates single cells into oil-based picodroplets. In the droplet, a FRET-based assay is able to quantify mAb secretion. Additionally, to provide additional assurance of clonality, many single cell sorting processes are coupled with a microscopic imaging technique at multiple timepoints during the singe cell deposition process. For example, coupling the Cyto-mine® technology with Solentim's Cell Metric plate imaging system enabled high levels of both probability (>99%) and assurance of clonality (Pybus et al., 2022). However, imaging steps often require manual verification, often by two trained scientists, to confirm single cell origin and select candidate clones for subsequent colony picking. Additionally, although many systems incorporate a productivity assay into the single cell cloning step, growth rate of clones cannot be assessed prior to selection.

A recent advancement in the field of single cell cloning has been the development of the Berkley Lights Beacon® (Beacon®) system. The Beacon® system integrates microfluidics and imaging techniques to provide real-time monitoring of the growth and productivity dynamics of individual cells (Figure 1.8C-E) (Berkley Lights, 2023). Upon loading of the cell line onto the instrument, a light-controlled deposition process sorts individual cells into one of the 1,758 nanopens on the microchip (Figure 1.8C-D). Subsequent perfusion of the chip with fresh medium facilitates the removal of waste products from the chip, allowing cells to be cultured in the nanopens for up to 5 days while maintaining high viability. Over the 5-day culture period,

continual cell counts can be acquired, and mAb productivity can be assessed using spotlight assays (Figure 1.8E). A combination of validation steps and cell-imaging enables cells to be single-cell sorted with a clonality assurance of >99%. By continuously monitoring the growth and productivity of individual cells in real-time, clones that are likely to exhibit favourable characteristics for manufacturing can be identified and exported into a 96 well plate for further screening. In comparison with FACS-based approaches, the Beacon® system has demonstrated the ability to generate clonal cell lines with similar specific productivities in a shorter timeframe and with the screening of fewer clones, underscoring its efficiency and effectiveness in CLD (Le et al., 2018).

1.9.4 Targeted Integration

Targeted integration (TI) strategies provide an alternative method for CLD, offering the potential to significantly reduce timelines (Kelley, 2020). As previously discussed, the traditional method for generating a high-producing recombinant cell line for manufacture is to randomly integrate the plasmid (Section 1.7). However, the location at which a gene integrates into the genome has a significant impact on transcription (referred to as "position effects"), meaning that random integration of a transgene produces a heterogeneous population of high and low producers. Therefore, lengthy screening processes are required to isolate and identify high producer clones which retain stable productivity during LTC. Alternatively, TI should allow transgenes to be directed solely to regions that promote transcription and are resistant to silencing.

To integrate a transgene at a target locus, a double-stranded break (DSB) in the target DNA sequence must first be generated. A range of genome editing tools have been developed over the past 20 years (such as ZFNs, transcription activator like effector nucleases (TALENs)), but the most commonly used method for such purposes is the CRISPR (Clustered regularly interspaced short palindromic repeats) /Cas9 (CRISPR associated system protein 9) system (Figure 1.9A). The CRISPR/Cas9 system consists of two components: a single guide RNA (sgRNA) and a CRISPR-associated endonuclease (Cas protein).

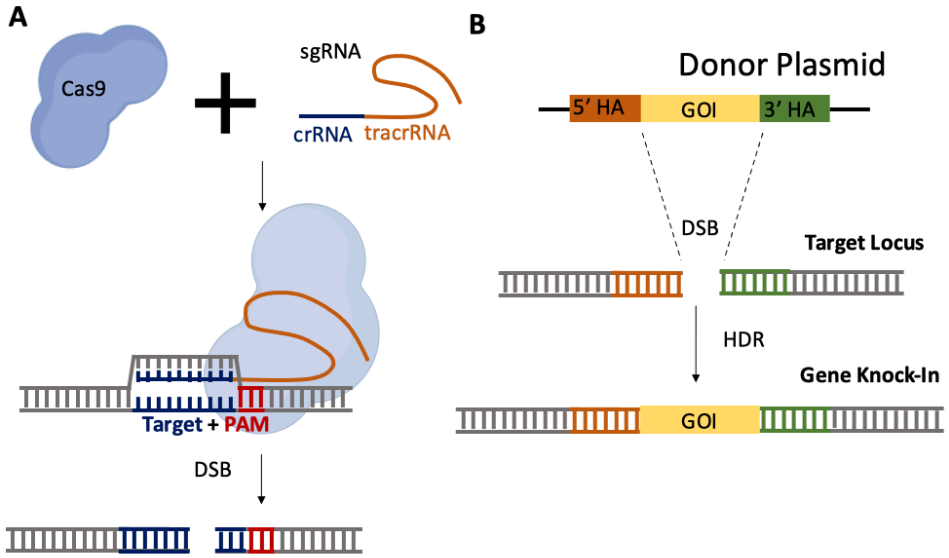


Figure 1.9 - CRISPR/Cas9 for Targeted Gene Integration.
(A) CRISPR/Cas9 is the most used method for the generation of targeted DSBs. **(B)** The DSB can be repaired via 2 pathways: NHEJ or HDR. The HDR pathway, which uses homologous donor DNA to repair DNA damage, can be used for insertion of a GOI into a specific locus.

The sgRNA is a fusion of two RNA molecules: a targeting RNA molecule (CRISPR RNA, crRNA) and a scaffold RNA (trans-activating crRNA, tracrRNA). Target recognition occurs via base pairing of the crRNA to the target genomic sequence, which must be immediately adjacent to a protospacer adjacent motif (PAM). The PAM sequence serves as a binding signal for Cas9, which generates a blunt DSB between the 3rd and the 4th nucleotide upstream from the PAM site (Figure 1.9B). Once a DSB is formed, endogenous cell machinery can repair the DNA via 2 main pathways: non-homologous end-joining (NHEJ) or homology-directed repair (HDR). NHEJ-mediated repair is useful if the intent is to knockout a target gene, as it is prone to generating indel errors.

However, for targeted transgene integration, precise repair via HDR is required, which uses homologous donor DNA to repair DNA damage (Figure 1.9B). To utilise HDR for targeted gene insertion, a DNA repair template

(containing the sequence to be inserted, flanked by segments of DNA homologous to the blunt ends of the cleaved DNA) is delivered into the cell. However, one of the drawbacks of all nuclease-mediated TI strategies (including CRISPR/Cas9) is the risk of aberrant recombination events around the cut site and off-target effects. Therefore, site-specific integration strategies which minimise off-target integration (and the subsequent negative effects) would be preferred.

Site-specific recombination (SSR) provides an alternative method to homologous recombination-based approaches for TI of transgenes. During SSR, a recombinase recognises and brings together two short (<50bp) target recombination sequences from different DNA strands. The strands are then cleaved at specific phosphodiester bonds within the target sites and ligated in a new arrangement, forming recombinants. In contrast with homologous recombination, SSR does not require the formation of a DSB and thus does not rely on endogenous repair pathways. Multiple SSR technologies have been applied in CHO cells, including the tyrosine (Flp-FRT and Cre-loxP) and serine (φC31 and Bxb1) recombinase-based systems. However, it is the serine integrase-based Bxb1 system that has the highest integration efficiency (Figure 1.10A) (Jusiak et al., 2019). Each recombinase shows high levels of specificity for their specific recombination site, ensuring that the off-target activity of recombinases is low. However, due to the lack of naturally occurring recombination sequences in mammalian cells, recombination sequences must be artificially integrated into the desired locus prior to SSR. Although prior integration of so-called “Landing Pads” (LPs) adds an extra step to the engineering process, once the master cell line has been established, it can be used as a platform for repeated integration of transgenes into the same locus.

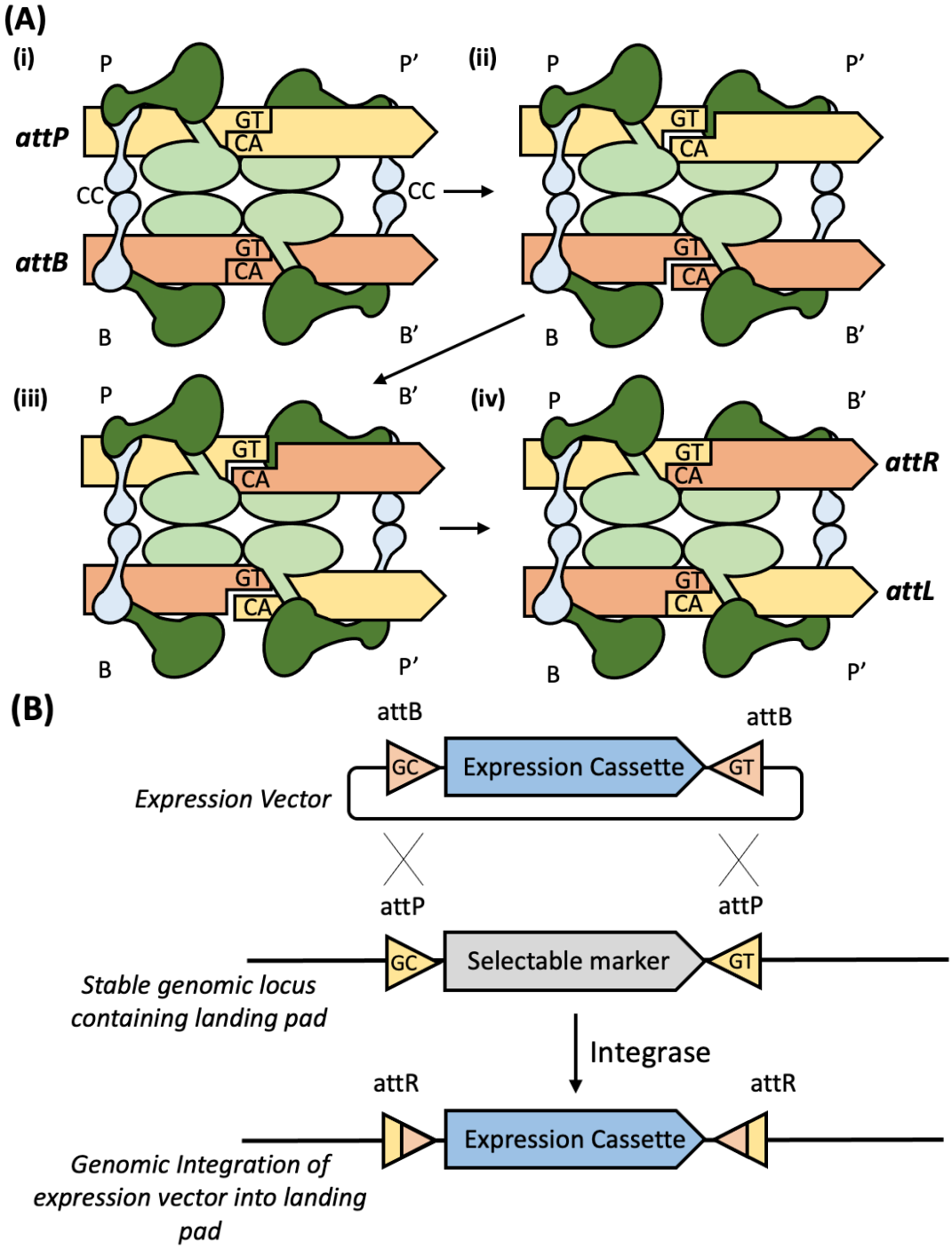


Figure 1.10 - SSR Mediated by a Serine Integrase

(A) Mechanism of serine integrases (figure adapted from Li et al., 2018; Merrick et al., 2018; Rutherford and Van Duyne, 2014). (i) Serine integrase dimers bind to attP and attB sites. Association between the 2 *Int-attB* and *Int-attP* dimers leads to the formation of a tetrameric complex, bringing together the 2 *att* sites. The interaction is stabilised by interactions between coiled-coil (CC)

(Continued on next page)

(Figure 1.10 continued) motifs (ii) All 4 strands are cleaved at the central dinucleotide (generating half-sites with 3′ overhangs that are 2 base pairs long) and the *Int* subunits become covalently attached to their corresponding DNA half-sites. (iii) 180° rotation of the B' and P' *Int* subunits, in relation to the other B and P *Int* subunits, causing the *attP* and *attB* half-sites to swap places. (iv) New attR and attL sites are formed upon re-ligation of DNA half-sites. The coiled-coil motif of the serine integrase is pale blue, *attP* half-sites are pale yellow, and *attB* half-sites are pale orange. **(B)** Integration into a LP via RMCE using Serine Integrases. A LP, with a selection marker flanked by two heterospecific attB sites (with different central dinucleotides), is integrated into a stable genomic locus. An expression vector is subsequently designed with an expression cassette flanked by two heterospecific attP sites, corresponding to the attB sites of the LP. Upon co-transfection of the expression vector with the appropriate serine integrase, integrase-mediated recombination between the corresponding attP sites results in the expression cassette replacing the selectable markers in the LP.

Improvements to LP systems have come with the development of recombinase-mediated cassette exchange (RMCE) technology (Figure 1.10B). By mutating the central overlapping sequence (the 8-bp core) of FRT and loxP sites, or the central dinucleotide of Bxb1 and φC31 recombination sites, heterospecific recombinase target sites have been generated. Such mutations ensure that recombination between two heterospecific sites in *cis* is blocked, but two of the same sites can interact with one another in *trans*. As a result, a LP can be designed which consists of a selection marker flanked by two heterospecific recombination sites. Upon transfection of the master cell line with a donor cassette, with co-expression of the appropriate integrase, the substrate DNA integrates into the LP in a unidirectional manner, replacing the markers.

1.10 Project Aims

Given the persistent challenges in mAb production stability and the emerging difficulties associated with novel modalities during CLD, it was hypothesised that further optimisation of CLD processes could improve the efficiency and reliability of CHO cell-based recombinant protein production. This optimisation was to be achieved via two main avenues: (1) improving expression vector systems and (2) implementing early predictive methods for assessing cell line production stability.

1.10.1 A Targeted Integration Strategy for Expression Vector Component Comparison (Chapter 2)

Given the vast array of expression vector optimisation work that has been conducted, and the potential value of product specific expression vector optimisation strategies, we first sought to develop a system for expression vector component comparison. It was hypothesised that the most efficient and effective method for rapidly comparing components in the same genetic context, thus minimising the impact of independent factors such as transgene copy number and integration site, would be to use a targeted integration strategy.

Chapter 2 Objectives

1. Identify target sites in the CHO cell genome conducive to stable production of mAb (i.e. resistant to silencing) at high product concentrations.
2. Design a landing pad for site-specific recombination, facilitating targeted integration of expression vector components.
3. Employ CRISPR/Cas9 technology to integrate the landing pad at the desired target loci within the CHO cell genome.
4. Perform comprehensive cell line analysis to confirm the correct integration of the landing pad at the target locus and assess for any off-target integration events.

5. Validate the effectiveness of the landing pad design by integrating constructs into the landing pad via recombinase-mediated cassette exchange.
6. Conduct a proof-of-concept study comparing different constitutive promoters to assess the targeted integration strategy's ability to distinguish between various promoter strengths effectively.

1.10.2 Developing a Method for Early Prediction of Production Instability (Chapter 3 and Chapter 4)

Incorporating early predictive methods for cell line stability could considerably reduce CLD timelines by enhancing confidence in cell line production stability. Therefore, a system was envisaged that used the Beacon® system to rapidly identify early indicators of instability in recombinant CHO cell lines. It was hypothesisd that the Beacon would be able to identify a subpopulation of low producing fast growing cells in the unstable cell line at an early generation. However, before this system could be tested, model recombinant CHO cell lines, which displayed different mAb production stabilities, were to be developed and characterised.

Chapter 3 Objectives

1. Characterise the production stability of recombinant CHO cell lines over 60 generations to distinguish between stable and unstable productivity profiles.
2. Utilise growth, metabolite, gene copy number, and mRNA expression analyses to elucidate the underlying factors contributing to production instability in the identified unstable cell line.

Chapter 4 Objectives

1. Optimise the conditions for efficient single-cell cloning using the Beacon® system, including load and bulk media optimisation.
2. Analyse the stable and unstable cell lines using the Beacon® system to identify differences at a single-cell population level.

3. Begin developing a predictive model for cell line stability based on data obtained from the Beacon®, aiming to enhance future predictions of cell line performance and stability.

Chapter 2 - Developing a Targeted Integration System for Expression Vector Component Comparison

2.1 Chapter Summary

Expression vector engineering is a key factor in driving improvements in Chinese Hamster Ovary (CHO) cell production. In this chapter, a landing pad (LP) system was developed to enable expression vector components to be compared in identical genetic contexts. The LP was targeted to 4 genomic loci (Fer1l4, LemD2, Mrpl4 and Cdk2ap2) in the Apollo™ X host cell line genome using CRISPR/Cas9 (clustered regularly interspaced short palindromic repeats/CRISPR associated protein 9). After a series of elimination steps designed for selecting successful clones, the Fer1l4_c8 clone was kept for further analysis. In Fer1l4_c8, the LP was shown to have successfully integrated into the Fer1l4 locus and enabled mCherry-based constructs to be integrated into the LP via cassette exchange. However, it was also hypothesised that there was an additional copy of mNeonGreen (mNG) in the cell line, as part of a non-functioning LP. While some silencing of the mNG transgene in the LP of Fer1l4_c8 occurred over 60 generations, the PCR studies led to hypotheses that this silencing was most likely linked to the LP at the additional genetic locus, rather than silencing at the Fer1l4 locus. To demonstrate the functionality of the LP system, the strength of various constitutive promoters was tested by quantifying mCherry expression levels. The system developed in this chapter will help facilitate the development of future expression vectors and enable systematic identification and optimisation of components to enhance CHO cell productivity.

2.2 Introduction

Fed batch cultures of CHO cells are used for the production of many biopharmaceuticals, particularly monoclonal antibodies (mAbs). Alongside improvements in media composition and cell line engineering methods, expression vector engineering strategies have contributed to mAb titres reaching up to 10 g/L (Kelley, 2020, 2009; Wurm, 2004). However, a growing number of next generation biologics, such as fusion proteins and bi-specific antibodies, are difficult to produce in sufficient quantities using CHO cells (Brinkmann and Kontermann, 2017; Lam et al., 2017; Romanova and Noll, 2018). Additionally, the loss of productivity during long term culture remains a problem in CHO cell-based bioproduction (Dahodwala and Lee, 2019). To overcome such challenges, the next generation of expression vector optimisation strategies will need to optimise existing components (such as promoters, untranslated regions (UTRs) and epigenetic regulatory elements) and generate novel components. It is well established that the standard components of an expression vector are a selection marker (driven by a weak promoter) and a mAb expression cassette, which contain strong promoters, 5'-UTR, signal peptides, codon-optimised light chain (LC) and heavy chain (HC) genes, and 3'-UTRs (containing a polyadenylation sequence) (Kaufman, 2000). Optimisation of each individual component has yielded incremental increases to therapeutic protein production but there are still many potential avenues for expression vector improvement (previously discussed in Chapter 1, section 1.9.1.

A key challenge for expression vector development is the ability to rapidly test and compare components independent of factors such as transgene copy number and integration site. Therefore, a system that can alleviate these dependencies will allow systematic identification and optimisation of expression vector components. The factors that should be considered while evaluating an expression vector comparison system are:

(1) Efficiency - If large numbers of expression vector compositions are to be compared, it is important that this can be done rapidly.

(2) Application – Expression vectors that perform well in the expression vector component comparison system should also perform well when randomly integrated into the CHO cell, as part of cell line development (CLD).

(3) Uniformity – When screening expression vector components, they should be compared in the same genetic context. This means that expression vectors should ideally be integrated at the same genomic locus/loci with equal copy numbers, to ensure that any increases in productivity are only due to expression vector components.

In this section, current expression vector component comparison systems (transient transfection, stable clones and stable pools generated by random integration (RI) of expression vectors) are evaluated and their limitations are discussed.

2.2.1 Limitations of Transient Transfections

A quick and simple transient transfection screen is often used as the first step for comparing large numbers of expression vector improvements (Wang et al., 2022; Yang et al., 2022). Cells are transfected with the candidate expression vector and productivity is measured after 2-5 days, enabling rapid screening of large numbers of expression vector components. However, the value of the information obtained from using such methods is reduced by the inherent limitations of transient transfection techniques. In transient transfections, gene expression is measured from plasmid DNA that has not integrated into the cellular genome. Consequently, expression vector component comparisons are performed before cells with stably integrated expression vector have been selected and isolated. As a result, the expression patterns observed in transiently transfected cell populations are not sustained over multiple generations of cell division. Therefore, the impact of expression vector components on the long-term stability of recombinant gene expression cannot be assessed. Additionally, expression vector components which perform well in transient systems do not always perform well in stable systems. For example, exchanging the BGH poly A for a SV40

poly A sequence in the expression vector failed to significantly increase eGFP expression in a transient transfection screen, but the change led to a significant increase in eGFP expression when compared in a stable integration system (Wang et al., 2022).

There are a few reasons which could account for differences in results between the transient and stable screens. Firstly, product titre can be impacted by transfection efficiency. In transient systems, cell populations with a higher proportion of cells expressing the plasmid of interest will produce higher quantities of recombinant protein than those with a lower transfection efficiency. Alongside factors such as cell health, number of cell passages, contamination, DNA quality and quantity and transfection reagent, transfection efficiency can also be impacted by the size and sequence of the expression vector (Fus-Kujawa et al., 2021). For example, in a previous study with HEK293 cells, eight plasmids containing different promoters driving eGFP expression showed varying transfection efficiencies, ranging from 40% to 80% (Dou et al., 2021). Additionally, transfection efficiency can also be impacted sequence-specific factors. For successful gene expression, the DNA must be transported from the cytoplasm to the nucleus, a process facilitated by the microtubule network and dynein motor (Bai et al., 2017). Interestingly, promoters containing cAMP Response Element-Binding Protein binding domains, such as the CMV or cauliflower mosaic virus 35S promoter, efficiently bound to microtubules, potentially aiding nuclear transfer, while others lacking these domains showed poor binding (Badding et al., 2012). Additionally, stably expressing clones can exhibit different production bottlenecks to cells that have been transiently transfected with the expression vector. For example, a single amino acid mutation in the HC of a mAb reduced production in both stable and transient systems (Mason et al., 2012). Further analysis of mRNA, gene copy number, intracellular antibody content, and secreted antibody, revealed that transcriptional bottlenecks were the primary cause for the drop in titre for the stable system, whereas translational/post-translational mechanisms were the limiting factor in the transient system.

2.2.2 Limitations of Stable Clones Generated by RI

To ensure that components which work well in the screening system also function well in stable recombinant cell lines used for biomanufacturing, an effective expression vector comparison tool could closely mimic CLD by stably integrating the transgene into the genome. Additionally, by stably expressing the transgene, the impact of the expression vector on expression stability may also be assessed. The most commonly used CLD method is to select high producing, fast growing CHO cell clones from a heterogenous population, generated by RI of the transgene (Wurm, 2004). Therefore, an expression vector comparison system could be created by randomly integrating candidate expression vectors into the host cell line and comparing the productivity and stability of CHO cell clones. However, the problem with this strategy is that RI of the transgene produces a heterogenous population of cells, in terms of transgene copy number and insertion site(s). Hence, the isolation of clones from this heterogenous transfected pool of cells for each candidate of the expression vector component does not ensure similar integration patterns in all the candidates, creating different genetic contexts. Both copy number and insertion site(s) have a large impact on recombinant protein production and production stability. This is often termed “position effects”. Therefore, differences in productivity and stability between different candidates of expression vector component could thus be a result of transgene copy number and/or insertion site(s), in addition to component features. As a result, extracting useful information for expression vector component comparison from clones with RI is difficult. Additionally, generating high producing clones for each expression vector combination is time consuming (~3 months) and resource intensive. Therefore, alternative screening strategies which enable rapid and effective comparison of expression vector components, yet still closely mimicking CLD methods, are required.

2.2.3 Limitations in Comparing Stable Pools

To avoid limitations associated with comparing CHO cell clones or transient transfections, a common method for evaluating novel expression vector components is to compare heterogenous pools of cells, produced by RI of the expression vector. Following transfection of candidate expression vectors and subsequent antibiotic selection, heterogenous populations are allowed to recover and compared using fed-batch culture. Stable heterogenous pools can be generated much quicker than stable clones because cloning steps (e.g., via limited dilution cloning) are not required. In theory, heterogenous cell pools (produced by RI of different expression vectors) should contain a similar distribution of transgene copy number and integration sites. Therefore, any variations in productivity (or long-term production stability) between different heterogenous pools should thus be a result of expression vector improvements, rather than differences in site of integration or copy number. However, this is not necessarily true for all expression vector components. For example, ubiquitously acting chromatin opening elements (UCOEs) originate from promoter regions of ubiquitously expressed housekeeping genes and consist of methylation-free CpG island fragments (Neville et al., 2017). Positioning UCOEs upstream of the promoter usually leads to a significant increase in the average expression level of the transfected cell population (Saunders et al., 2015). However, when the highest producing clones of the heterogenous population are isolated, the impact of having the UCOE on recombinant protein production is variable (Betts et al., 2015; Saunders et al., 2015). For example, following MTX-mediated amplification, the no-UCOE clones showed higher expression levels than the UCOE-clones (Betts et al., 2015). Therefore, in certain cases, just because an expression vector component performs well in an heterogenous pool screen, it does not necessarily mean that it will improve the productivity and stability of the highest producing clones.

2.2.4 Considerations for Designing an Expression Vector Component Comparison System: Towards Targeted Integration

Although the previously discussed systems for expression vector component comparison have their advantages, it is clear that the biopharmaceutical industry would benefit from an additional comparison system which reduces the impact of non-expression vector component factors on recombinant gene expression (such as transfection efficiency, site of integration or copy number). In contrast with RI strategies, targeted integration (TI) allows a transcriptionally active genomic locus (that promotes high and stable transgene expression) to be selected and targeted. By targeting an expression vector to a specific locus (with a gene copy number of 1), novel components could be compared in the same genetic context. As previously discussed (Chapter 1, section 1.9.4), site-specific recombination (SSR) provides a suitable approach for this purpose, because it enables rapid targeting of expression vectors to specific loci whilst minimising off-target integration (and the subsequent negative effects). SSR is usually achieved via the integration of a LP, which consists of a selection marker flanked by two heterospecific recombination sites, into the target locus using CRISPR/Cas9. Upon transfection of the LP cell line with a donor cassette, with co-expression of the appropriate integrase, the substrate DNA integrates into the LP in a unidirectional manner, replacing the markers. This is often termed “recombinase-mediated cassette exchange (RMCE)”.

So far, the use of LPs in CHO cells has centred around the development of an alternative system to current RI strategies. With this in mind, LPs have been used to direct transgenes to transcriptionally active regions, also known as “hot spots”, of the genome (Gaidukov et al., 2018; Huang et al., 2007; Kim et al., 2008; Kito et al., 2002; Zhang et al., 2015; Zhou et al., 2010). As previously discussed (Chapter 1, Section 1.7), the traditional method for generating a high-producing recombinant cell line for manufacture is to randomly integrate the plasmid. However, the location at which a gene integrates into the genome has a significant impact on transcription (referred

to as "position effects"), meaning that RI of a transgene produces a heterogeneous population of high and low producers. Therefore, lengthy screening processes are required to isolate and identify high producer clones which retain stable productivity during long term culture.

Alternatively, host cell lines with LPs at genomic hotspots should enable transgenes to be targeted solely to regions that promote transcription and are resistant to silencing. A commonly used method for producing LP hosts has been to randomly integrate the LP into CHO cells and screen the resultant clones for ones that stably express high levels of a marker protein, such as GFP. Upon co-transfection of the parental clone with the appropriate recombinase, donor vectors are targeted to the LP, enabling consistent and predictable production of the protein of interest. More recently, whole genome resequencing (Zhang et al., 2015) or genome walking (Gaidukov et al., 2018) have been used to identify the site of LP integration in the high producing master clones. Such studies have informed the design of single guide RNAs (sgRNAs) and homology arms (HAs) for LP cell lines to be reproduced via CRISPR/Cas9-mediated integration (Gaidukov et al., 2018). Alternatively, rational approaches for hotspot identification have been proposed, based on multi-omics data and *in silico* models (Hilliard and Lee, 2021; Lee et al., 2019).

An example of a hotspot that has been identified through RI screens is *Fer1l4*. In this case, a mAb expression cassette (flanked by heterospecific FRT (flippase recognition target) sites) was randomly integrated into a CHO-GS (glutamine synthetase) cell line (Zhang et al., 2015). After single-cell sorting of the transfected cells, clone 11A7 was identified as a single integration clone which stably produced high-levels of mAb over 220 generations. *Fer1l4* was identified as the site of integration when the cell line was sequenced. To produce the master cell line for integration of future mAb transgenes, a donor vector containing puromycin acetyl transferase and thymidine kinase selection markers (flanked by the appropriate FRT sites) was integrated into clone 11A7 via RMCE, replacing the mAb transgenes.

More recently, the Bxb1 integrase LP system was targeted to Fer1l4 using CRISPR/Cas9 (Inniss et al., 2017). Additionally, Fer1l4 has recently been used as the locus for a GS adapted system (Feary et al., 2021). In this system, the LP (containing a GFP transgene gene flanked by FRT recombination sites) was integrated into the Fer1l4 region of a GS-deficient CHO cell line. Successful RMCE of the donor vector, which contains a promoterless GS expression cassette, removes the glutamine dependency. Together, these studies have shown that Fer1l4 is a genomic locus which enables stable production of mAb transgenes, can be targeted by CRISPR/Cas9 and is amenable to RMCE.

LP systems have also been used for expression vector comparison in CHO cells. For example, altering subunit dosage by adding extra copies of the LC transgene to the expression vector was shown to improve mAb productivity, due to improvements in antibody assembly (Carver et al., 2020). Additionally, optimisation of various expression vector components (promoter, polyA) significantly improved production (Sergeeva et al., 2020a). However, both of these studies were performed in multi-LP cell lines, with the aim of improving titres from TI systems. To produce a system which is made purely for the comparison of expression vector components, it makes sense to use a cell line with a single LP at a single site. There are multiple reasons for this. Firstly, the efficiency of RMCE reduces as the number of LP copies increases (Gaidukov et al., 2018). A reduced efficiency of integration may increase the time it takes for the cell population to recover after transfection, thus increasing the time it takes to screen vector improvements. Additionally, integration of an expression vector into a multi-LP cell line will yield a pool with a range of transgene copy numbers and will require fluorescence-activated cell sorting (FACS) to isolate the cells that have integrated the expression vector into all the LPs. Alternatively, in a single LP cell line, all the cells that have successfully undergone RMCE will survive antibiotic selection and will have the same number of transgene copies.

2.2.5 Proposed Expression Vector Comparison Strategy

Inspired by much of the prior work that has been discussed, an expression vector comparison system with the following properties was envisioned. The platform would be based on a LP integrated into a pre-selected locus in the genome of a mammalian cell line. The genomic LP would exhibit stable expression during long term culture and serve as a site for introducing different expression vectors by SSR.

2.3 Materials

2.3.1 Antibiotics for Molecular Biology

- **Ampicillin:** Stock solutions of 100 mg/ml in 70% EtOH were prepared and filtered using a 0.22 μM filter. Stock diluted to a working concentration of 100 μg/ml.
- **Kanamycin:** Stock solutions of 50 mg/ml in dH20 were prepared and filtered using a 0.22 μM filter. Stock diluted to a working concentration of 50 μg/ml.

2.3.2 Reagents and Solutions for Molecular Biology

- **Agarose (Ultra-Pure):** purchased from Invitrogen (11553277).
- **Potassium acetate (CH3COOK):** purchased from Sigma Aldrich (236497)
- **Rubidium chloride (RbCl):** purchased from Sigma Aldrich (215260)
- **Calcium chloride (CaCl2):** purchased from Thermofisher scientific, (12685077).
- **Manganese chloride (MnCl2)**: purchased from Sigma Aldrich (1375127).
- **Glycerol (C3H8O3)**: purchased from Sigma Aldrich (G9012).
- **Acetic acid (CH3COOH)** purchased from Sigma Aldrich, Cat. No. A6283.
- **MOPS**: purchased from Sigma Aldrich, Cat. No. M1254.
- **Potassium hydroxide (KOH):** purchased from Thermofisher scientific, Cat. No. 10575355.

2.3.3 Buffers

- **TFB1 buffer:** 30 mM CH3COOK, 100 mM RbCl,10 mM CaCl2, 50 mM MnCl2,15% glycerol. pH adjusted to 5.8 with CH3COOH
- **TFB2 buffer:** 10 mM MOPS, 75 mM CaCl, 10mM RbCl, 15% glycerol. pH adjusted to 6.5 with KOH
- **TAE 1X:** 40 mM TRIS, 20 mM Acetate, 1 mM EDTA.

- **Dulbecco's phosphate-buffered saline (DPBS):** purchased from Thermofisher scientific, Cat. No. 14040141

2.3.4 Molecular Biology Kits and Enzymes

- **NEBuilder® HiFi DNA Assembly Cloning Kit:** purchased from NEB (E5520)
- **E.Z.N.A.® Plasmid Mini Kit I (Qspin):** purchased from Omega Bio-Tek (D6942-01)
- **Zymoclean Gel DNA Recovery Kit:** purchased from Zymo (D4006)
- **Esp3I:** purchased from NEB (R0734S)
- **BbsI:** purchased from NEB (R0539S)
- **T4 ligase buffer (10x):** purchased from NEB (B0202S)
- **T4 ligase:** purchased from NEB (M0202S)

2.3.5 Bacterial Culture Media

- **LB:** 10 g/L tryptone, 5 g/L yeast extract, 10 g/L NaCl
- **LB agar:** 10 g/L tryptone, 5 g/L yeast extract, 10 g/L NaCl. 20 g/L agar

2.3.6 Ordered/ Gifted Constructs

Table 2.1 - Ordered/ Gifted Constructs

Code	Construct Name	Description	Notes
Addgene plasmid #42230	pX330-U6-Chimeric_BB-CBh-hSpCas9	sgRNAs cloned into this construct for CRISPR/Cas9 mediated integration of LP	a gift from Feng Zhang (Cong et al., 2013; Ran et al., 2013)
JDcon_1	p2-p12_CAGp-connDE-Bxb1attBGA-mNG-p2A-BSDR-	Part 2 of LP assembly (see Figure 2.2)	ordered from GeneArt on a

Code	Construct Name	Description	Notes
	connJK-polyA-Bxb1attBGT		kanamycin backbone
JDcon_2	p14-p25_connnr-pCMV-mRuby2-polyA-connwz	Part 4 of LP assembly (see Figure 2.2)	ordered from GeneArt on a kanamycin backbone
JDcon_3	20AARWGP_017_Expression_vecto_pMA-RQ_(AmpR)	Plasmid containing attP sites required for integrating into LP. Used as a backbone to clone various mCherry-based constructs into.	ordered from GeneArt on an ampicillin backbone
-	pEIF1p_V5-Bxbl_pA	pEF1α-Bxb1 encoding the Bxb1 serine integrase, used for integration of constructs into LP	a gift from Dirk Jan Kleinjan

2.3.7 Assembled Constructs

Table 2.2 - Assembled Constructs

Code	Construct Name	Description
JDcon_4	px330_cdk2ap2_gRNA	**Cdk2ap2 sgRNA** cloned into pX330-U6-Chimeric_BB-CBh-hSpCas9
JDcon_5	px330_fer1l4_gRNA	**Fer1l4 sgRNA** cloned into pX330-U6-Chimeric_BB-CBh-hSpCas9

Code	Construct Name	Description
JDcon_6	px330_mrpl4_gRNA	**Mrpl4 sgRNA** cloned into pX330-U6-Chimeric_BB-CBh-hSpCas9
JDcon_7	px330_lemd2_gRNA	**LemD2 sgRNA** cloned into pX330-U6-Chimeric_BB-CBh-hSpCas9
JDcon_8	Cdk2ap2_assembled_landing_pad	**(Cdk2ap2)5HA-CAGp-attB(GA)-kozak-mNG-p2A-BSDR-polyA-attB(GT)- (Cdk2ap2)3HA-CMVp-mRuby2** (See section 2.5.3 for details)
JDcon_9	Fer1l4_assembled_landing_pad	**(Fer1l4)5HA-CAGp-attB(GA)-kozak-mNG-p2A-BSDR-polyA- attB(GT)-(Fer1l4)3HA-CMVp-mRuby2** (See section 2.5.3 for details)
JDcon_10	Mrpl4_assembled_landing_pad	**(Mrpl4)5HA-CAGp-attB(GA)-kozak-mNG-p2A-BSDR-polyA- attB(GT)-(Mrpl4)3HA-CMVp-mRuby2** (See section 2.5.3 for details)
JDcon_11	LemD2_assembled_landing_pad	(LemD2)5HA-CAGp-attB(GA)-kozak-mNG-p2A-BSDR-polyA- attB(GT)-(LemD2)3HA-CMVp-mRuby2 (See section 2.5.3 for details)
JDcon_12	hygR-p2A-mcherry	**attP(GA)-kozak-hygR-p2A-mcherry-polyA-attP(GT)** promoterless hygro-p2A-mcherry construct inserted into the LP in section 2.5.12

Code	Construct Name	Description
JDcon_13	Mcherry-p2a-hygR	**attP(GA)-kozak-mcherry-p2A-hygR-polyA-attP(GT)** promoterless hygro-p2A-mcherry construct inserted into the LP in section 2.5.16

2.3.8 Tissue Culture Media

- **Cloning Media - Condition 1:** CD DG44 CHO media (Gibco™, Cat no. 12610010), 6mM L-Glutamine (Gibco™, Cat no. A2916801), 2x GlutaMAX™ (Gibco™, Cat no. 35050087), 0.18% Pluronic™ F-68 Non-ionic Surfactant (Gibco™, Cat no. 24040032), 1x Antibiotic-Antimycotic (Gibco™, Cat no. 15240096), 1x commercially available animal component free (ACF) medium supplement
- **Cloning Media - Condition 2 and 3:** CD DG44 CHO media (Gibco™, Cat no. 12610010), 6mM L-Glutamine (Gibco™, Cat no. A2916801), 2x GlutaMAX™ (Gibco™, Cat no. 35050087), 0.18% Pluronic™ F-68 Non-ionic Surfactant (Gibco™, Cat no. 24040032), 1x Antibiotic-Antimycotic (Gibco™, Cat no. 15240096), 10% Fetal Bovine Serum (Sigma Aldrich, Cat. No. F7524)
- **Media for routine subculture of the Apollo™ X host cell line:** CD DG44 CHO media (Gibco™, Cat no. 12610010), 0.18% Pluronic™ F-68 Non-ionic Surfactant (Gibco™, Cat no. 24040032), 8mM L-Glutamine (Gibco™, Cat no. A2916801)
- **Media for routine subculture of LP cell lines:** CD DG44 CHO media (Gibco™, Cat no. 12610010), 0.18% Pluronic™ F-68 Non-ionic Surfactant (Gibco™, Cat no. 24040032), 8mM L-Glutamine (Gibco™, Cat no. A2916801), 4 μg/ml Blasticidin (BSD) (Invivogen, Cat no. ant-bl-05)
- **Media for RMCE selection:** CD DG44 CHO media (Gibco™, Cat no. 12610010), 0.18% Pluronic™ F-68 Non-ionic Surfactant (Gibco™, Cat no. 24040032), 8mM L-Glutamine (Gibco™, Cat no. A2916801), 500 μg/ml Hygromycin

2.3.9 CRISPR/ Cas9 sgRNA Sequences

Table 2.3 - CRISPR/ Cas9 sgRNA Sequences

Target Gene	Spacer Sequence	Source of Sequence
Fer1l4	CATCCCCCCAGAGCACACAG	Benchling sgRNA Design Tool
LemD2	CCCTGTTATCTGCCACATGC	Benchling sgRNA Design Tool
Mrpl4	TTACAGGCCTCTGAGCCCAG	WO2022123242A1 - CHO CELL MODIFICATION (Rosser and Kleinjan, 2022)
Cdk2ap2	GTCACCTTGCCTCATCTCTG	WO2022123242A1 - CHO CELL MODIFICATION (Rosser and Kleinjan, 2022)

2.3.10 Primers for Diagnostic PCRs

Table 2.4 - Primers for Diagnostic PCRs

Name	Sequence	Application
283_mruby_fwd	GACATTCTTGCCACGTCGTTCATGTATG	Forward primer targeting mruby2 transgene for the amplification of the LP plasmid backbone (see section 2.5.7)
288_amp_rev	AGTTCTGCTATGTGGCGCGGTATTATC	Reverse primer targeting the ampicillin transgene for the amplification of the LP plasmid backbone (see section 2.5.7)

Name	Sequence	Application
236_fer1l4_fwd_3	CGGCTCACATCAGCCCCTTAGAAC	External Fer1l4-locus specific forward primer for PCR amplification of 5' LP region (see section 2.5.8)
305_fer1l4_5prrev_1	ACTAATACGTAGATGTACTGCCAAGTAGGAAAGTC	Internal LP-specific (annealing to the CAGp) reverse primer for PCR amplification of 5' LP region (see section 2.5.8)
294_lppcr_1	GGATAACATGGCCTCTCTCCCAGC	Internal LP-specific (annealing to the mNG transgene) forward primer for PCR amplification of internal LP region (see section 2.5.8)
295_lppcr_2	CAGCATTATTAGCCTTCCCACACGTAGC	Internal LP-specific (annealing to the BSDR transgene) reverse primer for PCR amplification of internal LP region (see section 2.5.8)
273_lp_pcr_fwd	GCTACGTGTGGGAAGGCTAATAATGCTG	Internal LP-specific (annealing to the BSDR transgene) forward primer for PCR amplification of 3' LP region (see section 2.5.8)
237_fer1l4_rev_4	GCATCCCGAGTGCTTGAAGACCTATC	External Fer1l4-locus specific reverse primer for PCR amplification of 3' LP region (see section 2.5.8)
391_ferl14lppcr_fwd	GACTTTCCTACTTGGCAGTACATCTACGTATTAGT	Internal LP-specific (annealing to the CAGp) forward primer for PCR amplification of 5' attB(GA) site for sequencing (see section 2.5.8)

Name	Sequence	Application
244_mrpl4_fwd_3	TGGTGACATGGCTCAATAGCAGAAATTCC	External Mrpl4-locus specific forward primer for PCR amplification of 5' LP region (see section 2.5.10)
306_fer1l4_5prrev_2	GTTACTATGGGAACATACGTCATTATTGACGTCAATG	Internal LP-specific (annealing to the CAGp) reverse primer for PCR amplification of 5' LP region of Mrpl4 (see section 2.5.10)
253_mrpl4_rev_5	CCTGTGTGCAGTGCTCTCAGAGG	External Mrpl4-locus specific forward primer for PCR amplification of 3' LP region (see section 2.5.10)
392_lppcr_fwd	TGCGTTGGGACTTAGCCTTTAGTGAAC	Forward primer targeting the CAGp for diagnostic PCR of LP integration (see section 2.5.13)
393_lppcr_rev	TGGTATCCGGAGCCATCTACCATGG	Reverse primer targeting the mNG transgene for diagnostic PCR of LP integration (see section 2.5.13)
446_hygro_rev	CGCGCATATGAAATCACGCCATGTAG	Reverse primer targeting the hygromycin resistance transgene for diagnostic PCR of LP integration (see section 2.5.13)
447_mcherry_fwd	CTACAACGTCAACATCAAGTTGGACATCACC	Forward primer targeting the mCherry transgene for diagnostic PCR of LP integration (see section 2.5.13)

2.4 Methods

2.4.1 Preparation of Competent Cells

NEB Top10 cells were streaked out onto a Streptomycin LB agar plate (50 µg/ml), and the next day, a single colony was picked to inoculate a 5 ml culture of Streptomycin LB (50 µg/ml). Cultures were shaken at 200 rpm, 37°C overnight and the next morning diluted 1:400 in 500 ml Streptomycin LB. Diluted cultures were then returned to 37°C with 200 rpm shaking and monitored until OD_{600} had reached 0.48. Once cultures had reached an OD_{600} of 0.48, they were immediately transferred to chilled 50 ml falcon tubes and incubated on ice for 10 minutes. The cells were then pelleted from culture by centrifugation for 5 minutes at 3750 g in a precooled centrifuge at 4°C. The supernatant from this spin step was then discarded, and cells were resuspended in 40 ml ice-cold TFB1 buffer per 100 ml original culture and the suspension was incubated on ice for 5 minutes. After incubation, cells were pelleted once more by centrifugation for 10 minutes at 1350 g, 4°C. The supernatant from this step was discarded and replaced with 4 ml ice-cold TFB2 per 100 ml culture. The resuspended cells were then incubated on ice for 15 minutes. Finally, 200 µl aliquots of cells were transferred into Eppendorf tubes on ice, snap-frozen in liquid nitrogen and stored at -80°C.

2.4.2 Bacterial Transformation

Competent cells were thawed on ice for 15 minutes and then incubated with 10 pg - 100 ng DNA for 20 minutes. Cells were then subjected to a 45 second heat shock at 42°C and returned to the ice for 1 minute. 150 µl S.O.C. medium was added to each tube of cells, and tubes were then transferred to a shaker at 37°C 200 rpm. Cells were allowed to recover for 20 minutes for plasmids encoding ampicillin resistance and 1 hour for plasmids encoding kanamycin resistance and plated onto LB-agar containing the appropriate antibiotic.

2.4.3 Plasmid DNA Preparation

3 ml or 10 ml overnight LB cultures containing appropriate antibiotics were inoculated with a single colony of *Escherichia coli* and incubated for >16

hours shaking at 200 rpm and 37°C. Plasmid DNA was extracted using an E.Z.N.A.® Plasmid Mini Kit I, (Q-spin) (Omega) according to manufacturer's instructions and eluted in 50 µl of the provided elution buffer.

2.4.4 Plasmid and Sequence Verification

All new plasmids were subjected to diagnostic restriction digests using NEB restriction enzymes and run on a 1% agarose TAE gel. Those which generated an expected banding pattern were sent for sequencing at GENEWIZ Sequencing. Any PCR products for interrogation were also sent to GENEWIZ Sequencing after gel extraction and cleanup.

2.4.5 EMMA Assembly of LPs

The LP plasmids were assembled from modified EMMA parts (ordered from IDT and GeneArt, see section 2.5.3 for more information) using a modified version of the EMMA assembly protocol (Martella et al., 2017) which omits the PlasmidSafe ATPase digestion step that the original publication suggests. 1-2.5 µl of each assembly was transformed into 50 µl Top10 cells.

2.4.6 Gibson Assembly of Promoterless mCherry-hygro/ hygro-mCherry Plasmids

The constructs Gibson assembly cloning method, NEBuilder® HiFi DNA Assembly Cloning Kit (NEB, E5520) was used to assemble pre-digested and PCR amplified DNA fragments following manufacturer's protocols.

2.4.7 Maintaining Mammalian Cell Lines

The Apollo™ X CHO host cell line was maintained in CD DG44 CHO media (Gibco™, Cat no. 12610010), 0.18% Pluronic™ F-68 Non-ionic Surfactant (Gibco™, Cat no. 24040032) + 8mM L-Glutamine (Gibco™, Cat no. A2916801). Viable cell density and viability were determined using a Countess® II FL Automated Cell Counter (Life Technologies). Cells were sub-cultured at a seeding density of 0.2 x 10^6 cells/ml every 2-3 days. LP cell lines were maintained using the same media and subculture procedure, with BSD (Invivogen, Cat no. ant-bl-05) added at a concentration of 4 µg/µl.

2.4.8 Calculating Generation Number

Generation number was calculated using the equation below:

$$\text{Generation number} = \text{Previous generation number} + \frac{(\ln(\text{viablecellcount}) - \ln(0.5))}{\ln(2)}$$

2.4.9 Generation of Cas9 Edited Cell Lines

One day prior to transfection, Apollo™ X host cells were seeded at 0.5 x 10^6 cells/ml. On the day of transfection, cell density was adjusted to 1 x 10^6 cells/ml and 1 ml of cells were transfected using a total amount of 1 µg of DNA (900 ng of LP plasmid and 100 ng of sgRNA plasmid) packaged with 3 µg of PEI. 72 hours later, cells were spun down and resuspended in 2 ml of BSD-treated media in a 6-well plate. Cells were passaged every 2 days by spinning cells down at 180g for 5 minutes and resuspending in 2 ml fresh BSD-treated media.

2.4.10 FACS

Prior to cell-sorting 5x10^6 cells were spun down at 180g and resuspended in 500 µl of DPBS. FACS sorting was carried out using BD FACS Aria IIIu 4-laser/11 detector Cell Sorter (The University of Edinburgh Institute of Immunology & Infection research Flow Cytometry Core Facility).

2.4.11 Diagnostic PCRs

To amplify genomic DNA, 25 µl reactions of Q5 polymerase were prepared with 5 µl Q5 5x Buffer, 0.2mM dNTP, 0.5M each primer, ddH20 and 50ng template gDNA per reaction. Annealing temperatures were calculated with the NEB Tm Calculator. Thirty-five amplification cycles were carried out for each reaction, and reactions were run on a 2% TAE gel.

2.4.12 Flow Cytometry

All flow cytometry experiments were carried out using the BD Fortessa instrument. 0.5-1 ml of cells were spun down at 180g and resuspended in

200ul of DPBS in 96-well plates. During flow cytometry analysis, cells were first gated by size using forward and side scatters (SSC-A against FSC-A), and singlets were gated using forward scatters (FSC-A against FSC-W). All flow cytometry data analysis was carried out using FlowJo and RStudio.

2.4.13 RMCE-Integration of Expression Cassettes

One day prior to transfection, cells were subcultured at 0.2 x 10^6 cells/ml. On the day of transfection, cell density was adjusted to 0.25 x 10^6 cells/ml and 0.5 ml of culture was transfected with 500 ng of DNA (250 ng of the donor vector and 250 ng of the pEF1α-Bxb1 plasmid) packaged with 1.5 µg of PEI. After 3 days, the cells were spun down at 180g for 5 minutes and resuspended in 1 ml of hygromycin-treated media. Cells were passaged every 3 days by spinning cells down at 180g for 5 minutes and resuspending in 1 ml fresh hygromycin-treated media.

2.5 Results

2.5.1 Identification of a Suitable Cell Line for LP integration

FUJIFILM Diosynth Biotechnologies's (FDB) Apollo™ X DG44-CHO host cell line was selected as the host cell line for the TI system. It is important that a CHO cell line is used for expression vector screening because the activity of certain expression vector components (such as promoters) can be cell line specific. For example, promoter design strategies have taken advantage of differences in gene expression between cell types, to design promoters which exhibit strong activity in a desired cell type, but show low activity in other cell types (Johnson et al., 2022). Therefore, to ensure the expression vector components that perform well in the screening system also perform well in FDBs manufacturing cell line, it is important that the system is developed in CHO cells. The proline-deficient CHO-K1 cell line was established in 1968 (Kao and Puck, 1968), approximately 10 years after the first CHO cell line was isolated from an ovarian biopsy of an adult Chinese hamster (Puck et al., 1958). The DG44 cell line was generated later by deleting both alleles of the DHFR gene from the first isolated CHO cell lines (not the K1 cell line) using ionising radiation (Urlaub et al., 1983). Although the K1 and DG44 cell lines are derived from a common ancestor, they exhibit significant phenotypic and genetic differences. For example, transcriptomic and proteomic studies revealed 2,875 mRNAs and 304 proteins that were differentially expressed in CHO-K1 compared with CHO-DG44 (Lakshmanan et al., 2019). Additionally, in the development of the Apollo™ X host cell line, a directed evolution approach, during which CHO-DG44 host cell lines were cultured in a 2L continuous chemostat culture for 51 days, was used to improve the properties of the previous host cell line (Pybus et al., 2020). This likely led to further genomic, transcriptomic and proteomic changes (Weinguny et al., 2020). Therefore, to ensure that novel expression vectors are compared in a similar genetic, transcriptomic and proteomic background to the cell line that they will be applied in, it made sense to use FDB's Apollo™ X host cell line. However, at the time of starting the experiment, there was no sequence data available for the Apollo™ X host cell line.

Instead, public genomic data sets for CHO cells had to be used, adding to the difficulty of targeting specific regions of the genome using CRISPR/Cas9.

2.5.2 Identification of a Suitable Locus for LP Integration

The next step in the development of a TI system is to identify stable loci which are resistant to silencing (to ensure stable expression of the transgene) and amenable to RMCE. As previously discussed, an important locus in the development of LP systems in CHO cell lines has been *Fer1l4* (see section 2.2.4). The *Fer1L4* locus was identified by randomly integrating a mAb expression vector into a CHO-K1 cell line, and screening the top 6 best-performing clones based on productivity and growth characteristics (Zhang et al., 2015). The GS CHOK1SV 11A7 clone was shown to contain a single integration site at the Fer1l4 locus and exhibit stable cell growth and productivity over 7 months (220 generations) (Zhang et al., 2015). Additionally, the Fer1l4 locus has subsequently been successfully targeted using CRISPR/Cas9 and shown to be amenable to RMCE (Inniss et al., 2017). As a good starting point for this study, and to provide a point of comparison for other target sites, the decision was made to target the Fer1l4 locus in the Apollo™ X host cell line's genome.

As well as Fer1l4, it was decided to target a LP to 3 additional loci in the Apollo™ X host genome. This decision was made for multiple reasons. Firstly, as previously discussed, the genome sequence of Apollo™ X host cell line was not available at the time. Due to the potential for the DNA sequence of certain genes to differ between CHO cell lines it made sense for multiple loci to be targeted as a precautionary measure. Secondly, as previously discussed, TI systems have been proposed as an alternative to RI methods for CLD. However, cell lines containing 1 LP (with 1 copy of the LC and HC) express lower quantities of antibody than cell line with multiple copies of the HC and LC (Shin and Lee, 2020a). In RI cell lines, titres of 3-8 g/L are consistently achieved (Kelley et al., 2018). To increase antibody expression, whilst keeping the advantages of TI systems (predictable and stable production), gene copy numbers have been increased by targeting

LPs into multiple stable loci, with titres reaching approximately 3 g/L in certain cell lines (Carver et al., 2020). Combining a multi-LP platform with a multi-copy mAb expression vector resulted in a linear increase in mAb titre up to nine mAb copies (Gaidukov et al., 2018). Therefore, it was envisaged that, if all loci could be successfully targeted in this study, the next step in the process could be to sequentially add all LPs to a single cell line, generating a cell line capable of producing similar quantities of antibody to a RI cell line.

For the additional loci, *LemD2*, *Cdk2ap2* and *Mrpl4* were chosen as a continuation from the work of Caroline Wardrope and Dirk-Jan Kleinjan (Rosser Lab, University of Edinburgh), who identified these as high-expression loci (Rosser and Kleinjan, 2022) (Figure 2.1). In their work, a LP containing a GFP expression cassette and selectable marker was designed and randomly integrated into an adherent CHO-K1 cell line using a transposase. Selection was conducted using antibiotics or by using Laser Enabled Analysis and Processing, a scanning cytometry method which uses laser ablation to isolate fluorescent adherent cells in the plates they are growing in by selectively ablating non-fluorescent cells (Szaniszlo et al., 2006). Following selection, 2 clones (CHO-B4 and CHO-G6) displaying high GFP expression were selected and shown to exhibit stable GFP expression during long term culture (70-120 generations). Further analysis of clone CHO-B4 using qPCR and inverse PCR suggested there had been a single LP integration in the genomic region upstream of *LemD2* gene on chromosome 1. Further analysis of clone CHO-G6 using qPCR and inverse PCR suggested there had been LP integration at 2 genomic loci, an intron in the *Cdk2ap2* gene and an intron in the *Mrpl4* gene. As well as providing a suitable locus for an expression vector comparison system, it was thought that any additional work would help with further characterisation of the loci. However, it must be noted that the development of the expression vector comparison tool was the primary aim of the project and would take priority over characterisation studies.

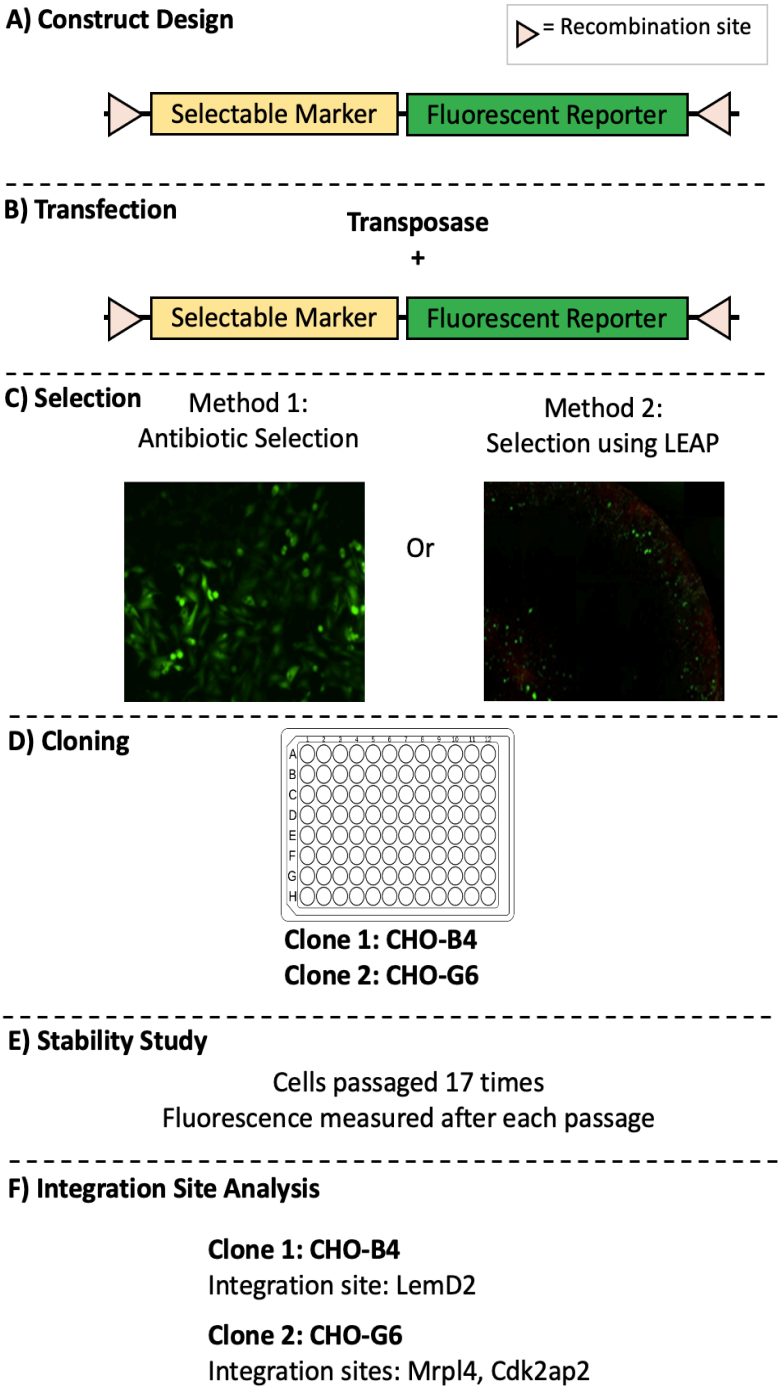


Figure 2.1 - Previous Work Identified 3 Candidate Loci for LP Integration.

(A) A LP construct was designed, containing fluorescent and selection markers flanked by recombination sites. **(B)** Co-transfection of the LP and transposase into adherent CHO-K1 resulted in RI of the LP plasmid. **(C)** The transfected cell population was enriched for fluorescent cells via

(Continued on next page)

(Figure 2.1 continued) two methods. Method 1) antibiotic selection. Method 2) Laser Enabled Analysis and Processing (LEAP) was used to remove the non-fluorescent cell population. **(D)** After the cell population had recovered, pools were cloned using limiting dilution cloning. **(E)** The fluorescence of clones was measured to check for stable production over long-term culture (17 passages). **(F)** Southern blots, qPCR and Inverse PCR was used to assess clones of interest for LP copy number and integration site. 2 cell lines were identified in this screen: CHO-B4 (which had a single integration site at LemD2) and CHO-G6 (which had 2 integration sites at Mrpl4 and Cdk2ap2).
The data from Figure 2.1 was gathered by Caroline Wardrope and Dirk-Jan Kleinjan (Rosser Lab, University of Edinburgh), which was conducted prior to the start of James Donaldson's PhD.

2.5.3 Design and Assembly of the LP

CRISPR/Cas9 was chosen as the method for integrating the LPs at the desired genomic loci. Firstly, sgRNAs were designed to effectively target the loci of interest and the corresponding HAs were identified. To aid such work, the 2018 release of the CHO-K1 genome sequence (RefSeq Assembly GCF_000223135.1) was extracted from the CHO Genome Viewer (https://chogenome.org) and the sequences of target loci were analysed on the DNA sequence editing software, Benchling. Due to a previous study that had successfully targeted *Cdk2ap2* and *Mrpl4* in a different (industrially-relevant) CHO cell line using CRISPR/Cas9, we already had access to the sgRNA sequences for these loci (Rosser and Kleinjan, 2022). By aligning the gRNA sequences to the relevant locus of the CHO-K1 genome, the 5'- and 3'-HA sequences could thus be identified by extracting the ~1kb regions upstream (5'-HA) and downstream (3'-HA) of the PAM sequence (see appendix 3 for sequences). For *LemD2* and *Fer1l4*, Benchling's CRISPR design tool was used to identify the top "on-target" scoring sgRNAs that would effectively target upstream of *LemD2* and the coding region of *Fer1l4*.

Four CRISPR/Cas9 plasmids were produced (one for each locus) by cloning a sgRNA spacer sequence targeting the locus into “plasmid 1 - pX330-U6-Chimeric_BB-CBh-hSpCas9”. This plasmid encodes SpCas9 and has a Bbs1 Type IIs restriction site upstream of a tracrRNA sequence for introducing locus-specific sgRNA spacer sequence that is expressed by the pU6 promoter.

Next, a LP was designed for integration into the target loci (Figure 2.2A). The Bxb1 serine integrase system has previously been shown to be the most efficient system for SSR (Jusiak et al., 2019). Additionally, Bxb1 RMCE systems have been built by pairing the wild type *attP* sequence (with a GT central dinucleotide) with its orthogonal Bxb1 *attB* mutant sequence (with a GA central dinucleotide) (Inniss et al., 2017). Inspired by these results, a LP was designed with an mNG reporter gene and BSD resistant (BSDR) marker, flanked by the Bxb1 *attB* GA mutant sequence and wild-type *attB* GT sequence. The CAG (cytomegalovirus enhancer fused to the chicken beta-actin promoter) promoter driving mNG-p2A-BSDR was positioned outside of the Bxb1 sites, creating a promoter trap for a promoterless reporter construct (such as a mAb expression vector) to be integrated into the LP upon co-transfection with the Bxb1 recombinase. This ensures that expression of the integrated transgene only occurs upon exchange with mNG, and will not be expressed if randomly integrated into the genome (Grav et al., 2018). Flanking the LP are 5’- and 3’ HAs for integration of the construct into the target locus by homology-directed repair (HDR). To screen against RI of the LP construct, a mRuby2 reporter gene (driven by a CMV promoter) was included outside of the HAs (Grav et al., 2018). Following transfection with the sgRNA-Cas9 and LP constructs, cells which have successfully undergone HDR will thus be $mNG^{+}mRuby2^{-}$ (i.e. fluoresce green but not red), while RI will result in cells that still express mRuby2 ($mNG^{+}mRuby2^{-}$). Incorporating sgRNA target sites (with an adjacent PAM) at the outer flank of each HA has been shown to increase CRISPR/Cas9-mediated knock-in efficiency in CHO-K1 cells (Shin and Lee, 2020b). Such alterations should

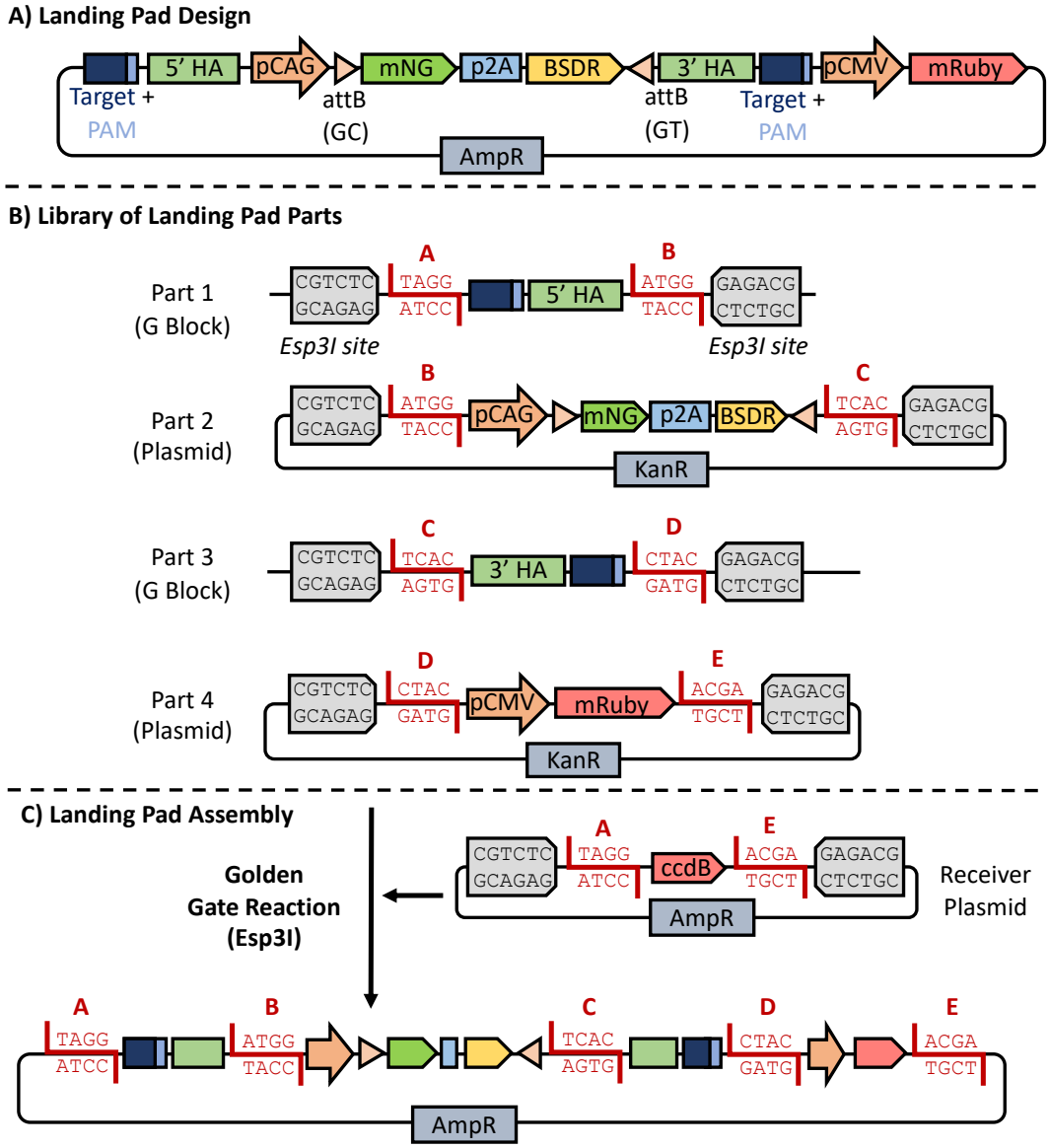


Figure 2.2 - Design and Assembly of the LP for TI.

(A) Design of the LP construct. An mNG reporter gene and BSD resistant (BSDR) marker is flanked by the Bxb1 attB GA mutant sequence and wild-type attB GT sequence for recombinase-mediated cassette exchange. Incorporating sgRNA target sites (with an adjacent PAM) should cause the Cas9-sgRNA to cut and linearise the DNA to prevent integration of the backbone (Target + PAM). Homology arms were designed to target the landing pad to the target locus (HA). The CAG promoter (CAGp) driving mNG-p2A-BSDR was positioned outside of the Bxb1 sites, creating a

(continued on next page)

(Figure 2.2 continued)

promoter trap for a promoterless reporter construct to be integrated into the LP. To screen against RI of the LP construct, a mRuby2 reporter gene (driven by a CMV promoter) was included outside of the Has. **(B)** A library of EMMA parts was ordered as plasmids (from GeneArt) and G-blocks (from IDT). The design methodology enables additional HAs (targeting additional sites) to be designed, ordered as G-blocks and easily swapped into the assembly step. **(C)** The LP was cloned into an EMMA backbone using golden gate assembly.

cause the Cas9-sgRNA to cut both the genome and integrating vector. Successful Cas9-mediated vector cutting will linearise the DNA and prevent integration of the vector backbone sequence (Shin and Lee, 2020b). Assembly of the LP was achieved using the Extensive Mammalian Modular Assembly (EMMA) system (Martella et al., 2017). EMMA uses a Golden-Gate based modular cloning method (with an accompanying toolkit of biological parts) to build mammalian expression vectors (Martella et al., 2017). To assemble the designed LP using the standard EMMA toolkit, 22 parts would need to be assembled into the plasmid backbone, a number which is close to the maximum number of parts that can be assembled using the EMMA method. Therefore, a simplified method was established by collating multiple parts into 2 plasmids (part 2 and part 4) and flanking them with EMMA overhangs (Figure 2.2B). The HAs (flanked by the corresponding EMMA overhangs) could subsequently be designed as G-blocks (part 1 and part 3).

2.5.4 Integration of the LP into the Chosen Loci

To integrate the LP into target loci, the LP plasmid (900ng) was co-transfected with the corresponding sgRNA plasmid (100ng) into Apollo^{TM} X host cells. To enrich for cells that had successfully integrated the LP into the genome, cells were subjected to increasing concentrations of BSD for 13-18 passages (Figure 2.3A). The passage number for each cell line was

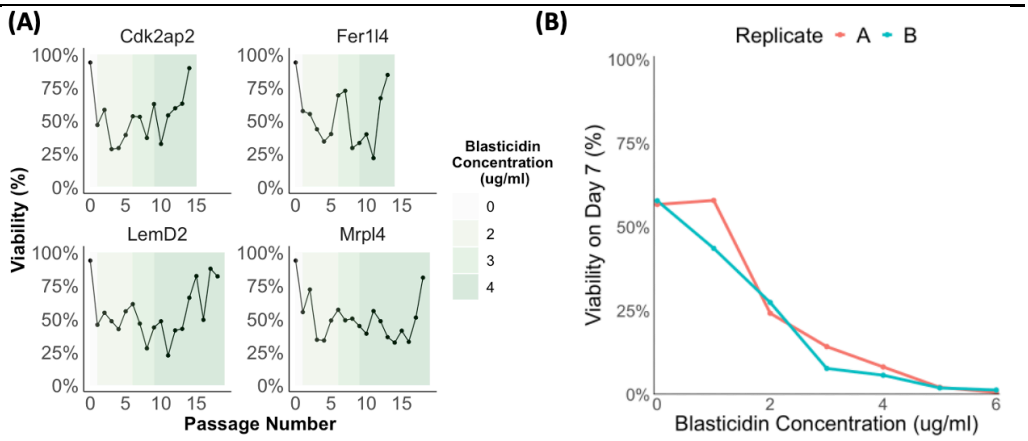


Figure 2.3 - Antibiotic Selection of the Transfected Population.

(A) Cell line viability of the transfected cell populations over a period of 13-18 passages. Apollo^{TM} X host cells were co-transfected with the LP plasmid and the appropriate sgRNA. After transfection, the transfected cells were subjected to increasing concentrations of BSD. **(B)** Viability of Apollo^{TM} X host cells following 7 days of treatment with different concentrations of BSD. Each point represents the mean viability of 2 replicates, performed concurrently.

dependent on the time taken for the transfected population to recover to >80% viability following antibiotic selection. Once the cell population had recovered, a cell bank was made before taking cells forward for FACS. To reduce the risk of selection leading to transgene amplification in the genome (a practice that is common in MTX-mediated selection during CLD), the concentration of BSD was steadily increased from 1 µg/ml to 4 µg/ml (Figure 2.3A). 4 µg/ml had previously been shown to kill ~95% of cells over a period of 7 days (Figure 2.3B).

Following antibiotic selection of the transfected cell population, FACS was used to single-cell sort mNG^{+}mRuby2^{-}

cells (

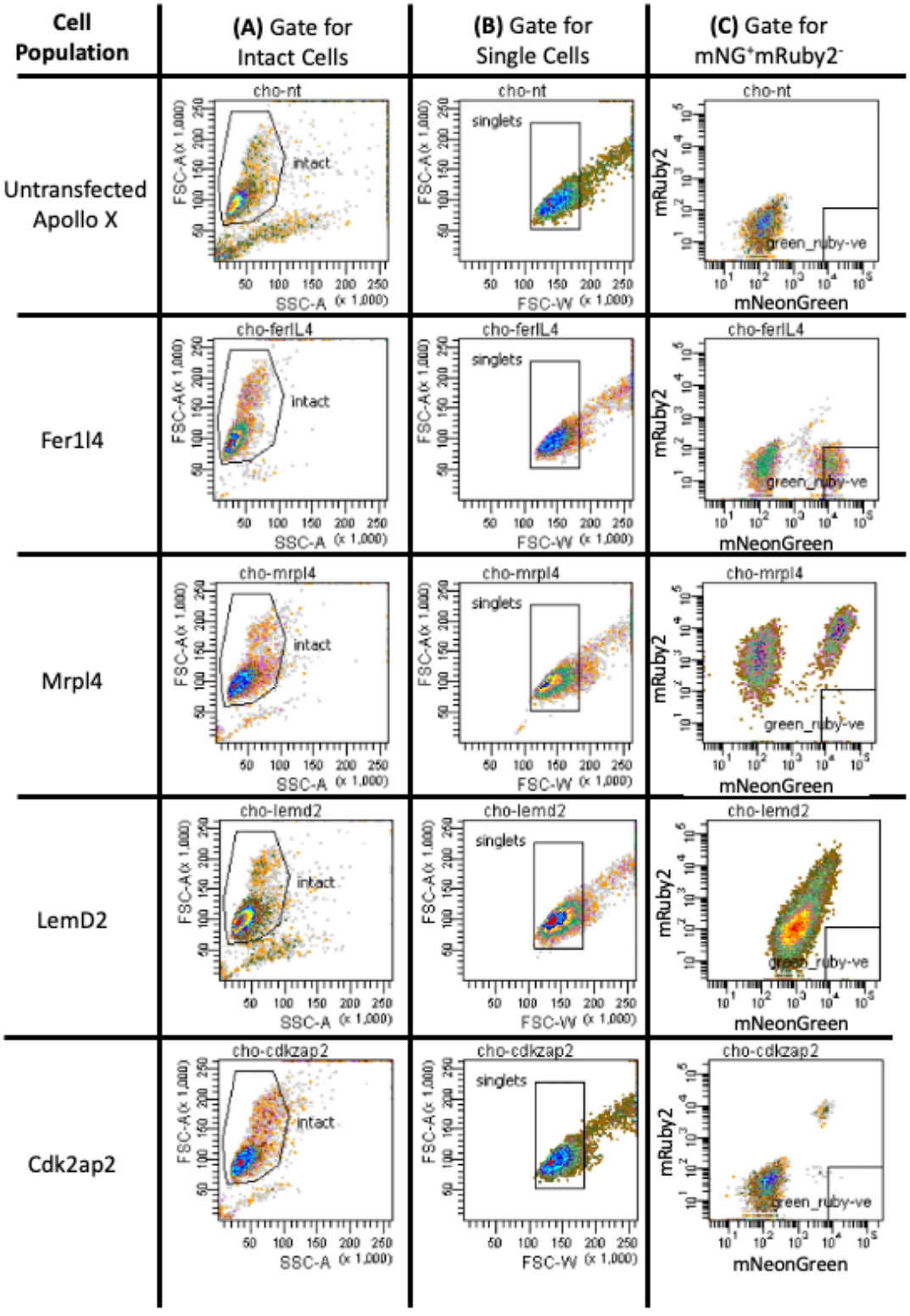


Figure 2.4). For the *Fer1l4* cell line, 3 populations of $mNG^{+}mRuby2^{+}$, $mNG^{+}mRuby2^{-}$ cells and $mNG^{-}mRuby2^{-}$ cells could be identified in the FACS analysis. These populations were presumed to reflect a mixture of events occurring during

the transfection process. The presence of $mNG^{+}mRuby2^{+}$ cells was presumed to be a result of RI of the LP plasmid into the genome. The presence of $mNG^{+}mRuby2^{-}$ cells was presumed to be a result of the HDR and thus the population that was subsequently isolated to generate the LP cell line. The presence of healthy $mNG^{-}mRuby2^{-}$ cells suggested that the concentration of BSD had not been high enough to kill all non-transfected cells. Due to space constraints in the incubator, the kill curve was produced by culturing cells in static culture. Clearly, this is not the optimal conditions for suspension cell growth and thus it is likely that, as a result, the cells were more sensitive to BSD treatment. This is illustrated by cells with 0% BSD having a viability of <60% after 7 days (Figure 2.3B). However, it was estimated that there were enough transfected cells in the enriched population for FACS selection of positive clones and thus the non-transfected population was not necessarily a problem. Additionally, the presence of the non-transfected population (following selection) suggested that the cells had not been under intense selection pressure. This could reduce the risk of LP duplication events. However, for the other cell lines, the

mNG⁺mRuby2⁻ population was less easily identifiable (

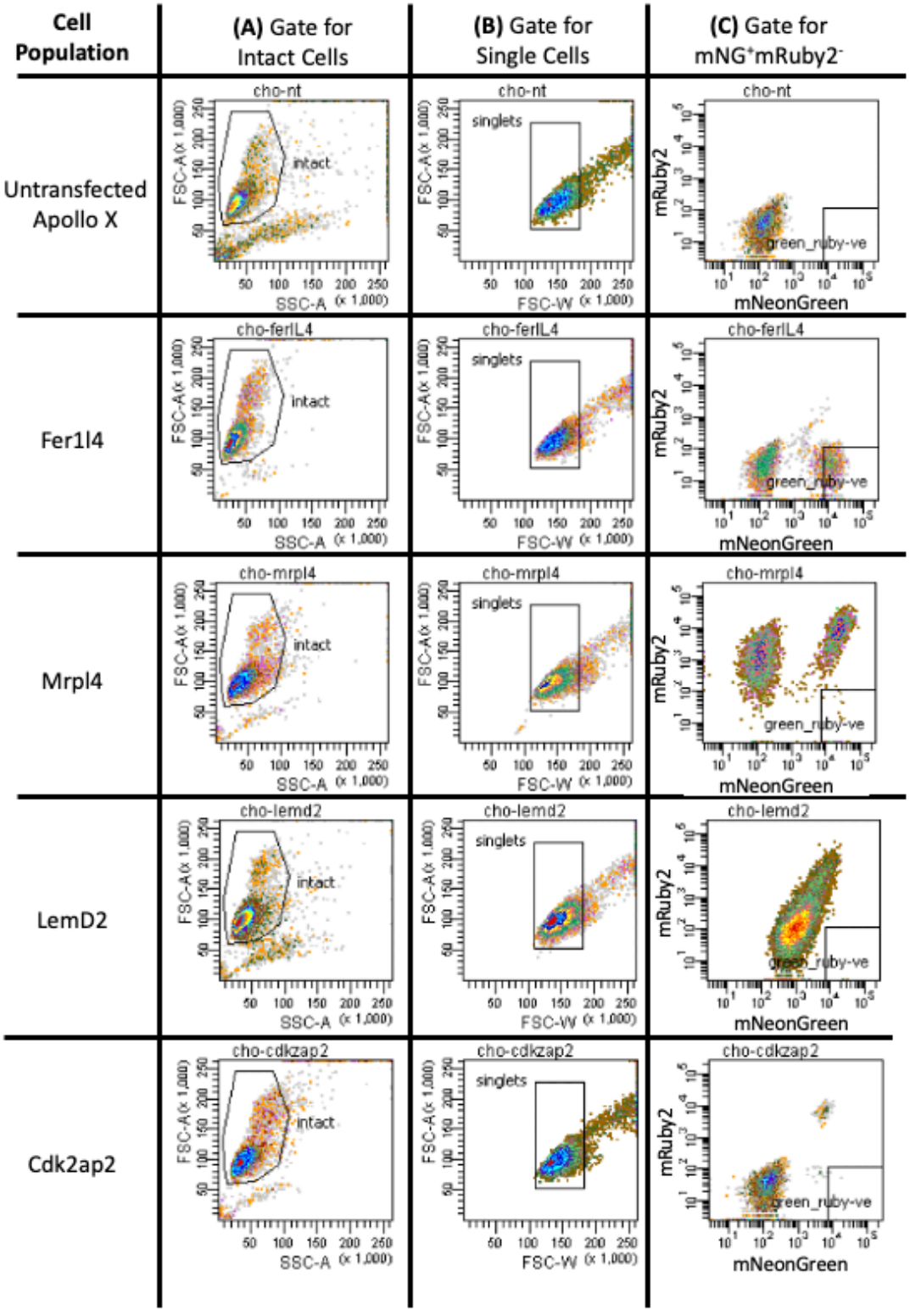


Figure 2.4). Therefore, the gating strategy used to isolate the mNG⁺mRuby2⁻ population in the Fer1l4 cell line was also used for the other LPs. In the Fer1l4 cell line, 25% of the singlet cell population occupied the mNG⁺mRuby2⁻ quadrant. However, in the LemD2, Mrpl4 and Cdk2apcell

lines, only 0.1%, 1.9% and 0.5% of the singlet cell population occupied the $mNG^{+}mRuby2^{-}$ quadrant, respectively.

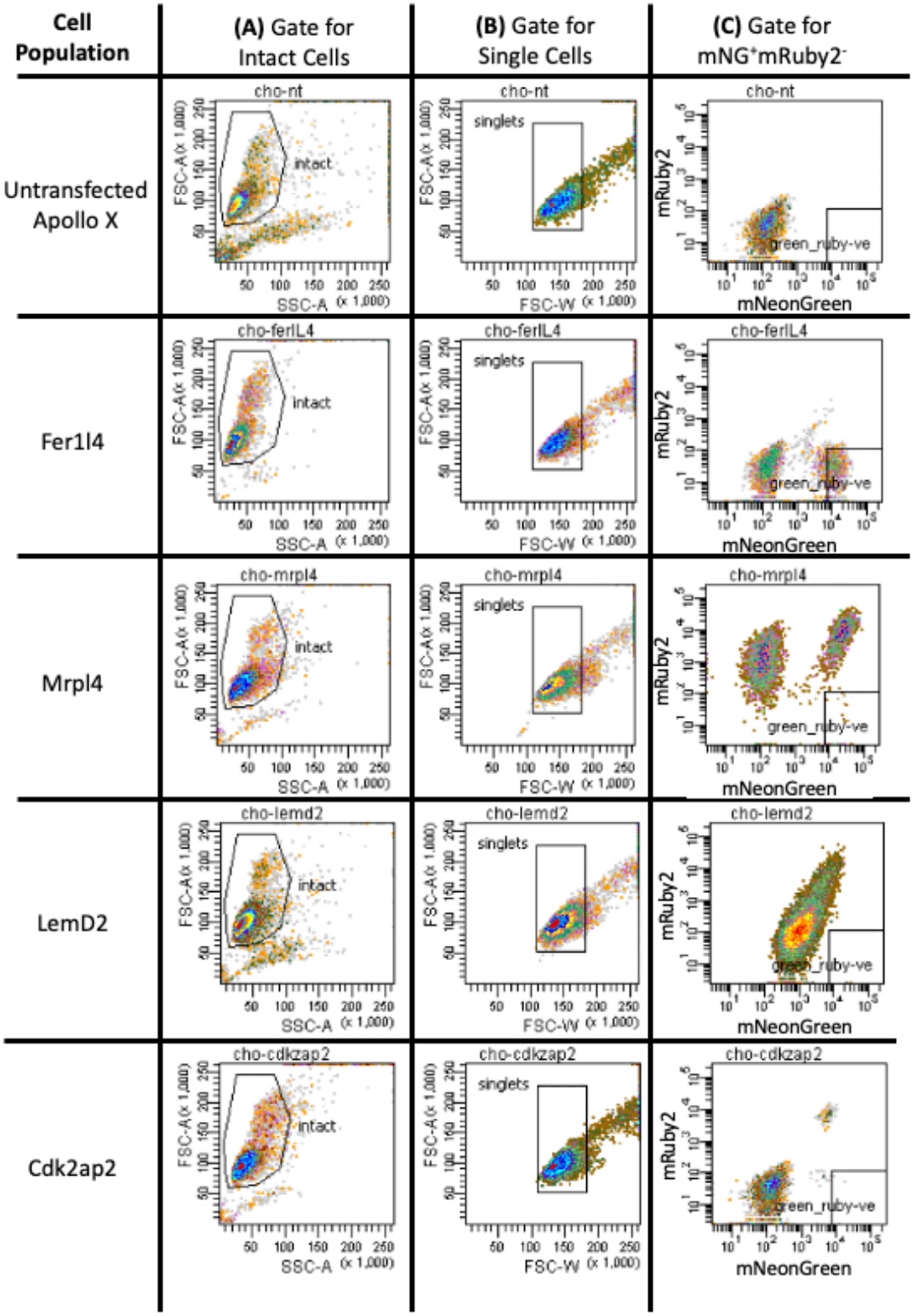


Figure 2.4 - Gating Strategy used for FACS Single Cell Sorting of the Transfected Populations

The transfected cell populations were single-cell sorted using FACS to select cells that have successfully integrated the landing

(Continued on next page)

(Figure 2.4 continued)

pad via CRISPR/Cas9. **(A)** Using the forward scatter area (FSC-A) vs side scatter area (SSC-A), intact cells were selected. **(B)** Using the forward scatter area (FSC-A) vs forward scatter width (FSC-W), single cells were selected. **(C)** mRuby2 expression was plotted against mNG expression. Using the untransfected Apollo X cell line (expressing neither mNG or mRuby2), the $mNG^{-}mRuby2^{-}$ population was identified. Using the Fer1l4 cell line, the $mNG^{+}mRuby2^{-}$ cell population was identified and a gate which selected $mNG^{+}mRuby2^{-}$ cells was generated. The same $mNG^{+}mRuby2^{-}$ gate was used for the Mrpl4, LemD2 and Cdk2ap2 cell lines. Cells in the $mNG^{+}mRuby2^{-}$ gate were single cell sorted into 96 WPs.

For each LP, $mNG^{+}mRuby2^{-}$ cells were single-cell sorted into three 96-well plates, containing different media compositions/ feeding regimes (Table 2.5). For condition 1, 200 µl of cloning medium (see section 2.3.8. for details) plus 1x ACF Supplement was added to each well of a 96 well-plate. ACF Supplement is a serum-free and animal component-free supplement which increases the cloning efficiency of CHO cells. After single-cell sorting with FACS, the cells were left to grow (37°C, 5% CO_2) for 2 weeks, in accordance with FDB's standard procedure for cloning recombinant Apollo™ X-derived cell lines. In an attempt to reduce the cost of future single-cell cloning experiments, fetal bovine serum (FBS) was tested (as an alternative to ACF Supplement) in conditions 2 and 3. For condition 2, 200 µl of cloning medium plus 10% FBS was added to each well of a 96-well plate and cells were left to grow (37°C, 5% CO_2) for 2 weeks. For condition 3, 50 µl of cloning medium plus 10% FBS was added to each well of a 96-well plate. After

single-cell sorting, 50 µl of additional cloning media (with 10% FBS) was added to each well of plates with condition 3 every 2-3 days to prevent nutrient limitation. After 2 weeks, the percentage of wells with a confluency >33% in each plate was noted. Unfortunately, no wells showed signs of cell growth in any of the plates with FBS (plates 2 and 3). Therefore, for all future

Table 2.5 - Outgrowth of Cloned LP Cells following Single-Cell Sorting.

Cell Line	Percentage of Wells >33% Confluent 2 weeks after sorting (%)		
	Condition 1 200 µL medium (1:40 ACF supp.)	**Condition 2** 200 µL medium (with FBS)	**Condition 3** 50 µL medium (with FBS)
Fer1l4	52	0	0
Mrpl4	63	0	0
Cdk2ap2	34	0	0
LemD2	35	0	0

single-cell sorting experiments, 1x ACF Supplement was used in the cloning media.

2.5.5 Expansion and Selection of LP Clones

The next step in the LP CLD process was to select 3 monoclonal cell lines per target site to be taken forward for further analysis and cryopreservation (Figure 2.5). Two weeks after FACS, 24 clones from each target site were progressed from the 96-well plate to a 24-well plate (Figure 2.5A). Next, 0.5 ml of cell culture was removed from each well of the 24-well plate for flow cytometry analysis. To ensure the selected cell lines were growing well, cell lines showing less than 1000 single cell events (such as Cdk2ap2_a7) during flow cytometry were discarded (Figure 2.5B).

Although FACS should have removed all mRuby2-expressing cells (signalling RI of the LP plasmid), an additional screening layer was added to remove cell lines (such as LemD2_c11) with more than 1% of the cell population

showing mRuby2 expression (Figure 2.5C). 16 out of the remaining 22 LemD2 cell lines and 5 out of the remaining 17 Cdk2ap2 cell lines were removed at this point. This was not surprising as these 2 cell lines were the cell lines with the lowest percentage of cells in the mNG^{+}mRuby2^{-}gate (

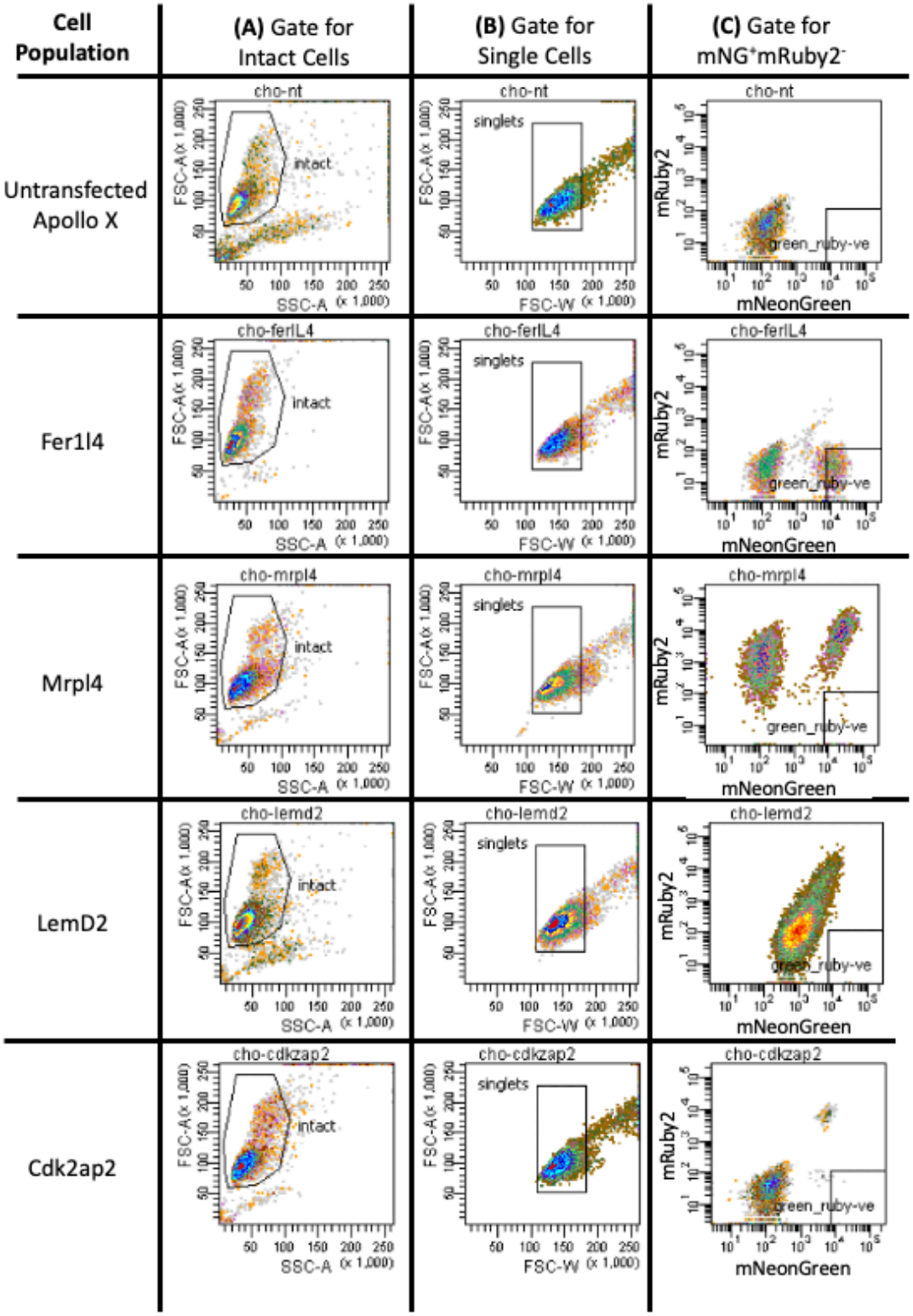


Figure 2.4).

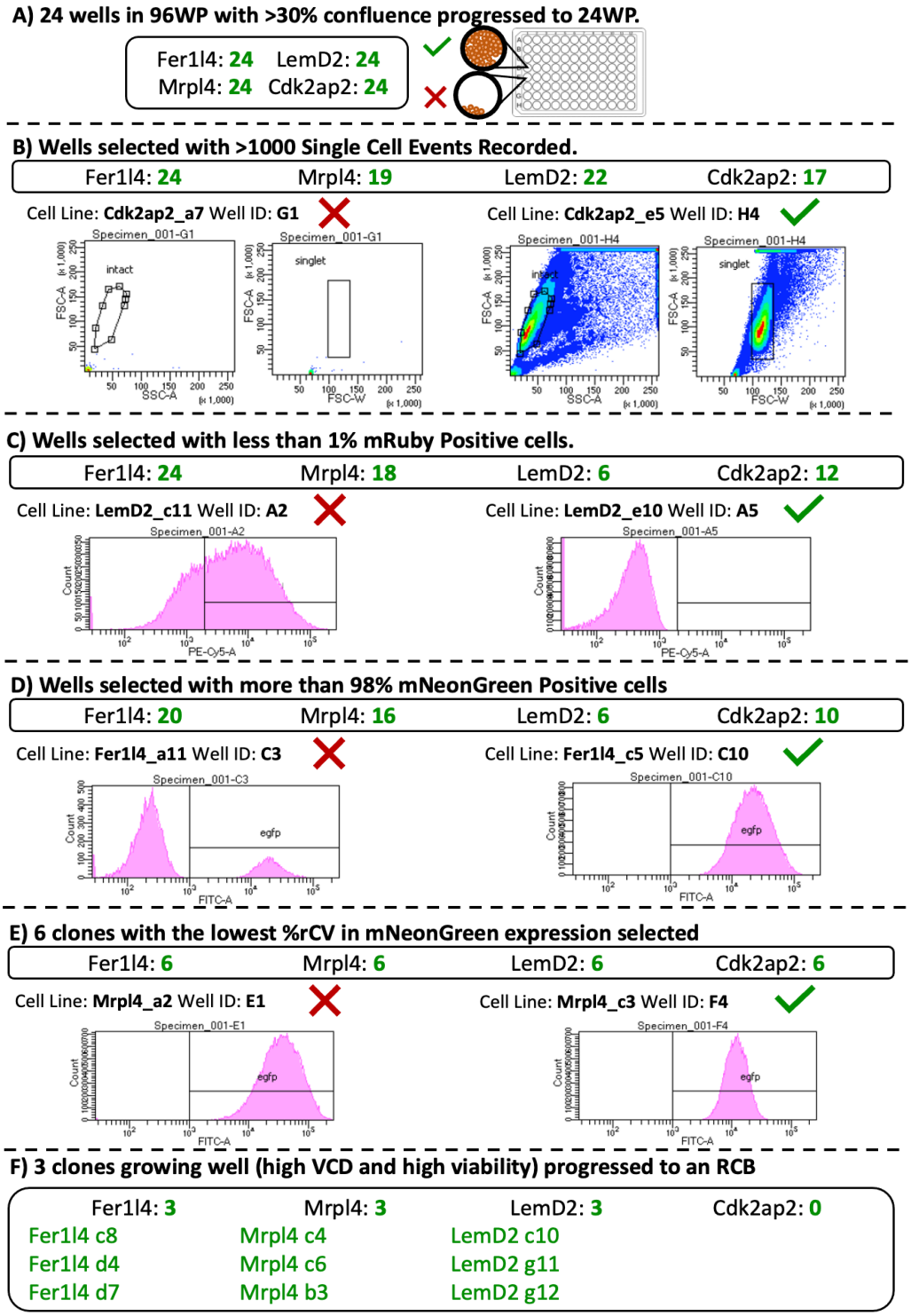


Figure 2.5 - Selection of Clones at the 24-well plate Stage

After the clones had been progressed from the 96-well plate to the 24-well plate stage, they were analysed by flow cytometry and clones of interest were selected for further analysis. **(A)** For each LP target, 24 clones that were growing well (>30% confluent)

(continued on next page)

(Figure 2.5 continued) from 24 wells of a 96WP were transferred to a 24WP. **(B)** The clones were subsequently analysed using flow cytometry to select 3 clones per target to be progressed to the next stage of analysis. To ensure the selected clones were growing well, cell lines showing less than 1000 single cell events during flow cytometry were discarded. **(C)** To prevent selection of clones which had randomly integrated the LP, clones with more than 1% of the cell population showing mRuby2 expression were discarded. mRuby2 expression was measured using the PE-Cy5-A channel. **(D)** To prevent the selection of cell populations containing mNG^- cells (e.g. through gene loss/ epigenetic silencing), clones with more than 2% of the cell population showing loss of mNG expression were discarded. mNG expression was measured using the FITC-A channel. **(E)** In an attempt to select clones that had successfully integrated a single copy of the LP, the clones with the lowest variation in mNG expression were selected. **(F)** Following flow cytometry, clones which maintained successful growth were progressed to RCB.

To prevent the selection of cell populations containing mNG^- cells (e.g. through gene loss/ epigenetic silencing), cell lines (such as Fer1l4_a11) with more than 2% of the cell population showing loss of mNG expression were discarded (Figure 2.5D). Four out of the remaining 24 Fer1l4 cell lines were removed at this point. This could indicate that the Fer1l4 locus may be more susceptible to silencing than the other target sites. Six cells lines with the lowest robust Coefficient of Variation (rCV) were then selected per target site to be progress from the 24-well plate to a 6-well plate (Figure 2.5E).

It is known that gene expression is a stochastic process (Nijhout, 2013; Raj and Oudenaarden, 2008). Therefore, due to inherent randomness in transcription and translation, even if all cells contain a single copy of the LP at the same location, there will be cell-to-cell variations in mRNA, protein

levels and thus mNG fluorescence. The source of such noise in gene expression is thought to be both extrinsic (affecting the expression of all transgene copies in a cell equally due to cell-to-cell variations in the numbers of RNA polymerases or ribosomes) and intrinsic (affecting each copy of the transgene independently due to the inherent randomness of transcription and translation) (Raj and Oudenaarden, 2008). With this in mind, it was hypothesised that cell lines with a lower variation in mNG expression were more likely to contain a single copy of the LP than cell lines with a wider variation in mNG expression. 3 cell lines per target site were then selected (which were growing well in the 6-well plate) to be scaled up again for research cell banks to be generated (Figure 2.5F).

Unfortunately, due to space constraints in the incubator, all other cell lines had to be discarded at this point. Additionally, the Cdk2ap2 cell lines were discarded because all 6 cell lines were not growing well (low viability and low growth rate). The poor growth of the Cdk2ap2 cell lines was consistent with other data in the CLD process. For example, Cdk2ap2 showed the lowest percentage of wells with >33% confluency at the 96-well plate stage (Table 2.5) and the lowest number of wells with >1000 single cell events at the 24-well plate stage (Figure 2.5).

2.5.6 Flow Cytometry Analysis of the Selected Clones

The level of mNG protein expression for the selected clones was determined by flow cytometry (Figure 2.6A). Compared with the other target loci, the LemD2 cell lines showed the highest mean mNG expression with the lowest variation in mNG expression (Figure 2.6B and Figure 2.6C). The Mrpl4 locus showed the lowest mean mNG expression level. The Fer1l4 cell lines showed the greatest variation in mNG expression (Figure 2.6B and Figure 2.6C).

2.5.7 PCR Analysis Reveals Integration of the Plasmid Backbone in the LemD2 Cell Line

To determine whether the LPs had integrated successfully via HDR, the first step was to assess the cell lines for plasmid backbone integration. Genomic

integration of plasmids via NHEJ can cause insertion of backbone fragments. Integration of the bacterial plasmid backbone (alongside the mammalian transgenes) into the genome is associated with unstable transgene expression (Chen et al., 2008, 2004; Wang et al., 2018). It has been proposed that transgene silencing is caused by formation of repressive heterochromatin on the plasmid DNA backbone, which then spreads and inactivates the transgene (Chen et al., 2008). Genomic DNA was extracted from each of the 9 clonal cell lines that had progressed through the previous screening steps. A diagnostic PCR was designed that would yield a product if the LP plasmid backbone was present in the clonal cell line (Figure 2.7A). When the amplification products from the genomic DNA were visualised on an agarose gel, a strong band for each of the LemD2 clones indicated that some backbone integration had occurred (Figure 2.7B). For this reason, the LemD2 clones were not progressed for further analysis. In contrast, the other cell lines were not deemed to have integrated the plasmid backbone, and were thus taken forward for further analysis.

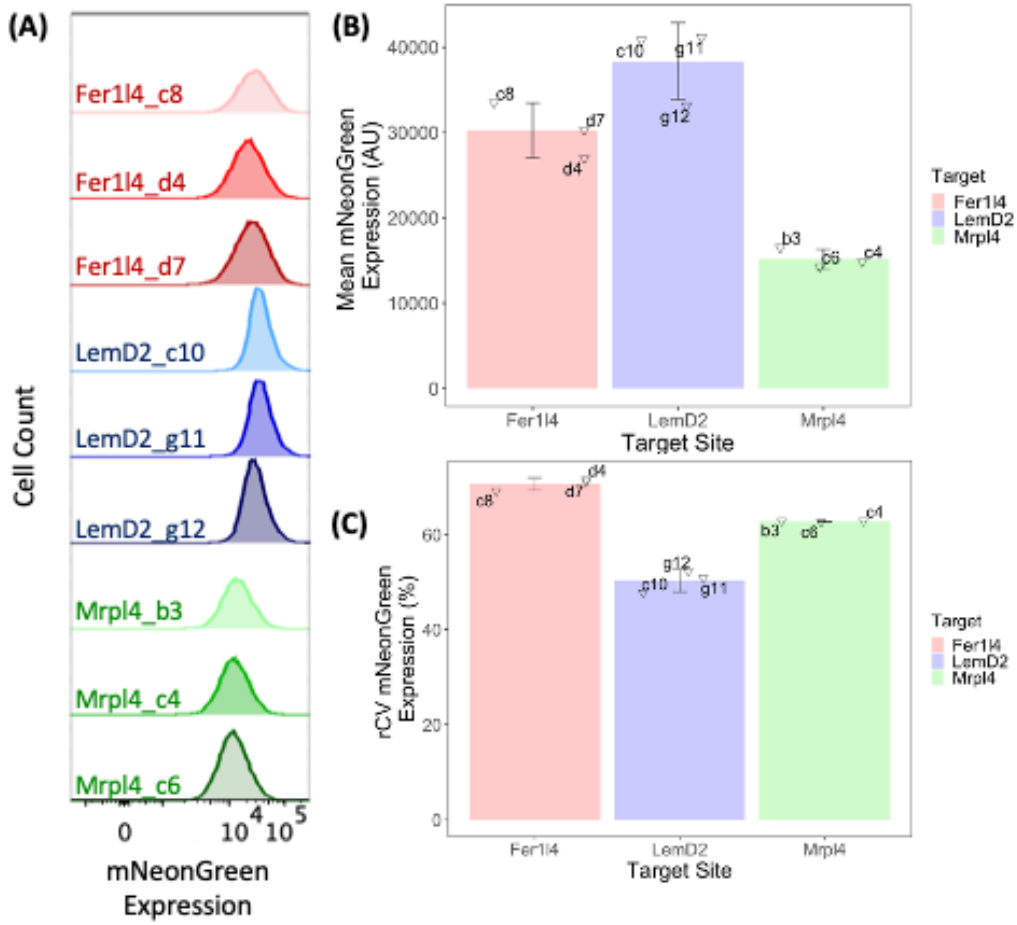


Figure 2.6 - Analysis of LP Clones using Flow Cytometry. **(A)** Histogram showing the mNG expression levels from the 3 selected clones from each locus. **(B)** Mean mNG fluorescence was determined by flow cytometry for the 3 selected clones. The mean of the 3 mNG fluorescence means was taken and plotted as a bar graph. Error bars show the standard deviation of the 3 means. Triangles show the individual values. **(C)** The robust coefficient of variance for mNG fluorescence was determined by flow cytometry for the 3 selected clones. The mean of the 3 mNG fluorescence rCVs was taken and plotted as a bar graph. Error bars show the standard deviation of the 3 rCVs. Triangles show the individual rCVs.

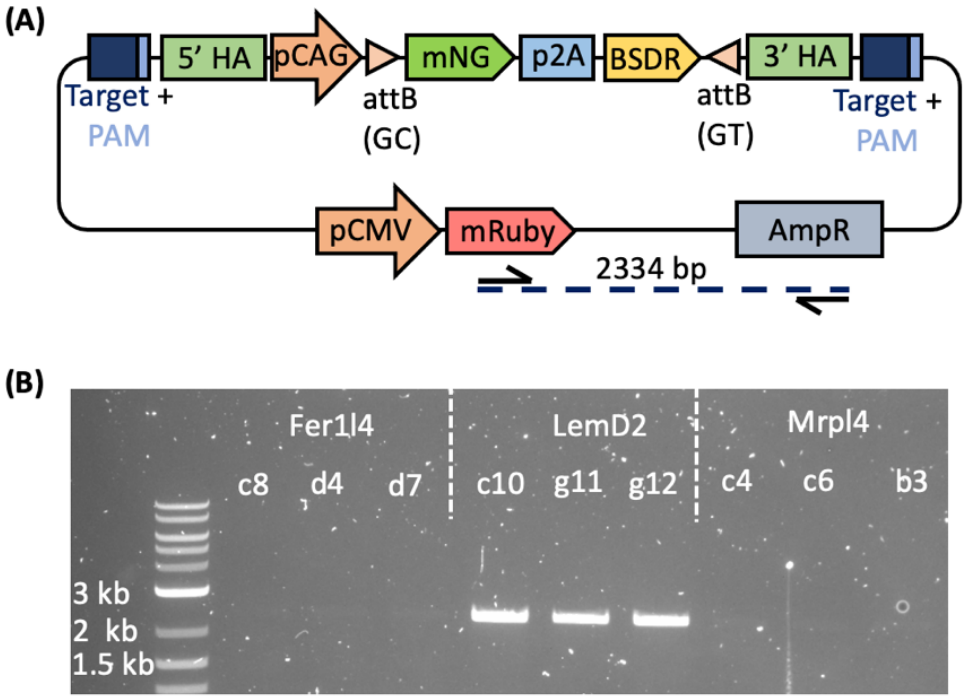


Figure 2.7 - Diagnostic PCRs reveal integration of the LP backbone in the LemD2 cell lines.

(A) Design of the diagnostic PCR reaction. A forward primer (283_mruby_fwd) was designed to target the mRuby2 transgene and a reverse primer (288_amp_rev) was designed to target the ampicillin resistance gene. The expected PCR product was 2334 bp long. **(B)** The PCR products were visualised on an agarose gel.

2.5.8 Further Analysis of the Fer1l4 c8 Cell Line Indicates Successful LP Integration at the Fer1l4 Locus

The Fer1l4_c8 cell line was the first cell line to be chosen for further analysis. To understand whether the LP had integrated into the target locus, two sets of diagnostic PCR reactions were designed that would yield products if the anticipated genome insert junctions from an HDR knock-in were present. Each PCR reaction involved a "LP"-targeted primer, which annealed within the integrated LP cassette, and a corresponding "Fer1l4-specific" primer, which annealed to the genome either upstream or downstream of the 5'- or 3' HA respectively (Figure 2.8A).

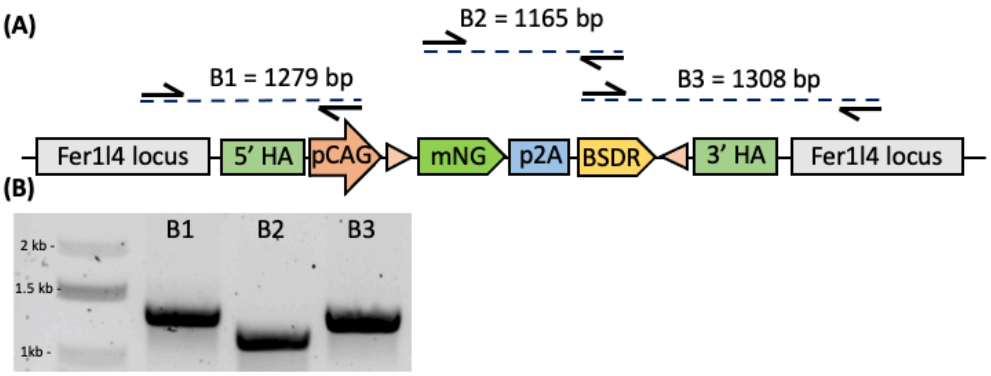


Figure 2.8 - Diagnostic PCRs for LP Integration at the Fer1l4 Locus.

(A) Design of the diagnostic PCR reaction. For band 1 (B1), a forward primer (236_fer1l4_fwd_3) was designed to target the Fer1l4 locus (upstream of the LP) and reverse primer (305_fer1l4_5prrev_1) was designed to target the CAG promoter (pCAG) of the LP. For band 2 (B2), a forward primer was designed to target the mNG transgene (294_lppcr_1) and a reverse primer was designed to target the BSDR transgene (295_lppcr_2). For band 3 (B3), a forward primer was designed to anneal to the mRuby2 gene (273_lp_pcr_fwd) and a reverse primer was designed to anneal to the ampicillin resistance gene (237_fer1l4_rev_4). **(B)** Agarose gel of the PCR products from the above reactions. **(C)** PCR reactions were designed to amplify the attB sites of the LP.

If integration into the Fer1l4 locus had been successful, the anticipated amplicon sizes would be 1279bp for the 5' junction and 1308bp for the 3' junction. An additional PCR reaction was included to amplify a region inside of the LP as a positive control (1165bp amplicon size). When the amplification products from Fer1l4_c8 gDNA were visualised on an agarose gel, they indicated that on-target integration had occurred (Figure 2.8B). Subsequent DNA extraction and Sanger sequencing of the 3'-band indicated that the band that was visualised corresponded to the desired amplicon (see appendix 3). An additional PCR reaction was designed to amplify regions of the LP which contain the attB sites, to check that the recombination sites had

remained intact during the LP integration process. Subsequent DNA extraction and Sanger sequencing of the PCR products indicated that the attB sites were intact (see appendix 3). Therefore, integration into the LP should be possible.

2.5.9 A Stability Study for the Fer1l4 c8 Cell Line

As previously stated, the LP should be integrated at a site that is resistant to epigenetic silencing and gene loss, enabling stable expression of the inserted transgene over time. In industry, a recombinant cell line is generally deemed unstable if it fails to retain >70% volumetric productivity titre over a period of 60 generations (Dahodwala and Lee, 2019). Therefore, a stability study for mNG expression was conducted over the same 60 generation period. A vial of Fer1l4_c8 cells was thawed and split into 6 cultures (3 cultures with BSD in the media and 3 cultures without) and grown for 60 generations. Unfortunately, 1 culture from each condition (1 culture +BSD and 1 culture -BSD) was lost during this period due to a large drop in viability (Figure 2.9A). This could have been due to a number of reasons. Due to space restraints in the incubator, the cells were grown in 2 ml cultures in 6-well plates, increasing the risk of pipetting error when sub-culturing cells. Additionally, as previously mentioned, cells were shaken at 120 rpm, which is not an ideal culture condition for CHO cell lines. Even with the two remaining cell lines, viability was inconsistent over 60 generations (Figure 2.9A). However, the growth rate of the remaining cell line was relatively consistent over 60 generations, both with and without BSD (Figure 2.9C). Surprisingly, the cells cultured with BSD reached generation 60 quicker than cells without BSD (Figure 2.9C). At FDB, 2 replicates for each selection pressure condition is deemed enough to assess stability. Therefore, the decision was made to continue the study with the 4 remaining cultures (2 with BSD and 2 without BSD).

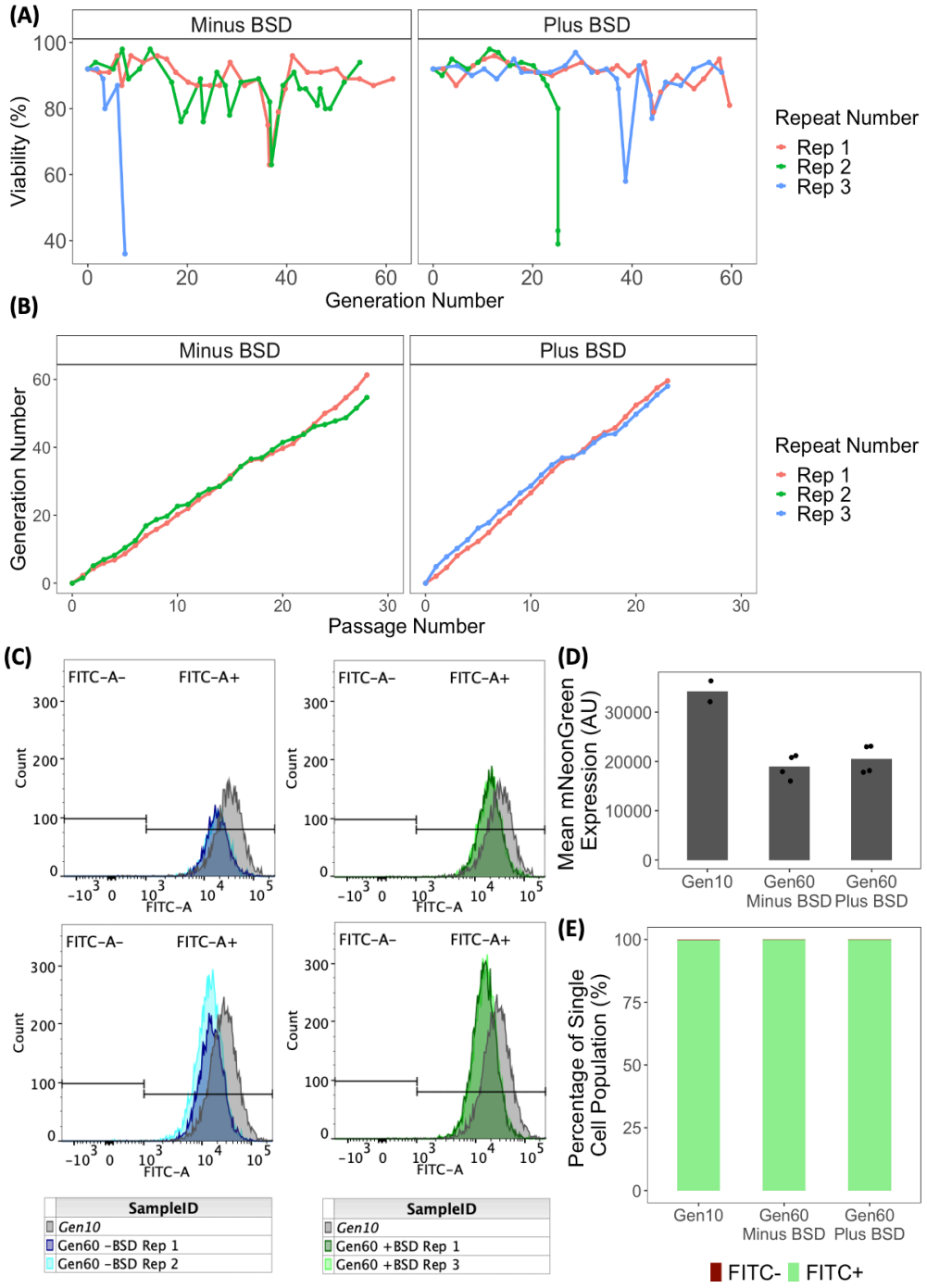


Figure 2.9 - The Fer1l4_c8 Cell Line Shows Unstable Production of mNG over 60 Generations.

(A) Viability of the Fer1l4 cell line. A vial of Fer1l4_c8 cells was thawed and split into 6x 2 ml cultures (3 with BSD and 3 without BSD in the media). Every 2-3 days, the cell density was adjusted to $2x10^5$ cells/ml. **(B)** The Fer1l4_c8 cells were grown for 60

(Continued on next page)

(Figure 2.9 continued) generations. **(C)** When the cells reached generation 60, a fresh vial of Generation 0 Fer1l4_c8 cells were thawed and passaged 3 times. The generation 0 and generation 60 cell lines were compared using flow cytometry. **(D)** Bar chart to show mean mNG expression at different generation numbers. Bars represent the mean of 2 biological replicates with fluorescence measured on consecutive days. Points represent individual mean mNG expression values. (E) Bar chart to show the percentage of mNG^+ cells vs mNG^- cells at generation 10 vs generation 60. Bars represent the mean of 2 biological replicates with fluorescence measured on consecutive days.

When the cells reached generation 60, a vial of Generation 0 Fer1l4_c8 cells was thawed, passaged 3 times (approximately 10 generations), and mNG expression was quantified using flow cytometry (Figure 2.9C). Over 60 generations, there was a significant drop in mNG expression (Figure 2.9D). This result contrasted with other studies that had showed Fer1l4 to be a stable locus for recombinant protein production. However, the percentage of mNG^- cells remained negligible (Figure 2.9E).

2.5.10 The Mrpl4_c4 cell line Showed Poor Growth Over 60 Generations and PCR Analysis is Inconclusive for LP Integration

Whilst the Fer1l4_c8 cell line was being characterised, a similar study was also performed on the Mrpl4_c4 cell line. Firstly, a diagnostic PCR reaction was conducted on the genomic DNA of the Mrpl4_c4 cell line to determine whether the LP had successfully integrated at the Mrpl4 locus, with primers designed to target upstream, downstream and within the LP (Figure 2.10A). When the amplification products from the genomic DNA were visualised on an agarose gel, bands were present, but B2 (the internal LP band) was longer than expected. Additionally, the 5'- (B1) and 3'- (B3) bands were shorter than expected (Figure 2.10B). Subsequent DNA extraction and Sanger sequencing of the PCR products of the 5'- and 3' bands was unsuccessful (data not shown).

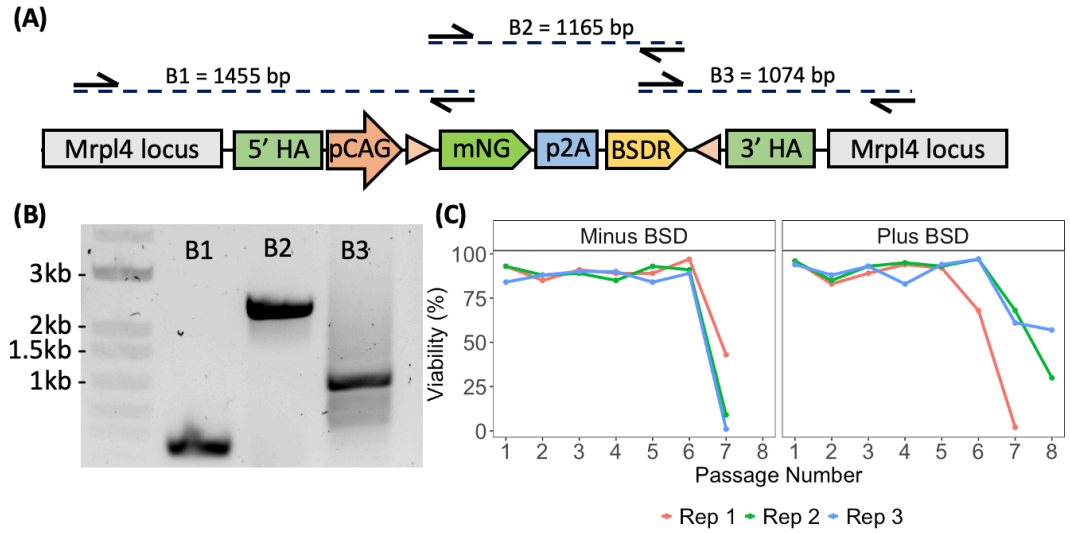


Figure 2.10 - Analysis of the Mrpl4 Cell Line.
(A) A diagnostic PCR reaction was designed to determine whether the LP had successfully integrated at the Mrpl4 locus. **(B)** The PCR products were run on an agarose gel. **(C)** The Mrpl4_c4 cell line was grown in culture over 7-8 passages.

Additionally, the Mrpl4_c4 cell line did not grow well in culture and viability dropped off significantly in all cultures after 6-8 passages (Figure 2.10C). Attempts were made to repeat the diagnostic PCR by designing multiple primers upstream and downstream of the LP (data not shown). However, when the PCR products were run on an agarose gel, no additional bands were visualised that appeared to be of correct size.

2.5.11 Further Analysis of the Fer1l4_c8 clone

At this point, it was decided to prioritise the analysis of the Fer1l4_c8 cell line, rather than performing any further troubleshooting analysis on the Mrpl4_c4 cell line or looking at additional banked cell lines. This decision was taken for multiple reasons. As previously discussed, the initial aim of the work was to produce multiple cell lines (exhibiting stable growth and mNG **production) that could be used for expression vector component comparison, with LPs at different locations. Therefore, novel expression vector components could be tested at different loci, reducing the impact of potential locus-**

specific effects on gene expression. However, it was known from the start that successfully targeting multiple loci would be a challenge, due to the lack of available sequence data for the Apollo^{TM} X genome, challenging both HA and sgRNA design, as well as primer design for diagnostic PCRs. For example, it was uncertain whether the unclear results from the diagnostic PCRs from the Mrpl4_c4 analysis was due to improper primer design or unsuccessful LP integration at the Mrpl4 locus. In the interest of time and given that CRISPR/Cas9 mediated integration at the Fer1l4 locus gave the highest proportion of mNG^{+}/mRuby2^{-} cells upon FACS

cell sorting (

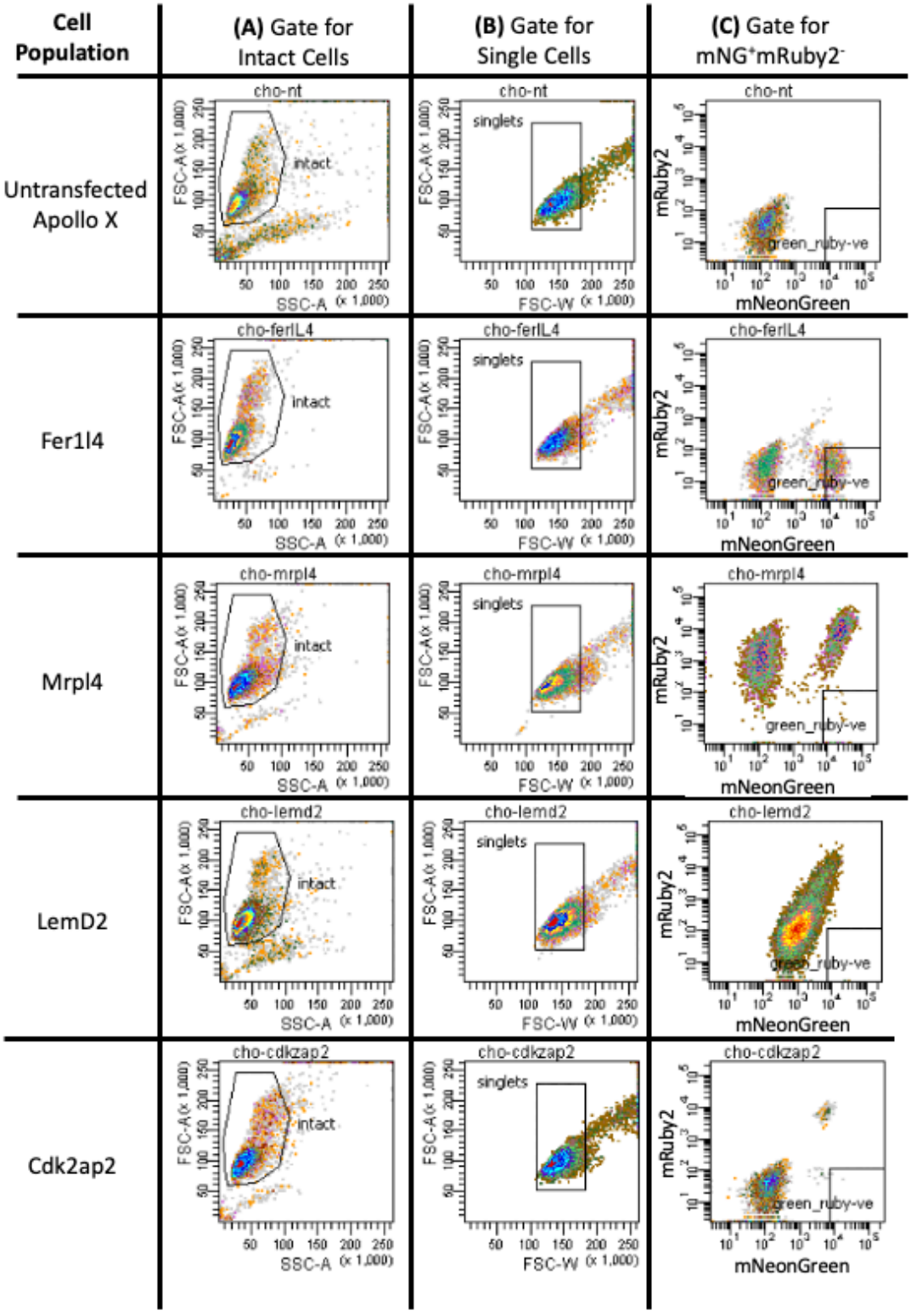


Figure 2.4) and that the diagnostic PCRs for the Fer1l4_c8 cell line showed successful integration at the Fer1l4 locus, it was reasoned that the banked cell lines with a LP in the Fer1l4 locus (Fer1l4_c8, Fer1l4_d4, Fer1l4_d7)

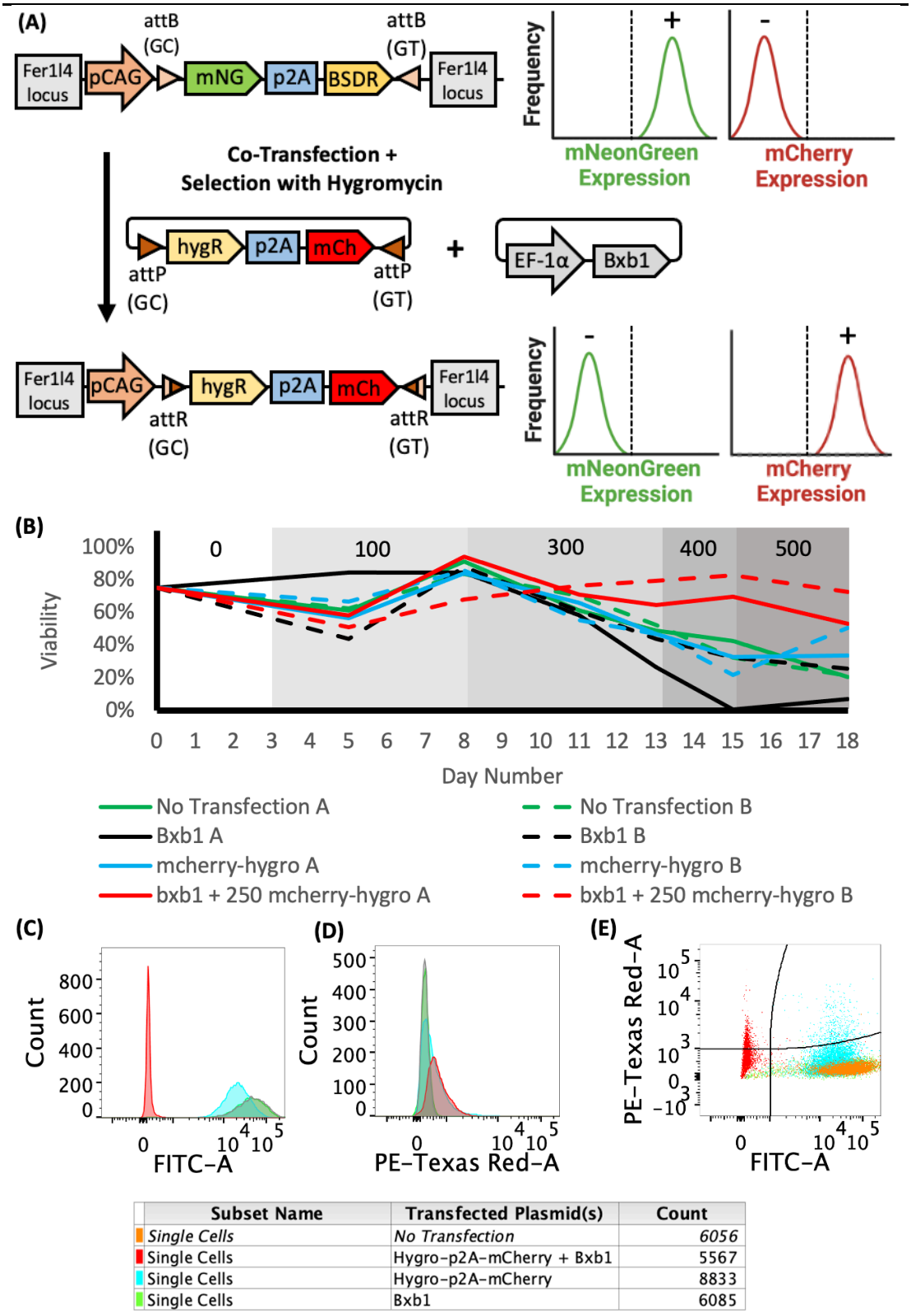
were the cell lines to pursue. Therefore, the Mrpl4 cell lines were removed from any further analysis.

The stability study of the Fer1l4_c8 cell line showed unstable mNG expression (Figure 2.9D). To characterise the other Fer1l4 LP cell lines (Fer1l4_d4 and Fer1l4_d7), another stability study could have been conducted. However, it takes approximately 10 weeks to grow cells from generation 0 to generation 60, and poor growth in the shaking incubator had been a common theme in the previous experiments. Given that Fer1l4_d4 and Fer1l4_d7 showed very similar levels of mNG expression to Fer1l4_c8, it is likely that these cell lines could also exhibit unstable mNG expression. At this point, it was unknown whether the LP system was functional in the Apollo^{TM} X host cell line. Therefore, it was reasoned that a better usage of time would be to analyse the Fer1l4_c8 clone rather than analyse other Fer1l4 clones.

2.5.12 Bxb1 Mediated Cassette Exchange in the Fer1l4_c8 Clone

As a first priority, the functionality of the LP at the Fer1l4 locus needed to be determined. If constructs could not be integrated into the LP, efforts to characterise the cell line (or any additional cell lines) would be futile. Firstly, an attPGC_hygR-p2A-mCherry_attPGT cargo vector was constructed, containing a hygromycin resistance gene (hygR) and an mCherry fluorescent reporter (construct referred to as hygR-p2A-mCherry) (Figure 2.11A). Since the LP design relies on a promoter trap, displacement of the mNG expression cassette from the CAG promoter by a cargo vector should generate exclusively mNG^{-} cells. When a promoterless selection and fluorescent marker (i.e hygR-p2A-mCherry) is present in the cargo, only mNG^{-}mCherry^{+} cells should be generated. Due to the hygromycin-p2a-mCherry cassette being promoterless, RI of the plasmid in the cell line should not yield hygromycin resistance or mCherry expression (Figure 2.11A). Therefore, only integration of the cargo vector into the LP should yield mCherry expression.

To integrate the hygro-p2A-mCherry construct into the LP, Fer1l4_c8 cells were co-transfected with the hygro-p2A-mCherry construct and a pEF1α-Bxb1 plasmid. Following transfection, cells were given 3 days to recover before antibiotic selection. Over 18 days, the concentration of hygromycin was gradually increased from 100 µg/ml to 500 µg/ml (Figure 2.11B). Following enrichment with hygromycin, the transfected cells were analysed using flow cytometry. When Fer1l4_c8 cells were transfected with both the hygro-p2A-mCherry construct and a pEF1α-Bxb1 plasmid, a large drop in mNG expression was seen (Figure 2.11C). However, this drop in mNG expression did not occur in the untransfected population or when Fer1l4_c8 cells were transfected with either pEF1α-Bxb1 or hygro-p2A-mCherry. This suggested that it was Bxb1-mediated integration of the hygro-p2A-mCherry construct into the LP (displacing the mNG expression cassette from the CAG promoter in the LP) that was causing the loss of mNG expression. However, integration of the hygro-p2A-mCherry construct into the LP only led to a small increase in mCherry expression (Figure 2.11D). Although, in theory, RI of the promoterless hygro-p2A-mCherry construct should not lead to mCherry expression, there appears to be a population of $mNG^{+}mCherry^{+}$ cells, when the Fer1l4_c8 cell line was transfected with just the hygro-p2A-mCherry construct (Figure 2.11E). This was thought to be due to geomic integration of the transgene near endogenous promoters/ enhancers.



Subset Name	Transfected Plasmid(s)	Count
Single Cells	*No Transfection*	*6056*
Single Cells	Hygro-p2A-mCherry + Bxb1	5567
Single Cells	Hygro-p2A-mCherry	8833
Single Cells	Bxb1	6085

Figure 2.11 - Successful Integration of a Promoterless Hygro-P2A-mCherry onstruct.

(A) Schematic diagram of integrating a donor vector into the LP.

(Continued on next page)

(Figure 2.11 continued) Prior to transfection, the LP cell line will express mNG. However, upon integration of the mCherry expression cassette via RMCE, mNG expression will be lost and the cell line will express mCherry instead. **(B)** Viability of the cell line over time as the transfected cells were exposed to increasing concentrations of hygromycin. Only upon transfection with both hygro-p2A-mCherry and Bxb1 should the cell line be able to conduct RMCE. Therefore, in all other conditions, the transfected cells should die when hygromycin is added to the culture media. **(C)** Change in mNG expression (measured using the FITC-A channel) following transfection with different constructs. **(D)** Change in mCherry expression (measured using the PE-Texas Red-A channel) following transfection with different constructs. **(E)** Scatter plot showing mCherry expression (measured using the PE-Texas Red-A channel) vs mNG expression (measured using the FITC-A channel).

2.5.13 Diagnostic PCRs Confirm That Integration into the LP has been Successful

To confirm whether the loss of mNG expression (and small increase in mCherry expression) upon co-transfection of the Fer1l4_c8 cell line with pEF1α-Bxb1 and hygro-p2A-mCherry was a result of integration into the LP, a series of diagnostic PCRs were performed. Genomic DNA was extracted from both the Fer1l4_c8 host cell line and the Fer1l4_c8 cell line following co-transfection with hygro-p2a-mCherry and pEF1α-Bxb1. The following diagnostic PCRs were designed (Figure 2.12A). Firstly, primers were designed to amplify the LP. For band 1, primers targeted the pCAG promoter (forward) and the mNG transgene (reverse). For band 2, primers targeted the BSDR transgene (forward) and the Fer1l4 genomic DNA downstream of the LP (reverse). Secondly, primers were designed to amplify the LP after successful integration of the hygro-p2a-mCherry had occurred. For band 3, primers targeted the pCAG promoter (forward) and the mCherry transgene (reverse). For band 4, primers targeted the mCherry transgene (forward) and the Fer1l4 genomic DNA downstream of the LP (reverse).

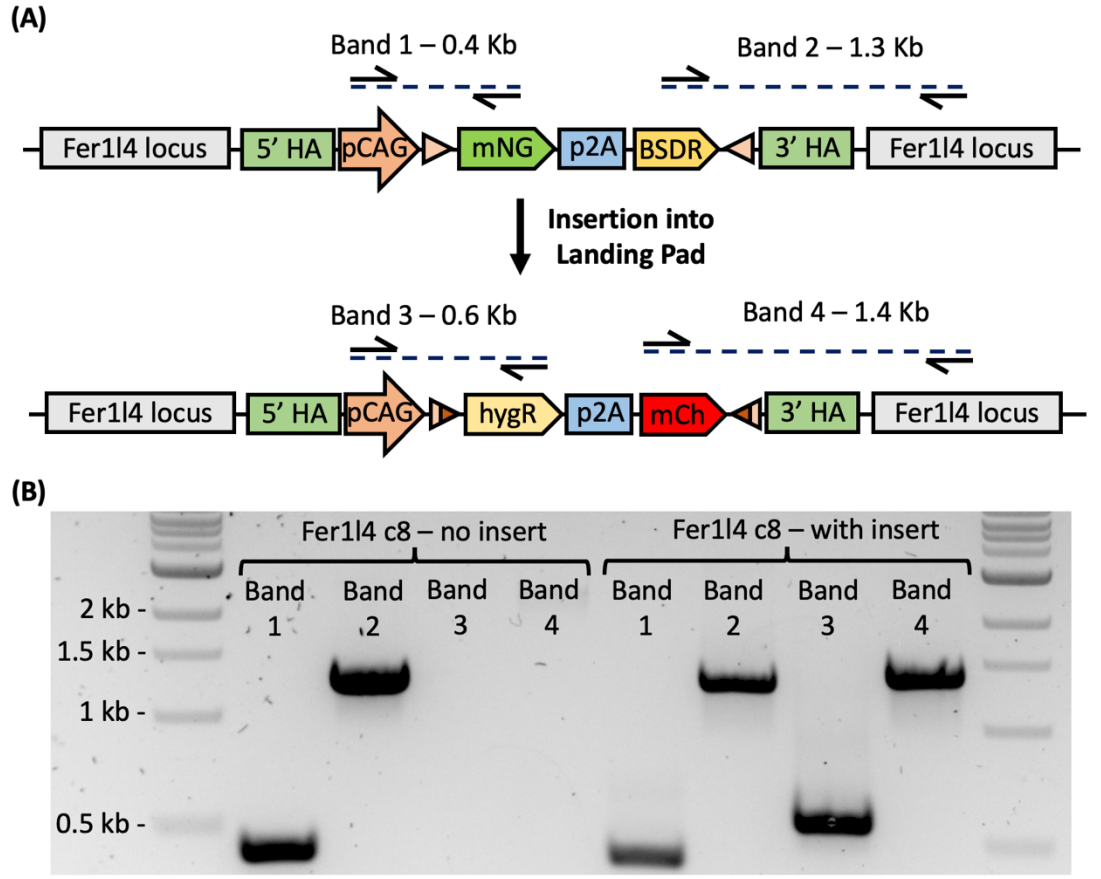


Figure 2.12 - Diagnostic PCRs Show Successful Integration of the hygR-p2A-mCherry Construct into the LP

(A) Design of diagnostic PCRs for integration into the LP. For B1 and B3, the same forward primer (392_lppcr_fwd), targeting the CAG promoter was used. For B1, a reverse primer (393_lppcr_rev) was designed to target the mNG transgene. For B3, a reverse primer (446_hygro_rev) was designed to target the hygromycin resistance transgene. For B2 and B4, the same reverse primer (237_fer1l4_rev_4), targeting the Fer1l4 locus, was used. For B2, a forward primer (273_lp_pcr_fwd) was designed to target the BSDR transgene. For B4, a forward primer (447_mcherry_fwd) was designed to target the mCherry transgene. **(B)** The PCR products were run on an agarose gel.

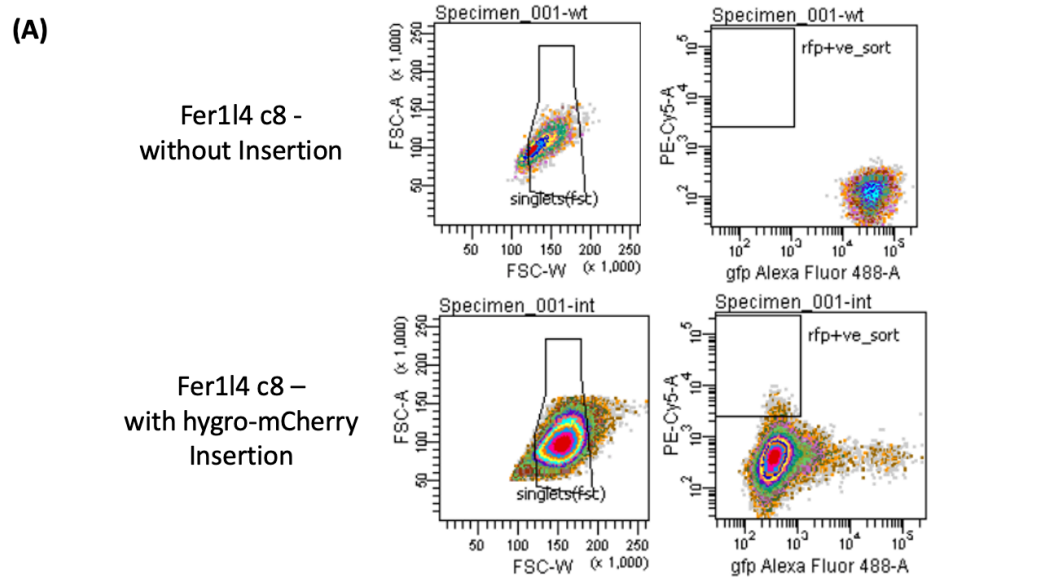
When the amplification products from the genomic DNA of the Fer1l4_c8 cell line (without transfection) were visualised on an agarose gel, bands 1 and 2 were present in the expected locations (Figure 2.12B). As expected, there were no bands present for bands 3 and 4. However, when the amplification products from the genomic DNA of the Fer1l4_c8 cell line (following transfection with hygro-p2a-mCherry and pEF1α-Bxb1) were visualised, all 4 bands (bands 1-4) were present in the expected locations (Figure 2.12B). The presence of bands 3 and 4 suggested that integration of the hygro-p2a-mCherry construct into the LP had occurred successfully. However, the presence of bands 1 and 2 suggested 2 possible scenarios had occurred. Firstly, and perhaps the most likely, Fer1l4_c8 host cells were still present in the transfected population. Alternatively, there were additional LPs in the cell line (either at off-target loci or on the other allele of the chromosome) that were giving rise to bands 1 and 2. Therefore, to remove any Fer1l4_c8 host cells from the population, it was decided that the diagnostic PCR analysis should be conducted on single cell clones (see next section 2.5.14).

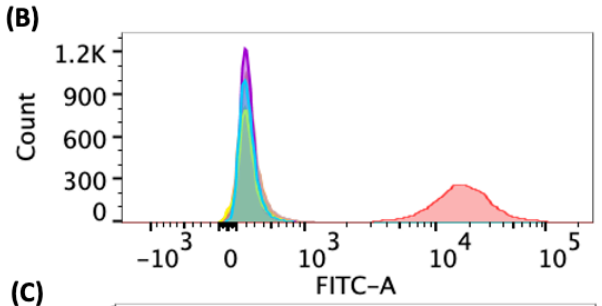
To fully confirm that integration into the LP had occurred, the *Att* recombination sites of the Fer1l4_c8 cell line (before transfection), hygR-p2a-mCherry plasmid and Fer1l4_c8 cell line (after transfection) were sequenced (see appendix 3). As previously discussed, the LP should contain AttB recombination sites. Upon co-transfection of the cell line with a construct flanked by the appropriate AttP sites and the appropriate integrase, recombination of the adjacent Att sites should yield AttR sites. The 5'-AttB (GC) site of the LP had previously been sequenced when diagnostic PCRs were conducted for the Fer1l4_c8 cell line (see section 2.5.8 for more details) (appendix 3). Sequencing of the 5'-AttP (GC) site of the hygR-p2a-mCherry plasmid showed that the AttP site was the expected sequence (appendix 3). Subsequent DNA extraction and Sanger sequencing of the PCR products of bands 3 and 4 confirmed that there had been successful integration of the mCherry construct into the LP (appendix 3).

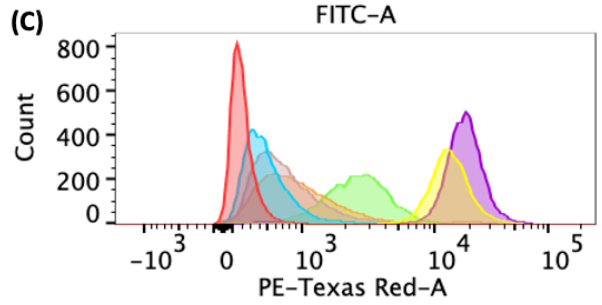
2.5.14 FACS Single Cell Sorting and Subsequent Diagnostic PCRs Suggest there is an Additional, Non-Functioning LP in the Fer1l4_c8 Cell Line

To isolate cells that had successfully integrated the hygR-p2a-mCherry plasmid and were expressing mCherry, FACS single cell sorting was used (Figure 2.13A). Firstly, a gating strategy was used to isolate the single cells in the population. Using Fer1l4_c8 host cells as a reference point, the $mCherry^{+}mNG^{-}$ population could be identified and sorted into a 96 WP. After single-cell sorting with FACS, the cells were grown at 37°C and 5% CO_2 for 2 weeks, in accordance with FDB's standard procedure for cloning Apollo™ X-derived cell lines. 5 clones were progressed from the 96 WP to a 24 WP and were subsequently analysed using flow cytometry. As expected, all 5 clonal cell lines were negative for mNG expression (Figure 2.13B). However, interestingly, there was a much wider variation in mCherry expression for each clone (Figure 2.13C). This variation was thought to be due to additional RI of the hygR-p2a-mCherry plasmid.

The genomic DNA of each clone was subsequently extracted and the same diagnostic PCR was performed as before (Figure 2.12A). Following integration of the hygR-p2a-mCherry, if the cell line contained 1 copy of the LP, one would expect to see bands 1 and 2 disappear. This is because there should be no mNG or BSDR transgene for the internal primers to anneal to, as the single mNG-p2A-BSDR cassette in the LP would have been displaced by the hygR-p2a-mCherry construct. Instead, bands 3 and 4 would be present because the hygR primer (band 3) and mCherry primer (band 4) can now anneal. However, this result was not seen when the PCR products were visualised on an agarose gel (Figure 2.13D). The band 2 (1.3kb) that was seen in the previous agarose gel (Figure 2.12B), when the genomic DNA of the transfected pool was analysed, had disappeared when the same diagnostic PCR was performed on the isolated clones. However, band 1 (0.4Kb) was still present.







	Subset Name	Count	SampleID
	Single Cells	*10058*	*No Transfection*
	Single Cells	10432	Hygro-mCherry (Pool)
	Single Cells	8701	Hygro-mCherry 2G6
	Single Cells	11650	Hygro-mCherry 2D10
	Single Cells	11909	Hygro-mCherry 1G4
	Single Cells	10587	Hygro-mCherry 1B8
	Single Cells	9513	Hygro-mCherry 1D8

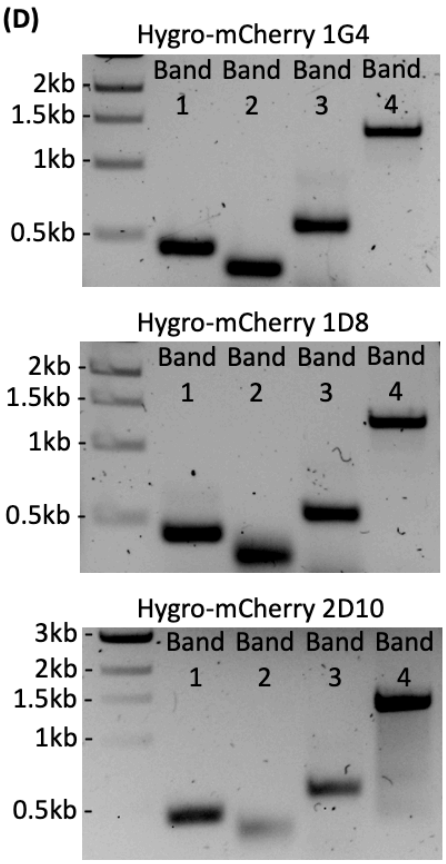


Figure 2.13 - Single Cell Sorting of hygR-P2A-mCherry Cells and Subsequent PCR Analysis

(A) FACS single-cell sorting of the Fer1l4 c8 cell line following co-transfection with the hygro-p2A-mCherry and Bxb1 plasmid. **(B)** Flow cytometry analysis of mNG expression (measured using the FITC-A channel) of various clones that *(Continued on next page)*

(Figure 2.13 continued) had been isolated using FACS. **(C)** Flow cytometry analysis of mCherry expression (measured using the PE-Texas Red-A channel) of various clones that had been isolated using FACS. **(D)** The same PCR reactions as before (see section 2.5.13) were conducted and the PCR products were run on an agarose gel.

The presence of band 1 and the disappearance of band 2 was thought to be due to an additional, non-functioning copy of the LP in the cell line. This hypothesis can be explained by the following logic. Even if there are any additional copies of the LP in the Fer1l4_c8 genome, band 2 will most likely not be observed because successful integration into the Fer1l4 LP will prevent the forward primer annealing to the BSDR transgene, and the reverse primer for this PCR product anneals to the Fer1l4 genome (outside of the LP). This means that any LP integrated elsewhere in the genome would not be amplified in this PCR reaction. Therefore, band 2 acts as an indicator for successful integration at the Fer1l4 locus only.

However, for band 1, both the forward and reverse primer anneal within the LP (the forward primer anneals to the pCAG promoter and the reverse primer anneals to the mNG transgene). If the additional LP was a functional LP, one may expect that certain clones will have successfully integrated the hygR-p2a-mCherry construct into the additional LP, yielding a higher mCherry transgene copy number. It could be argued that this gave rise to the varying levels of mCherry expression shown by the isolated clones (Figure 2.13). With this in mind, one would expect that clones which have successfully integrated the mCherry expression cassette at multiple LP loci will show higher expression of mCherry than cells that have only integrated the mCherry transgene at the Fer1l4 locus. However, there are multiple reasons why this is unlikely to be the case in this situation. Firstly, 3 clones (1G4, 1D8 and 2D10) show varying levels of mCherry expression (Figure 2.13B), but each clone shows almost identical mNG expression (Figure 2.13C). This suggests that the boosted levels of mCherry expression in clone 2D10 is not

because of integration into additional LPs but is instead a result of RI of the construct. Secondly, if the boosted level of mCherry expression in clone 2D10 was due to integration at any additional LPs, one may expect band 1 to disappear, as a result of successful integration of the hygR-p2a-mCherry construct into all additional LPs. However, band 1 persists in all clones (Figure 2.13D).

This leaves an alternative scenario which could account for the presence of band 1 in the diagnostic PCRs of the clones (Figure 2.13D). If the additional LP is not a functioning version of the LP, the hygR-p2a-mCherry plasmid will not be able to integrate into it, displacing the mNG transgene. Therefore, band 1 will persist in all clones, because both forward and reverse primers can anneal to this additional non-functioning LP. As was shown during flow cytometry, mNG expression will also be equal in all clones, regardless of mCherry expression, because there will be no difference between clones in the number of LPs with mNG displaced (Figure 2.13B). However, in the Fer1l4_c8 cell line, integration of the hygR-p2a-mCherry plasmid into the still leads to complete loss of mNG expression in the cell line (Figure 2.13B). If there was an additional, non-functioning LP in the cell line (expressing mNG), one would expect that mNG expression would not be completely lost following integration of the hygR-p2a-mCherry plasmid into the Fer1l4-locus LP, even if the additional LP is located in a poorly expressed region of the genome. Therefore, the question remains as to why this is not the case for the Fer1l4_c8 cell line.

Given that in other studies, the Fer1l4 locus has been shown to be a site that enables stable transgene expression, but the Fer1l4_c8 cell line exhibited loss of mNG expression over 60 generations (see section 2.5.9), it was hypothesised that the additional copy(ies) of mNG in the cell line were being silenced. This could explain why integration into the LP at the Fer1l4 locus of Fer1l4_c8 leads to complete loss of mNG expression in the cell line (Figure 2.13B). To pursue this idea further, the mNG expression of the Fer1l4_c8 cells used in the transfection study (named Fer1l4_c8_Trans), was analysed

using flow cytometry. The mNG expression of Fer1l4_c8_Trans was overlaid with the mNG expression of generation 10 and generation 60 Fer1l4_c8 cells from the previous stability study (Figure 2.9). Interestingly, the mNG expression of the Fer1l4_c8_Trans was much closer to the mNG expression of generation 60 Fer1l4_c8 cells than the generation 10 Fer1l4_c8 cells (Figure 2.14). Therefore, it was reasoned that it is the additional copies of the mNG transgene that are being silenced over time, rather than the mNG cassette in the Fer1l4 LP.

2.5.15 Retrospective Analysis of Fer1l4 Clones

From the previous Fer1l4_c8 analysis, it was believed that the cell line contained a functional copy of the LP at the Fer1l4 locus plus an additional copy (or multiple copies) of a non-functioning LP at a different locus/ loci. The additional copy of an mNG transgene was thought to boost mNG initially, but over time this additional copy was silenced. Therefore, to identify clones with a single copy of the LP (with no additional copies of mNG) we decided to look back at the initial flow cytometry data that was performed on the Fer1l4 clones, after the clones had been progressed from the 96-well plate to a 24 WP (see section 2.5.5). During the Fer1l4 clone selection process, 6 Fer1l4 clones were selected that showed >1000 Single Cell Events, <1% mRuby2+ cells, >98% mNG+ cells and the lowest %rCV in mNG expression. The mNG expression and mRuby2 expression of these clones are shown in Figure 2.15. Each of the 6 clones show almost no mRuby2 expression (Figure 2.15A). However, there are 2 clear groups of clones for mNG expression. Group 1 (Fer1l4_b9 and Fer1l4_c9) clearly displays lower expression of mNG than group 2 (Fer1l4_c8, Fer1l4_d4, Fer1l4_d7, Fer1l4_d11). It is likely that the clones showing lower levels of mNG expression (Fer1l4_b9 and Fer1l4_d4) are the single integration clones. However, due to space restraints in the incubator at the time, Fer1l4_b9 and Fer1l4_c9 unfortunately had to be discarded before a cell bank could be made.

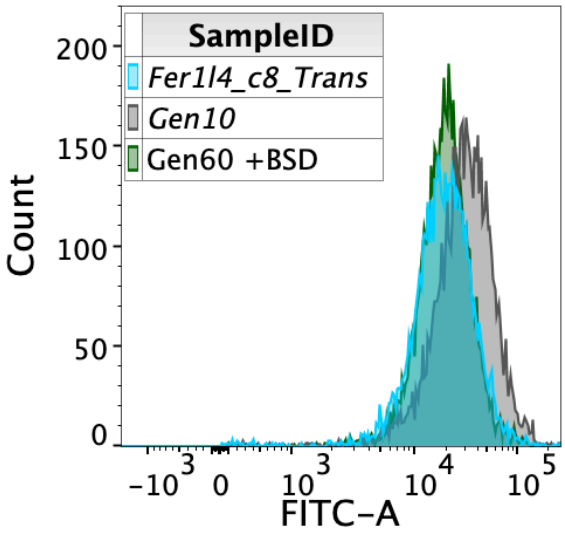


Figure 2.14 - Potential Silencing of Additional mNG Copies. Comparison of mNG expression (measured using the FITC-A channel) between the Fer1l4_c8 cell line that had been transfected with the hygR-p2a-mcherry and Bxb1 plasmid (Fer1l4_c8_Trans), with the mNG expression of the Fer1l4_c8 host cell line at generation 10 (Gen10) and generation 60 (Gen60).

Therefore, it was decided that, although the Fer1l4_c8 was not the ideal cell line for expression vector component comparison, due to the presence of additional copies of mNG in the genome, it did not make sense to go back and spend time analysing the Fer1l4_d4 and Fer1l4_d7 cell lines that had been progressed to research cell bank.

2.5.16 Optimising Integration of Constructs into the LP

Once successful integration into the LP had been confirmed, the next step was to optimise the integration process. In the previous experiment (see section 2.5.12), a hygR-p2A-mCherry construct was inserted into the LP of Fer1l4_c8 cells. The displacement of the mNG transgene from the LP led to a large drop in mNG expression but only a negligible increase in mean mCherry expression. Therefore, in an attempt to increase mCherry expression, the hygR and mCherry were swapped so that mCherry was 5’ of

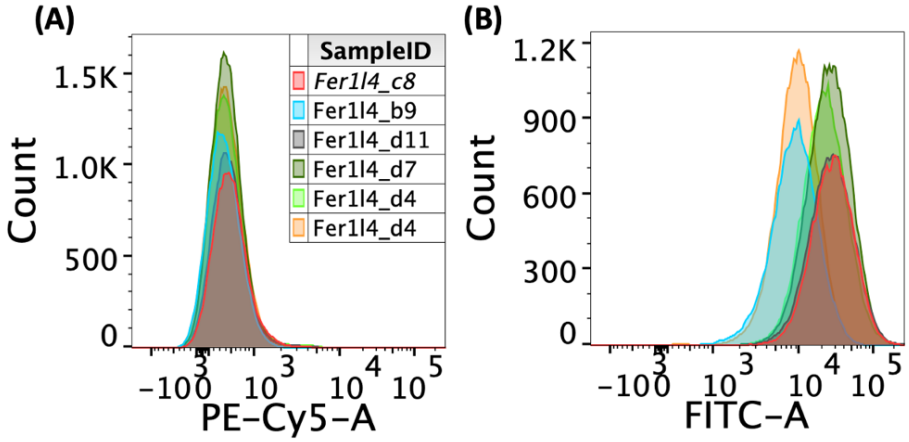


Figure 2.15 - Retrospective Analysis of Alternative Fer1l4 Clones Suggests that other Clones were the Single Integration Clones.

(A) Comparison of mRuby2 expression (measured using the PE-Texas Red-A channel) between different LP clones. **(B)** Comparison of mNG expression (measured using the FITC-A channel) between different LP clones.

hygR, producing a promoterless mCherry-p2A-hygR construct (Figure 2.16A).

To determine the optimum concentration of BSD for antibiotic selection, Fer1l4 c8 cells were cultured in different concentrations of hygromycin. To mimic the transfection process as closely as possible, 0.5 ml of Fer1l4_c8 cells were seeded at $2.5x10^5$ cells/ml into a 24 WP. After 3 days (the length of time the cells are usually left to recover following transfection), the cells from each well were spun down and resuspended in 1 ml of fresh media containing different concentrations of hygromycin. Cells were passaged a further 2 times (once every 3 days) using the same method. 3 days after the 3rd passage, cell viability was measured.

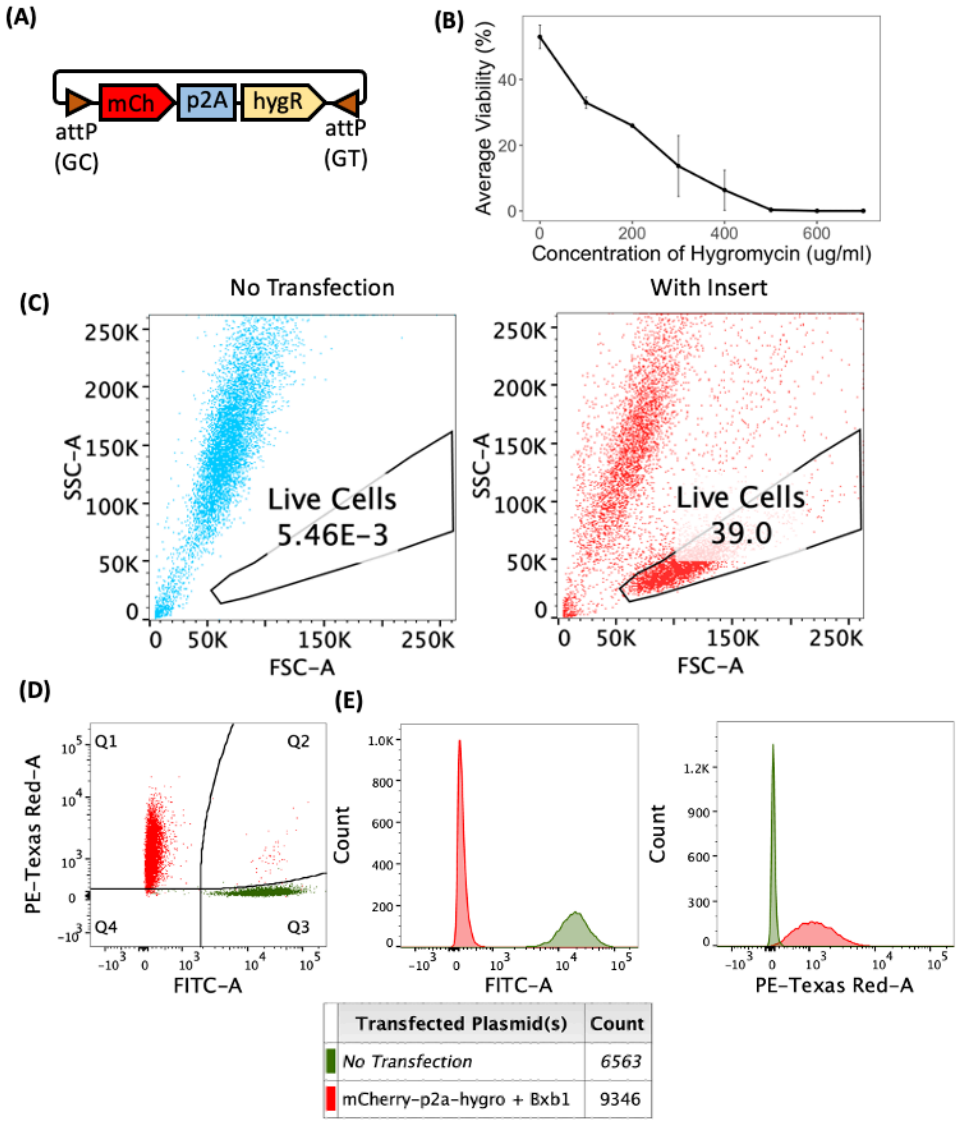


Figure 2.16 - Insertion of an mCherry-p2A-hygro Construct into the LP

(A) Design of the promoterless mCherry-p2A-hygR plasmid for integration into the LP. **(B)** Kill curve showing the impact of different concentrations of hygromycin on the viability of Fer1l4_c8 cells after 3 passages. **(C)** Fer1l4_c8 cells were co-transfected with the promoterless mCherry-p2A-hygR plasmid and a Bxb1 plasmid. The Fer1l4_c8 host cell line and the transfected cell population were subjected to 3 passages with 500 µg/ml of hygromycin in the media. The figure shows the difference in the percentage of live cells between the Fer1l4_c8 cell line with no

(Continued on next page)

(Figure 2.16 continued) transfection vs the Fer1l4_c8 cell line following selection with hygromycin. **(D)** Scatter plot for mCherry vs mNG expression for the untransfected Fer1l4_c8 population vs the Fer1l4_c8 population with mCherry-p2A-hygR integrated. **(E)** Difference in mNG expression (left) and mCherry expression (right) for the untransfected Fer1l4_c8 population vs the Fer1l4_c8 population with mCherry-p2A-hygR integrated.

500 µg/ml was the lowest concentration of hygromycin that was shown to kill >99% of cells and was thus selected as the concentration to be used for further experiments. However, it must be noted that at 0 µg/ml, average cell viability still dropped to 53%. Following co-transfection of Fer1l4_c8 cells with the promoterless mCherry-p2A-hygR construct and pEF1α-Bxb1, cells were left to recover for 3 days before antibiotic selection was started. Using the same antibiotic selection method as was used for the kill curve (3 passages with 500 µg/ml of hygromycin), the cells were analysed using flow cytometry. In the untransfected population 0.00546% of cells survived the selection process. This was in contrast with the cells that had been co-transfected with the mCherry-p2A-hygR construct and pEF1α-Bxb1, of which 39% cells survived the selection process (Figure 2.16C). Flow cytometry analysis of the transfected population showed that there had been a clear loss of mNG expression and an increase in mCherry expression upon integration of the promoterless mCherry-p2A-hygR construct (Figure 2.16D and E).

2.5.17 An Initial Exploration into Comparing the Strength of Different Promoters Using the Fer1l4_c8 cell line

The original aim of the study described in this chapter was to develop a system that could be used to compare novel expression vector components. Therefore, at this point, it was decided to do a quick and simple comparison of existing constitutive promoters using the LP system, with promoters driving mCherry expression (Figure 2.17). Firstly, 5x constructs were built which consisted of a promoterless hygromycin resistance gene and a promoter driving mCherry expression (Figure 2.17A).

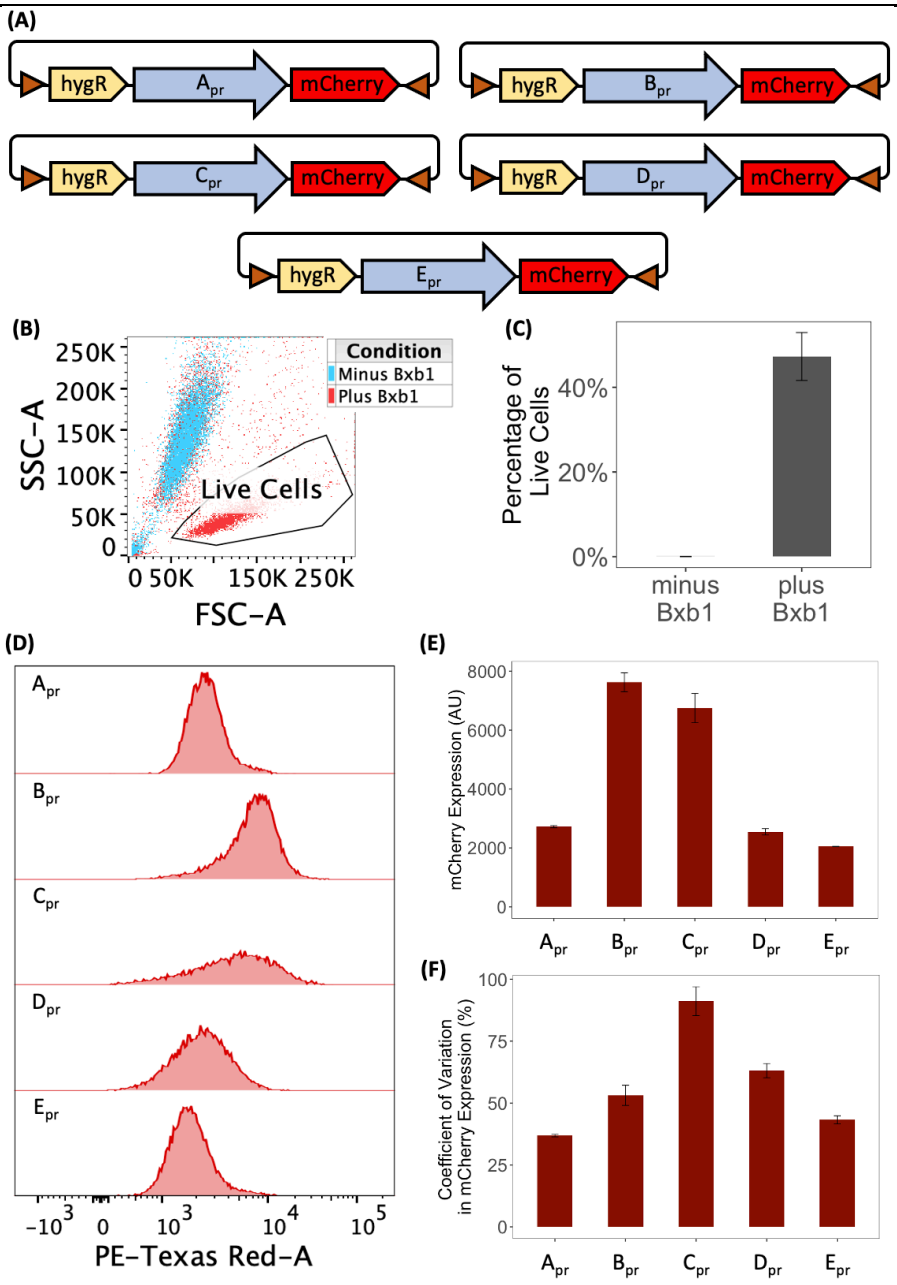


Figure 2.17 - Initial Comparison of Constitutive Promoters in the LP Cell Line

(A) 5 promoter-mCherry (A-E_{pr}) constructs were assembled. **(B)** A flow cytometry gating strategy was used to compare the viability of cells following hygromycin selection after transfection with pA-mCherry with or without co-transfection of pEF1α-Bxb1. **(C)** Bar chart comparing the viability of cells following hygromycin selection after transfection with pEF1α-mCherry with or without

(Continued on next page)

(Figure 2.17 continued) co-transfection of pEF1α-Bxb1. **(D)** Histograms of mCherry expression (measured using the PE-Texas Red-A channel) of the promoter-mCherry cell lines **(E)** mean mCherry expression (measured using the PE-Texas Red-A channel) of the promoter-mcherry cell lines. Bars show the mean expression of 3 replicates. Error bars show the standard deviation from 3 replicates. **(F)** Coefficient of variation in mCherry expression of the promoter-mcherry cell lines. Bars show the mean coefficient of variation of 3 replicates. Error bars show the standard deviation for coefficient of variation from 3 replicates.

Since the LP relies on a promoter trap, only integration of the construct into the LP should yield hygromycin resistance. The promoters were selected from a previous FDB study (A-E_{pr}, see Figure 2.17). In accordance with the methods used previously for integration into the LP, transfection of Fer1l4_c8 cells with 250 ng of the promoter-mCherry construct and 250 ng of pEF1α-Bxb1 was followed by antibiotic selection with hygromycin. Flow cytometry was used to check the percentage of live cells present in the population following antibiotic selection (Figure 2.17B). When Fer1l4_c8 cells were transfected with the A_{pr}-mCherry construct (without pEF1α-Bxb1), >99% of cells died during the transfection process (Figure 2.17C).

Flow cytometry was also used to analyse the mCherry expression of the transfected cell populations (Figure 2.17D). The highest mCherry expression was from cells that had been transfected with the B_{pr} and C_{pr} promoters (Figure 2.17E). Interestingly, both promoters (B_{pr} and C_{pr}) exhibited higher mCherry expression than A_{pr}, the promoter that FDB currently use for CLD. However, cells that had been transfected with C_{pr} also showed a much greater variation in mCherry expression than A_{pr}, as shown by the higher coefficient of variation. Cells transfected with A_{pr} showed the lowest coefficient of variation in mCherry expression.

2.6 Discussion

The aim of this chapter was to develop a system that could be used to compare expression vector components (outlined previously in section 1.10). To do this, a LP was to be designed, into which candidate vectors could be integrated via cassette exchange using site specific recombination. *Fer1l4*, *Cdk2ap2*, *Mrpl4* and *LemD2* were chosen as target loci for the LP to be integrated into using CRISPR/Cas9. From the data of previous studies, it was anticipated that these loci should be conducive for stable transgene expression (Rosser and Kleinjan, 2022; Zhang et al., 2015). The aim was to integrate a single LP copy into the cell line at a stable locus. This has previously been achieved in other studies that have managed to integrate a LP as a single copy into the CHO genome (Gaidukov et al., 2018; Inniss et al., 2017). Single copy integration of a LP would enable novel expression vector components to be compared at the same stable genomic locus with equal copy number (something that is not possible when comparing clones or transfected pools obtained by RI, or when using transient transfection).

Unfortunately, the *Mrpl4*, *LemD2* and *Cdk2ap2* cell lines had to be dropped for various reasons (discussed further in section 2.6.2). Instead, the clonal Fer1l4_c8 cell line was chosen for further analysis. Diagnostic PCRs showed that the LP had successfully integrated into the *Fer1l4* locus. However, further analysis indicated that there were additional non-functioning LP copies in the cell line, suggesting that some off-target integration had occurred (discussed further in section 2.6.1). Despite this setback, various mCherry-based constructs were shown to successfully integrate into the LP and a strategy for optimising the integration of constructs into the LP was developed. This confirmed the data from previous studies (Inniss et al., 2017), that the Bxb1 RMCE system is an effective system for integrating constructs into the CHO cell genome via SSR. As an initial proof-of-concept for comparing expression vector components using the LP system, 5 constitutive promoters were selected and compared, confirming the ability of the LP system to distinguish between various promoter strengths effectively.

2.6.1 Further Analysis of the Fer1l4_c8 Cell Line

As previously discussed, the presented results indicate that the Fer1l4_c8 cell line contains a functioning LP at the Fer1l4 locus, but also contains additional copies of mNG (as part of what is thought to be a non-functioning LP) at additional loci. Previously, LPs have been targeted to specific regions of the CHO cell genome using CRISPR/ Cas9 via homologous recombination. This was followed by diagnostic PCR, southern blot and droplet PCR to confirm that the LP had integrated as a single copy (Gaidukov et al., 2018; Inniss et al., 2017). Given that the diagnostic PCRs already confirmed that there were multiple LP copies in Fer1l4_c8, it was decided that the additional experiments (usually reserved to confirm single site integration) were not needed. To fully characterise the Fer1l4_c8 cell line, a series of additional steps could be conducted. A useful next step for understanding the degree of off-target integration would be to sequence the cell line. For example, it would be interesting to know if the sites in the genome that contain additional LPs are susceptible to silencing. This could be achieved by looking at existing CHIP-seq datasets, as well as transcriptomic data sets.

Additionally, in a stability study for Fer1l4_c8, it was shown that mNG expression decreased over a period of 60 generations (see section 2.5.9). This was in opposition to previous studies which had shown Fer1l4 to be a stable locus for transgene expression (Zhang et al., 2015). As a result, it was hypothesised that the loss of mNG expression witnessed over 60 generations could be due to silencing of additional copies of mNG in the cell line. Although previous studies have managed to integrate single copies of a LP into the CHO genome, confirmed via southern blots and dPCR, it would be interesting to determine how stable the number of integration sites are. For example, previous studies have shown extensive chromosomal rearrangements during long term culture (Vcelar et al., 2018).

Despite the clear limitations of Fer1l4_c8, the cell line is still of use. mCherry-based expression vector cassettes were successfully integrated into the LP

via RMCE, indicating that the LP design functions effectively in the Apollo™ X host cell line. This was the first study to integrate a LP into the Apollo™ X host cell line and thus, prior to the study, it was not known whether this LP design would function in the cell line. Additionally, as previously discussed, it is thought that off-target copies of mNG in Fer1l4_c8 are part of non-functioning LPs. As a result, it is thought that Fer1l4_c8 contains only 1 functioning copy of the LP. Therefore, for expression vector component comparison, expression vectors will still be compared at a single copy number and at the same genomic locus. The silencing of the mNG transgene over time is concerning and could complicate the interpretation of future studies. However, integrating constructs into the landing pad in the Fer1l4_c8 cell line at the same generation number should negate some of these issues.

2.6.2 Future Avenues for LP Integration at Stable Loci in Apollo™ X Host Cells for Expression Vector Component Comparison

Due to the Fer1l4_c8 cell line containing additional copies of mNG, and the difficulties that were had targeting the other loci (Mrpl4, Cdk2ap2 and LemD2), it is likely that, if an optimised cell line is to be generated for expression vector comparison, much of the work conducted in this chapter will need to be repeated (with a few key changes to the design of the experiment). Therefore, in this section, the challenges of targeting a single copy of a LP to specific loci are evaluated and alterations that could be made to future experimental design strategies are discussed. From the start, this study was challenged by multiple factors, which may have impacted the successful integration of LPs into the Apollo™ X host cell line. Fortunately, many of these problems have now been alleviated and thus understanding these challenges could aid the design of future experiments.

Firstly, the size of the incubator used to culture the cell lines was a challenge for experimental design. For example, it is often recommended that multiple sgRNAs should be designed for each gene of interest, because the activity and specificity of sgRNAs can be unpredictable. However, to reduce the

number of cell lines in the incubator, only 1 sgRNA could be tested per cell line. Additionally, after single-cell sorting, quick decisions had to be made as to which cell lines to keep (and progress to RCB) and which cell lines to discard. This selection step was performed by using flow cytometry to analyse clones at the 24-well plate stage. Clones were selected that showed >1000 Single Cell Events, <1% mRuby2^{+} cells, >98% mNG^{+} cells and the lowest %rCV in mNG expression (Figure 2.5). However, this quick selection step may have led to the single integration Fer1l4 clone being discarded (see section 2.5.15. for more details).

Designing experiments for LP integration was also challenged by the lack of available DNA sequence data for the Apollo™ X host cell line. As previously discussed, at the time of starting the experiment, there was no sequence data available for the Apollo™ X host cell line. Instead, public genomic data sets for CHO cells had to be used, adding to the difficulty of designing sgRNAs for targeting specific regions of the genome using CRISPR/Cas9. This may have impacted the success of the experiments for 2 reasons- designing sgRNAs for targeting genomic loci and primers for diagnostic PCRs. Additionally, it was uncertain whether the unclear results of the diagnostic PCRs from the Mrpl4_c4 analysis was due to improper primer design or unsuccessful LP integration at the Mrpl4 locus.

An early indication of how successful the CRISPR/Cas9 integration strategy may have been in the results obtained

from FACS single-cell sorting (

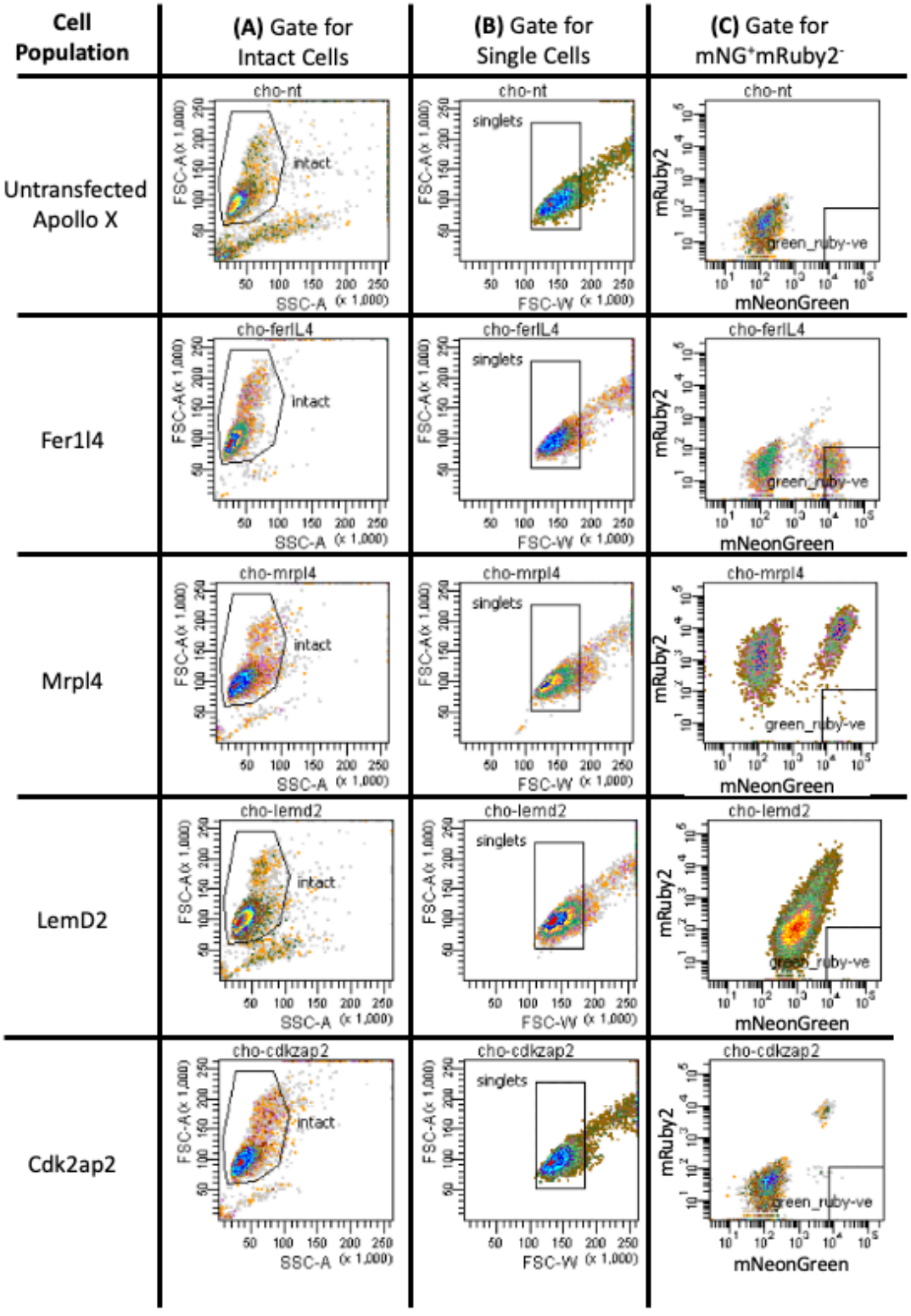


Figure 2.4). Using FACS, $mNG^{+}mRuby2^{-}$ cells (indicating successful CRISPR/Cas9 mediated LP integration) were selected for single-cell sorting. In the Fer1l4 cell line, 25% of the single cell population occupied the $mNG^{+}mRuby^{-}$ quadrant. However, less than 2% of the single population

occupied this quadrant for each of the other cell lines (LemD2, Mrpl4 and Cdk2ap2), suggesting that CRISPR/Cas9 mediated integration (via HDR) had been a rare event. One reason for this may have been the low transfection efficiency. For LP integration into the Apollo™ X host cell line, 1 ml of cells (at $1x10^6$ cells/ml) were transfected with 1ug DNA at a 1:9 ratio of sgRNA plasmid to LP plasmid. When comparing transfection reagents, the transfection method used for the LP integration method yielded a transfection efficiency of 5-20% when transfecting the cells with a single plasmid (data not shown). Therefore, it is likely that the transfection efficiency for the sgRNA plasmid was extremely low and thus sgRNA design may not have been the issue. Later optimisation of the transfection process, by culturing cells at a lower cell density prior to transfection, improved transfection efficiency to ~20-30% (data not shown). As a result, future experiments may benefit from transfecting the Apollo™ X host cell line using electroporation, which can yield transfection efficiencies of >95% in CHO suspension cells (Steger et al., 2015). However, the equipment required for electroporation was unavailable at the time.

A significant challenge throughout the experiments presented in this chapter was the poor growth of LP cell lines. The Cdk2ap2 cells were discarded due to poor growth after the single cell sorting stage. Mrpl4_c4 was discarded due to poor growth in the stability study. The standard operating settings for culturing suspension CHO cells in a shaking cell culture incubator should be as follows: 37°C, 125 RPM (at 25mm throw), 5% CO_2 and ≥80% humidity. Converting the shaking speed from RPM to RCF ($RCF = 1.118 \times 10^{-5} \times RPM^2 \times radius$, where the radius is half of orbit in cm) gives an operating setting of 0.218 ($= 1.118 \times 10^{-5} \times 125^2 \times 1.25$) for shaker speed. However, the CO_2 Resistant Shaker (ThermoFisher, Cat No. 88881102) that was used for these experiments has an orbit of 19mm and a maximum speed of 120 RPM (without breaking). Using the same calculation this gives an RCF value of 0.15 ($=1.118 \times 10^{-5} \times 120^2 \times 0.95$). It is important that suspension cells are shaken at the correct speed to ensure optimum growth and efficient gas exchange. Therefore, as the cells weren't shaken at a high enough speed, it

is highly likely that this could have impacted growth, leading to the loss in viability/ poor growth seen in the Cdk2ap2 and Mrpl4 cell lines. Fortunately, a new incubator was purchased later in the PhD meaning that the Fer1l4_c8 cells could be grown in this incubator during the experiments in which mCherry constructs were integrated into the LP. During these experiments, the growth and viability of the cell lines were much more consistent (data not shown), suggesting that the incubator played a significant role in the early experiments.

2.6.3 Expression Vector Component Comparison Using the Fer1l4_c8 Cell Line

Based on the analysis presented in section 2.6.1, the Fer1l4_c8 cell line was considered usable for any future expression vector component comparison experiments, despite its limitations. In this chapter, the Fer1l4_c8 cell line was used for an initial comparison of constitutive promoters. However, it was decided not to pursue this line of study further until novel promoters were developed. The focus was then intended to shift towards designing novel expression vector components, utilising transcriptomic data from fed-batch cultures of recombinant Apollo™ X recombinant cell lines. Unfortunately, the availability of this transcriptomic data was delayed (due to challenges in genome annotation), causing a shift in the research approach to explore alternative avenues for optimising CHO cell-based mAb production (see next chapter for more detail).

2.6.4 Future Avenues for Expression Vector Comparison

As previously discussed, expression vectors for mAb production consist of a selection marker (driven by a weak promoter) and a mAb expression cassette, which contain multiple different components including strong promoters, 5'-UTR, codon-optimised LC and HC genes, and 3'-UTRs. Traditionally, the individual components of the expression vector have been optimised to boost mAb production. Usually, the result of such work is a platform vector backbone that is subsequently used to generate cell lines for the manufacturing of all future mAb products. Although a platform-based approach to expression vector design has worked previously, there are now

a growing number of alternative product motifs (such as multispecific proteins and toxic proteins) entering the biologics market which pose their own individual challenges for production. With in mind, it is likely that a product-specific strategy for expression vector development will need to be established.

To implement a product-specific optimisation strategy for expression vectors, it will be vital that two things can be achieved. Firstly, a screening method must be developed that is capable of quickly comparing large numbers of expression vector components. In this chapter, a LP system has been presented as a potential solution for this purpose. Secondly, new vector component compositions must be quick and easy to assemble. Traditionally, DNA assembly approaches have relied on restriction digestion and ligation of DNA fragments using type II restriction enzymes. However, such strategies can be incredibly time consuming, requiring extensive planning to ensure the appropriate enzyme restriction sites are available and often relying on multiple rounds of cloning and verification. More recently, improvements in DNA synthesis technologies mean that tailor-made vectors can be synthesised and ordered online. However, if large numbers of expression vector compositions are to be compared, both experiment cost and vector production times will quickly escalate. To circumvent such issues, multiple reusable DNA parts can now be assembled in a single reaction using a variety of modular assembly approaches that have recently been adapted from Golden Gate (such the extensible mammalian modular assembly toolkit (EMMA) (Martella et al., 2017)) and Gibson Assembly methods (such as GMAP (Akama-Garren et al., 2016)). Of these methods, the EMMA toolbox is of particular interest since it enables expression vectors to be assembled from a library of DNA parts (each flanked by a pair of BsmBI recognition sites, which generate user-defined 4bp overhangs upon cleavage) in a single reaction. Cloning variants of CHO expression vector components into standardised EMMA entry vectors, and combining them in different combinations, would quickly yield a library of expression vector combinations for testing. By combining this DNA assembly strategy with the LP method

that has been developed in this chapter, one can now begin to envisage a system in which expression vectors are tailor-made for each product, utilising combinations of the aforementioned expression vector components, maximising both production and expression stability.

Chapter 3 - Characterising the Production Stability of Two Monoclonal CHO Cell Lines

3.1 Chapter Summary

Chinese Hamster Ovary (CHO) cells are the preferred mammalian cell line for the production of many therapeutic proteins, but production instability (i.e., loss of monoclonal antibody (mAb) productivity) during long term culture (LTC) remains a significant problem. To ensure that an unstable cell line does not advance to the manufacture stage, the productivity of lead candidate cell lines, identified during cell line development (CLD), are evaluated over a period of ~60-70 generations, covering the period of time it takes to scale-up from cell bank to production volume. Such stability studies take approximately 3 months to complete, thus representing a significant bottleneck in the CLD timeline. To reduce timelines, methods which enable the early prediction of cell line stability are required (discussed further in Chapter 4). However, before any methods of predicting instability could be pursued, model stable and unstable production cell lines were required. In this chapter, 2 monoclonal CHO cell lines (32-121 and 32-124) producing the same mAb were identified (from a previous FUJIFILM Diosynth Biotechnologies (FDB) study) and the production stabilities of the cell lines were validated over ~60 generations. The 32-124 cell line demonstrated minimal loss of productivity when cultured in the presence (1%) or absence (13%) of methotrexate (MTX) over 60 generations. Conversely, the 32-121 cell line exhibited a more substantial drop in productivity, with a 24% decrease when cultured in the presence of MTX and a 27% decrease without MTX. Further analyses of growth characteristics, specific productivity (Q_p), metabolite profiles, gene copy number, and transgene mRNA expression were subsequently conducted to investigate the underlying cause of this productivity loss.

3.2 Introduction

CHO cells are a well-established mammalian cell line for the production of many therapeutic proteins, due to their ability to grow in suspension culture at high cell densities, adapt to a variety of growth conditions and carry out correct protein folding and post-translational modifications. Such favourable attributes have enabled CHO cell-based production system to achieve high product concentrations, up to 10 g/L, thus making them an attractive choice for biopharmaceutical manufacturing. However, unpredictable loss of mAb productivity (termed “production instability”) during LTC in certain cell lines remains a significant issue for CHO cell bioproduction. Loss of productivity during LTC has primarily been attributed to loss of transgene copy numbers due to chromosomal rearrangements (Chusainow et al., 2009; Kim et al., 2011), or epigenetic modifications (such as methylation or histone modifications) resulting in transgene silencing (Osterlehner et al., 2011; Veith et al., 2016). The inherent genomic plasticity of CHO cells, characterised by continuous changes in the genomic landscape due to chromosomal rearrangements and chromosome loss, can lead to increased DNA damage and genomic mutations, impacting the stability of CHO cell phenotype (Wurm and Wurm, 2017, 2021; Zhang et al., 2016). These genotypic changes may affect metabolism, global gene expression (Jamnikar et al., 2015; Li et al., 2015; Tzani et al., 2021), sensitivity to cell stress (Bailey et al., 2012; Chusainow et al., 2009), DNA repair mechanisms (Torres et al., 2023), and apoptosis susceptibility (Dorai et al., 2012). All of these may contribute to production instability during LTC of recombinant CHO cell lines. As a result, time-costly stability studies are required to ensure that an unstable cell line does not progress to the manufacture stage. This section outlines an industry definition of CHO cell stability and the limitations associated with current method for assessing production stability are evaluated.

3.2.1 Defining Instability

To ensure that biomanufacturing processes are reproducible and of high quality, the biomanufacturing industry has established criteria for production processes, including guidelines for production stability (Dahodwala and Lee,

2019). Following CLD, the cell culture must be scaled from cell bank to the volume required for seeding a production bioreactor. As a result, it is critical that the cell line chosen for its high productivity, product quality and growth rate during CLD retains these qualities over the time frame required to scale up to large-scale production. *In vitro* cell age is defined by the International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) as the "Measure of time between thaw of the master cell bank vial(s) to harvest of the production vessel measured by elapsed chronological time, by population doubling level of the cells, or by passage level of the cells when subcultivated by a defined procedure for dilution of the culture" (ICH Q5B, 1996). 60-70 population doublings (frequently termed "generations") is often accepted as the limit for *in vitro* cell age for a commercial CHO cell manufacturing process (Dahodwala and Lee, 2019). A cell line is usually defined as stable if it is able to retain >70% of volumetric productivity titre over this generation limit (Dahodwala and Lee, 2019). However, it is preferred that cell lines retain over 80% of volumetric productivity titre (Dorai et al., 2012). Therefore, at FDB, cell lines retaining 70-80% volumetric productivity over 60 generations are placed into an alert limit for potential instability. Additionally, the therapeutic should not exhibit any "clinically meaningful" changes in product quality attributes over the defined period, relative to the reference product (Dahodwala and Lee, 2019).

3.2.2 A Stability Study Characterises the Production Stability of the Cell Line, Prior to Manufacture

To prevent unstable cell lines from advancing to manufacture stages, the productivity of lead candidate cell lines, identified during CLD, are evaluated over a period of 60-70 generations. During such studies, cell culture samples are taken at various generation timepoints (e.g., Gen10, Gen25, Gen45, Gen60) and used to seed fed-batch cultures (FBCs) (Figure 3.1).

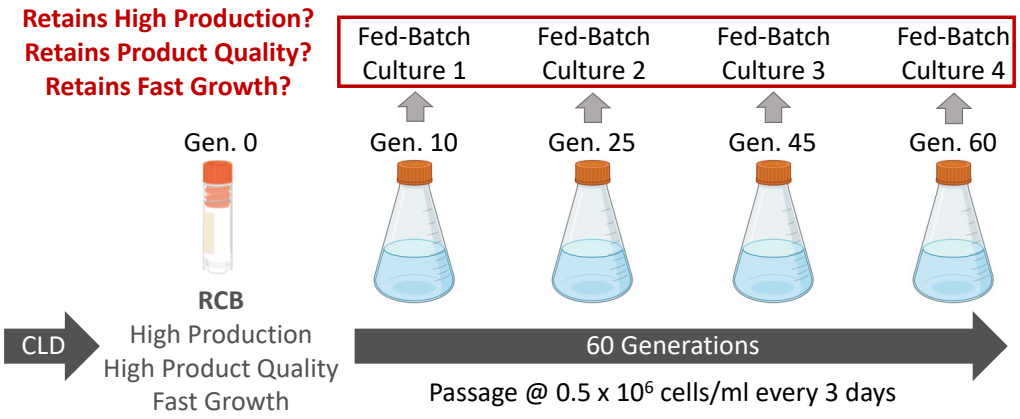


Figure 3.1 - The Stability Study.
Following CLD, research cell banks (RCBs) are produced for lead candidate cell lines that exhibit favourable growth characteristics and produce large quantities of product with the desired product quality attributes. The RCB is defined as generation 0. During the stability study, cell lines are grown under routine subculture for 60-70 generations after thawing. Cell culture samples are taken from the routine subculture at various generation timepoints (e.g., Gen10, Gen25, Gen45, Gen60) and used to seed FBCs. The product concentration, product quality and growth characteristics of the FBCs from different generation timepoints are compared and used to assess the stability of the cell line.

Cell lines that do not maintain >70% of their volumetric product concentration over the FBCs are usually identified as unstable and are not taken forward to the manufacture stage. However, this process can take up to 3 months, thus representing a significant bottleneck in the CLD timeline (Wurm and Wurm, 2017). The incorporation of early predictive methods for cell line stability could substantially reduce CLD timelines. Although a stability study may still be required for regulatory purposes, the predictive strategies could increase confidence in the stability status of a particular cell line. As a result, the stability study could be uncoupled from the critical path to manufacturing, enabling cell lines to progress to subsequent manufacturing stages before completion of the stability study. In Chapter 4, a method which uses the

Berkley Lights Beacon® system to compare cell lines with different stabilities is explored. However, before any methods of predicting instability could be pursued, a model stable and unstable production cell line was required.

3.3 Materials

3.3.1 Cell Culture Media

- **FDB-MAP subculture medium (+MTX):** FDB MAP (Merck, 87453C), 8 mM L-glutamine (Gibco, 25030-024), 175 nM Methotrexate (MTX) (Sigma, A6770)
- **FDB-MAP subculture medium (-MTX):** FDB MAP (Merck, 87453C), 8 mM L-glutamine (Gibco, 25030-024)
- **JM-05B FBC medium:** JM-05B (Irvine Scientific, 991439E), 8 mM L-glutamine (Gibco, 25030-024)3
- **FDB MAP freezing medium:** FDB MAP (Merck, 87453C), 10% DMSO (Sigma, D2438)

3.3.2 FBC Supplements

- **Cell Boost 7a** (GE, SH31026)
- **Cell Boost 7b** (GE, SH31027)
- **Glucose** (Sigma, G8769)

3.3.3 Molecular Biology Reagents

- E.Z.N.A.® Total RNA Kit I | Total RNA Extraction Kit (Omega, R6834-01)
- PureLink™ Genomic DNA Purification Kit (Invitrogen, K182000)
- LunaScript® RT Master Mix Kit (Primer-free) (NEB, E3025S)
- PowerUp™ SYBR™ Green Master Mix (Applied Biosystems, A25741)

3.4 Methods

3.4.1 Stability Study Design

A stability study was conducted on two selected cell lines. One vial from the cryopreserved RCB per cell line (generation 0) was thawed into FDB-MAP subculture medium (-MTX), following the method outlined in Section 3.4.2. After thawing, cells were subcultured every 3 days for the duration of the study following the method outlined in Section 3.4.3. During the first subculture, the culture was split into two flasks, one with MTX and one without MTX. At the second subculture post-revival, three cultures were established from each of the parent cultures (Figure 3.2). These three cultures were maintained as independent parallel lineages, denoted as lineage A, lineage B, and lineage C. Therefore, for each cell line, three cultures were maintained using FDB MAP growth medium without MTX, and three cultures were maintained using FDB MAP growth medium with MTX. Each lineage was maintained under routine subculture until the generation number of the cells in each culture reached at least 60 generations beyond that of the RCB.

A FBC shake flask screen was conducted at four different generation timepoints for each cell lineage: "early" (approximately 10 generations beyond the RCB), "early-mid" (approximately 25 generations beyond the RCB), "late-mid" (approximately 45 generations beyond the RCB) and "late" (≥ 60 generations beyond the RCB). The "early-mid", "late-mid" and "late" cultures were derived from cell culture samples taken from the routine subculture flasks at the appropriate timepoints, during the stability study. The "early" cultures were established by reviving vials of the RCB when the “mid” cultures reached approximately generation 50 and utilised for the FBC shake flask screen once they reached approximately 10 generations beyond the RCB. Subsequently, the FBC shake flask screens from the "early" and “late” cultures were conducted concurrently. At each generation timepoint, two FBCs were set up from two of the triplicate lineages (lineage A and lineage

B) from the “early”, “mid” and “late” cultures for each medium condition (plus or minus MTX).

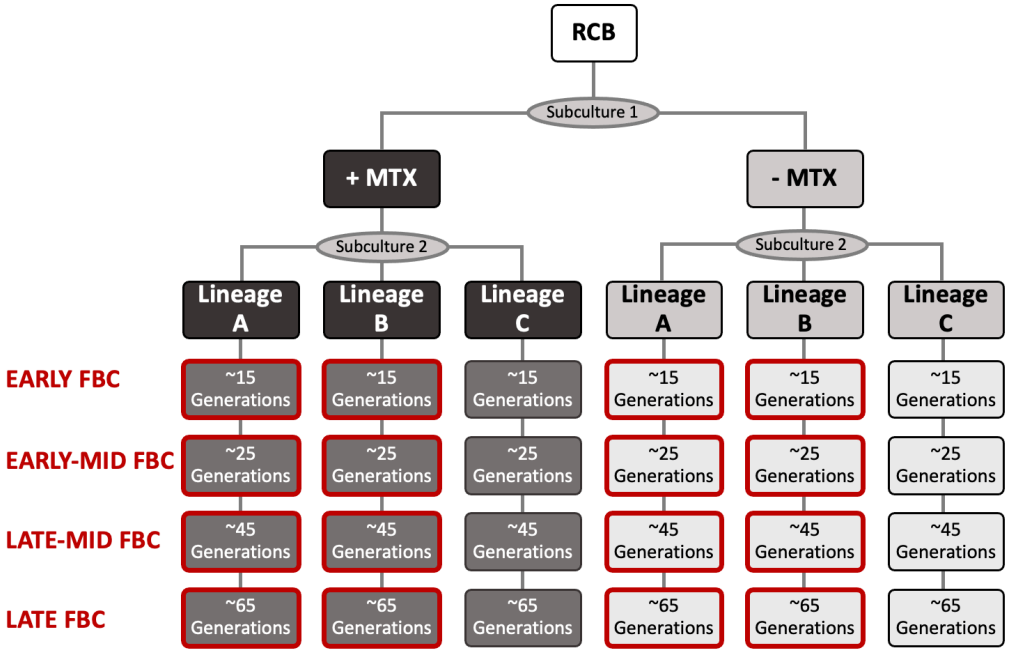


Figure 3.2 - Design of the Stability Study

A stability study was designed to validate the stability of the 32-121 and 32-124 cell line. A generation 0 RCB was thawed for each cell line into subculture media which did not contain MTX. At the first subculture, the flask was split into 2 flasks: 1 with MTX and 1 without. At the second subculture, the flasks were further split into 3 lineages. The 3 lineages were maintained in routine culture over a period of 60 generations. FBCs were conducted on lineages A and B, at “early” (~generation 15), “early-mid” (~generation 25), “late-mid” (~generation 45) and “late” (~generation 60) timepoints. 2 technical replicates were conducted for each cell line lineage.

3.4.2 Revival of Cell Lines from Cryopreserved Stocks

A cryopreserved vial containing 1.5 x 10^6 cells in 1.5 ml of freeze medium was thawed by partially immersing it in a 37°C water bath. The contents of the vial were then transferred to a 250 ml shake flask. Subsequently, 20 ml of FDB MAP growth medium was added drop-wise over 1-2 minutes. Viable cell density (VCD) and viability were determined using a Vi-CELL XR™ Cell Viability Analyser (Beckman Coulter). The VCD was adjusted to 0.5 x 10^6 cells/ml, and the cell line was transferred to an incubator set at 37°C, 5% CO2, 125 rpm shaking with a 25 mm throw.

3.4.3 Routine Subculture of Recombinant Apollo™ X Cell Lines (32-121 and 32-124)

Recombinant Apollo™ X CHO cells were sub-cultured every 3 days at a seeding density of 0.5 x 10^6 cells/ml. The Vi-CELL XR™ Cell Viability Analyser (Beckman Coulter) was used to determine cell concentration and viability. A suitable volume of cell culture and growth medium was transferred to 250 ml shake flasks with a working volume of 60 ml, ensuring the desired seeding density. The shake flasks were incubated at 37°C, 5% CO2, 80% humidity, and agitated at 125 rpm with a 25mm throw.

3.4.4 Cryopreservation of Cell Banks

At regular intervals, cryopreserved vials were obtained for each cell line. The cell count at subculture was used to calculate the volume of cell suspension required to achieve a concentration of 1.5 x 10^7 viable cells per cryovial. Following centrifugation at 180 x g for 10 minutes, the cell pellets were resuspended in freezing medium. Aliquots of 1.5 ml cell suspension were then aliquoted into 1.8 ml cryovials. The cryovials were subjected to controlled-rate freezing at -80°C for a minimum of 24 hours using a CoolCell or 'Mr Frosty' cell freezer. Subsequently, the cryopreserved vials were transferred to the vapour phase of a liquid nitrogen dewar for long-term storage. This cryopreservation procedure ensures the preservation and availability of cell lines for subsequent experiments and analysis in the stability study.

3.4.5 Stability Study: FBC Shake Flask Screen

The cultures were seeded at 0.5 x 10^6 cells/ml in one 250 ml Thomson flask in a total volume of 60 ml JM-05B fed-batch medium. The flasks were incubated in a shaking incubator set at 37 °C, 5% CO_2 in air and 125 rpm shaking with a 25 mm throw. Flasks were sampled daily from day 2 onwards and VCD and extracellular metabolite levels determined. VCD was determined using the Vi-CELL™ XR Cell Viability Analyser and extracellular metabolite levels determined using the YSI 2950 Biochemistry Analyser (YSI Life Sciences). Samples for product concentration analysis were also taken every other day from day 4 onwards and stored at -80°C. The cultures were fed daily from day 2 onwards, using a proprietary feeding regime.

3.4.6 Product Concentration Analysis

Cell culture samples were collected from each FBC flask on days 4, 6, 8, 10, 12, and 14. After centrifugation at 180g for 10 minutes, the supernatant was collected, and the cell pellets were retained for mRNA expression analysis (see section 3.4.9). Product concentration of the supernatant sample was quantified using Protein A affinity Ultra-Performance Liquid Chromatography (UPLC). The UPLC system used was a Waters Acquity H class bio with a TUV detector and Empower software, with UV detection set at 280 nm. The column used for the analysis was a Poros A/20 2.1 x 30 (Applied Biosystems, Cat. No. 2100100). To determine the product concentration, a standard curve was generated by analysing known concentrations of a generic purified antibody. The sample's product concentration was then calculated with reference to the standard curve.

3.4.7 Calculations

3.4.7.1 Generation Number

Generation number was calculated using the equation below:

$$\text{Generation number} = \text{Previous generation number} + \frac{(\ln(\text{viablecellcount}) - \ln(0.5))}{\ln(2)}$$

3.4.7.2 Time Integral of Viable Cell Density (IVC)

The time integral of viable cell density (IVC; x 10^6 cells·h/ml) was calculated using the method described by Renard et al., 1988. Firstly, the area under the growth curve [cell time] for adjacent data time points was calculated using the following equation:

$$\text{Cell Time} = \frac{X_0+X_1}{2} \times (t_1 - t_0)$$

Where X_0 is the VCD at the 1st cell culture sample point, X_1 is the VCD at the 2nd cell culture sample point, t_0 is the elapsed time at the 1st time point and t_1 is the elapsed time at the 2nd time point. IVC is then calculated as the sum of the individual cell times across the entire growth profile.

3.4.7.3 Relative Product Concentration

To maintain confidentiality of the product concentration measurements obtained during the FBCs, the graphs were presented using relative product concentration values. In this approach, each product concentration measurement was divided by the day 14 product concentration obtained from the early generation FBC of the 32-124 Plus MTX Lineage A (Replicate 1) cell line. Consequently, the day 14 32-124 Plus MTX Lineage A (Replicate 1) cell line was assigned a relative product concentration value of 1 for the early generation FBC. This normalisation method allowed for the comparison and visualisation of the relative performance of different samples without revealing the proprietary product concentration values. By using relative product concentration, the focus of the analysis remained on the relative changes and trends observed among the samples rather than the absolute values of the product concentration measurements.

3.4.7.4 Relative Specific Productivity

Methods for calculating mAb production per cell and time unit were employed, utilising two distinct approaches as outlined below:

Method 1 – Relative Single Point Specific Productivity (sp-Q_p)

A single point specific productivity (sp-Q_p) for specific timepoints was calculated by dividing the product concentration at that timepoint with the IVC

at the same timepoint. For example, because cell count data was unavailable, sp-Q_p was used to measure day 14 Q_p for the Ambr®15 bioreactor study (see section 3.5.2). Additionally, sp-Q_p was used to analyse variations in mAb production per cell throughout the FBC (see section 3.5.10). Relative sp-Q_p values were obtained by normalising each sp-Q_p measurement against the sp-Q_p value derived from the early generation FBC of the 32-124 Plus MTX Lineage A (Replicate 1) cell line.

Method 2 – Specific Productivity Rate (Q_p-rate)

The specific production rate (Q_p-rate) for the entire 14-day FBC was determined using linear regression analysis of the product concentration against the IVC (Figure 3.3). The slope of the regression line represents the specific production rate, denoted as Q_p-rate. Relative Q_p-rate values were obtained by dividing each Q_p-rate measurement by the Q_p-rate value derived from the early generation FBC of the 32-124 Plus MTX Lineage A (Replicate 1) cell line.

3.4.7.5 Relative Metabolite Concentration

To maintain confidentiality of the metabolite concentration measurements obtained during the FBCs, the graphs were presented using metabolite concentration values. In this approach, each metabolite concentration measurement was divided by the respective day 2 metabolite concentration obtained from the early generation FBC of the 32-124 Plus MTX Lineage A (Replicate 1) cell line. Consequently, the day 2 32-124 Plus MTX Lineage A (Replicate 1) cell line was assigned a relative metabolite concentration value of 1 for the early generation FBC for lactate, glucose, glutamine and glutamate.

3.4.8 Linear Regression Analysis to Assess Cell Line Stability

Cell line stability was assessed by monitoring changes in growth (IVC and maximum VCD) and productivity (product concentration and Q_p) attributes

over a period of 60 generations after thawing. The primary method for assessing production stability was to determine the percentage change in

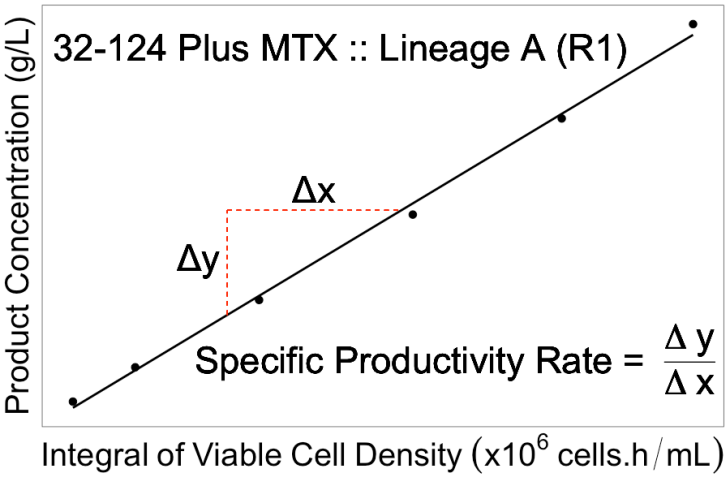


Figure 3.3 – Calculating Specific Productivity Rate using Linear Regression Analysis of the Product Concentration against IVC

each attribute between the early and late FBC. Linear regression analysis was also used to complement the percentage change analysis for the product concentration and Q_p attributes. Firstly, the relationship between the chosen parameter and increasing generation number was explored. If the data series appeared to be linear upon initial examination, a linear regression line was fitted with generation number as the independent variable (x) and using the generation number as the independent variable (x) and the corresponding growth or productivity parameter as the dependent variable (y). The null hypothesis was formulated, assuming that the slope of the fitted line was equal to zero (indicating no significant change in the parameter with increasing generation number) and tested at a 5% significance level. The t-statistic was computed by dividing the slope of the fitted line by the standard error of the slope. A critical t-value was determined based on the degrees of freedom and a significance level of p=0.05. If the calculated t-statistic exceeded the critical t-value, the null hypothesis was rejected, signifying a statistically significant change in the parameter. Conversely, if the calculated t-statistic fell below the critical t-value, the null hypothesis was not rejected.

3.4.9 Quantitative real-time PCRs

3.4.9.1 Sample Preparation

Genomic DNA for Gene Copy Number Analysis

At various generation timepoints, on the day of subculture, a culture volume equivalent to 5 x 10^6 cells was collected from the routine subculture flask. The cells were then centrifuged for 10 minutes at 180g, and the supernatant was discarded, retaining the cell pellet. Genomic DNA was isolated from the cell pellets using the PureLink™ Genomic DNA Purification Kit (Invitrogen, K182000).

cDNA for mRNA Expression Analysis

1 ml samples from day 6 of each FBC flask were extracted. The cells were then centrifuged for 10 minutes at 180g, and the supernatant was discarded, retaining the cell pellet. RNA was extracted from the cell pellets using E.Z.N.A.® Total RNA Kit I | Total RNA Extraction Kit (Omega, R6834-01). 1000 ng of RNA was converted to cDNA using the LunaScript® RT Master Mix Kit (Primer-free) (NEB, E3025S) kit.

3.4.9.2 Threshold Cycle (CT) Determination

Primer sets were designed to target the transgenes of interest and the reference genes. Primer efficiency was assessed using the standard curve method. A standard curve was generated by preparing serial dilutions of genomic DNA or cDNA. The RT-qPCR reactions were performed using StepOne™ Real-Time PCR System, and the threshold cycle (CT) values were recorded for each dilution. Linear regression analysis was applied to calculate the slope of the CT values against the log of the genomic DNA/ cDNA concentration. The primer efficiency (E) was then determined using the equation below:

$$\textit{Primer Efficiency} = 10^{\frac{-1}{Slope}} - 1$$

An efficiency of 100% would indicate perfect amplification, but values between 90-110% were considered acceptable for reliable quantification. Additionally, to assess the goodness of fit, the coefficient of determination (R2) was required to be greater than 98%.

Samples from each cell line across 60 generations were then analysed using RT-qPCR to obtain the CT values for the genes of interest (DHFR, heavy chain, and light chain) and the reference genes. The relative gene copy number for each gene of interest was then calculated using the delta CT method, which involves normalising the CT values of the target genes to the CT values of the reference genes.

3.5 Results

3.5.1 A previous FDB study had Produced and Analysed the Stability of 48 Monoclonal Cell Lines Expressing the same mAb

Before any methods of predicting instability could be pursued (see Chapter 4), a model stable and unstable production cell line was required. In a previous study, FDB employed a standard CLD protocol to generate 46 monoclonal cell lines, each producing the same mAb (an IgG4k). During this process, the expression vector (which contains heavy and light chain transgenes as well as a DHFR transgene for MTX-mediated selection) was randomly integrated into Apollo™ X DG44-CHO host cells to produce a heterogenous population of cells. 48 high-producing monoclonal cell lines were subsequently identified by isolating and screening clones. To determine the production stability of the clones, the cell lines were grown for 60 generations and cell culture samples were collected at generation 10 and 60. It is important to note that the Ambr®15 study was only conducted on cell lines that had been grown in the presence of MTX. These cell culture samples were used to seed a single FBC per cell line that were run using an Ambr®15 bioreactor system (Run 1 – Generation 10 cell culture samples, Run 2 – Generation 60 cell culture samples) and the percentage change in product concentration was subsequently measured (Figure 3.4A). It is worth

emphasising that the approach described here differs from the standard setup typically employed by FDB for their stability studies.

Interestingly, the majority of the cell lines exhibited an increase in product concentration between generation 10 and 60 (Figure 3.4A). This was unusual and not a common characteristic of recombinant CHO cells during LTC. A particularly striking example of this was that 3 cell lines exhibited an increase in product concentration of more than 100% between generation 10 and 60. Therefore, it was hypothesised that the large changes in productivity characteristics between generation 10 and 60 were at least partly due to conditions within the Ambr®15 bioreactor system itself, not just changes in cell characteristics during LTC.

Establishing the root cause of the unusual productivity results could help with the selection of cell lines for further study. Firstly, it was thought that changes in culture conditions within the Ambr®15 bioreactor system between generation 10 and 60 may have contributed to improvements in cell growth during the generation 60 run, supporting large increases in product concentration for many of the cell lines. Two measurements of growth performance were assessed: maximum VCD and IVC. Maximum VCD refers to the highest VCD that is reached during the FBC. IVC quantifies the effective working time for a dynamic viable cell concentration within a specific time frame, similar to how man-hours are calculated in workforce management (Hamilton, 2021). However, the mean IVC (Figure 3.4B) and maximum VCD (Figure 3.4C) for all cell lines showed no significant difference between generation 10 and 60. With this in mind, it was thought that the unusual product concentration results may have been specific to certain culture stations, perhaps due to a problem in one of the culture stations during the generation 10 run. The cell lines grown in culture station 3 showed a small but insignificant increase (at the 95% confidence level) in mean IVC between generation 10 and 60 (Figure 3.4D) and was the only culture station to show a significant increase in product concentration between generation 10 and 60 (Figure 3.4E). However, when the cell line

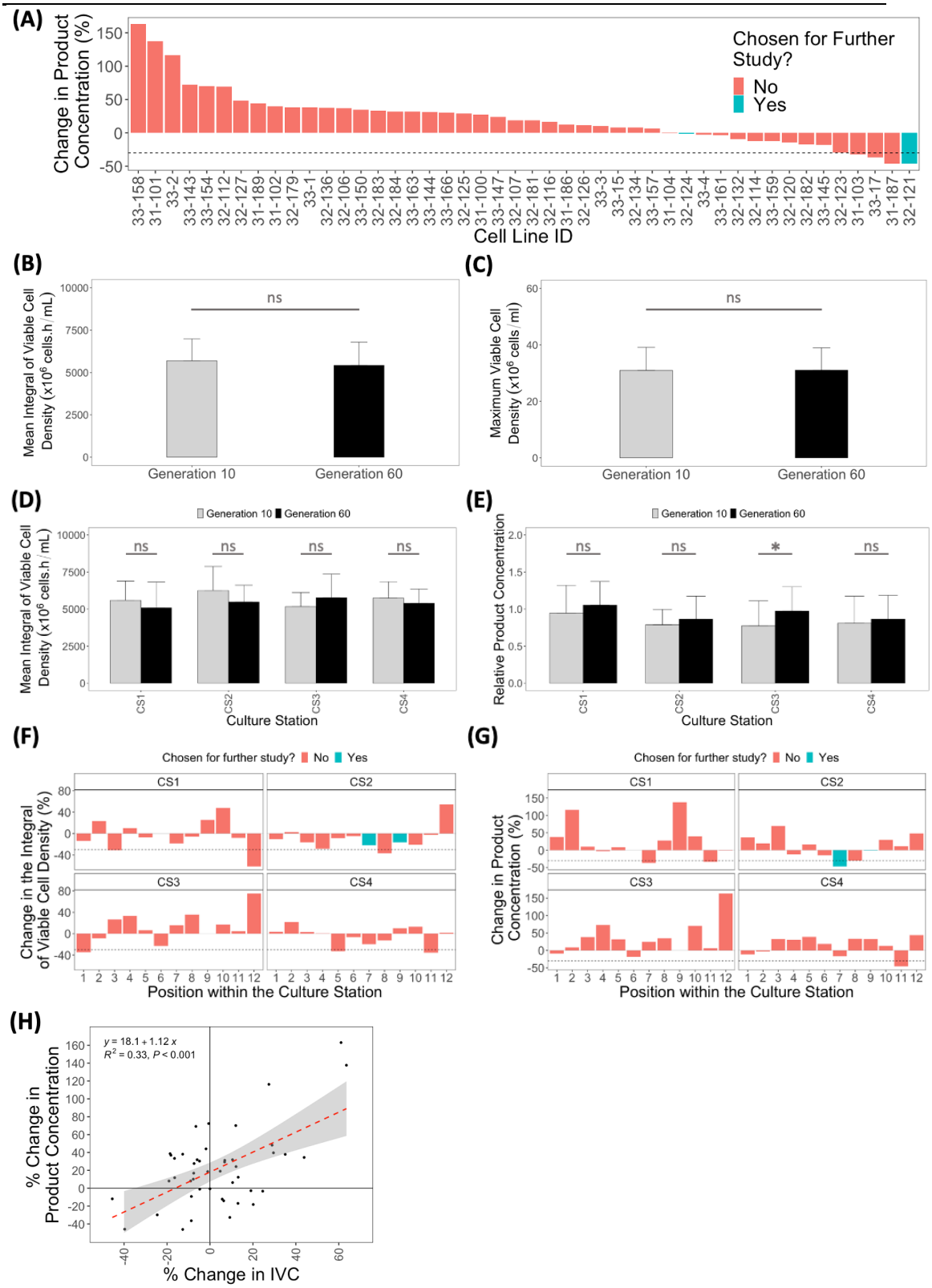


Figure 3.4 - Stability Analysis of 48 Monoclonal Cell Lines over 60 Generations Using the Ambr®15 Bioreactor System

(Continued on next page)

(Figure 3.4 continued) Stability analysis of 46 monoclonal cell lines expressing the same mAb over 60 generations performed by FDB. Two Ambr®15 bioreactor runs were performed at generation 10 and 60. **(A)** The percentage change in product concentration for the 46 cell lines. Dashed black line indicates a drop in productivity of 30%. **(B)** The difference in mean IVC between generation 10 and generation 60. Data represented as the mean IVC + standard deviation of the 46 cell lines. **(C)** The difference in average maximum VCD between generation 10 and generation 60. Data represented as the mean maximum VCD + standard deviation of the 46 cell lines. **(D)** The mean IVC (+ standard deviation) for each culture station (CS1/2/3/4) at generation 10 and generation 60. **(E)** The mean relative product concentration (+ standard deviation) for each culture station at generation 10 and generation 60. **(F)** The percentage change in IVC for each cell line, grouped by culture station (CS1/2/3/4) and position within the culture station (1-12). Dashed black line indicates a drop in IVC of 30%. **(G)** The percentage change in product concentration for each cell line, grouped by culture station and position within the culture station. Dashed black line indicates a drop in productivity of 30%. **(H)** The correlation between the percentage change in IVC and percentage change in product concentration between generation 10 and 60. Dashed red line represents the linear regression line and shaded area represents the 95% confidence level interval for predictions from the linear model. For graphs **(B)**, **(C)**, **(D)** and **(E)** $^{ns}P > 0.05$, $*P \leq 0.05$, $**P \leq 0.01$.

The data from Figure 3.4 was gathered in 2018 by FDB as part of a previous study, which was conducted prior to the start of James Donaldson's PhD. Data from FDB was re-analysed to inform future stability study experiments by James Donaldson.

data was grouped according to culture station, there was not a clear relationship between the culture station position and change in product concentration (Figure 4.4F) or IVC (Figure 4.4G), suggesting that any relationship between culture station position and change in titre is weak. For example, in culture station 3, the 33-158 cell line exhibited an increase in IVC of 75% between generation 10 and 60, but the 32-132 cell line exhibited a decrease in IVC of 35% between generation 10 and 60. Discussions with FDB also suggested that they have experienced variations in both productivity and growth characteristics between duplicate small-scale bioreactors, even within the same culture station, in previous studies (data not shown). Therefore, it is unlikely that the unusual product concentration results can be solely attributed to changes at a single culture station. This was important because, if the unusual product concentration results could be attributed to changes at a single culture station, the cell lines from that particular culture station could be excluded from further investigation. Interestingly, there was a significant correlation ($p < 0.001$) between the percentage change in IVC and the percentage change in product concentration between generation 10 and 60 (Figure 4.4H). However, the percentage change in IVC only accounted for 43% of the variation in percentage change in product concentration, suggesting that there are other factors that impacted the percentage change in product concentration (such as cell line stability).

3.5.2 The 32-121 and 32-124 Cell Lines were Selected for Further Investigation

Although the changes in product concentration of certain cell lines in the Ambr®15 bioreactor stability study experiments were unlikely to solely be due to changes in cell line characteristics over 60 generations, and instead partly due to the variability of Ambr®15 bioreactor runs, it was decided that this data would still be used to select a “stable” and “unstable” cell line for further investigation. A stability study, which incorporated both biological and technical replicates, would then be conducted to fully validate the stability of the chosen cell lines.

With this in mind, cell lines 32-124 and 32-121 were chosen for further study. In the FDB study, cell line 32-124 exhibited a drop in titre of 1.1% between generation 10 and 60, while cell line 32-121 exhibited a drop in titre of 46.1% (Figure 3.4A). As previously discussed in section 3.2.1, a cell line is considered unstable if it exhibits a drop in product concentration >30% over the defined stability study period. Based on this criterion, cell line 32-124 was deemed to be stable, while 32-121 was classified as unstable. At generation 10, the 32-121 cell line produced a higher day 14 product concentration (relative product concentration = 1) than the 32-124 cell line (relative product concentration = 0.87) (Figure 3.5A). To estimate the rate of protein expression per cell and time unit, sp-Q_p was calculated on day 14 by dividing the day 14 product concentration by the day 14 IVC (see section 3.4.7.4 for calculation). Although the 32-121 cell line exhibited a drop in sp-Q_p between generation 10 and 60, the 32-124 cell line exhibited an increase in sp-Q_p (Figure 3.5B). Both cell lines showed similar growth characteristics during the generation 10 and generation 60 FBCs (Figure 3.5C and Figure 3.5D). At generation 10, both cell lines showed a similar IVC (32-121 = 6539 x 10^6 cells.hr/ml, 32-124 = 6757 x 10^6 cells.hr/ml) and maximum VCD (32-121 = 37 x 10^6 cells/ml, 32-124 = 35 x 10^6 cells/ml). At generation 60, both cell lines exhibited a similar drop in IVC (32-121 = 5118 x 10^6 cells.hr/ml, 32-124 = 5644 x 10^6 cells.hr/ml) and maximum VCD (32-121 = 33 x 10^6 cells/ml, 32-124 = 34 x 10^6 cells/ml).

3.5.3 Designing a Study to Validate the Stability of the 32-121 and 32-124 Cell Lines

To validate the stability of the chosen cell lines, the following study was devised (see section 3.4.1 for more details). A vial for each of the 32-121 and 32-124 cell lines was thawed and used to seed 6 flasks per cell line. 3 flasks were maintained in medium containing MTX (Lineages A-C, + MTX) and 3 flasks were maintained in medium without MTX (Lineages A-C, -MTX) (Figure 3.2). The flasks were grown for a period of 60 generations after thaw. FBC screens were conducted at four different generation timepoints for 2 out of the 3 cell lineages (lineages A and B).

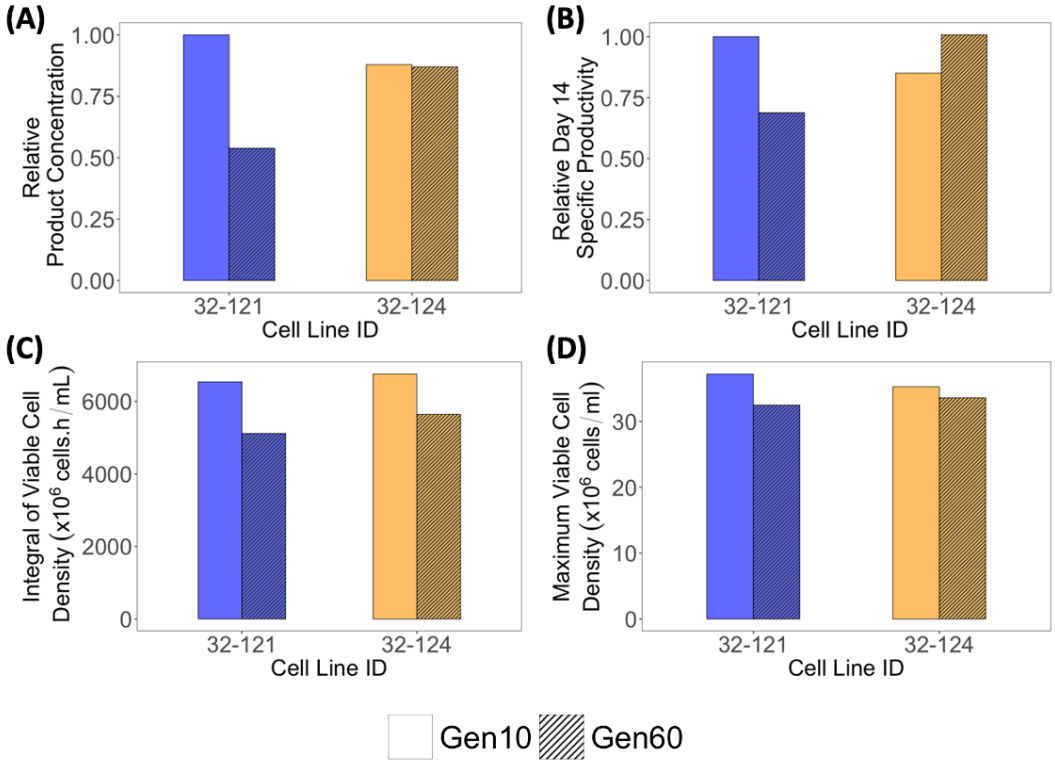


Figure 3.5 - Further Analysis of the 32-121 and 32-124 Cell Lines

Change in the **(A)** relative day 14 product concentration, **(B)** day 14 sp-Q_p, **(C)** IVC and **(D)** maximum cell density of the 32-121 and 32-124 cell lines between generation 10 and 60 during the Ambr®15 bioreactor system stability study. A single replicate was conducted for each cell line.

The data from Figure 3.5 was gathered in 2018 by FDB as part of a previous study, which was conducted prior to the start of James Donaldson's PhD. Data from FDB was re-visited and re-analysed to inform future stability study experiments by James Donaldson.

The timepoints of the FBCs were "early" (approximately 15 generations beyond the RCB), "early-mid" (approximately 25 generations beyond the RCB), "late-mid" (approximately 45 generations beyond the RCB) and "late" (≥ 60 generations beyond the RCB). It is important to highlight that all FBCs were conducted in the absence of MTX, regardless of whether the routine subcultures were conducted in the presence of MTX.

3.5.4 Changes in Growth Rate during Routine Subculture over 60 Generations

The cell lines were maintained under routine subculture, with or without MTX, for >60 generations after thaw. All cell lines reached 60 generations by 21 subcultures (Figure 3.6A). The growth rate of the 32-124 cell line was largely unaffected by the presence or absence of MTX, with all cultures requiring 19 subcultures to reach 60 generations. However, the 32-121 cell line showed a small difference when cultured with or without MTX. Cultures with MTX took 21 subcultures, while the cultures without MTX took 20 subcultures to reach 60 generations.

Previous studies have identified cell lines which exhibit a decrease in Q_p but an increase in growth rate following LTC in the absence of selection (Baik and Lee, 2017; Dorai et al., 2012). This is often attributed to a low producing, fast growing subpopulation of cells outcompeting the rest of the population. To assess how the growth rate of the cell lines changed over 60 generations during routine subculture, the mean generation number increase per subculture of the 2 cell lines for each condition was calculated at different subculture phases (subcultures 3-9/ subcultures 10-16, subcultures 17-23) (Figure 3.6B). During routine subculture, cell density was adjusted to $0.5x10^6$ cells/ml every 3 days (detailed in section 3.4.3). If a culture had a doubling time of 24 hours, the VCD would double each day, resulting in three population doublings over a 3-day period, thus adding three generations to the total generation number. The only cell line that exhibited a generation number increase of less than 3 per subculture (indicating a doubling time longer than 24 hours) was the 32-121 cell line during subcultures 3-9 when cultured with MTX. For both cell lines, MTX addition led to a decrease in the

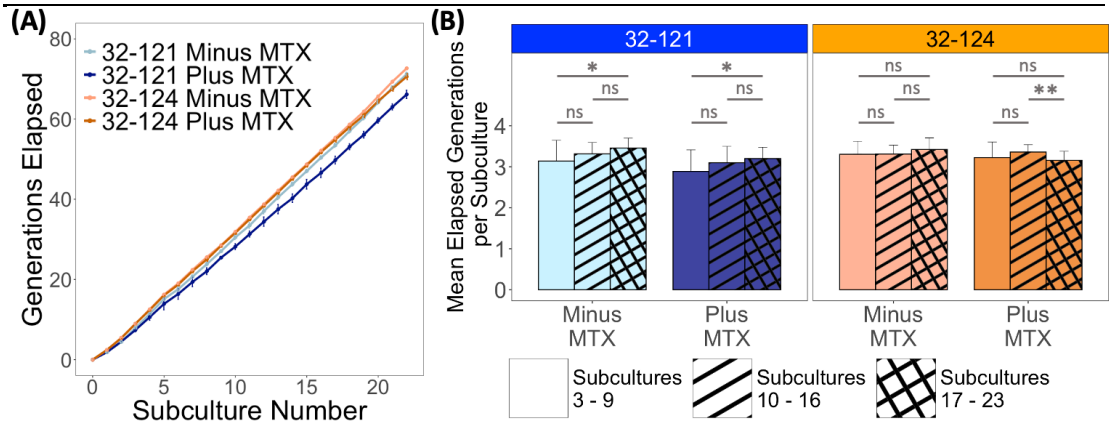


Figure 3.6 - Growth Characteristics of Cell Lines under Routine Subculture conditions over 60 Generations

(A) Increase in generations elapsed with each subculture number. **(B)** Change in mean generations elapsed per subculture for each cell line. Data represented as the mean increase in generation number per subculture + standard deviation of the 3 cell lineages for each condition across 7 subcultures (subcultures 3-9/ subcultures 10-16, subcultures 17-23). $^{ns}P > 0.05$, $*P \leq 0.05$, $**P \leq 0.01$ using ANOVA.

overall mean generation number increase per subculture across the 60 generations in both the 32-121 cell line (+MTX = 3.06, -MTX = 3.30, p = 0.001) and 32-124 cell line (+MTX = 3.25, -MTX = 3.34, p = 0.05). Notably, the mean generation number increase per subculture for the 32-121 cell line (both with and without MTX) increased between subcultures 3-9 and 17-23. However, the 32-124 cell line did not exhibit a significant increase in growth rate between subcultures 3-9 and 17-23.

Cell culture samples were collected at specific subcultures to initiate the FBCs: subculture 5 (early FBC, approximately 15 generations after thaw), subculture 8 (early-mid FBC, approximately 25 generations after thaw), subculture 14 (late-mid FBC, approximately 45 generations after thaw), and subculture 21 (late-mid FBC, approximately 65 generations after thaw). The details of the cell culture sample collection are provided in Table 3.1.

Table 3.1 – Cell culture samples were taken from the routine subculture flasks at specific subculture numbers to seed early, early-mid, late-mid and late FBCs.

Generation	Subculture Number	Fed-Batch Culture Generation Number							
		32-121				32-124			
		- MTX		+ MTX		- MTX		+ MTX	
		Line A	Line B	Line A	Line B	Line A	Line B	Line A	Line B
Early FBC	5	14.5	14.6	12.3	12.7	15.7	16.1	14.4	14.3
Early-Mid FBC	8	24.5	24.7	23.7	22.8	25.8	26.0	25.8	26.0
Late-Mid FBC	14	44.1	43.2	40.3	41.3	45.5	45.7	45.5	45.5
Late FBC	21	68.2	67.3	63.3	63.6	69.1	69.6	68.0	67.8

3.5.5 The 32-121 Cell Line Exceeded the Alert Limit for Cell Line Instability

The primary method for establishing the production stability of a cell line is to compare the day 14 product concentrations from the early and late FBCs (Figure 3.7). The mean day 14 product concentration was significantly higher in the early FBCs (both with and without MTX) for the 32-124 cell line compared with the 32-121 cell line (Figure 3.7A). The 32-124 cell line demonstrated minimal loss of productivity when cultured in the presence of MTX (1%). However, a significant drop in mean day 14 product concentration was observed between the early and late FBCs for the 32-121 cell line (plus and minus MTX) and the 32-124 cell line (minus MTX) (Figure 3.7A). In the absence of MTX, the 32-124 cell line exhibited a 13% drop in mean product concentration over 60 generations. The observed mean percentage drop in product concentration for the 32-121 cell line, both with MTX (24%) and without MTX (27%), exceeded the alert limit for cell line instability (Figure 3.7B). This indicates a substantial decrease in product concentration and raises concerns about the stability of the cell line under the given conditions. Although the production stability of the 32-121 cell line falls into the alert limit

for FDB, it must be noted that according to the commonly used definition of production instability, which defines an unstable cell line as one which fails to retain >70% volumetric productivity over 60 generations, the 32-121 cell line would be characterised as stable (Dahodwala and Lee, 2019).

For both cell lines, the percentage decrease in product concentration between the early and late generation FBCs was greater when cells were grown without MTX, compared with those which has been grown in the presence of MTX (Figure 3.7B). This was unsurprising, given that MTX functions as a selection agent in this context. The expression vector, which contains the HC and LC expression cassettes, and was integrated into Apollo^{TM} X host cells to generate the 32-121 and 32-124 recombinant cell lines, also contains a DHFR transgene. Therefore, DHFR expression should be closely coupled with mAb expression. DHFR is involved in the reduction of dihydrofolate to produce tetrahydrofolate, a vitamin required for the synthesis of hypoxanthine and thymidine. During routine subculture, cell lines are grown in FDB MAP subculture media, which lacks hypothanthine or thymidine. Therefore, only cells that have successfully integrated and expressed the DHFR gene can survive. The addition of MTX, a DHFR inhibitor, adds an extra layer of selection. Cells which lose transgene expression (e.g. via gene loss or epigenetic silencing of the transgene) will likely exhibit a decrease in DHFR expression. A drop in DHFR expression may mean that cells do not produce enough DHFR to overcome the inhibitory effect of MTX. Therefore, cells that are cultured in the presence of MTX should exhibit a smaller drop in productivity during LTC than those which are cultured in the absence of MTX.

Product concentration increased exponentially with time during the FBC (Figure 3.7C). For the 32-121 cell line (plus and minus MTX) and the 32-124 cell line (minus MTX), the difference in product concentration between the early and late FBCs increased with the number of days elapsed in the FBC (Figure 3.7C).

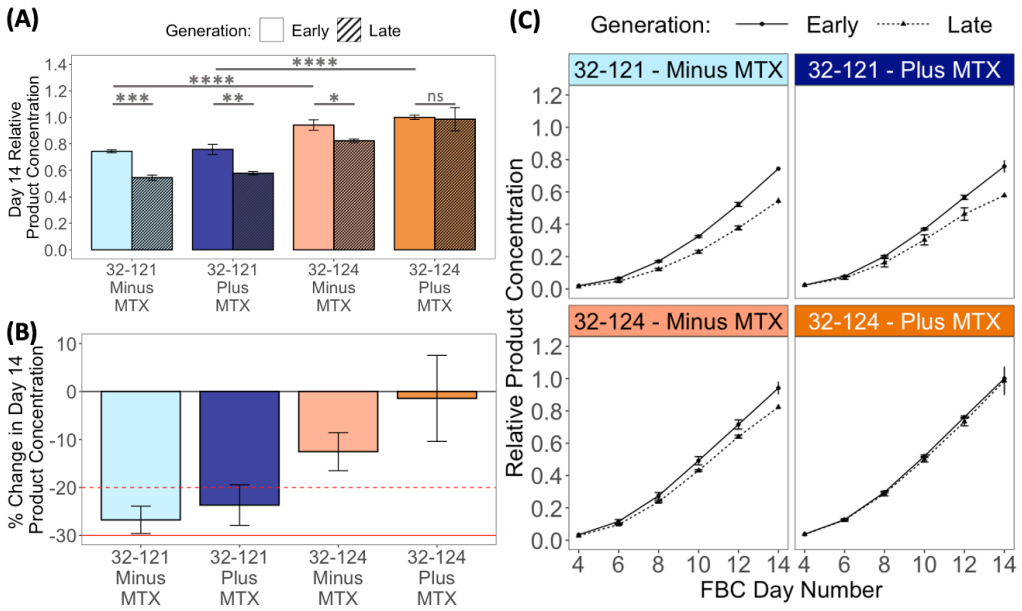


Figure 3.7 - Difference in Product Concentration Between the Early and Late FBC for the 32-121 and 32-124 Cell Lines

(A) Relative day 14 product concentration for the early and late FBCs, for each cell line and condition. Bars represent the mean day 14 product concentration of the two independent lineages performed in duplicate. Error bars represent ± standard deviation of the mean. $^{ns}P > 0.05$, $*P \leq 0.05$, $**P \leq 0.01$, $***P \leq 0.001$, $****P \leq 0.0001$. **(B)** Percentage change in day 14 relative product concentration between the early and late FBCs. Dashed red line (-20%) indicates the alert limit for cell line instability. Solid red line indicates the cut off point for cell line instability. Error bars represent the combined standard deviations of the early and late generation product concentration measurements, calculated using the delta method. **(C)** Line graphs showing change in product concentration over time for each condition. Points represent the mean product concentration of the two independent lineages performed in duplicate. Error bars represent ± standard deviation of the mean.

This observation implies that the productivity of the early generation cell lines remained consistently higher than the late generation cell lines throughout the FBC, and there was no distinct point in the FBC where the two lineages diverged in terms of productivity. The consistent difference in productivity across the entire FBC duration suggests that the measurements taken on day 14 were not isolated occurrences but rather indicative of a sustained productivity difference between the early and late generations in these cell lines.

3.5.6 Linear Regression Models were used to Complement the Analysis of Production Instability

To complement the analysis of the early and late FBCs, the relationship between the generation number of the culture at day 0 of the FBC and the resultant day 14 product concentration of the FBC was assessed using linear regression models (Figure 3.8A). This analysis aimed to determine whether productivity changes over time occurred suddenly or exhibited a slow, gradual decrease.

Consistent with the previous analysis (see section 3.5.5), the 32-124 cell line (plus MTX) did not demonstrate a significant relationship between generation number and day 14 product concentration (at the 95% confidence level). In contrast, for the 32-121 cell line (plus and minus MTX) and the 32-124 cell line (minus MTX), a statistically significant linear relationship (at the 95% confidence level) was observed between day 14 product concentration and generation number. However, the slope of the regression lines for the 32-121 cell line (-0.004 for both with and without MTX) was greater than the slope of the regression line for the 32-124 cell line (minus MTX), indicating a faster decline in productivity for the 32-121 cell line. Using the regression lines to measure the loss of productivity between generation 5 and 65, the percentage drop in product concentration for the 32-121 cell line exceeded the alert limit for cell line instability (Figure 3.8B).

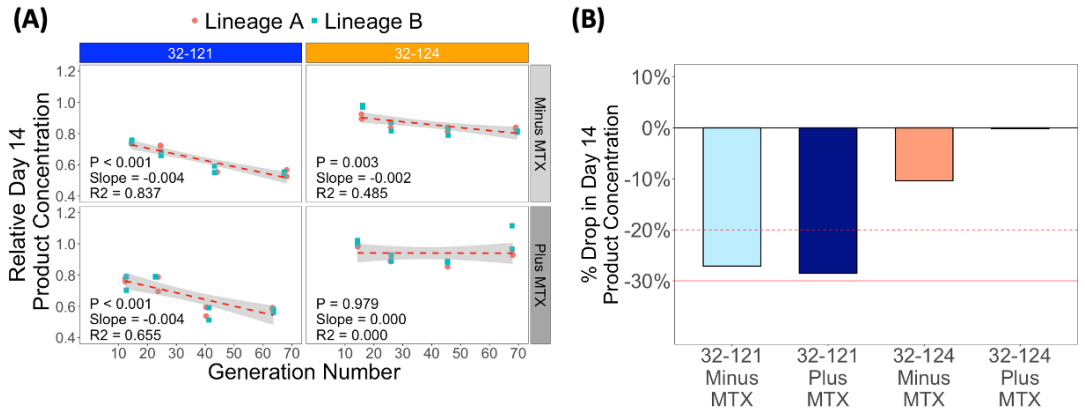


Figure 3.8 - Change in day 14 Product Concentration over 60 generations for Regression Line Analysis of Production Stability

(A) The day 14 product concentration from early, early-mid, late-mid, and late FBCs for each condition and lineage. The generation number is derived from the culture generation at day 0 of the FBC. Linear regression models were generated to represent the relationship between generation number and product concentration, and were plotted as dashed red lines. The shaded area indicates the 95% confidence interval for predictions derived from the linear model. **(B)** The drop in product concentration between generation 15 and 60, as predicted by the linear regression model. Dashed red line (-20%) indicates the alert limit for cell line instability. Solid red line indicates the cut off point for cell line instability.

Despite the statistical significance, the R^2 values of the linear models were relatively low for the 32-121 cell line (plus and minus MTX) and the 32-124 cell line (minus MTX) (0.655, 0.837 and 0.485 respectively) (Figure 3.8A). This suggests that the relationship between generation number and product concentration may not follow a strict linear pattern. Notably, a distinct pattern emerged in the 32-121 cell line, where there was a clear differentiation between the day 14 product concentration in the early and early-mid FBCs compared with the late-mid and late FBCs. This observation implies that the

change in productivity mainly occurs during the period between the early-mid and late-mid FBCs. Moreover, an unexpected dip in day 14 FBC product concentration was observed between the early/ late FBCs and the middle generation (early-mid and late-mid) FBCs in the 32-124 cell line (both with and without MTX), which likely influenced the R2 values. Additionally, the product concentration of the late-mid FBC was lower than that of the late generation FBC for the 32-121 cell line (plus MTX). Such a phenomenon is unusual, as cell lines typically do not recover productivity after experiencing a drop. Given that the early and late generation FBCs are conducted concurrently, it was thought that something had happened during the early-mid or late-mid generation FBCs that caused a reduction in productivity. This is investigated further in section 3.5.7.

3.5.7 The Relationship Between Changing Growth Characteristics and Productivity over 60 Generations

After assessing the production stability of the cell lines over 60 generations under different conditions (with and without MTX selection), the focus shifted towards identifying the factors which contributed to the decline in productivity that the 32-121 cell line exhibited. Growth rate is a fundamental parameter that directly impacts the overall productivity of a cell line during the FBC process.

Daily measurements of VCD and viability were conducted after day 2, enabling growth characteristics to be assessed across the early, early-mid, late-mid and late generation FBCs (Figure 3.9A). For each generation and condition, the 32-124 cell line maintained a viability >75% for the duration of the FBCs. In contrast, the viability of the 32-121 cell line dropped more rapidly during late phase culture, with most cultures falling below 70% viability by day 14.

From the growth curves, the maximum VCD was identified (Figure 3.9B) and the IVC was calculated (Figure 3.9C) for each cell line and condition. Surprisingly, no significant differences were observed in the maximum VCD

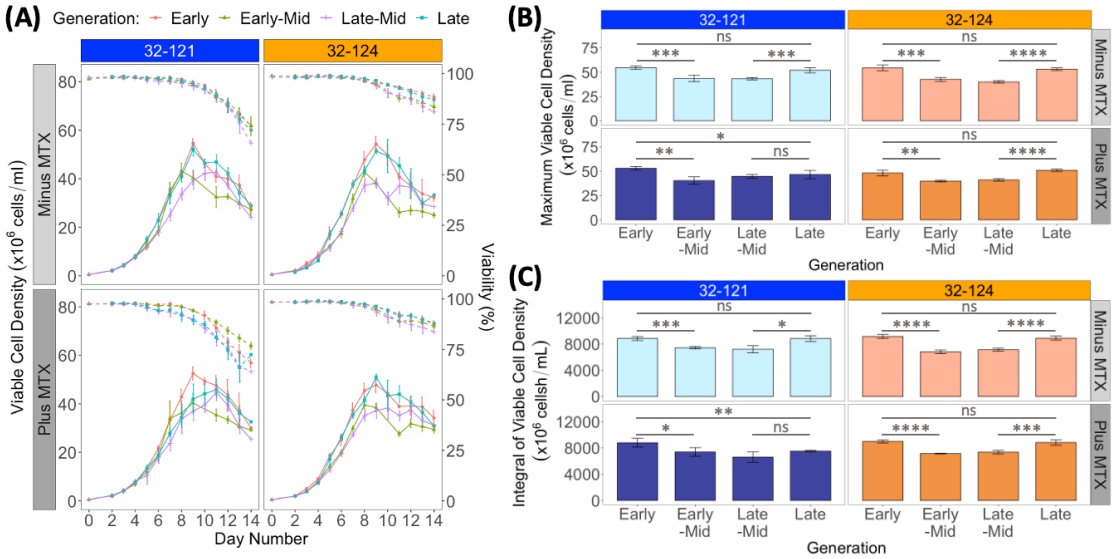


Figure 3.9 - Change in Growth Characteristics over 60 generations

(A) Early, early-mid, late-mid and late generation FBC growth curves for each condition and cell line. Dashed lines indicate change in percentage viability over the 14-day FBC. Solid lines indicate VCD over the 14-Day FBC. The **(B)** maximum VCD and **(C)** IVC for the early, early-mid, late-mid and late generation FBCs for each cell line and condition. Bars represent the mean of the two independent lineages performed in duplicate. Error bars represent ± standard deviation of the mean. $^{ns}P > 0.05$, *P ≤ 0.05, **P ≤ 0.01, ***P ≤ 0.001, ****P ≤ 0.0001 using ANOVA.

or IVC between the early and late generation FBCs for the 32-121 cell line (minus MTX) and the 32-124 cell line (plus and minus MTX). This contrasted with the growth data obtained from routine subculture, during which the 32-121 cell line exhibited a significant increase in growth rate when cultured without MTX over 60 generations. The growth characteristics of a cell line during FBC can be influenced by various factors, including the higher cell densities in FBC compared with routine subculture and the accumulation of toxic metabolites such as lactate. Changes that occur in the cell line during LTC may influence its ability to cope with such stresses, potentially counteracting the effects of an increasing growth rate during routine

subculture. Additionally, these results imply that the changes in product concentration observed in the 32-121 and 32-124 cell lines, when grown for 60 generations without MTX, were primarily influenced by changes in the average productivity per cell rather than decreases in growth rate.

In contrast, the 32-121 cell line (when cultured with MTX) exhibited a significant drop in both maximum VCD ($p < 0.05$) and IVC ($p < 0.01$). This result highlights the potential impact of MTX selection on cell growth characteristics and suggests that the change in growth profile of this cell line may have contributed to the reduction in mAb productivity over 60 generations.

An additional observation was the significant reduction in both the maximum VCD (Figure 3.9B) and IVC (Figure 3.9C) from the early/ late FBCs to the early-mid/ late-mid FBCs for the majority of cell lines and selection conditions. As discussed in section 3.4.1, the early and late FBCs were conducted concurrently in the same incubator, using the same batch of Cell Boost 7a, Cell Boost 7b and JM-05B. However, due to incubator space constraints, the early-mid and late-mid generations FBCs were conducted in a different incubator. Despite the incubators being set to the same settings, it is possible that subtle differences between the two incubators may have led to a drop in IVC and maximum VCD during the middle generation FBCs. Hence, it is plausible that this additional variable contributed to the variance in product concentration and the "U" shape pattern observed in the product concentration versus generation number graph of the 32-124 cell line (both with and without MTX) (Figure 3.8). Therefore, further analysis of any changes in cell culture attributes over the 60-generation period must take this into consideration.

3.5.8 The Drop in Q_p-Rate for the 32-121 Cell Line (Minus MTX) exceeded the Alert Limit for Cell Line Instability

As previously discussed, it was thought that the drop in the observed day 14 product concentration for the 32-121 cell line over 60 generations, when

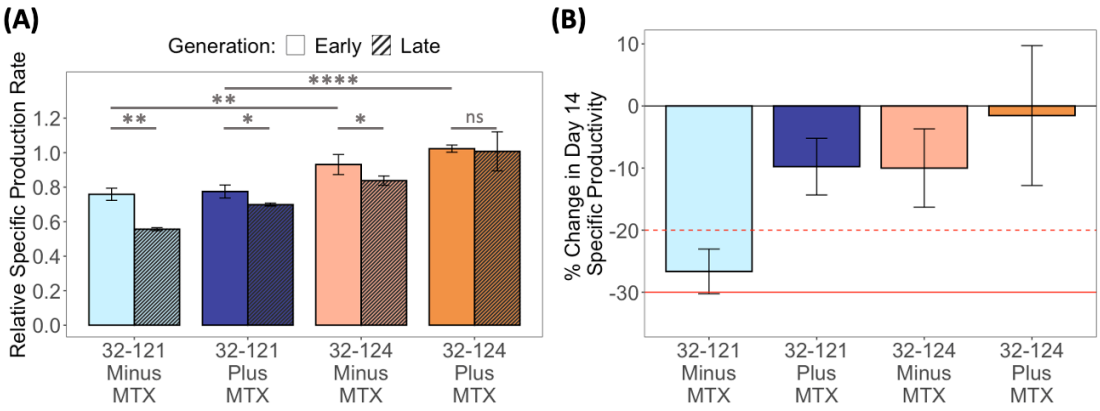


Figure 3.10 - Difference in Q_p- Rate between the Early and Late FBC for the 32-121 and 32-124 Cell Lines

(A) Relative Q_p-rate for the early and late FBCs, for each cell line and condition. Bars represent the mean Q_p-rate of the two independent lineages performed in duplicate. Error bars represent ± standard deviation of the mean. $^{ns}P > 0.05$, *P ≤ 0.05, **P ≤ 0.01, ***P ≤ 0.001, ****P ≤ 0.0001. **(B)** Percentage change in day 14 relative product concentration between the early and late FBCs. Dashed red line (-20%) indicates the alert limit for cell line instability. Solid red line (-30%) indicates the cut off point for cell line instability. Error bars represent the combined standard deviation of the early and late product concentration measurements, calculated using the delta method.

cultured without MTX, was primarily due to a drop in the average productivity per cell, rather than changing growth characteristics. To test this, the Q_p-rate of the 14-day FBCs was determined through linear regression analysis of product concentration against IVC (detailed in section 3.4.7.4). A comparison was subsequently made between the Q_p-rate of the early and late FBCs (Figure 3.10). Similar to the product concentration data, a significant decrease in mean day 14 Q_p-rate was observed between the early and late FBCs for the 32-121 cell line (plus and minus MTX) and the 32-124 cell line (minus MTX) (Figure 3.10A). In agreement with the initial hypothesis that the loss of productivity was due to a drop in the average productivity per cell rather than changes in growth rate, the mean percentage drop in Q_p-rate for

the 32-121 cell line without MTX (27%) exceeded the alert limit for cell line instability (Figure 3.10B). However, for the 32-121 cell line cultured with MTX, the drop in Q_p-rate was only 10%, comparable to the decrease observed in the 32-124 cell line (minus MTX) (Figure 3.10B). Taken together with the growth data for the 32-121 cell line (plus MTX), loss of productivity is likely to be due to a combination of changes in growth profile and Q_p-rate.

Additional comparisons can be made between the results of the flask-based stability study and the initial Ambr®15 bioreactor data. Consistent with the product concentration results of the flask-based stability study, the mean Q_p-rate was significantly higher in the early FBCs (both with and without MTX) for the 32-124 cell line compared with the 32-121 cell line (Figure 3.10A). Again, this was in contrast with the results from the previous Ambr®15 bioreactor data (Figure 3.5B), where the 32-121 cell line exhibited a higher day 14 sp-Q_p during the generation 10 FBC than the 32-124 cell line. The 32-124 cell line exhibited an increase in sp-Q_p from generation 10 to 60 during the Ambr®15 bioreactor study (Figure 3.5B). However, contrary to these findings, the flask-based study did not show a similar increase in Q_p-rate for the 32-124 cell line (with or without MTX) (Figure 3.10B).

Regression line models were also used to assess the relationship between generation number and Q_p-rate (Figure 3.11). For the 32-121 cell line (minus MTX), a statistically significant linear relationship (at the 95% confidence level) was observed between Q_p-rate and generation number (Figure 3.11A). Using the regression lines to measure change in Q_p-rate between generation 5 and 65, the percentage drop in Q_p-rate for the 32-121 cell line (minus MTX) exceeded the alert limit for instability (Figure 3.11B). However, despite the statistical significance, the R^2 value of the linear model was also relatively low (0.698), suggesting that factors other than the generation number (such as changing culture conditions disrupting cell growth for the middle generation FBCs) contributed to the variance in Q_p-rate.

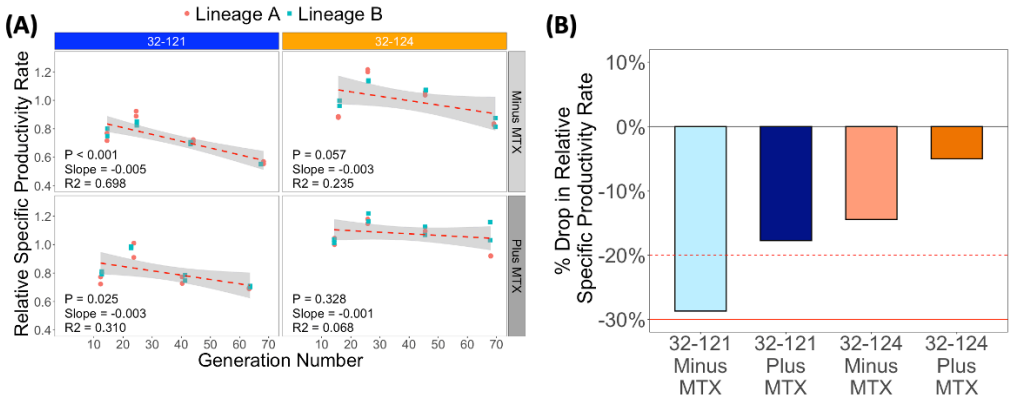


Figure 3.11 - Regression line analysis for the change in Q_p-Rate over 60 generations

(A) The **Q_p**-rate from early, early-mid, late-mid, and late FBCs for each condition (+/- MTX) and cell line (32-121/ 32-124). The generation number is derived from the culture generation at day 0 of the FBC. Linear regression models were generated to represent the relationship between generation number and Q_p-rate, and were plotted as dashed red lines. The shaded area indicates the 95% confidence interval for predictions derived from the linear model. (B) The drop in Q_p-rate between generation 15 and 60, as predicted by the linear regression model. Dashed red line (-20%) indicates the alert limit for cell line instability. Solid red line indicates the cut off point for cell line instability.

3.5.9 Comparing Flask FBC Results with the initial Ambr®15 Study

The results of the flask FBCs were also compared with the results from FDB's initial Ambr®15 study (Figure 3.5A). As previously noted (section 3.5.1), the Ambr®15 study was only conducted on cell lines that had been grown in the presence of MTX. Therefore, comparisons were only made with the flask FBC data from cell lines that had been cultured with MTX. It must be noted that the Q_p measurements are represented as a sp-Q_p for the Ambr®15 study and Q_p-rate for the flask study. Given that changes in Q_p were being assessed rather than absolute values, this was deemed acceptable. The comparison between the two studies is summarised in Table 3.2.

Table 3.2 – Comparison between the initial Ambr®15 and flask FBC study

Cell Line	Day 14 Product Concentration		Specific Productivity		Maximum VCD		IVC	
	Ambr ®15	Flask FBC	Ambr ®15	Flask FBC	Ambr ®15	Flask FBC	Ambr ®15	Flask FBC
32-121	-46%	-24%	-31%	-10%	-13%	-12%	-22%	-15%
32-124	-1%	-1%	+18%	-2%	-5%	-6%	-16%	-2%

The table provides a summary of the percentage changes in mean day 14 product concentration between the early and late generation FBCs, as assessed by the initial Ambr®15 and flask FBC studies.

The 32-124 cell line exhibited a similar drop in product concentration between the early and late generation FBCs for both the Ambr®15 and flask studies, indicating minimal productivity loss during LTC. However, some differences between the two studies were observed. Notably, the 32-124 cell line exhibited an 18% increase in sp-Q_p during the Ambr®15 study, but a 2% decrease in Q_p-rate during the flask study. Increases in Q_p during LTC are unusual and thus the Ambr®15 results were surprising. In both studies the 32-124 cell line exhibited a modest drop in maximum VCD (5% for the Ambr®15 study and 6% for the flask study) between the early and late generation FBCs, suggesting that the cell line exhibited only small changes in growth characteristics during LTC. However, the 32-124 cell line exhibited a much greater decrease in IVC (16%) in the Ambr®15 study between the early and late generations runs, compared with the flask study (2%). This large drop in IVC may have contributed to the increase in the calculated sp-Q_p (calculated by dividing the day 14 product concentration by the day 14

IVC) for the Ambr®15 run, observed during the late generation FBC. It is plausible that this notable reduction in IVC was influenced by errors in cell counting, particularly on days when the maximum VCD was not reached, during the late generation Ambr®15 study of the 32-124 cell line. The high cell concentrations reached during FBCs, coupled with the need for sample dilution for accurate cell counting, may have introduced variability in the daily cell counts, resulting in the significantly lower IVC observed in the late generation Ambr®15 study compared to the early generation study.

For the 32-121 cell line, a much larger drop in both day 14 product concentration (Ambr®15 = - 46%, Flask = -24%) and Q_p (Ambr®15 = 31%, Flask = -10%) was observed between the early and late generation FBCs in the Ambr®15 study than in the flask study. This distinction is important because, according to the conventional definition of production instability, which labels cell lines with a product concentration decrease exceeding 30% over 60 generations as unstable, the 32-121 cell line would be defined as unstable in the Ambr®15 study and stable in the flask study. The 32-121 cell line exhibited a relatively similar percentage decrease in maximum VCD between the early and late generation runs for both the Ambr®15 (-13%) and flask study (-12%). Although there was a slightly greater decrease in IVC in the Ambr®15 study (-22%) compared with the flask study (-15%) over 60 generations, the differences in product concentration results between the two studies are likely to be due to variations in productivity per cell rather than growth characteristics. Given the complexity of mammalian cell genome, and the many factors which can impact gene expression (e.g. transcription, translation, secretion etc), it is plausible for the same clonal cell line to exhibit distinct productivity characteristics during different LTC experiments, even with slight disparities in culture conditions. However, due to the variability associated with the results from the Ambr®15 study, and the lack of repeats, it was decided that the results from the flask study should be used as the basis for further experiments.

3.5.10 Drop in sp-Q_p was not Specific to the Early or Late Phase for the 32-121 Cell Line

In a previous study by Bailey et al. (2012), a suspension GS-CHO cell line displayed a 40% decline in productivity over 60 generations, but no significant reduction in the IVC. As expected, the Q_p of the cell line decreased by ~30% over a period of 60 generations. Notably, when analysing Q_p in the exponential phase of batch culture (days 0-7), the decline was relatively modest (~13% decrease). However, during the decline phase of batch culture (days 9-15), Q_p dropped by around 50%. It was anticipated that similar investigations into phase-specific changes in Q_p could yield further insights into potential sources of productivity loss in the 32-121 cell line. For example, in *Bailey et al. (2012)*, the drop in Q_p during late phase culture was thought to be a result of an increased sensitivity to cellular stress, exemplified by increased mRNA expression of the stress inducible gene GADD153.

The average sp-Q_ps for the early (days 4-8) and late (days 10-14) phases were calculated and compared for the early and late FBCs, following the method described in section 4.4.6.3 (Figure 3.12). For the 32-121 cell line (plus and minus MTX), there was a significant drop in sp-Q_p between the early and late generation FBCs for both the early and late phases, suggesting that the source of the decreases in sp-Q_p was not phase-specific (Figure 3.12A). Specifically, the 32-121 cell line without MTX exhibited a drop of 28% in sp-Q_p during the exponential phase and 27% during the stationary/ death phase (Figure 3.12B). The 32-121 cell line with MTX showed a decrease of 10% in sp-Q_p during the exponential phase and 8% during the stationary/ death phase (Figure 3.12B).

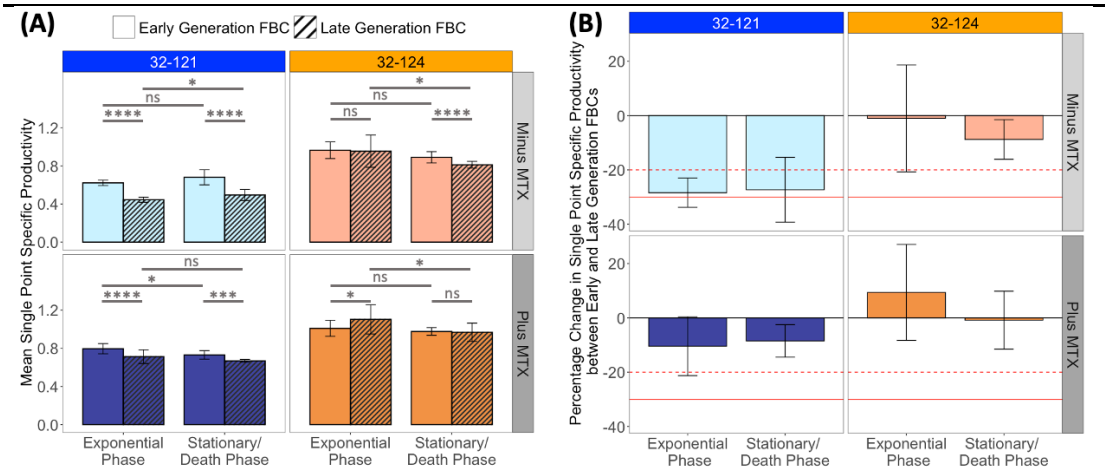


Figure 3.12 - Comparison of phase-specific sp-Q_p

(A) Comparison of mean sp-Q_p for the exponential phase vs stationary/ death phase for each cell line and condition during the early and late generation FBCs. The mean sp-Q_p was calculated for each day of the FBC that a product concentration sample was taken (day 4, 6, 8, 10, 12 and 14). The mean sp-Q_p for the exponential phase (days 4-8) and stationary/ death phase (days 10-14) for each cell line and condition were subsequently calculated for the early and late FBCs. Error bars represent ± standard deviation of the mean. $^{ns}P > 0.05$, $*P \leq 0.05$, $**P \leq 0.01$, $***P \leq 0.001$, **** $P \leq 0.0001$. (B) Percentage change in sp-Q_p between early and late generations for the exponential and stationary/ death phases. Dashed red line (-20%) indicates the alert limit for cell line instability. Solid red line (-30%) indicates the cut off point for cell line instability. Error bars represent the combined standard deviation of the early generation and late generation sp-Q_p measurements, calculated using the delta method.

In contrast, changes in sp-Q_p between the early and late generation FBCs in the 32-124 cell line (minus MTX) displayed a phase-specific aspect. For the 32-124 cell line (minus MTX), there was a significant change in sp-Q_p between the early and late generation FBCs at the stationary phase (9%), but not during the exponential phase. However, for the 32-124 cell line (plus MTX) there was no significant decrease in sp-Q_p for either phase between the early and late generation FBCs.

3.5.11 Analysis of FBC Metabolite Data for the 32-121 and 32-124 Cell Lines

Previous studies have demonstrated that the metabolism of CHO cells can undergo significant alterations during LTC, potentially influencing cell growth and productivity (Torres et al., 2023). To investigate these changes, the extracellular concentrations of glucose, lactate, glutamine, and glutamate were monitored daily during the FBCs of the 32-121 and 32-124 cell lines (Figure 3.13). It must be noted that the daily metabolite measurements were conducted before the culture was supplemented with the proprietary feeding regime that was conducted daily.

The extracellular concentrations of the metabolites were compared with the growth curves of the FBCs. As discussed previously (see section 3.5.7), the 32-121 cell line (minus MTX) and the 32-124 cell line (both plus and minus MTX) exhibited a very similar growth profile over the 14-day FBC at an early and late generation (Figure 3.13A). However, the 32-121 cell line, when cultured with MTX, exhibited a drop in the maximum VCD between the early and late generation FBCs (Figure 3.13A). In agreement with this, the 32-121 cell line, when cultured with MTX, started to decrease in viability from day 5 in the late generation FBC, whereas viability started to fall from day 8 in the early generation FBC (Figure 3.13B).

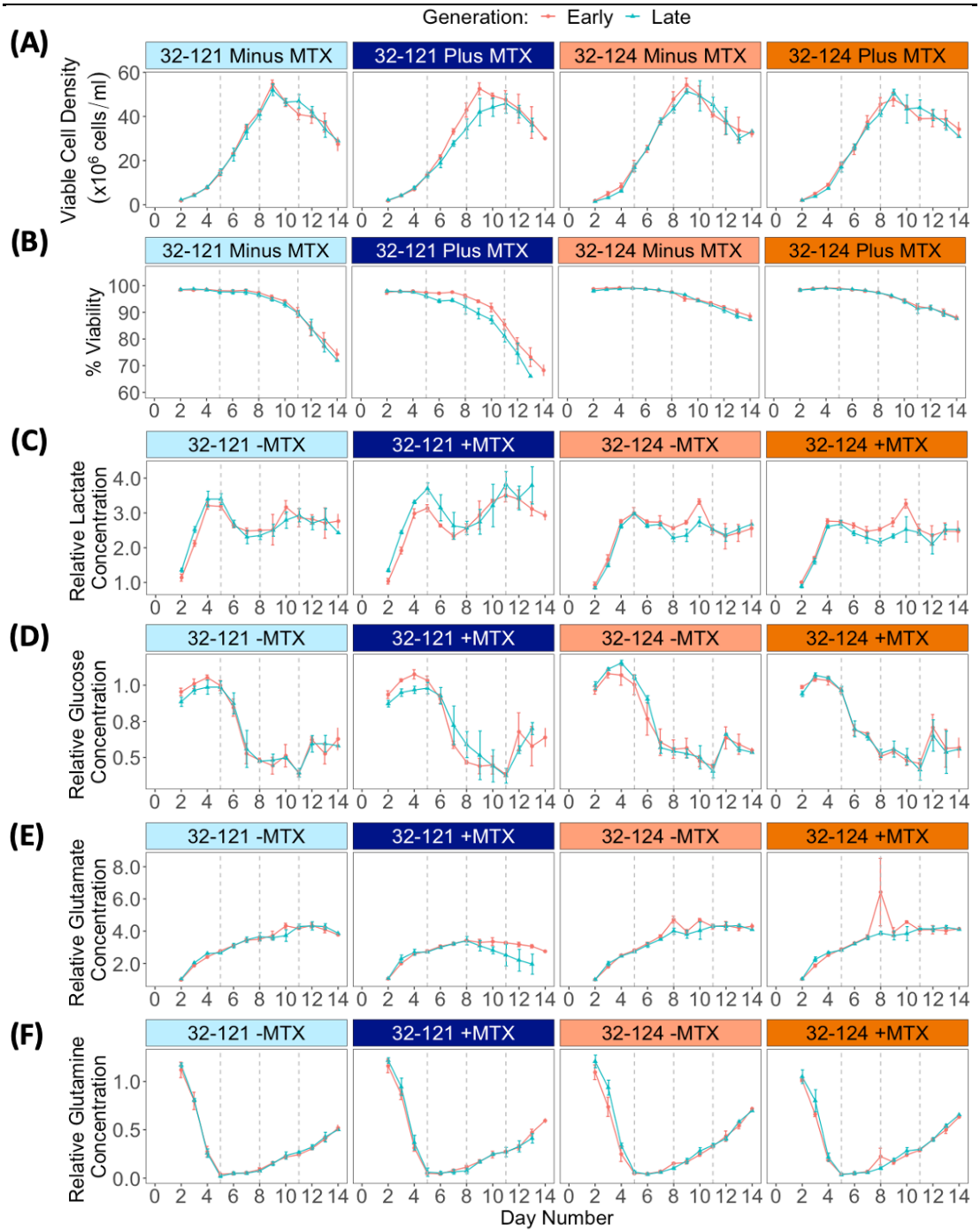


Figure 3.13 - Metabolite analysis of the 32-124 and 32-121 cell lines

Change in **(A)** VCD and **(B)** % viability over 14 days for the early and late generation FBCs for each cell line. From day 2 onwards, extracellular **(C)** lactate, **(D)** glucose, **(E)** glutamine and **(F)** glutamate levels were determined. Data points represent relative values relative to the respective metabolite measurements from

(continued on next page)

(Figure 3.13 continued) day 2 of the early generation 32-124 Plus MTX Lineage A (Replicate 1) cell line FBC. Error bars represent ± standard deviation of the mean obtained from 4 measurements obtained from 2 lineages, each with 2 technical replicates.

The first 8 days of all FBCs were characterised by exponential growth (Figure 3.13A). During the first 5 days of the FBCs, the concentration of lactate increased rapidly, and the concentration of glutamine decreased rapidly, in all cell lines (plus and minus MTX) for both the early and late generation FBCs. In contrast, the concentration of glucose stayed relatively constant at approximately 5 g/L, indicating that the feeding strategy was relatively effective at maintaining glucose concentrations. Interestingly, lactate peaked at a significantly higher concentration ($p<0.001$) on day 5 in the late generation FBC than the early generation FBC for the 32-121 cell line (plus MTX). This increase in peak lactate concentration corresponded with the start of the viability decrease. The 32-121 cell line (minus MTX) exhibited a small increase in day 5 lactate concentration for the late generation FBC compared with the early generation FBC, but this increase was not significant at the 95% confidence interval ($p = 0.07$).

After day 5, glucose levels started to decrease in all cell lines (both with and without MTX). In the 32-121 cell line (both with and without MTX), lactate levels also started to decrease after day 5. The switch from lactate production to lactate consumption following glutamine depletion (even when glucose remains present in the media) has been observed previously (Ghorbaniaghdam et al., 2014; Nolan and Lee, 2011; Wahrheit et al., 2014). However, lactate concentration stayed relatively stable in the 32-124 cell line (both with and without MTX) after day 5. The glutamine concentration profile was almost identical for the early and late generation FBCs for both cell lines (plus and minus MTX) over the course of the 14-day period. Glutamine levels gradually increased after day 5 as a result of feeding. This was thought to have led to an excess of glutamine in the media.

Growth rate slowed down in all cell lines (plus and minus MTX) after day 8. Between days 8 and 11, the concentration of lactate increased again in the 32-121 cell line (plus MTX) FBC, likely as a result of increased cell death (Pan et al., 2017). However, the second lactate concentration peak (measured on day 11) was not significantly higher for the “32-121 plus MTX” FBC than the “32-121 Minus MTX” FBC ($p = 0.18$). Glutamine concentration increased gradually after day 8 in all cell lines (plus and minus MTX).

Interestingly, glutamate was consumed at a faster rate by the 32-121 cell line (plus MTX) in the late generation FBC than the early generation FBC. In contrast, the 32-121 (minus MTX) and 32-124 cell lines (plus and minus MTX) exhibited a very similar glutamate concentration profile during the early and late generation FBCs.

3.5.12 Design and Setup of Gene Copy Number Experiments using Real-Time Quantitative PCR

Data from the FBCs had shown a significant drop in day 14 product concentration between the early and late generation FBCs in the 32-121 cell line (both with and without MTX). By analysing changes in Q_p and growth characteristics between the early and late generation FBCs, it was suggested that the loss of productivity in the 32-121 cell line (minus MTX) was due to a drop in the average productivity per cell, rather than a change in growth characteristics. In contrast, the loss of productivity exhibited by the 32-121 cell line (plus MTX) was thought to be due to a combination of a decrease in the average productivity per cell and a change in growth attributes (e.g. lower maximum VCD and lower IVC). In agreement with this, the 32-124 cell line (minus MTX), which exhibited a very similar drop in Q_p-rate (~10%) and no significant change in growth characteristics between the early and late generation FBCs, only exhibited a drop in day 14 product concentration of ~12% between the early and late generation FBCs.

To elucidate whether loss of transgene copies was an underlying cause of productivity loss in the 32-121 cell line, quantitative PCR (qPCR)

experiments were performed on the genomic DNA of the cell lines, to determine whether there were any changes in gene copy number during LTC. Given that the primary interest of the experiment was to assess changes in gene copy number, and the high number of samples that needed to be analysed, it was decided that the delta-CT method should be used to quantify gene copy number relative to 1 or 2 reference genes (ideally with known copy numbers), rather than attempting to quantify absolute gene copy number.

ActB, encoding the cytoskeletal protein β-Actin, β2-microglobulin, encoding a molecular chaperone, and Cog1, encoding a subunit of the conserved oligomeric Golgi (COG) protein complex, are commonly used reference genes in CHO cell gene copy number experiments and were thus initially chosen as reference genes for this experiment. At the time of the study, the genomes of the Apollo™ X cell line and a recombinant Apollo™ X-derived 222-107 cell line (expressing a mAb) had been partially sequenced but not fully assembled. To estimate gene copy numbers of these three reference genes, BLAST searches against the Apollo™ X host and 222-107 genomes were conducted by Edinburgh Genomics. The BLAST results suggested that B2M, ActB and Cog1 were located at a single position in the Apollo™ X host genome, indicating a copy number of 1. In the 222-107 cell line, Cog1 and B2M were also found at a single position in the genome but ActB was detected at two locations. However, based on the locations at which ActB annotations were found, it was suspected that the second location was on a sequence that is incorrectly assembled. As a fully assembled genome was not available at this point, it was difficult to fully determine the copy numbers of each reference gene. However, it was assumed that these reference genes would likely retain relatively stable copy numbers during LTC, given their common usage in such studies. As a result, they could be used to assess the changes in relative transgene copy number over time. For each of the selected reference genes, 3 primer sets were designed to amplify a 200bp region of the coding sequence of the gene, covering an intron-exon boundary to avoid amplification of any contaminating RNA. Primer sets

amplifying different 200bp regions of the DHFR, heavy chain and light chain transgenes (which do not contain introns) were also designed. Cell culture samples of 5 x 10^6 cells were taken at different generation timepoints and the genomic DNA was extracted. Standard PCR reactions, using the OneTaq Polymerase, were conducted on a genomic DNA sample with the different primer sets and the resultant PCR products were run on an agarose gel to confirm that successful amplification had occurred and that single 200 bp PCR products were being generated. 2 primer sets, which generated a single 200 bp amplicon, were selected for each gene target (Figure 3.14A and C). Given that a single 200 bp amplicon was successfully produced for the ActB and Cog1 genes, and that all primer sets for the B2M gene failed to generate a single 200 bp amplicon, B2M was dropped as a potential reference gene. The amplification efficiencies of the primer sets were calculated by determining the CT at serial dilutions of genomic DNA (Figure 3.14B and D). A single primer set, with an appropriate amplification efficiency (>90% and < 110%) and high R^2-value (>98%) was subsequently selected for each gene target.

3.5.13 Gene Copy Number Analysis Suggests the Loss of Transgene Copies was Unlikely to be the Primary Cause of Productivity Loss in the 32-121 Cell Line

Relative copy numbers of the DHFR, heavy chain and light chain transgenes were subsequently assessed using Cog1 (Figure 3.15) and ActB (Figure 3.16) as reference genes. For the 32-121 cell line (minus MTX), the linear regression model did not indicate a significant relationship between generation number and DHFR, heavy chain or light chain transgene copy number (for both lineages A and B) at the 95% confidence level when using either Cog1 or ActB as a reference gene. However, at the 90% confidence level, a significant decrease in DHFR transgene copy number was observed over 60 generations for lineage B when using Cog1 as a reference gene. However, this relationship was not significant at the 90% confidence level for lineage A, or when using ActB as a reference gene for either lineage A or B.

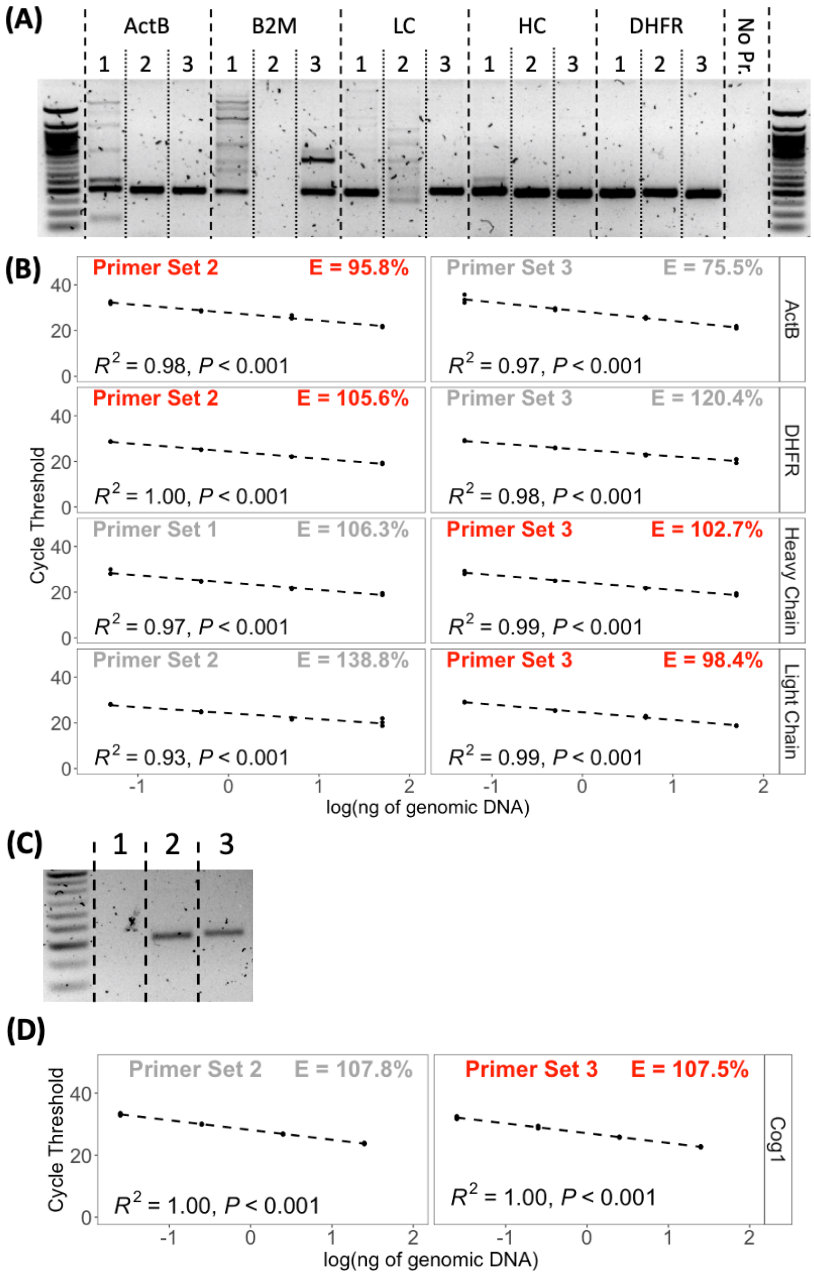


Figure 3.14 - Selection of Primers for Gene Copy Number Analysis

(A) Primer sets for the genes ActB, B2M, DHFR, Heavy chain, and light chain were screened for their amplification specificity. A standard PCR reaction was performed, and the resulting PCR products were separated on an agarose gel. The presence of a single band at 200bp indicated successful primer specificity. **(B)** From the previous screening, two primer sets were chosen for each target gene (ActB, DHFR, Heavy chain, and light chain).

(Continued on next page)

(Figure 3.14 continued) The primer efficiency of these selected primer sets was assessed by preparing serial dilutions of genomic DNA and subjecting them to RT-qPCR. The CT values were measured, and primer efficiency was calculated by determining the slope of the regression line for Ct values versus the logarithm of genomic DNA concentration. The primer sets selected for further use are highlighted in red **(C)** The same method as (A) but for Cog1. **(D)** The same method as (B) but for Cog1. The primer set selected for further use is highlighted in red.

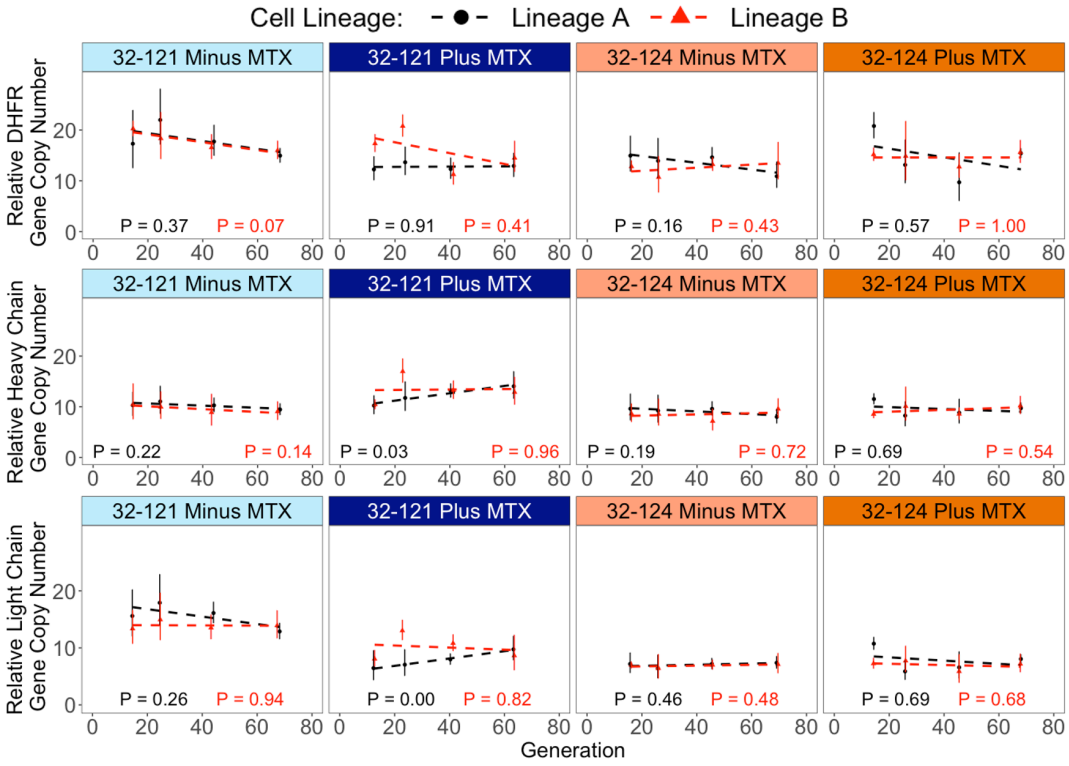


Figure 3.15 - Gene Copy Number Analysis Using Cog1 as a Reference Gene

Relative gene copy number analysis performed by RT-qPCR. The gene copy number for each target gene (DHFR, heavy chain, and light chain) was normalised to the gene copy number of the reference gene, Cog1, calculated using the delta CT method. The data points represent the mean $2^{-\Delta Ct}$ values (± standard deviation) calculated from triplicate measurements for each lineage.

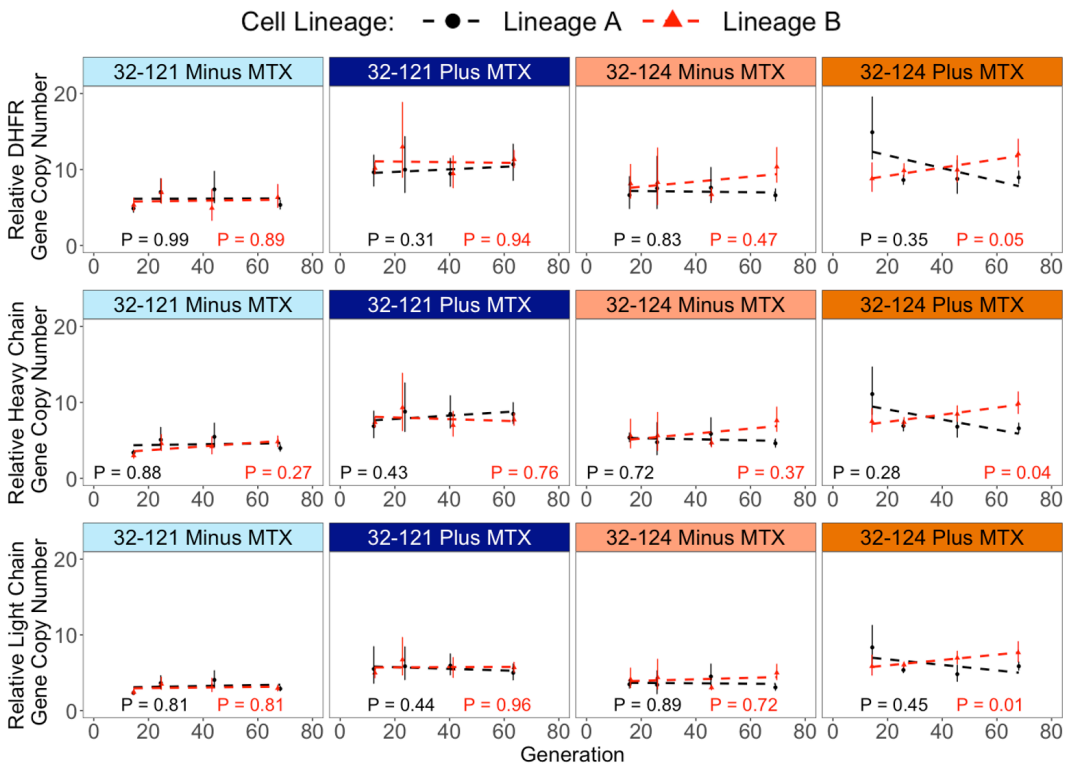


Figure 3.16 - Gene Copy Number Analysis Using ActB as a Reference Gene

Relative gene copy number analysis performed by RT-qPCR. The gene copy number for each target gene (DHFR, heavy chain, and light chain) was normalised to the gene copy number of the reference gene, ActB, calculated using the delta CT method. The data points represent the mean $2^{-\Delta Ct}$ values (± standard deviation) calculated from triplicate measurements for each lineage.

For the 32-121 cell line (plus MTX), the linear regression model revealed a significant positive relationship between both heavy chain and light chain transgene copy number (for lineage A) and generation number at the 95% confidence level when using Cog1 as a reference gene, suggesting transgene amplification in the cell line. However, the relationship between DHFR transgene copy number (for lineage A) and generation number was not significant at the 95% confidence level. Additionally, the linear regression model did not indicate a significant relationship between generation number and DHFR, heavy chain, or light chain transgene copy number for both lineages A and B at the 95% confidence level when using ActB as a reference gene.

For the 32-124 cell line (minus MTX), the linear regression model did not show a significant relationship between generation and the number of DHFR, heavy chain, or light chain transgene copy numbers for both lineages A and B at the 95% confidence level when using either Cog1 or ActB as a reference gene. For the 32-124 cell line (plus MTX), the linear regression model revealed a significant positive relationship between DHFR, heavy chain, and light chain transgene copy number (for lineage B) and generation number at the 95% confidence level when using ActB as a reference gene, suggesting transgene amplification in the cell line. However, the same relationships were not significant at the 95% confidence level for lineage A. Additionally, the linear regression model did not indicate a significant relationship between generation and the number of DHFR, heavy chain, or light chain transgene copy numbers for both lineages A and B at the 95% confidence level when using Cog1 as a reference gene.

However, contradictions between the results from different reference genes, especially in the 32-121 cell line, made interpretation of the data challenging. To complement the linear regression analysis, the percentage fold change in mean relative gene copy number between early and late generations was assessed using each reference gene (Figure 3.17A).

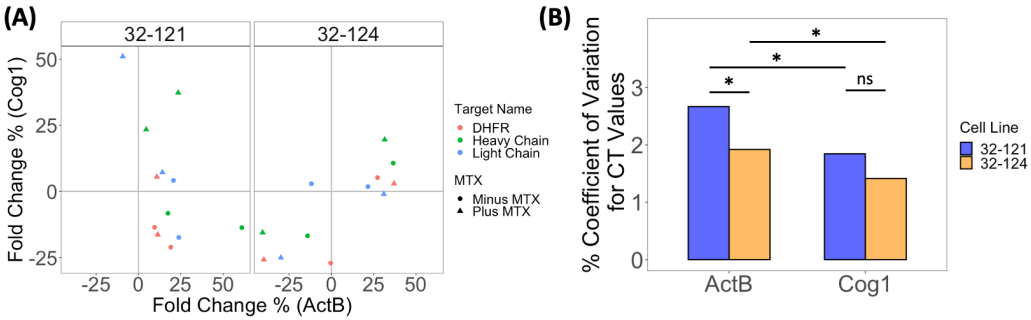


Figure 3.17 – Reference Gene Comparison for Gene Copy Number Analysis

(A) The fold change in relative gene copy number of the DHFR, heavy chain and light chain transgenes between the early and late generation measurements, measured relative to ActB gene copy number (x-axis) versus Cog1 gene copy number (y-axis). The data points represent measurements from the 32-121 and 32-124 cell lines across the early, early-mid, late-mid, and late generations. **(B)** Coefficient of variation of all CT values for the ActB and Cog1 genes in the 32-121 and 32-124 cell lines across the early, early-mid, late-mid, and late generations.

In the 32-124 cell line, results were consistent between the two reference genes, but in the 32-121 cell line, many samples showed divergent trends. For example, many of the samples which increased in transgene copy number, relative to the ActB reference gene, decreased in copy number relative to the Cog1 reference gene in the 32-121 cell line. The coefficient of variation of CT values was significantly higher for ActB than Cog1 in both the 32-121 and 32-124 cell lines, indicating that ActB may be more susceptible to gene rearrangements than Cog1, leading to variable gene copy numbers and unexpected variations in the results (Figure 3.17B). Interestingly, the higher coefficient of variation for ActB in the 32-121 cell line compared to the 32-124 cell line suggested that the 32-121 cell line might be more prone to DNA rearrangements, potentially contributing to its production instability.

The fold change analysis revealed no clear relationship between selection (MTX or no MTX), transgene (DHFR, heavy chain, or light chain), or cell

lineage (A or B) with gene amplification or gene loss for either cell line. Additionally, most genes exhibited a fold increase or decrease in copy number of less than 25% (relative to at least one of the reference genes) for both cell lines. Based on the results from the linear regression analysis, which indicated no significant decrease in transgene copy number in either cell line over 60 generations, and the lack of visible differences in the fold change analysis between the two cell lines, it was concluded that gene loss was not the primary cause of instability in the 32-121 cell line.

3.5.14 Design and Setup of Transgene mRNA Expression Quantification Using Real-Time Quantitative PCR

Having concluded that transgene copy number loss was unlikely to be the primary cause of productivity loss in the 32-121 cell line, it was decided that transgene mRNA expression should be analysed to determine whether transgene silencing is having a significant impact. To pursue this, cDNA samples from day 6 of the FBCs were analysed using quantitative reverse transcription PCR (qRT-PCR). Previous analysis had suggested that the drop in Q_p for all cell lines was not phase-specific (see section 3.5.10). Therefore, day 6 was chosen as cells were still at a high viability and growing exponentially. ActB, a widely used reference gene for gene expression studies in CHO cells, was initially selected as the reference gene for these experiments. To amplify ActB mRNA, and avoid amplification of genomic DNA, a primer set spanning two exons of the ActB gene was designed. Given that the DHFR, heavy chain and light chain transgenes do not contain introns, the same primer sets used for gene copy number analysis were suitable for cDNA amplification. Amplification efficiency tests were conducted on serial dilutions of a cDNA sample (Figure 3.18). As required, the primer sets for the heavy chain, light chain and ActB genes had an amplification efficiency of 90-110% and a high R^2-value (>98%). Initially, the calculated primer efficiency for the DHFR gene was calculated at 51% with an R^2 of 90%. However, it was believed that at high cDNA concentrations, there could be some off target amplification, giving rise to a lower CT than expected. Given that gene expression analysis was conducted with 0.1ng of cDNA

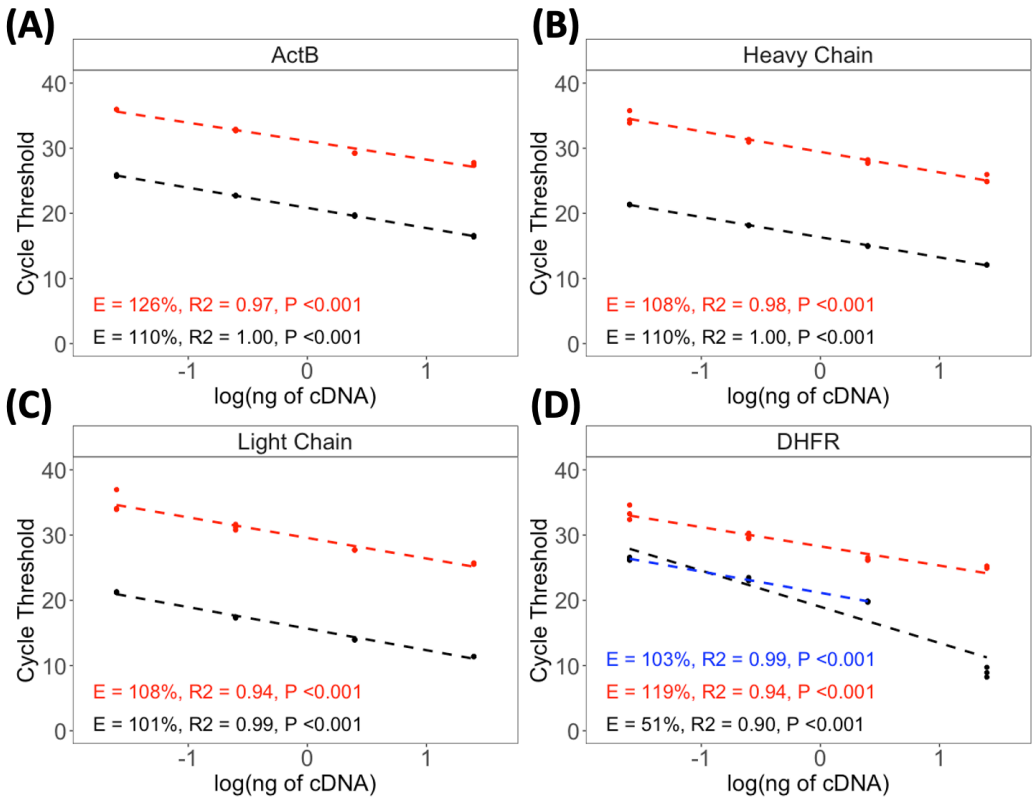


Figure 3.18 – Amplification Efficiencies of Primer Sets for Gene Expression Analysis

The primer efficiency of the selected primer sets for **(A)** ActB **(B)** Heavy Chain **(C)** Light Chain **(D)** DHFR were assessed by preparing serial dilutions of cDNA and subjecting them to RT-qPCR. The CT values were measured, and primer efficiency was calculated by determining the slope of the regression line for Ct values versus the logarithm of genomic DNA concentration.

sample, it was decided that the high cDNA concentration measurement could be removed from the analysis. With this measurement removed, the DHFR primer set was calculated to have a primer efficiency of 103% with an R^2 of 98%. A "no reverse transcriptase" control was included during cDNA synthesis to assess genomic DNA contamination. Although there should be no DNA amplification in the no reverse transcriptase control if there is no genomic DNA present in the RNA sample, some amplification was observed

for all genes. However, the difference in CT values between the RT(+) and RT(−) reactions was at least 5 for all genes and cDNA dilutions, meeting the accepted limit for accurate estimation of gene of interest (GOI) expression with <3% of the total signal originating from genomic DNA (Laurell et al., 2012). For all genes and cDNA dilutions, the difference in CT between the RT(+) and RT(−) reactions was at least 5. It was assumed that the level of genomic DNA contamination was consistent across RNA samples, to avoid multiple qRT-PCR runs and minimise expenses.

3.5.15 A Significant Relationship Between the CT Of Actb and Generation Number was Observed

An appropriate reference gene for gene expression studies should exhibit consistent expression across various conditions. Initially, ActB was selected due to its common use as a reference gene in CHO cell studies, including LTC experiments (Bailey et al., 2012; Lakshmanan et al., 2019; Reinhart et al., 2019; Sommeregger et al., 2013). However, in this study, ActB's CT values showed a significant positive correlation with generation number in all cell lines (both with and without MTX), indicating a decrease in ActB expression during LTC (Figure 3.19). Recent research also supports the notion that ActB can display variable expression (Bahr et al., 2009; Brown et al., 2018; Ma et al., 2020; Zboray et al., 2015), suggesting that it may not be the most suitable reference gene for gene expression studies.

3.5.16 Screening Reference Genes for Gene Expression Studies

To identify a more suitable reference gene for gene expression analysis, a literature search was conducted to identify genes with proven stable expression during both LTC and FBCs. Previous studies used transcriptomic data to select reference genes that exhibit minimal expression changes across various conditions. The stability of these reference genes was further validated using RT-qPCR and ranked using algorithms such as geNorm, NormFinder, and BestKeeper, along with analysis of CT value ranges.

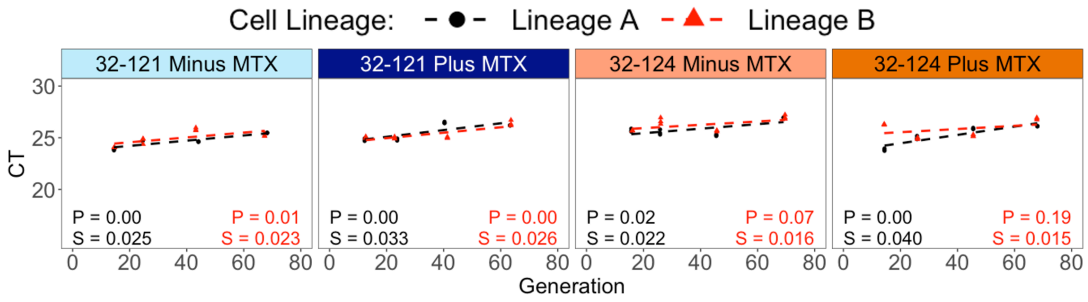


Figure 3.19 - CT values for the ActB gene Exhibit a Significant Linear Relationship with Generation Number.

Plot displays the individual CT values for the ActB gene at various generation timepoints. Linear regression analysis was performed, and regression lines were drawn for each lineage. The p-value (p) and slope (S) were calculated to assess the significance of the relationship between CT values and generation number.

Although transcriptomic data from LTC of a recombinant Apollo™ X cell line was not available, RNA-Seq data from a recent FDB FBC transcriptomic study was included in the decision making process. During the previously conducted study, FBCs were conducted and RNA samples were collected on various days by FDB. RNA samples were then analysed using RNA-Seq by Edinburgh Genomics. Among the 26 genes identified from the literature, 21 were successfully annotated in the Apollo™ X RNA-Seq data. The genes were ranked based on mean counts per million (CPM), to ensure the reference gene is expressed at similar levels to the genes being studied. Additionally, they were ranked based on coefficient of variation and fold change between consecutive samples. Five genes (Aprt, Mmadhc, Gnb1, Rps16, and Fkbp1a) were selected for further investigation due to their favourable rankings in both the literature and the Apollo™ X transcriptomic data (summarised in Table 3.3).

Table 3.3 – Selection of Reference Genes From The Literature and Apollo™ X Transcriptomic Data

	Paper 1 (Ma et al., 2020)		Paper 2 (Brown et al., 2017)	Paper 3 (Bahr et al., 2009)	Apollo X Transcriptomics					
Target Gene	LTC Ranking	FBC Ranking			Mean CPM	Mean CPM Ranking	CV%	CV% Ranking	Max FC	Max FC Ranking
Akr1a1	1	1								
Gpx1	2	9			55.0	17	81.8	21	-2.5	21
Aprt	**3**	**7**			**58.8**	**16**	**4.5**	**3**	**-0.1**	**3**
Mmadhc	**4**	**12**	**4**		**103.9**	**12**	**5.2**	**4**	**-0.2**	**4**
Atp5mg	5	3			152.4	8	7.3	7	0.2	5
Eif3i	6	5		5	578.2	4	10.8	10	-0.3	10
B2m	7	4		10	182.4	7	31.0	19	1.0	19
Gnb1	**8**	**6**	**1**		**640.1**	**3**	**1.7**	**1**	**0.1**	**2**
Hprt	9	11			65.8	14	8.7	9	-0.2	7
Pgk1	10	10			576.8	5	12.5	11	-0.4	12
Eif3k	11	14			117.8	10	7.0	5	-0.3	9
Rps16	**12**	**2**			**1081.4**	**2**	**7.3**	**6**	**-0.3**	**8**
Gapdh	13	17	5	8	72.8018815	13	15.5665	13	-0.4	11
Gusb	14	13			141.519536	9	18.0436	17	-0.5	16
Actr5	15	15		2	47.7367257	19	15.7435	14	-0.6	17
Fkbp1a	**16**	**16**	**2**		**222.658**	**6**	**8.192**	**8**	**0.2**	**6**
Actb	**17**	**18**	**6**	**9**	**2572.8**	**1**	**15.79**	**15**	**-0.5**	**15**
Hirip3	18	20		1	61.8387485	15	44.1212	20	-1.4	20
Atp5f1	19	8								
Pabpn1	20	19		3	106.976964	11	16.7047	16	-0.5	14
Tmed2			3							
Pkar1a			5							
Pgam1			6							
Vezt				4	35.1874634	20	3.24479	2	-0.1	1
Cog1				6	49.6842884	18	19.3181	18	-0.6	18
Yaf2				7	22.6262327	21	13.8815	12	-0.4	13

This table shows the rankings of candidate reference genes for reference gene selection from different studies and transcriptomic data: **Ma et al., 2020**: Evaluated 20 candidate genes in CHO cells with varying productivities under different conditions. Used geNorm, NormFinder, BestKeeper, and ΔCt methods for stability ranking in LTC and FBC. **Brown et al., 2017**: Assessed 8 candidate genes across distinct CHO cell lines in various culture phases. Ranked gene stability using GeNorm, BestKeeper, and NormFinder. **Bahr et al., 2009**: Identified candidate genes from microarray data of IgG-producing CHO cells. Ranked genes based on CT value standard deviation. **Apollo™ X Transcriptomics**: Ranked genes from RNA-seq data of recombinant Apollo™ X cell line during FBC based on CPM, coefficient of variation, and fold change (FC).

Aprt encodes a protein which is involved in purine metabolism (Valaperta et al., 2014), Mmadhc encodes a protein which is involved in vitamin B12 metabolism (Banerjee et al., 2009), Gnb1 encodes the guanine nucleotide-binding proteins protein subunit β1 (Da Silva et al., 2021), Rps16 encodes the ribosomal subunit protein S16 (Alqahtani and Jansen, 2021) and Fkbp1a encodes an immunophilin protein which plays a role in immunoregulation and basic cellular processes such as protein folding and trafficking (Ozdemir Kutbay et al., 2020). Interestingly, these selected genes exhibited improved expression stability compared to ActB, as reported in the studies by Ma et al., 2020, and Brown et al., 2017. However, it is worth noting that the CPM value for ActB was much higher than all 5 selected genes during the Apollo™ X transcriptomic study. To increase the sensitivity of the experiment, so that subtle differences in gene expression can be detected, it is often recommended that for accurate quantification of gene expression, the reference and target genes should exhibit similar expression levels (Aithal and Rajeswari, 2015; Foquet and Song, 2020; Silver et al., 2006; Vandesompele et al., 2002).

After selecting the candidate reference genes, three primer sets per target were designed to span two exons. Standard PCR reactions were performed with the different primer sets on cDNA samples using OneTaq Polymerase, and the resulting PCR products were visualised on an agarose gel to confirm successful amplification of single 200 bp products (Figure 3.20A). For primer efficiency testing, eight primer sets were evaluated simultaneously using serial dilutions of a cDNA sample, including two sets each for Aprt, Mmadhc, and Gnb1, and one set each for Rps16 and Fkbp1a (Figure 3.20B). Each primer set successfully met the desired criteria of 90-110% amplification efficiency and an R^2 of >98%. The most highly expressed reference genes were selected based on the lowest CT at 1 ng of cDNA, resulting in the selection of Fkbp1a (primer set 1), Gnb1 (primer set 3), and Rps16 (primer set 2) for further study.

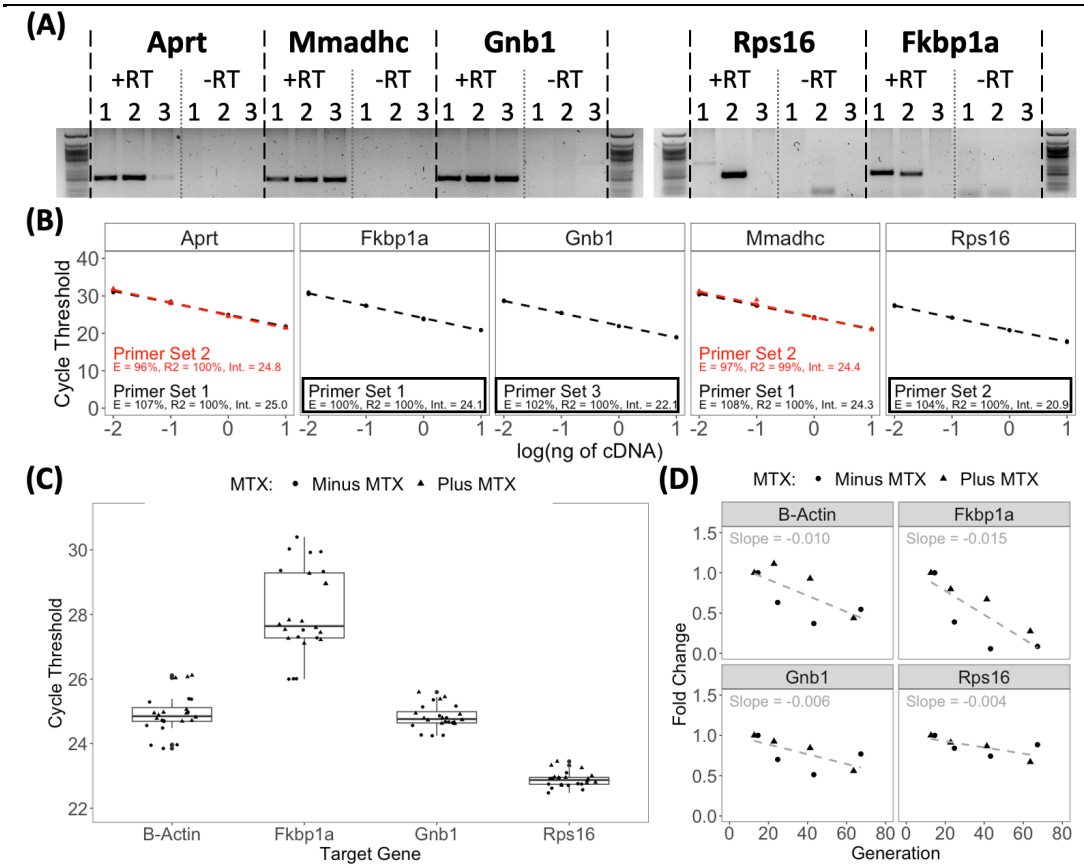


Figure 3.20 – Selection of Reference Genes For Gene Expression Analysis

(A) Primer sets for the cDNA of endogenous genes Aprt, Fkbp1a, Gnb1, Mmadhc, and Rps16 were screened for amplification specificity. Standard PCR reactions were performed on cDNA samples made from RNA with reverse transcriptase (+RT) and a no reverse transcriptase control (-RT). The resulting PCR products were separated on an agarose gel, and successful primer specificity was indicated by the presence of a single band at 200bp in +RT samples, while -RT samples helped identify any genomic DNA contamination. **(B)** Amplification efficiency assessment for selected primer sets. RT-**qPCR** was performed on serial dilutions of cDNA, and primer efficiency (E) was calculated using regression analysis of Ct values against the logarithm of cDNA concentration. R2 and y-intercept values are provided. **(C)** Boxplot showing CT values (Ct) for candidate reference genes at

(Continued on next page)

(Figure 3.20 continued) different generation timepoints (early, early-mid, late-mid, and late) for the 32-121 cell line with and without MTX. **(D)** Changes in expression levels of each reference gene over various generation timepoints were estimated by determining the fold change relative to the early generation expression level using the delta CT method. Fold changes expressed as $2^{-\Delta Ct}$.

The expression stability of the selected genes was assessed by measuring the variation in CT values of lineage A cDNA samples (plus and minus MTX) at different time points during the FBCs (early, early-mid, late-mid, and late) (Figure 3.20C). Genes with low variation in CT values were considered more stable. Rps16 exhibited the lowest variability in CT values across the different generation time points. To estimate changes in gene expression over time, a fold change was calculated by comparing the CT value at each time point with the CT value at the early generation using the delta CT method (Figure 3.20D). Interestingly, all genes showed downregulation over time when compared with expression at the early generation time point, perhaps due to a reduction in cell health with passage number. Rps16 showed the smallest decrease in expression and was thus chosen for future experiments.

3.5.17 Gene Expression Analysis of 32-121 and 32-124 Cell Lines

Gene expression analysis was performed on day 6 cell culture samples from the early, early-mid, late-mid, and late FBCs, using Rps16 as a reference gene (Figure 3.21). While comprehensive gene expression analysis throughout the FBC would be optimal, day 6 was selected to minimise the number of samples. This choice was made because day 6 corresponds to a period of high cell viability and exponential growth. Identifying a consistent trend in expression over the early, early-mid, late-mid, and late FBCs proved challenging in many cases.

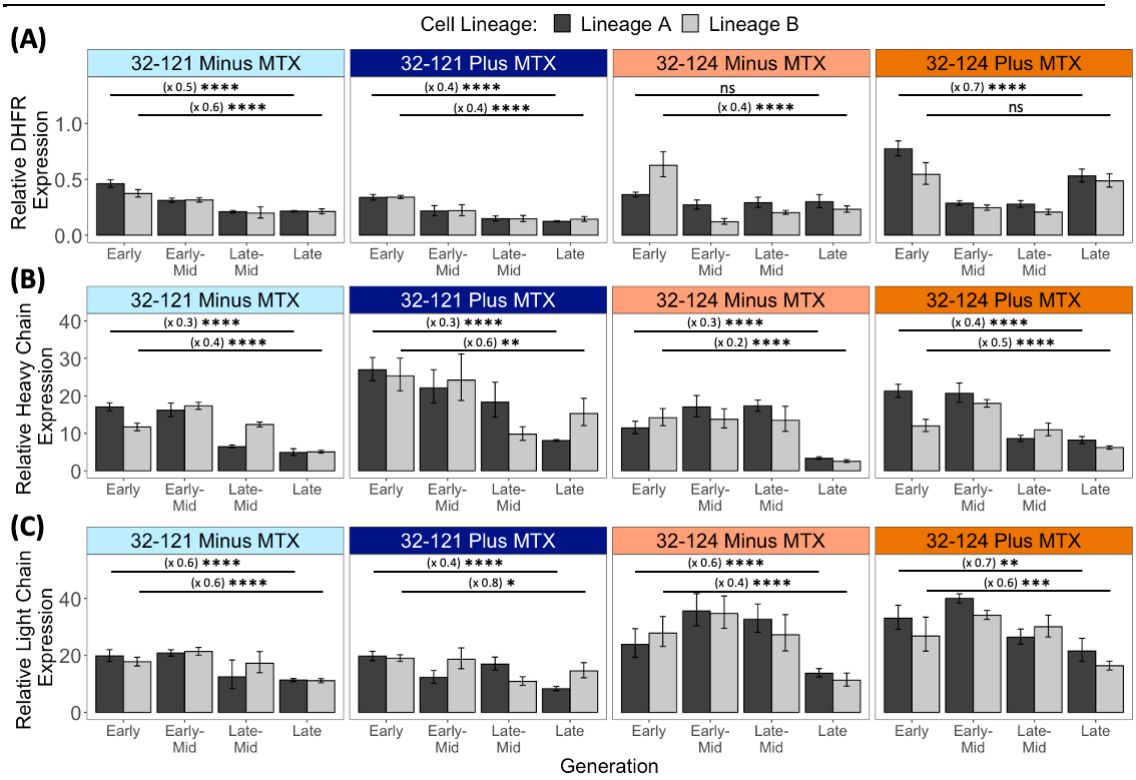


Figure 3.21 - mRNA Expression Analysis of the 32-121 and 32-124 Cell Lines

Relative mRNA expression analysis performed by RT-qPCR. The mRNA expression for each target gene (DHFR, heavy chain, and light chain) was normalised to the mRNA expression of the reference gene, Rps16, calculated using the delta CT method. The data points represent the mean $2^{-\Delta Ct}$ values (± standard deviation) calculated from triplicate measurements for each lineage. Fold change between the mean $2^{-\Delta Ct}$ values of early and late generation samples was calculated. $^{ns}P > 0.05$, $*P \leq 0.05$, $**P \leq 0.01$, $***P \leq 0.001$, $****P \leq 0.0001$ as calculated using a two sample Z-test.

For example, in lineage B of the 32-124 (plus MTX) samples, expression of the light chain increased during the early-mid FBC compared to the early FBC, but then decreased during the late-mid FBC when compared to the early FBC. While attempts were made to normalise mRNA content between FBCs by analysing 1000ng RNA per sample, variations in cell number (due to different growth rates between FBCs), especially in the mid-generation FBCs, may have contributed to the variability in the results. As a result, the main focus of the analysis was on comparing expression between the early and late generation FBCs.

In the 32-121 cell line, both with and without MTX, a clear decrease (approximately 2-fold between early and late generation FBCs) in DHFR expression was observed over 60 generations, between the early and late generation FBCs. In the 32-124 cell line (minus MTX), lineage A displayed relatively stable DHFR expression throughout the FBCs, with no significant difference between the early and late generation FBCs, while lineage B showed a significant 2-fold drop in DHFR expression between the early and late generation FBCs. Interestingly, a U-shaped relationship between generation and relative DHFR expression was observed in the 32-124 cell line, with higher expression during the early and late FBCs and lower expression during the early-mid and late-mid FBCs. The change in relative DHFR expression between the early and late generation was significant for lineage A but not for lineage B. As expected, the drop in relative DHFR expression was greater in the minus MTX cell line than the plus MTX cell line for the 32-124 cell line.

In all cell lines (both with and without MTX), there was a significant decrease in relative gene expression between the early and late generation FBCs for both heavy chain and light chain genes. Interestingly, the mean fold change decrease in expression was greater in the 32-124 cell line (x0.25 for heavy chain and x0.5 for light chain) than in the 32-121 cell line (x0.35 for heavy chain and x0.6 for light chain) under minus MTX conditions, despite the 32-121 cell line exhibiting a higher drop in productivity due to a decrease in Q_p.

However, for both gene copy number and mRNA expression, there was little difference between the 32-121 and 32-124 cell lines in the minus MTX conditions. For the plus MTX conditions, the 32-121 and 32-124 cell lines exhibited a similar fold change decrease in expression for the heavy chain (x0.45 for 32-121 and x0.45 for 32-124) and light chain (x0.6 for 32-121 and x0.65 for 32-124). Again, this was surprising because the drop in productivity that the 32-121 cell line exhibited was thought to be due to a combination of a drop in Q_p and changing growth characteristics (Figure 4.7). However, for both gene copy number and mRNA expression, there was little difference between the 32-121 and 32-124 cell lines in the plus MTX conditions, suggesting that other factors may contribute to the observed differences in productivity between the two cell lines.

3.6 Discussion

The objective of this chapter was to identify and characterise a model stable and unstable cell line for further investigation (as detailed in chapter 5). A previous FDB stability study using the Ambr®15 bioreactor system provided cell lines with different stability profiles over 60 generations, identifying 32-121 as a model unstable cell line and 32-124 as a model stable cell line. However, the data lacked robustness, due to the absence of replicates, prompting the design of a new stability study with both biological and technical replicates (see section 3.6.1 for experiment limitations and challenges). To comprehensively assess the stability of the 32-121 and 32-124 cell lines, cells were cultured for 60 generations under plus or minus MTX conditions, and FBCs were conducted on cell culture samples from various generations. Comparing day 14 product concentrations between early and late generation FBCs, it was observed that the 32-124 cell line exhibited a moderate loss of 1% with MTX and 13% without MTX. Conversely, the 32-121 cell line showed a more substantial productivity drop, 24% loss with MTX and 27% without MTX. Further analyses of growth characteristics, Q_p, metabolite profiles, gene copy number, and transgene mRNA expression were subsequently conducted to investigate the

underlying cause of the productivity loss in the 32-121 cell line (see section 3.6.2 for further discussion).

3.6.1 Experiment Limitations and Challenges of Experimental Design

Initially, the plan was to utilise data from the Ambr®15 bioreactor stability study to select cell lines for further investigation. However, a surprising finding emerged, as most of the cell lines in the Ambr®15 study exhibited an unexpected increase in product concentration between generation 10 and 60, which is not typical for recombinant CHO cells during LTC. Despite efforts to comprehend these unusual results, including evaluating the influence of culture station on growth and productivity, it was determined that the variability in conditions between individual bioreactors likely contributed to these unexpected outcomes.

At this point, the idea of developing my own recombinant cell lines was considered, but ensuring the creation of an unstable cell line posed a significant challenge. The literature presents a wide range (8-63%) of production instability occurrences in recombinant CHO cell lines, making it difficult to predict the proportion of cell lines which would have exhibited production instability (Dahodwala and Lee, 2019; Dorai et al., 2012; Leonard et al., 2009). In the Ambr®15 study, 9% (4 out of 46) of the cell lines in the previous Ambr®15 bioreactor study failed to retain 70% of productivity. However, considering that the most unstable cell line (32-121) from the Ambr®15 study exhibited a productivity loss of less than 30% (both with and without MTX) in the study presented in this chapter, it was hypothesised that the other cell lines that had previously been labelled unstable in the Ambr®15 bioreactor study, might not exhibit production instability. Developing my own cell lines and testing them all for instability would have been time-consuming, and the number of cell lines that could have been tested simultaneously was limited, with each stability study taking approximately three months to complete. For example, the Ambr®15 can run 48 FBCs at once, but to account for variability between bioreactor runs, testing in triplicate is necessary, limiting the maximum number of tested cell lines to 16. Testing

cell lines in both the presence and absence of MTX conditions, a common industry practice to investigate instability causes, reduces the number of cell lines that can be tested to 8. Given these limitations and the time constraints of the project, it was considered too risky to develop multiple cell lines in the hope of identifying at least one unstable cell line.

Instead, it was decided that the Ambr®15 data would still be used to select a "stable" and "unstable" cell line for further investigation. This decision was made on the basis that although the changes in product concentration of certain cell lines in the Ambr®15 bioreactor stability study experiments were unlikely to solely be due to changes in cell line characteristics over 60 generations, and instead partly due to the inherent variability of Ambr®15 bioreactor runs, it provided a good indication of the likely stability profile of the cell lines. A stability study, which included the correct biological and technical replicates, could then be performed to fully evaluate the stability profile of the selected cell lines. Conducting this study had two clear benefits. Firstly, the stability profile of the selected cell lines could be fully characterised, with a robust study. Secondly, the incorporation of early-mid and late-mid FBCs meant that the change in product concentration over time could be monitored and the dynamics of productivity loss in the cell lines could be better understood.

The experimental design for the stability study was challenged by the limited availability of the Ambr®15 bioreactor and the need for multiple timepoints to assess production changes over time. Due to these constraints, shake flasks were used instead of the Ambr®15 system. FDB safety requirements and workload limits recommended a maximum of 36 FBCs to be run concurrently by one person. To manage the workload effectively, FBC screens were performed at four different generation timepoints for two out of the three cell lineages (lineages A and B), with early and late generation FBCs conducted concurrently to reduce variation in conditions. This approach followed the standard FDB method for stability assessment, which involves subculturing two flasks with MTX and two without MTX per cell line, with single or

duplicate FBCs typically conducted for each flask. Conducting CHO cell stability studies in the presence and absence of MTX is important because selection agents such as MTX can often improve the production stability of the cell line. However, LTC in MTX conditions has been associated with an increase in mutation rate, which may result in variations in the amino acid sequence of the desired product (Guo et al., 2010; Zhang et al., 2016). To reduce toxicity concerns, the scale-up of cell culture volume in the absence of selection pressure (MTX) is thus preferred. By reducing the number of FBC flasks per cell line, the study aimed to strike a balance between workload management and experimental quality.

Following product concentration analysis of the FBCs, it was shown that the 32-121 cell line exhibited a loss of productivity of 27% with MTX and 24% without MTX. This was a much lower drop in productivity than was observed in the Ambr®15 bioreactor study. The initial aim of the study was to develop and characterise 2 cell lines: one that exhibits stable productivity and one that exhibits production instability. As previously discussed, a well-accepted definition of an unstable cell line is a cell line that fails to retain >70% of its productivity over 60 generations. Therefore, by this definition the 32-121 cell line is not defined as an unstable cell line. As a result, the stability of more cell lines from the Ambr®15 bioreactor study could have been tested. However, as previously discussed, the 32-121 cell line showed the greatest production instability in the Ambr®15 bioreactor study, and therefore it was perhaps unlikely that the other cell lines would be more unstable. Additionally, any future stability study would take a further 3 months, significantly adding to the timelines of the project.

Therefore, the decision was made to work with the available data and conduct further analyses of the 32-121 and 32-124 cell lines that were available. The 32-121 cell line exhibited a significant drop in productivity over 60 generations. Although 70% is widely cited as the cut off for instability, >80% can also be used as a cut off. FDB places cell lines which retain 70-80% of their productivity over 60 generations in an alert limit for instability. If

necessary, this means that the cell line can be considered to be stable, however, a cell line which exbibits greater production stability would be preferred and the stability of other parameters (such as product quality and growth characteristics) are taken into account when recommending the cell line as suitable to progress to GMP manufacture. The 32-121 cell line did not exhibit a significant drop in maximum VCD or IVC when cultured without MTX. When cultured with MTX, there was a significant drop in VCD and IVC but the drop was not greater than 30%. In the previous FDB study, product quality analysis was conducted on 12 out of the 46 tested cell lines (including 32-121) and it was concluded that mAb product quality (including charge heterogeneity and glycan profile analysis) was comparable between early and late generations for all cell lines.

3.6.2 Understanding the Cause of Loss of Productivity in the 32-121 Cell Line

After establishing the production stability of the cell lines over 60 generations, the focus shifted towards identifying the underlying causes of productivity loss in the 32-121 cell line by comparing it with the 32-124 cell line. These comparisons were expected to reveal key differentiators between the two cell lines, providing insights into the productivity decline observed in the 32-121 cell line.

As has also been previously reported in other studies (Baik et al., 2021), during routine subculture, the presence of MTX had a negative impact on the growth rate of both cell lines, with a decrease in the mean generation number increase per subculture over the 60 generations compared to when cultured without MTX. The presence of MTX during LTC also had a negative impact on the growth performance of the 32-121 cell line in the late generation FBCs. While no significant difference in growth performance was observed between the early and late generation FBCs without MTX, the 32-121 cell line that was cultured with MTX exhibited a significant decrease in growth performance (lower IVC and maximum VCD) between the early and late generation FBCs. In particular, the viability of the late generation FBC for the 32-121 cell line (plus MTX) started to decline after day 5, which contrasts

with the minus MTX 32-121 cell line, where a similar viability profile was observed between the early and late generation FBCs. The drop in growth performance observed in the 32-121 cell line (plus MTX) had a notable impact on its overall productivity. Despite the similarity in Q_p-rate decline between the 32-121 cell line (plus MTX) and the 32-124 cell line (minus MTX), there was a notable difference in day 14 product concentration.

Initially, it was hypothesised that the metabolic burden of DHFR gene expression might be responsible for the impact of MTX on the growth performance of the 32-121 cell line over 60 generations. For example, previous studies have suggested that the expression of transgenes in mammalian cells can have a negative impact on growth (Misaghi et al., 2014; Ong et al., 2019). However, gene expression analysis revealed that the association between the presence/ absence of MTX and DHFR expression was not particularly strong, with a significant amount of DHFR transgene silencing observed on day 6 FBC samples over 60 generations in both plus and minus MTX conditions. However, it must be noted that DHFR expression was only monitored on a single FBC day. An alternative hypothesis was based on reports that nucleotide shortage (as a result of MTX-mediated DHFR inhibition) leads to an increase in the frequency of double-stranded break formation in the DNA, which subsequently leads to an increase in chromosomal rearrangements (Baik et al., 2021). It has previously been suggested that the double-stranded breaks are a consequence of pauses in the DNA synthesis machinery during genomic DNA replication in S phase, resulting in replication fork collapse (Baik et al., 2021). The MTX-induced stress on the replication machinery could thus lead to a reduction in the growth performance of the population. However, even in the absence of MTX, various reports have suggested that unstable lines may show an increase susceptibility to cell stress (Bailey et al., 2012) and apoptosis (Dorai et al., 2012), whilst down-regulating cell cycle pathways (such as p53) (Torres et al., 2023). Such studies suggest that cell age can have a negative impact on the growth performance of cell lines over time.

However, the 32-121 cell line (both with and without MTX) showed an increase in growth rate following LTC. This was in contrast with the 32-124 cell line, which exhibited no significant change in growth rate between the early and late generation subcultures. An increase in growth rate following LTC is commonly observed in the absence of selection pressure (Baik and Lee, 2018; Dorai et al., 2012). This is often attributed to cells that lose productivity having a growth advantage over high-productivity cells, eventually dominating the population over time. However, low productivity cells are not expected to survive in plus MTX conditions due to insufficient DHFR production to counteract MTX inhibition. Therefore, the increase in growth rate following LTC in the presence of MTX was unexpected. An additional consequence of an increase in the frequency of chromosomal rearrangements due to MTX treatment is an increase in the chromosomal heterogeneity of the cell population (Baik et al., 2021). As a result, heterogeneity within the clonal population has been shown to increase, potentially selecting subpopulations with growth advantages over LTC (Baik et al., 2021; Wurm and Wurm, 2017). Consequently, the observed increase in growth rate over time in the 32-121 cell line could be linked to the presence of chromosomal heterogeneity induced by MTX treatment.

Given that the 32-121 cell line (minus MTX) did not exhibit a significant change in FBC growth characteristics over 60 generations, it was anticipated that productivity loss was primarily due to a drop in Q_p-rate. It was also expected that any gene loss or gene silencing that occurs in the 32-121 cell line (minus MTX) would not be present to the same extent in the 32-124 cell line or 32-121 cell line (plus MTX). To further investigate the cause of the productivity drop, RT-qPCR experiments were conducted to assess changes in gene copy number and mRNA expression over 60 generations. Previous studies have used similar techniques to establish either gene loss (Bandyopadhyay et al., 2019; Li et al., 2016) or transgene silencing (Osterlehner et al., 2011; Veith et al., 2016) as the primary cause of productivity loss during long term culture.

However, gene copy number analysis showed relatively little change in the 32-121 cell line, indicating that gene loss was not a primary cause of instability. A significant decrease in mRNA expression between early and late generation FBCs was observed for both the heavy and light chain in the 32-121 cell line (minus MTX). However, this was also observed to a similar degree in the 32-124 cell line and the 32-121 cell line (plus MTX). This suggested that transgene silencing was not a clear differentiator between the two cell lines and thus further investigations are required to understand the production stability of the cell line.

3.6.3 Future Avenues for Understanding the Production Instability of the 32-121 Cell Line

This study aimed to identify and characterise model stable and unstable CHO cell lines for further investigation. The production stabilities of two model cell lines, 32-121 and 32-124, were characterised in the presence or absence of MTX, but the underlying cause of instability in the 32-121 cell line could not be fully elucidated. Nevertheless, the data obtained in this study provides a solid foundation for future experiments.

The 32-121 cell line (plus MTX) offers a valuable model for assessing the impact of MTX on cell line instability. To gain a more comprehensive understanding of how MTX influences viability and productivity over extended culture periods, further investigations such as transcriptomic analysis and chromosomal characterisation are warranted. This may shed light on the potential increase in frequency of chromosomal rearrangements in the 32-121 cell line (plus MTX) compared to the 32-121 cell line (minus MTX). Additionally, studying subpopulations with different growth profiles within the population and conducting chromosome characterisation experiments could provide valuable insights, though the challenge of avoiding subcloning-induced changes would need to be addressed.

On the other hand, the 32-121 cell line (minus MTX) presents an intriguing scenario with no significant change in growth characteristics but a drop in productivity. As qRT-PCR experiments did not reveal significant changes in

mRNA expression relative to the 32-124 cell line or 32-121 cell line (plus MTX), it is essential to consider the complexity of gene expression. Various stages between mRNA expression, such as translation, protein folding, and secretion, contribute to protein production. Therefore, exploring different pathways related to protein synthesis and secretion through omics studies could provide valuable insights into the underlying instability.

In-depth investigations into the causes of productivity loss in the 32-121 cell line hold the potential to generate hypotheses for addressing the issue of instability in the CHO biomanufacturing industry. Future CLD strategies, such as targeted integration techniques for gene loss or the incorporation of epigenetic regulatory elements to reduce gene silencing, could be explored based on these findings. However, of particular interest is the idea of being able to identify early indicators of instability in the 32-121 cell line, that are not present in the 32-124 cell line. As previously discussed, the incorporation of early predictive methods for cell line stability could substantially reduce CLD timelines by increasing confidence in the stability status of a particular cell line. As a result, the stability study could be uncoupled from the critical path to manufacturing, enabling cell lines to progress to subsequent manufacturing stages before completion of the stability study. This idea is discussed further and a potential method for identifying such early indicators is presented in the next chapter.

Chapter 4 - Comparing two cell lines with different production stabilities at an early generation using the Berkley Lights Beacon® System

4.1 Chapter Summary

Despite substantial advancements in CHO cell-based monoclonal antibody production over the past two decades, production instability continues to pose a significant challenge for CLD timelines. Due to the unpredictability of CHO cell productivity loss, time-costly stability studies, which monitor productivity over 60-70 generations, must be conducted. Therefore, methods which enable early prediction of instability would be of great value to the CLD process. In the previous chapter, the stability of two recombinant CHO cell lines (32-121 and 32-124), producing the same monoclonal antibody, were characterised in a stability study. The 32-121 exhibited a significant loss of mAb productivity over 60 generations whereas the 32-124 cell line showed stable productivity. In this chapter, the two cell lines were grown up to generation 15 in the absence of MTX and subsequently analysed used the Berkley Lights Beacon® (Beacon®) system, which facilitates single-cell analysis of growth and productivity profiles within cell populations over a five-day period. The 32-121 cell line demonstrated an increased proportion of low-producing fast-growing cells when compared with the 32-124 cell line. Additionally, the 32-121 cell line exhibited greater variability in both growth rate and cell specific productivity within the cell population. These results are particularly interesting as they highlight differences between the two cell populations at an early generation, representing only a quarter of the total stability study duration. It is thus expected that these findings could serve as a foundation for the development of a predictive method for assessing the stability of cell lines generated during CLD.

4.2 Introduction

As previously discussed, the unpredictable decline in mAb productivity in certain recombinant CHO cell lines, known as "production instability," poses a significant challenge for the biologics manufacturing industry. While the exact causes of production instability are not fully established, factors such as transgene copy loss (Bandyopadhyay et al., 2019; Beckmann et al., 2012; Fann et al., 2000; Kim et al., 2011, 1998), epigenetic silencing (Chusainow et al., 2009; Marx et al., 2018; Moritz et al., 2016, 2015; Veith et al., 2016), increased sensitivity to stress (Bailey et al., 2012) and increased susceptibility to apoptosis (Dorai et al., 2012) have been associated with productivity loss during long term culture (LTC). To prevent unstable cell lines from advancing to manufacturing stages, the productivities of lead candidate cell lines (identified during CLD) are assessed over ~60-70 generations, equivalent to the number of population doublings required to scale from research cell bank to production volume (Dahodwala and Lee, 2019). Generally, a cell line is considered stable if it maintains >70% of its volumetric productivity beyond this generation threshold (Dahodwala and Lee, 2019). However, such stability studies take approximately 3 months to complete, forming a substantial bottleneck in CLD timelines. Incorporating early predictive methods for cell line stability could considerably reduce development timelines by enhancing confidence in cell line stability (Dorai et al., 2012). Consequently, the stability study could be decoupled from the critical path to manufacturing, enabling cell lines to advance towards manufacturing stages before the stability study has been completed. In this section, the current limitations associated with employing omics strategies for predicting production stability are discussed and an alternative approach for real-time stability profile prediction, based on the identification of low-producing fast-growing cells within the cell population, is presented.

4.2.1 Predicting CHO Cell Production Instability using Sequencing and Omics is Inherently Difficult and Time Consuming

Considerable efforts have been dedicated to understanding the underlying causes of productivity loss in recombinant CHO cell lines (Barnes and Dickson, 2006; Dahodwala and Lee, 2019; O'Callaghan and Racher, 2015). This typically involves evaluating the production stability of various monoclonal cell lines that have been generated in parallel using the same CLD process. Cell lines are grown for 60-70 generations with samples taken at early and late generation timepoints. Downstream analyses often incorporate genomic, transcriptomic, epigenomic and/or metabolomic profiling to identify changes resulting from LTC that are unique to the cell lines exhibiting unstable productivity. These observations are subsequently used to build hypotheses for the underlying causes of productivity loss. A next step from such studies is to use the early generation omics profiles of the cell line to yield predictive cues about its future production stability. However, conducting such analyses in a manner that accurately predicts production stability whilst minimising analysis time poses a significant challenge, as will be discussed.

4.2.1.1 Predicting Production Stability from the Locus/Loci at which the Transgene Integrates

The genomic locus or loci at which the transgene integrates has been directly linked with the ability of a cell to stably produce high concentrations of mAb, termed "position effects". A common cause of instability, particularly in cell lines that have undergone MTX-mediated transgene amplification, is through the loss of transgene copies via chromosome rearrangements or deletions (Bandyopadhyay et al., 2019; Beckmann et al., 2012; Fann et al., 2000; Kim et al., 2011, 1998). Current CLD strategies typically use random transgene integration, a process characterized by uncontrolled integration of the transgene, which typically occurs at one or a few genomic loci (Bandyopadhyay et al., 2019; Clappier et al., 2023; Stadermann et al., 2022). However, several studies have indicated that the likelihood of transgene loss

is dependent on the specific locus at which it has integrated into (Bandyopadhyay et al., 2019; Li et al., 2016). For example, analysis of high producing and low producing subclones from a parental recombinant CHO cell line with three transgene integration sites demonstrated that loss of productivity in the low producing subclones was due to gene loss at a single locus, Rc3h1. This locus had experienced extensive amplification and structural rearrangements (Bandyopadhyay et al., 2019).

The inherent genomic instability of CHO cells has been highlighted by several karyotyping studies, which revealed varying chromosome numbers and copy numbers across different recombinant CHO cell lines (Cao et al., 2012; Derouazi et al., 2006; Martinet et al., 2007; Vcelar et al., 2018a, 2018b). However, the consensus among the studies has indicated that chromosomes 1, 2, and 9 consistently remain relatively stable in all tested cell lines (Bandyopadhyay et al., 2019). Therefore, analysis of the chromosome at which the integration has occurred could serve as a valuable indicator for predicting the potential stability profile of the cell line.

Epigenetic silencing of transgenes introduces an additional layer of complexity when attempting to predict the stability of CHO cells at early generations. Even if a transgene initially integrates into transcriptionally active euchromatin, silencing can still occur due to the spread of heterochromatin, a densely packed and transcriptionally repressive chromatin form (Cabrera et al., 2022). For instance, heterochromatin from an adjacent gene can spread to and silence the integrated transgene (Figure 4.1). To identify genomic regions conducive to stable transcription, high-throughput chromosome conformation capture and RNAseq were used to compare the epigenomes and transcriptomes of a monoclonal antibody-producing recombinant CHO cell line and its parental CHO-K1 host (Hilliard and Lee, 2021). The study revealed that 10.9% of the CHO genome is composed of three-dimensional chromatin structures that support transcriptional stability compared with the rest of the genome (Hilliard and Lee, 2021). Importantly, these regions aligned well with previously published

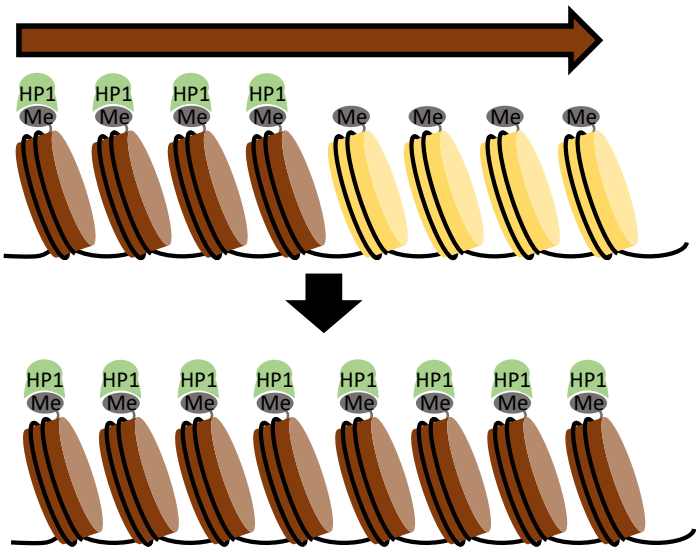


Figure 4.1 - The Spread of Heterochromatin
In heterochromatic regions of DNA, histone H3 lysine 9 (H3K9) methyltransferase (HMT) methylates histone 3 lysine 9 (H3K9me2/3). H3K9me2/3 is recognised by binding proteins, such as HP1. The spread of heterochromatin is caused by the recruitment of additional HMTs (via recognition of H3K9me2/3 or binding to HP1), which methylate adjacent nucleosomes (Wang et al., 2016).

CHO epigenome data. This suggests that recombinant cell lines with transgenes integrated into these identified loci are more likely to exhibit stable productivity during CLD.

As well as showing locus specific effects, epigenetic silencing is also dependent on the sequence of the expression vector. Some recombinant cell lines which display unstable productivity do not exhibit transgene copy loss but rather lose productivity via epigenetic silencing of the transgene, particularly when transgene expression is driven by the strong virally derived promoter, hCMV-MIE (Romanova and Noll, 2018). hCMV-MIE contains GC-rich regions that are susceptible to silencing via methylation. Efforts to predict production stability at an early generation timepoint based on the methylation status of a particular CpG island in the hCMV-MIE promoter

were only 75% effective due to the presence of both false negative (unmethylated but unstable) and false positive clones (methylated but stable) (Osterlehner et al., 2011). Removal of CpG dinucleotides from pCMV reduced promoter methylation but did not significantly improve long term expression stability due to the accumulation of repressive histone modifications (Ho et al., 2016). Interestingly, using chromatin immunoprecipitation (ChIP) to measure histone H3 acetylation, it was found that hyperacetylation of H3 at hCMV-MIE is a more effective indicator of production stability than DNA methylation (Moritz et al., 2016).

Additionally, the use of genome-wide scale models has been proposed as a method to identify genome regions conducive to stable transgene expression (Lee et al., 2019). These models can pinpoint genes crucial for cell functioning that are thus likely to maintain transcriptional activity and an open chromatin state during LTC. Consequently, random integration cell lines with transgenes integrated next to such genes would be expected to sustain stable transgene expression. However, a substantial challenge lies in recognizing instances when transgene integration might disrupt important cellular features, such as growth and productivity, which may impact the cell negatively over time. For example, integration at certain loci may interfere with the cell's ability to manage stress during fed-batch culture (Bailey et al., 2012).

4.2.1.2 Distal Factors which Impact Stability

The stability of transgene expression is not solely influenced by the site of transgene integration. Predicting production stability is further complicated by ongoing alterations in the genomic landscape of CHO cells during LTC, characterized by extensive chromosomal rearrangements, deletions, DNA damage, and an elevated mutation rate, leading to phenotypic changes (Baik et al., 2021; Baik and Lee, 2018, 2017; Bandyopadhyay et al., 2019). Moreover, DNA template mutations and transcriptional errors occur at a higher rate in CHO cells that have undergone LTC, contributing to phenotypic drift (Zhang et al., 2016). Therefore, in multiple cases, changes in

gene copy number or transgene silencing are not thought to be the primary factors behind productivity loss (Bailey et al., 2012; Dorai et al., 2012). Rather, various distal factors, affecting processes such as transcription, translation, protein folding, secretion, cell cycle regulation, and cellular stress responses, can influence productivity. Such changes are considerably more challenging to predict than locus-specific effects due to the interactions of a wide variety of endogenous genes which can impact the cellular phenotype. For instance, a decrease in productivity has been linked to heightened sensitivity to cellular stress, as indicated by the upregulation of GADD153 (Bailey et al., 2012). Alternatively, increased susceptibility to apoptosis, indicated by the upregulation of annexin V and caspase 3 was observed in another recombinant CHO cell line (Dorai et al., 2012).

Several transcriptomic studies have sought to identify genes whose expression patterns could serve as early markers for predicting production stability in CHO cell lines (Doolan et al., 2010; Jamnikar et al., 2015; Li et al., 2015; Torres et al., 2023; Tzani et al., 2021). For example, in a comparative analysis of two CHO cell lines producing IgG, one with stable and the other with unstable productivity, 213 genes that displayed differential expression at an early passage were identified (Torres et al., 2023). Similarly, another study was able to effectively distinguish between a set of stable and unstable recombinant CHO clones at an early generation by subjecting the expression profiles of five differentially expressed genes (E130203B14, BX842664.2/Hist1h3c, Ptpre, Cspg4, Fgfr2) to principal component analysis (Li et al., 2015). Four of these genes have known functions related to various cellular processes. Fgfr2 is associated with cell growth (Fan et al., 2018), Ptpre is involved in intracellular signaling (Granot-Attas and Elson, 2004), BX842664.2/Hist1h3c plays a role in DNA replication (López et al., 2019), and Cspg4 is important for cell migration and regulation (Ilieva et al., 2018). However, the degree to which these findings are representative of all stable/ unstable cell lines remains uncertain. The task of identifying genes that consistently exhibit upregulation or downregulation in cell lines displaying

unstable productivity, compared to those with stable productivity, is likely to pose a significant challenge.

4.2.1.3 Alternative Approaches for Predicting Cell Line Stability are Needed

In summary, forecasting production stability through early-generation omics data analyses poses inherent challenges. While certain predictions about production stability can be derived from the site of transgene integration, influenced by factors such as chromosome stability, chromatin structure, and the expression stability of nearby genes, these predictions are limited by the complex bioinformatics involved and the lack of knowledge regarding locus-specific transcriptional activity mechanisms (Hamaker and Lee, 2018). Consequently, predictions based on sequence inputs of transgene integration sites are not currently feasible. In addition to bioinformatic analyses, genomic hotspots have been identified by pinpointing the transgene integration site in recombinant clones (generated by random integration) which stably express high levels of the gene of interest (Dhiman et al., 2020). However, the extent to which the characteristics of these hotspots extend to the surrounding chromosomal regions remains uncertain (Hamaker and Lee, 2018). The impact of additional distal factors on production stability, such as global genomic and transcriptomic changes during LTC, adds complexity to prediction efforts. Therefore, given the existing knowledge gaps in our understanding of productivity loss mechanisms and the time-intensive nature of these analyses, it has become crucial to explore alternative approaches that can complement the current predictive methods for assessing cell line stability.

4.2.2 Identifying Low-Producing Fast-Growing Subpopulations in Early Generation Cell Lines

An intriguing aspect of CHO cell production instability stems from the presence of low-producing, fast-growing cell populations within the cell line. As previously discussed, cell populations, under selection-free

conditions, often exhibit increasing growth rates during extended culture, which is frequently linked to declining productivity. This phenomenon is often attributed to subpopulations of cells characterized by rapid growth and reduced productivity, gradually outcompeting the high-producing slow-growing subpopulation. A previous study has identified and isolated these low-producing subpopulations using fluorescence-activated cell sorting (FACS). However, FACS comes with inherent limitations, making it challenging to assess the growth characteristics and productivity of these subpopulations, especially in a high-throughput manner (Le et al., 2018).

Since this study, there have been significant advancements in single-cell profiling of recombinant cell populations (Tejwani et al., 2021). Of particular interest is the Beacon® system, which integrates microfluidics and imaging techniques to provide real-time monitoring of the growth and productivity dynamics of individual cells (Figure 4.2). Upon loading of the cell line onto the instrument, a light-controlled deposition process sorts individual cells into one of the 1,758 nanopens on the microchip (Figure 4.2A). Subsequent perfusion of the chip with fresh medium facilitates the removal of waste products from the chip, allowing cells to be cultured in the nanopens for up to 5 days while maintaining high viability. Over the 5-day culture period, continual cell counts can be acquired, and monoclonal antibody (mAb) productivity can be assessed using spotlight assays (Figure 4.2B). The Beacon® has previously been used as a tool for the early identification of high producing clones during cell line development. Following random integration of the transgene, the heterogenous pools are allowed to recover before being loaded onto the Beacon®. A combination of validation steps and cell-imaging enables cells to be single-cell sorted with a clonality assurance of >99%. By continuously monitoring the growth and productivity of individual cells in real-time, clones that are likely to exhibit favourable characteristics for manufacturing can be identified and exported into a 96 well plate for further screening. In comparison with FACS-based approaches, the Beacon® system has demonstrated the ability to generate clonal cell lines with similar specific productivities in a shorter timeframe and with the screening of fewer

clones, underscoring its efficiency and effectiveness in cell line development (Le et al., 2018).

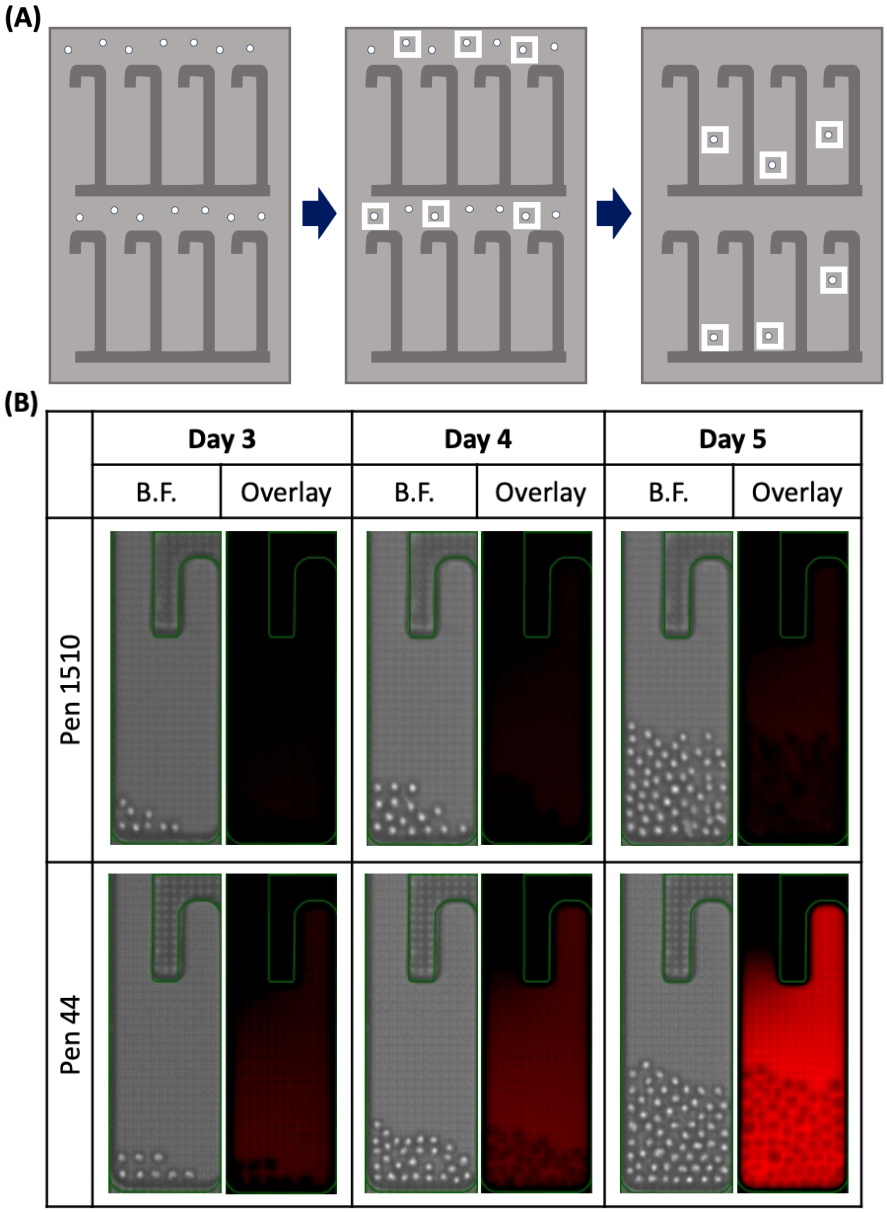


Figure 4.2 - The Beacon® System

(A) The Beacon® software automatically identifies single cells and directs them into nanopen chambers using optoelectropositioning (OEP™). **(B)** Images from 2 representative nanopens (a low producer and high producer) from the 32-121 cell line Beacon run (see subsequent sections for further description). Brightfield (BF) cell count images of individual nanopens were collected on days 3, 4 and 5. Secretion assays were conducted on days 3, 4 and 5 using Spotlight™ Human Fc assay reagent to measure mAb

secretion. Secretion assay images were overlayed with the brightfield images.

However, to our knowledge, the Beacon® has not been employed for the early prediction of production instability. In theory, the instrument should have the capability to identify subpopulations of cells within the overall population that exhibit distinctive growth and productivity characteristics. By conducting a comparative analysis between stable and unstable cell lines using the Beacon®, data on the population dynamics that are specific to unstable cell lines could be collected. This information could then be harnessed to formulate predictions regarding cell line instability. In this chapter, the Beacon® system is used to perform a comparison of the 32-121 and 32-124 cell lines, previously demonstrated to have differing production stabilities, revealing distinctions in population dynamics between the two cell lines at an early generation.

4.3 Materials

4.3.1 Cell Culture Media for Beacon® Runs

Bulk and Load media: CD OptiCHO™ Medium (Light Protected, Gibco, 12681011), 15% Conditioned Media, 5% commercially available animal component free (ACF) medium supplement, 4 mM L-glutamine (Gibco, 25030-024), 4mM GlutaMAX™ Supplement (Gibco, 35050061), 0.1% Pluronic™ F-68 Non-ionic Surfactant (Gibco™, 24040032), 2% B-27™ Plus Supplement (Gibco, A3582801)

4.3.2 Cell Lines

The two cell lines analysed in this chapter ("32-121" and "32-124") were previously generated in the same CLD process (during a previous FDB study) using random integration of the expression vector and produce the same monoclonal antibody (an IgG4k). The production stabilities of the two cell lines were characterised in the previous chapter.

4.4 Methods

4.4.1 Cell Culture

See previous chapter for details on cell culture methods.

4.4.2 Analysis of Cell Lines using the Berkeley Lights Beacon® instrument

The recombinant CHO cell line was loaded onto a single OptoSelect™ Chip (Design 1750, Berkeley Lights, Emeryville, CA), during which single cells are deposited into individual pens with nanolitre volumes (nanopens), in accordance with procedures provided by Berkeley Lights Inc. OptoElectro positioning settings and scripts for loading and exporting cells were provided by Berkeley Lights Inc as part of the Opto™ CLD 2.0 Workflow.

Cells were cultured on the OptoSelect™ chips for 5 days. Repeated imaging and cell counting were conducted using the integrated 4X microscope and camera on the Beacon® instrument. Evidence of single-cell loading was

achieved by collecting an image of a single cell in a nanopen immediately after load operation. Cell count measurements were taken at 4-hour intervals after loading. Cell count images of individual nanopens were also collected on days 3, 4 and 5.

Secretion assays were conducted on days 3, 4 and 5 using Spotlight™ Human Fc assay reagent and assay scripts provided by Berkeley Lights Inc. Relative productivity measurements for individual nanopens were calculated by normalizing the assay data to the average productivity score detected for empty pens in the chip. Selected pens were exported using OEP™ to move cells out of pens followed by flushing off the chip into a 96-well microtiter plate prefilled with proprietary cloning media.

Caroline Wardrope operated the Berkley Lights Beacon® Optofluidic System in collaboration with the Edinburgh Genome Foundry. Dr. Caroline Wardrope exported the raw data files which were subsequently analysed by James Donaldson. This process was repeated for all Beacon® runs presented in this chapter.

4.4.3 Calculations

4.4.3.1 Doubling Time

Doubling time for each nanopen was calculated using the following equation:

$$\mathbf{T_d} = (\mathbf{t_{2-}}\,\mathbf{t_1}) \times \frac{\mathbf{Ln(2)}}{\mathbf{Ln}\left(\frac{\mathbf{N_2}}{\mathbf{N_1}}\right)}$$

$\mathbf{T_d} =$ Doubling time of the Population

$\mathbf{t} =$ time point

$\mathbf{N} =$ Number of cells

4.4.3.2 Cell Specific Productivity

Cell Specific productivity was calculated using the following equation:

$$\mathrm{Q_{p_t}} = \frac{\mathrm{AU\ Score_t}}{\mathrm{N_t}}$$

$\mathrm{Q_{p_t}} =$ Cell specific productivity for a nanopen at time, t

$AU\ Score_t = AU\ Score\ of\ the\ nanopen\ at\ time, t$

$N_t = Number\ of\ cells\ in\ the\ nanopen\ time, t$

4.4.4 Modelling Changing Cell Populations over 60 Generations

4.4.4.1 Starting Cell Population

Simple models to simulate how populations of cells may change over a period of 60 generations were generated. Each simulation began with 100 cells in an imaginary flask. The cell population consisted of 2 or 9 different subpopulations, with different doubling times and cell specific productivities.

Initially, cell growth was simulated for 3 days, during which continuous exponential growth was assumed for all subpopulations. To calculate the cell counts of each subpopulation at 12-hour time intervals during the 72-hour growth period, the exponential growth equation was used:

$$\mathbf{N_{S_t}} = \mathbf{N_{S_{t-12}}} \times \mathbf{e^{12r}}$$

$\mathbf{N_{S_t}} =$ Number of cells in the subpopulation at time, t (hours)

$\mathbf{r} =$ Growth rate of the subpopulation, calculated as $\frac{\mathbf{Ln(2)}}{\mathbf{T_d}}$

$\mathbf{T_d} =$ Doubling time of the subpopulation (hours)

Total cell count was determined by summing the cell counts of the individual subpopulations enabling the proportion of each subpopulation in the total cell population to be calculated. To replicate a routine subculture procedure, the total cell count was reduced to 100 cells after 72 hours. Following the subculture event, the cell counts were recalculated for each subpopulation, equal to the previously determined percentages of each subpopulation in the total population at the time of subculture. This approach ensured that the relative proportions of each subpopulation were maintained after the 100-cell reset. To keep track of the of the overall generation number of the cell population, the following equation was used:

$$\mathbf{G_{T_t}} = \mathbf{G_{T_{t-12}}} + \frac{\mathbf{Ln(NT_t)} - \mathbf{Ln(NT_{t-12})}}{\mathbf{Ln(2)}}$$

$\mathbf{G_{T_t}} =$ Generation number of the total population at time, t (hours)

$\mathbf{N_{T_t}} =$ Total Number of Cells in the Population at time, t (hours)

This process of simulating cell growth for 3 days, followed by a 100-cell reset, was repeated until the total population reached generation 60.

4.4.4.2 Batch Cultures

To simulate changes in antibody productivity of the cell population over 60 generations, batch cultures were modelled for the generation 0, 15, 30, 45 and 60 populations. For each simulated batch culture, 100 cells were seeded into an imaginary flask, using the same proportion of each subpopulation as the original population at that generation timepoint. Exponential growth was assumed throughout the batch culture. To assess the productivity of the cell population during the batch cultures, the IVC for each subpopulation was first determined by calculating the area under the growth curve the equation below:

$$\mathbf{IVC\ (cells/\ hour)} = \int_{\mathbf{0}}^{\mathbf{t}} [\mathbf{N_{S_{t=0}}} \times \mathbf{e^{rt}}]\ \mathbf{dt}$$

$\mathbf{N_{S_{t=0}}} =$ Number of cells in the subpopulation at time, $\mathrm{t} = 0$ hours

$\mathbf{t} =$ time of batch culture (hours)

$\mathbf{r} =$ Growth rate of the subpopulation, calculated as $\dfrac{\mathbf{Ln(2)}}{\mathbf{T_d}}$

$\mathbf{T_d} =$ Doubling time of the subpopulation (hours)

Having determined the IVC, the productivity of each subpopulation during the batch culture was subsequently calculated using the equation below:

$$\mathbf{Productivity} = \mathbf{IVC} \times \mathbf{Q_p}$$

$\mathbf{Q_p} =$ Cell specific productivity (mAb molecules/ cell/hour)

The total productivity for the cell population from the batch culture was calculated by summing the productivities of the subpopulations.

4.5 Results

4.5.1 A Simple Model to Illustrate a Population of Cells Consisting of 2 Subpopulations

To delve deeper into the concept of fast-growing low-producing subpopulations of cells, a hypothetical population of cells derived from a standard CLD process, producing the same mAb and undergoing routine subculture was considered. This population of cells consisted of two subpopulations: subpopulation A and subpopulation B (Figure 4.3A). Subpopulation A constituted 99% of the cell population and was characterized by a 24-hour doubling time and a specific production rate of 10 antibody molecules per cell per hour (10 mAbs/cell/hour). In contrast, subpopulation B represented just 1% of the cell population, with a 20-hour doubling time and a specific production rate of 4 antibody molecules per day (4 mAbs/cell/hour). Within the framework of these hypothetical conditions, the objective was to simulate changes in the cell population over 60 generations, a timeframe typical of CHO cell stability studies. Starting with an imaginary flask containing 100 cells, the cell population was allowed to grow for 3 days, followed by a resetting of the cell count to 100, emulating a standard subculture procedure. This reset mimicked a standard subculture procedure, involving a cell count measurement, transfer of 100 cells to a new flask, and addition of fresh media to achieve the desired volume. Consequently, the changing proportions of subpopulation A and subpopulation B were preserved with each 100-cell reset. This simulation was continued until the cell population had reached 60 generations (also known as population doublings).

Over 60 generations, subpopulation B would gradually outcompete subpopulation A (Figure 4.3B). Given the stated conditions, the population would reach 60 generations after 19 subcultures (Figure 4.3C). After approximately 34 population doublings, subpopulation B would start to account for a greater proportion of the cell population than subpopulation A. After 60 generations, subpopulation B would account for 96% of the population and subpopulation A would account for 4% of the population.

Additionally, the total cell count before the population is reset to 100 cells would increase before each passage, indicating an increase in growth rate over time (Figure 4.3D).

To replicate a CHO cell stability study, batch cultures were modelled at generations 0, 15, 30, 45, and 60. The simulation began with 100 cells in an imaginary flask, using the same proportion of subpopulations A and B as the original population at that generation timepoint. These cells were allowed to grow exponentially for 120 hours, emulating a batch culture (Figure 4.3E). The final product concentration for each batch culture was calculated by multiplying the Integral of Viable Cell Density (IVC) for each subpopulation by its specific productivity, then summing these results. Due to the increase in the proportion of subpopulation B in the population over time, the later generation batch cultures reached a higher cell count after 120 hours. However, the final product concentration declined over the 60-generation time period, indicating instability (Figure 4.3F). Notably, the product concentration showed minimal change between generations 0 and 15, with significant shifts occurring between generations 15 and 45. After 60 generations, the product concentration dropped below 70% of its initial value at generation 0. This meant that the hypothetical cell line had passed the threshold for being defined as an unstable cell line.

In light of the presented thought experiment, the expectation was that the Beacon® could be used to identify mixed subpopulations in cell populations exhibiting unstable productivity. While this thought experiment utilised hypothetical values for doubling time, productivity and percentage of the total cell count for two subpopulations in a hypothetical cell line, it provided a basis for further study. By incorporating experimentally derived parameters from stable and unstable cell lines using the Beacon® system, it was anticipated that a more robust model could be developed. Such a model could enhance our comprehension of CHO cell population dynamics, thus providing valuable insights for predicting production instability.

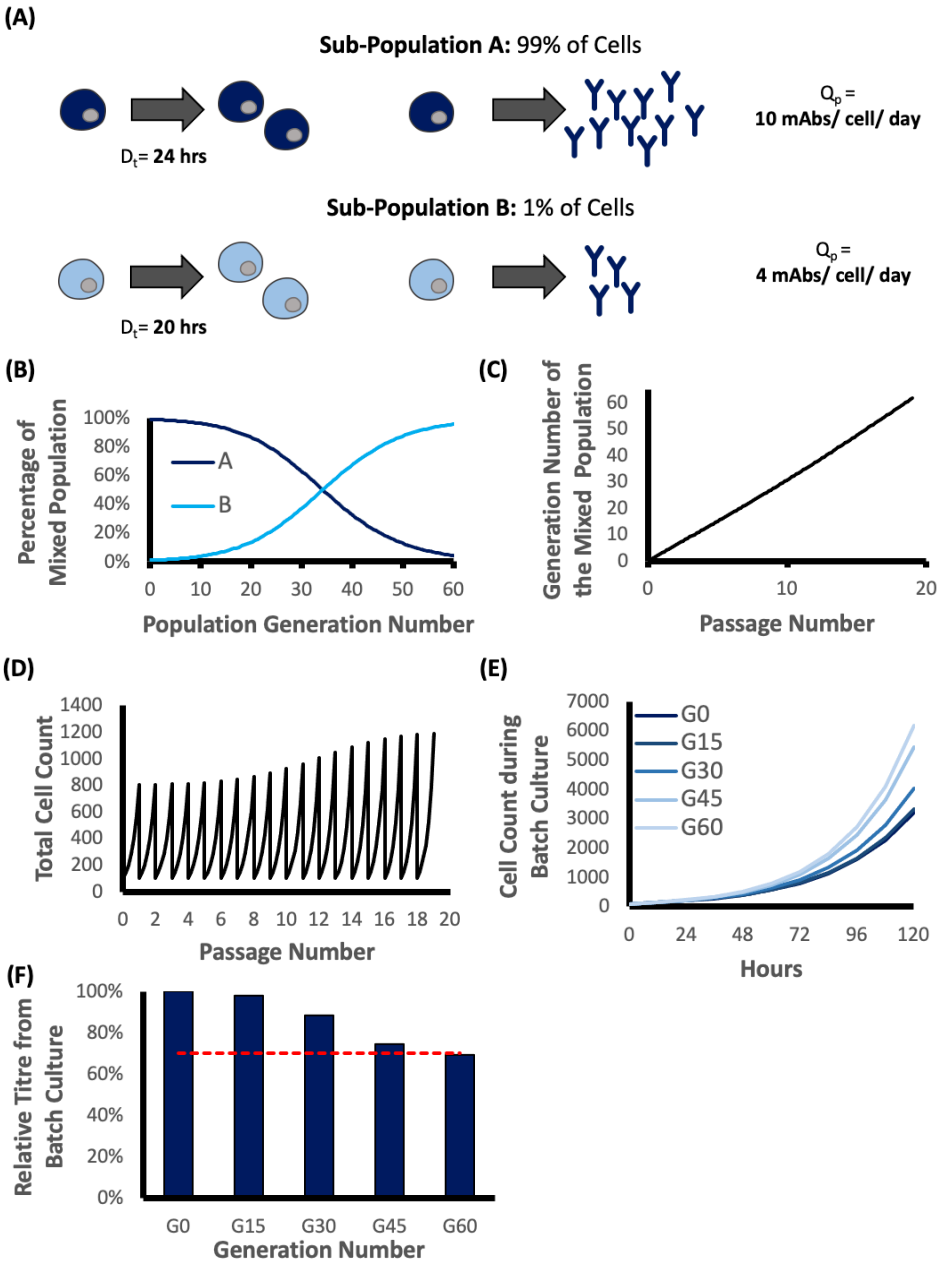


Figure 4.3 – Modelling Growth and Productivity Changes in a Mixed CHO Cell Population over 60 generations.

(A) A hypothetical population of 100 recombinant CHO cells, derived from a monoclonal population of antibody-producing cells, consists of 2 sub-populations with distinct growth rates and specific productivities: sub-population A and sub-population B. Sub-population A accounts for most of the population (99/100 cells), has a doubling time (Dt) of 24 hours and a specific productivity of 10 molecules of monoclonal antibody per cell per

(continued on next page)

(Figure 4.3 continued) day. In contrast, sub-population B represents a minor sub-population (1/100 cells), with a lower specific productivity of 4 molecules of monoclonal antibody per cell per day and a shorter doubling time of 20 hours. **(B)** The mixed cell population is modelled over a period of 60 generations under routine subculture. Every 3 days, a proportion of the cell population is removed, returning the mixed population count to 100 cells (modelling a standard passaging process). The changes in the proportion of each sub-population are monitored over time. **(C)** The hypothetical population is passaged under routine subculture until it reaches 60 generations. **(D)** The cell count of the mixed population is monitored over 19 passages. **(E)** A representative sample of 100 cells from the mixed population at different generations (0, 15, 30, 45, and 60) is used to seed 5-day batch cultures, and the growth curves of each population are plotted. **(F)** The integral of cell density for each batch culture is calculated and used to determine a titre, which is represented in the bar graph for each generation number. The dashed red line indicates 70% of the titre from the batch culture at generation 0, serving as a reference for comparison with subsequent generations.

4.5.2 Generation 15 was Chosen as the Early Timepoint for Comparing the 32-124 and 32-121 Cell Lines (Minus MTX)

The 32-121 and 32-124 cell lines were chosen for further analysis on the Beacon®. In the previous chapter, the 32-121 cell line showed an increase in growth rate over 60 generations during routine subculture in both the presence and absence of MTX. In contrast, the 32-124 cell line did not exhibit a significant increase under either selection condition. This was in agreement with other studies which have shown similar increases in growth rate during LTC, particularly in cell lines which exhibit unstable productivity. As previously discussed, this phenomenon is often attributed to subpopulations of cells characterized by rapid growth and reduced productivity, gradually outcompeting the wider cell population, leading to a decrease in productivity over time. It was anticipated that the Beacon® could be used to identify such populations in the 32-121 cell line, that are not present in the 32-124 cell line, giving an early indicator of instability.

Although the 32-121 cell line exhibited an increased growth rate during routine subculture, there were no significant changes in IVC or maximum VCD between early and late generation FBCs when cultured without MTX (see previous chapter). However, when cultured with MTX, the 32-121 cell line exhibited a significant decrease in both IVC and maximum VCD. The growth characteristics of a cell line during FBC can be influenced by various factors, including the higher cell densities in FBC compared with routine subculture and the accumulation of toxic metabolites such as lactate. Changes that occur in the cell line during LTC may influence its ability to cope with such stresses, potentially counteracting the effects of an increasing proportion of fast-growing low-producing cells in the population. Therefore, it was anticipated that this result would not eliminate the possibility of the existence of a fast-growing low-producing subpopulation in the 32-121 cell line.

The addition of MTX to culture media has been associated with an increased mutation rate, which can result in variations in the amino acid sequence of the desired product (Guo et al., 2010; Zhang et al., 2016). Consequently, to reduce toxicity concerns, scaling up cell culture volumes in the absence of MTX is preferred. In industry, a stability study is usually conducted on cell lines cultured in the presence and absence of MTX, to determine whether MTX is required to prevent productivity loss over 60 generations. To reduce CLD timelines, predictive strategies which enhance researchers' confidence in a cell line's stability without MTX will thus be particularly advantageous. Therefore, given that MTX appeared to have a negative impact on the growth of the cell line, the decision was made to conduct the Beacon® analysis on cells that had been cultured in the absence of MTX.

Generation 15 was chosen as the timepoint for analysing the cell lines using the Beacon® for multiple reasons. Firstly, generation 15 is too early to detect a decrease in titre using a traditional stability study approach because such methods typically require fed-batch cultures to be performed at generation 10, 25, 45 and 60. However, if a cell line exhibited a high probability of being stable during the generation 15 Beacon® screen, it could be progressed to subsequent manufacturing stages whilst the rest of the stability study is completed. Therefore, successful inclusion of the proposed method for early instability detection could enable analysis of candidate cell lines at generation 15 using the Beacon® system, representing a 75% reduction in the timeline of a typical 60-generation stability study. From a practical perspective, generation 15 was the earliest feasible timepoint for Beacon® analysis, since it allowed enough time for the generation 0 vials to be thawed, additional vials to be produced, and the cell lines to recover.

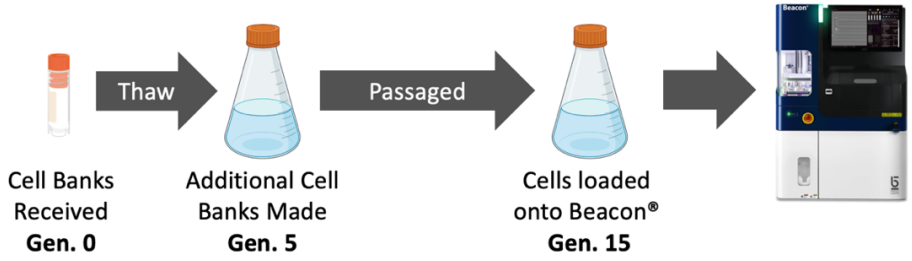


Figure 4.4 - Analysis of Generation 15 Cell Lines Using the Berkley Lights Beacon® Optofluidic System.
A cell bank (defined as generation 0) for each of the 32-121 and 32-124 cell lines were thawed and passaged to reach generation 15. The Berkley Lights Beacon® Optofluidic System was subsequently used to analyse the generation 15 cell lines.

With this in mind, the 32-121 and 32-124 cell lines were grown to generation 15 under routine subculture in the absence of MTX selection (Figure 4.4). Once the cells had reached generation 15, the cell lines were loaded onto the Beacon® for analysis (refer to section 4.4.2 for methodology details). Single cells were deposited into individual nanopens of the chip. The growth and productivity of the cells in each nanopen was monitored over a period of 5 days.

4.5.3 Optimisation of Cell Growth on the OptoSelect™ chip

Previously, successful growth of suspension CHO-K1 antibody producing cells on the OptoSelect™ chip has been achieved when cells were loaded and grown in normal growth media supplemented with 20% conditioned media (from previous experience of Dr. Caroline Wardrope and the Genome Foundry). Conditioned media, obtained by centrifuging the cell culture and filtering the supernatant, contains various secreted growth factors, nutrients and signalling molecules which can promote the survival and growth of a single cell colony. However, when generation 15 32-124 cells were loaded onto the chip using this media composition, the clones failed to divide and subsequently died (Figure 4.5A, run 1). For run 1, cells were transferred onto the chip from a culture flask on the day of passage with a viable cell density of 8.06 x 10^6 cells/ml. Due to the high cell density of the culture, it was

thought that the cells may have entered the stationary phase. To ensure that cells were in the exponential phase when loaded onto the chip, the experiment was repeated with the same media but with cells taken from the culture the day after a passage (Figure 4.5, run 2). Although the cells divided initially, on day 2 they started to swell and die.

Therefore, optimization of media composition was necessary to ensure successful growth of the Apollo X-derived recombinant cell lines in the Beacon® system. Due to the cost and time restraints of the study, it was not feasible to conduct a series of Beacon® runs with single changes to the media composition for each run. Instead, the decision was made to implement multiple changes (if necessary) between runs. Media optimization experiments were thus conducted on generation 15 32-124 cells.

Although the addition of B27 supplement and ACF (standard supplements for mammalian cell cloning) to the bulk and load media did not prevent cells from dying on day 2 (Figure 4.5A, run 3), improvements in on-chip growth were seen during days 0-2 in run 4 when the basal media was changed from FDB-MAP to OptiCHO and GlutaMAX™ was changed to L-glutamine (Figure 4.5A, run 4). FDB-MAP is FDBK's proprietary media for routine subculture of recombinant cell lines and is also used by FDBK for single-cell cloning. However, FDB-MAP is sensitive to light which can impact the outgrowth of clones. In contrast, OptiCHO was thought to be less sensitive to light, which may have led to improved growth during the early stages of the Beacon® run. However, cell growth slowed after day 2 and the total cell count for most pens on days 4 and 5 was less than 10. In the final attempt to promote continued on-chip cell growth over the 5 days, the concentration of ACF was increased from 2.5% to 5%, the media was protected from light, and a mixture of glutamine and GlutaMAX™ was used (Figure 4.5A, run 5). With these changes, the majority of pens showed successful growth over the 5 days of the Beacon® run (Figure 4.5B).

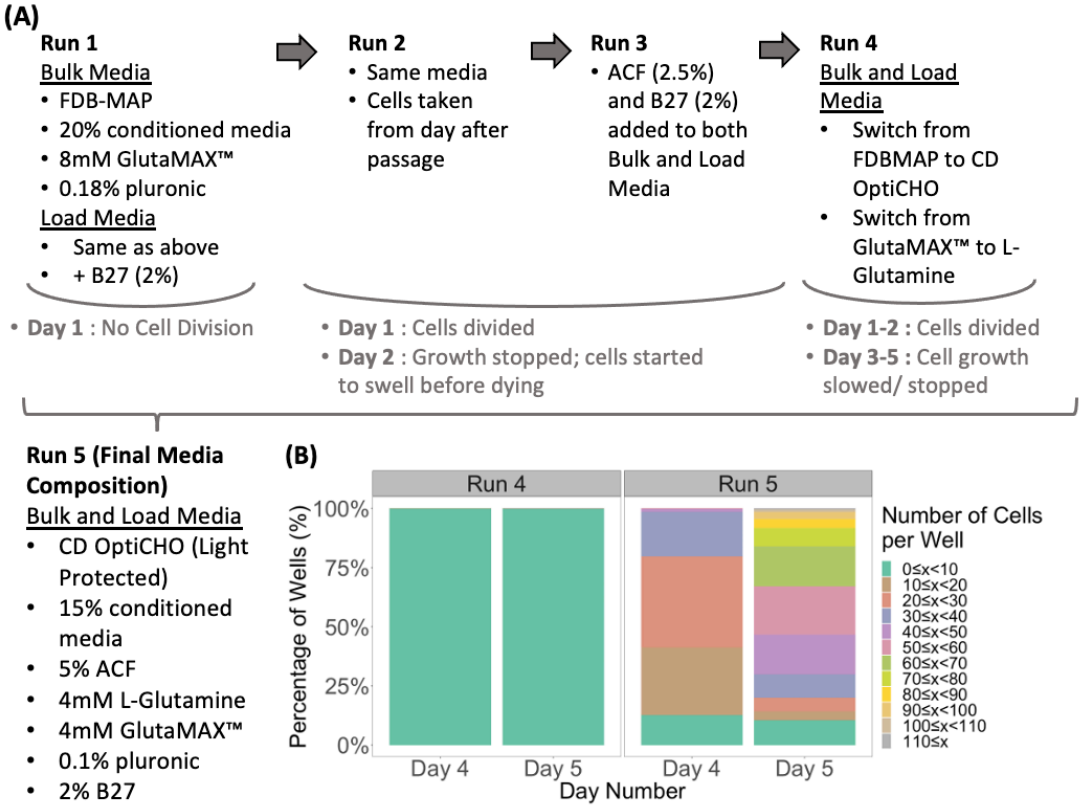


Figure 4.5 - Optimisation of Growth in the OptoSelect™ chip using the 32-124 Cell Line.

(A) Load and feed media for the OptoSelect™ chip were optimized over successive Beacon® runs. The load media is used to load the cells onto the OptoSelect™ chip. The feed media is used to feed the cells once on the OptoSelect™ chip. The data gathered from runs 1-4 was collated to inform the composition of the load and feed media for run 5. **(B)** Comparison of day 4 and day 5 cell counts between run 4 and run 5.

4.5.4 Filtering Process Prior to Beacon® Data Analysis

After media optimization, a second Beacon® run was performed on the 32-121 cell line, that had also been grown under routine subculture to generation 15, using the same optimized media conditions that were used for the 32-124 cell line. Before any comparisons were made between the generation 15 cell lines, the datasets were filtered to ensure the analysis was focused on nanopens that had exhibited successful growth (Figure 4.6). Each chip contains 1758 nanopens. Firstly, nanopens that were not loaded with 1 cell were excluded. Nanopens containing cells with a mean doubling time of >30 hours (i.e. very slow growing), which is much higher than the doubling time for a standard CHO cell culture (<24 hours), were also excluded from the analysis. This was to ensure that only the nanopens which exhibited successful growth were analysed. Day 4 was selected for the filtering process to ensure that the cells had not become overgrown in the nanopen but had had enough time to develop distinctive characteristics. All further analysis was conducted using this filtered dataset.

4.5.5 Comparison of Mean Doubling Time and Cell Specific Productivity Between the 32-121 and 32-124 Cell Lines

For an initial comparison between the 32-121 and 32-124 cell lines, the mean doubling times (Figure 4.7C) and cell-specific productivities (Figure 4.7D) on days 3, 4, and 5 were evaluated. For each cell line, the relevant measurement for each of the filtered nanopens was calculated, using the equations provided in section 4.4.3, and the overall mean for the entire chip was subsequently determined. This analysis aimed to provide a broad comparison of the behaviour of the two cell lines during their respective Beacon® runs.

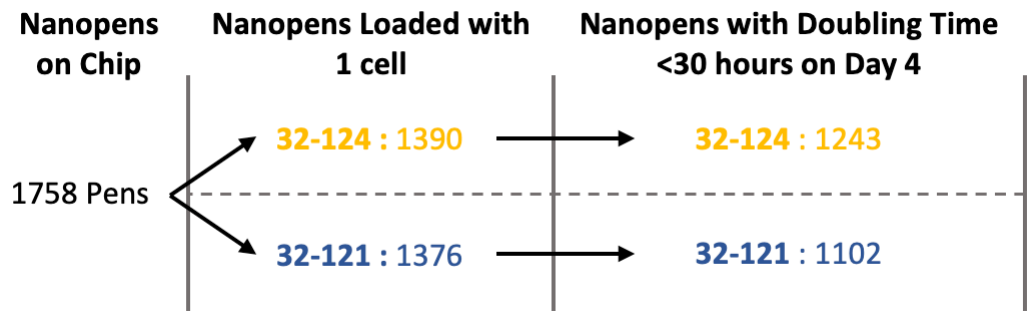


Figure 4.6 - Filtering Process Prior to Beacon® Data Analysis
Each chip contains 1758 nanopens. Prior to analysis the data was filtered to select nanopens which had been successfully loaded with 1 cell and exhibited a day 4 doubling time of less than 30 hours.

Both cell lines exhibited similar doubling times throughout the duration of the Beacon® culture (Figure 4.7A). However, on day 4, the 32-124 cell line demonstrated a slightly higher (yet statistically significant) doubling time in comparison with the 32-121 cell line, suggestive of slower growth. Furthermore, both cell lines displayed a marginal upswing in their doubling times between days 3 and 5, indicating slower growth. Overcrowding in the nanopens could have contributed to this trend. Interestingly, cell-specific productivity decreased during the 5-day Beacon® culture in both cell lines (Figure 4.7B), mirroring the trend observed in the FBC data (see previous chapter).

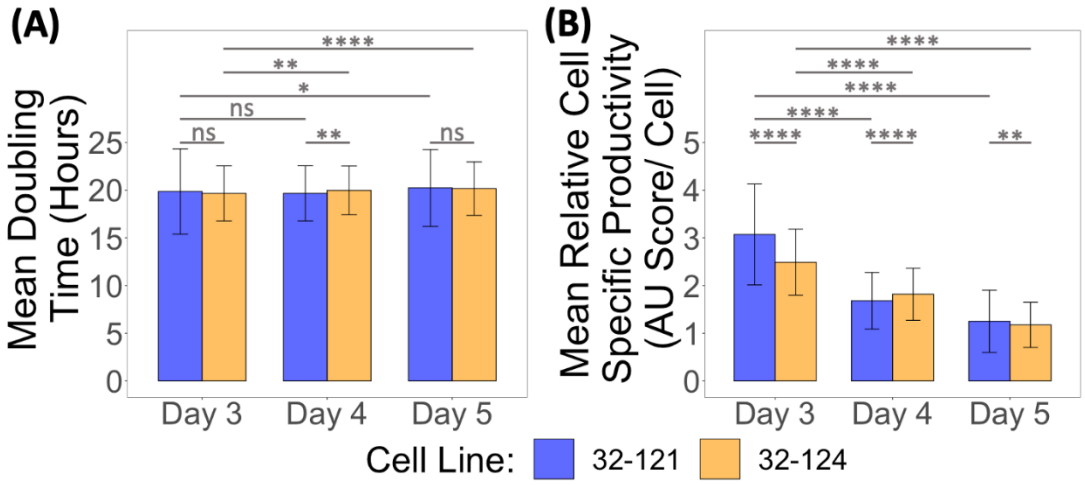


Figure 4.7 - Comparison of Mean Doubling Time and Cell Specific Productivity between the 32-121 and 32-124 cell lines

The mean (A) Doubling Time and (B) Relative Cell Specific Productivity across the selected nanopens on the chip (following the filtering process outlined in Figure 4.6) was calculated for each cell line. Data represented as the mean ± standard deviation across the nanopens. $^{ns}P > 0.05$, *P ≤ 0.05, **P ≤ 0.01, ***P ≤ 0.001, ****P ≤ 0.0001 using a Two Sample Z-Test which compares the means of two populations when the population standard deviations are known.

4.5.6 Beacon® Analysis Reveals an Increase in the Proportion of low-producing fast-growing cells in the 32-121 cell line on day 4

The next step in the analysis aimed to determine whether the Beacon® system could effectively identify subpopulations of cells with distinct growth rates and specific productivities within the broader cell population. As previously hypothesized, it was expected that the 32-121 cell line might contain a subpopulation of cells characterized by rapid growth but low productivity, potentially leading to a decline in overall mAb productivity over time.

To pursue this, the day 4 mean doubling time of the cells for each nanopen were plotted against the day 4 relative cell specific productivity for the same nanopen (Figure 4.8A). Each data point in Figure 5.6 thus represents a single nanopen. Initial observations revealed that neither cell line exhibited visibly distinct subpopulations within the broader population. Instead, the populations comprised of cells displaying a continuous spectrum of doubling times and specific productivities. Notably, the data indicated a strong correlation between doubling time and specific productivity ($p<0.0001$ for both cell lines). This finding was consistent with prior research, which suggested that the metabolic demands of recombinant protein production could adversely affect growth characteristics.

To further analyse the cell populations, the graph was divided into a 9-square grid, and the percentage of nanopens falling into each grid square (labelled A-I) was determined (Figure 4.8). The selection of grid parameters was based on the stability exhibited by the 32-124 cell line over a period of 60 generations, as discussed in the previous chapter. It was assumed that this cell line was unlikely to contain a subpopulation of low-producing, fast-growing cells that would significantly impact its long-term growth and productivity. With this in mind, grid square E, encompassed the majority of the 32-124 cell population (approximately 75%). In contrast, grid square C, which was used to identify low-producing, fast-growing cells, only contained 1.9% of the 32-124 cell population, and served as a reference point for comparison. Interestingly, the 32-121 cell line exhibited a higher proportion of cells in grid square C (10.9%) than the 32-124 cell line. This initial observation raised the possibility that the increased presence of this subpopulation within the broader population could serve as an early indicator of instability in the 32-121 cell line.

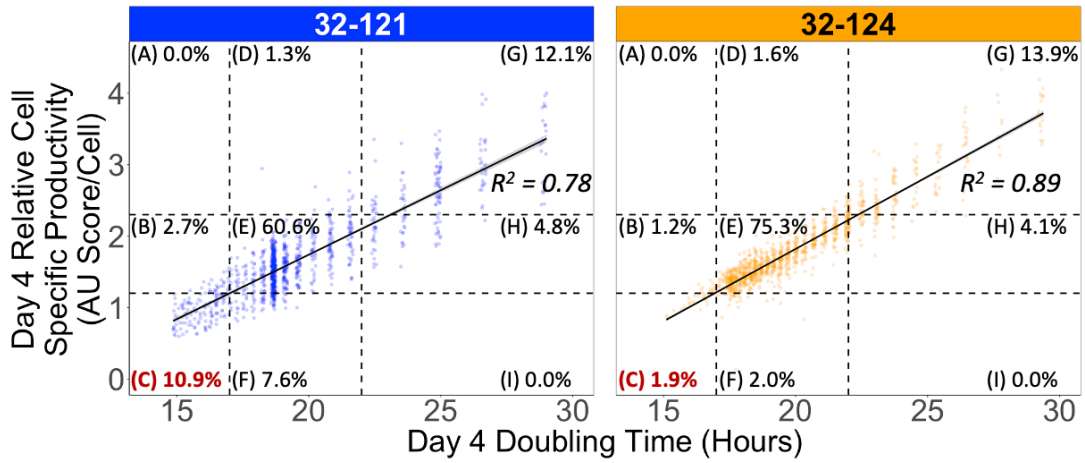


Figure 4.8 – Beacon® Analysis Reveals an Increase in the Proportion of low-producing fast-growing cells in the 32-121 cell line.

The association between doubling time and specific productivity was established through a linear regression model. Solid black line indicates the linear regression line. Each data point corresponds to an individual pen. The dashed black lines delineate the grid squares employed for segregating the cell population, labelled A-I. Grid square C represents the low producing fast growing subpopulation is highlighted in red.

4.5.7 The Beacon® Data Should be Normalised Prior to Analysis

For further comparison, the same graphs were constructed for the day 3 and day 5 datasets, employing the same gridlines as in the previous analysis (Figure 4.9). It is worth noting that the axes needed to be extended to accommodate cells with different doubling times and specific productivities on days 3 and 5 to day 4. Consistent with the day 4 data, a significant correlation between doubling time and cell-specific productivity was observed on days 3 and 5.

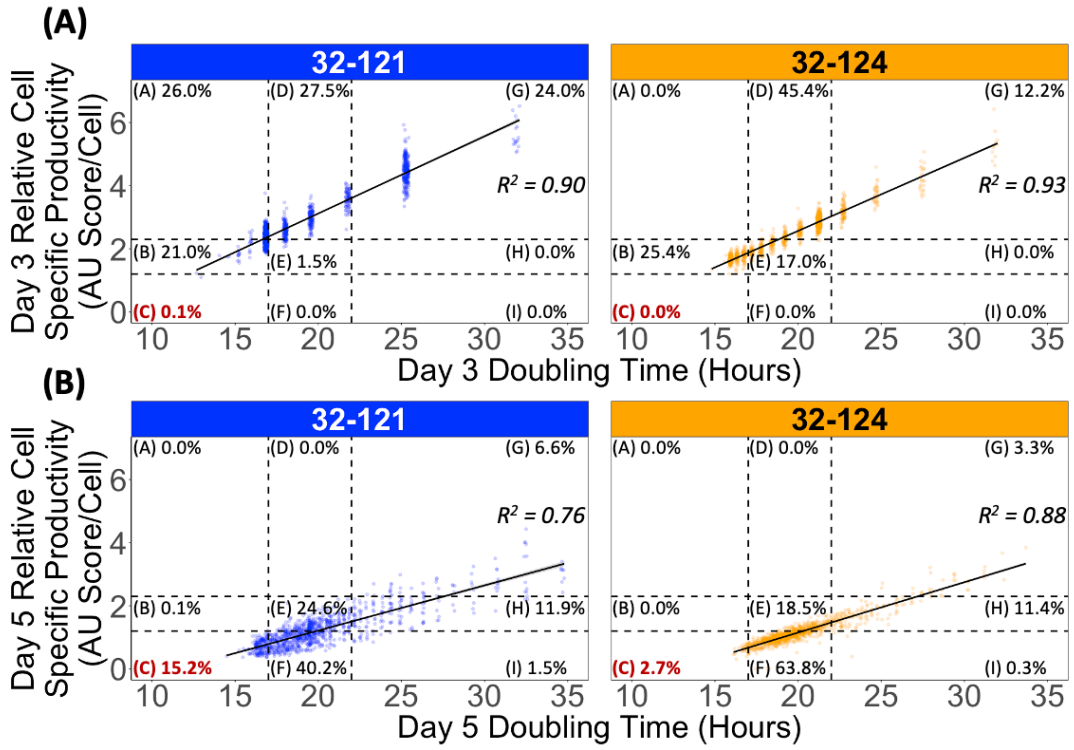


Figure 4.9 - Analysis of the day 3 and day 5 data reveals challenges for the Beacon® stability screening method.
The association between doubling time and specific productivity for the **(A)** day 3 and **(B)** day 5 data. The association between doubling time and specific productivity was established using a linear regression model. Solid black line indicates the linear regression line. Each data point corresponds to an individual pen. The dashed black lines delineate the grid squares employed for segregating the cell population, labelled A-I. The percentages indicate the proportion of cells in the grid square, labelled A-I. Grid square C represents the low producing fast growing subpopulation is highlighted in red.

Interestingly, the data from day 3 exhibited a less continuous distribution of doubling times compared with those of days 4 and 5. In contrast with days 4 and 5, during which the data points displayed a continuous range of doubling times and specific productivities, the datapoints from day 3 formed clusters with similar doubling times. This discontinuity was attributed to the shorter culture period, resulting in nanopens displaying similar cell counts and productivities. During FBCs, it often takes time for differences between cell lines to become apparent. For example, in the previous chapter, there were minimal differences in both growth (section 3.5.7) and volumetric productivity (section 3.5.5) profiles between the early and late FBCs for all cell lines during the first 4 days. However, after day 4, the growth and productivity profiles of the 32-121 cell line began to diverge. Similarly, it is likely that it takes time for cells within the population to develop distinct growth and productivity characteristics, as measured by the Beacon system, resulting in the continuous range of doubling times and specific productivities observed from day 4 onwards.

Although the day 4 data appeared the most suitable for further analysis, because the culture period was long enough for nanopens to exhibit diverging growth patterns but short enough to ensure nanopens had not overgrown, the day 3 and day 5 data raised some interesting challenges for future cell line assessments. In the initial day 4 analysis, gridlines were designed to encompass the majority of 32-124 cells in grid square E, with only a small percentage in grid square C. The aim was to use this grid as a reference point for identifying low-producing fast-growing subpopulations in other cell lines. This identification was based on observing whether there was a higher percentage of cells within grid square C, relative to the 32-124 cell line, in the cell line being tested. However, due to variations in the mean doubling time and specific productivity of both cell lines during the Beacon® culture, grid square E did not encapsulate the majority of 32-124 cells for either day 3 (17.0%) or day 5 (18.5%). This made it more challenging to identify subpopulations of low-producing fast-growing cells in the 32-121 population. Consistent with the day 4 data, the day 5 data showed an

increase in the proportion of low-producing fast-growing cells in the 32-121 population (15.2%) compared with the 32-124 population (2.7%). However, on day 3, only 0.1% of the 32-121 cells and 0% of the 32-124 cells fell into grid square C. In contrast with days 4 and 5, a higher proportion of the 32-121 cells (26.0%) occupied grid square A on day 3, indicating rapid growth and high cell-specific productivity. This is due to the 32-121 cell line having a significantly higher mean cell-specific productivity on day 3 than days 4 and 5.

To advance the development of a method for detecting early signs of production stability or instability in recombinant cell lines, future experiments may utilize the 32-124 cell line as a reference point to predict the production stability of other cell lines. However, if the cell line under examination exhibits a higher mean specific productivity (similar to what the 32-121 cell line demonstrated on day 3) or a longer mean doubling time compared to the 32-124 cell line, it is plausible that fewer cells will fall into grid square C during Beacon® analysis. Consequently, even if the cell line contains a substantial proportion of lower-producing, faster-growing cells in comparison with the overall population, the current Beacon® method may not detect them effectively. Conversely, in cases where there is a small but notable increase in cell-specific productivity and a modest decrease in mean doubling time, as observed in the 32-121 cell line on day 4, the cell population might shift slightly to the left. This could result in a higher subpopulation of cells falling into grid square C and being defined as low-producing and fast-growing. With these scenarios in mind, normalizing the data, achieved by dividing the relative cell-specific productivity or doubling time for each nanopen by the mean relative cell-specific productivity or doubling time for the entire chip, could enhance the comparability of different cell lines.

4.5.8 Normalisation of the Beacon® Data Confirms Previous Conclusions

Following the discussion in section 4.5.7, normalization of doubling time and relative cell-specific productivity was performed for the day 3, 4, and 5 data prior to plotting (Figure 4.10). In line with the previous analysis, the graph

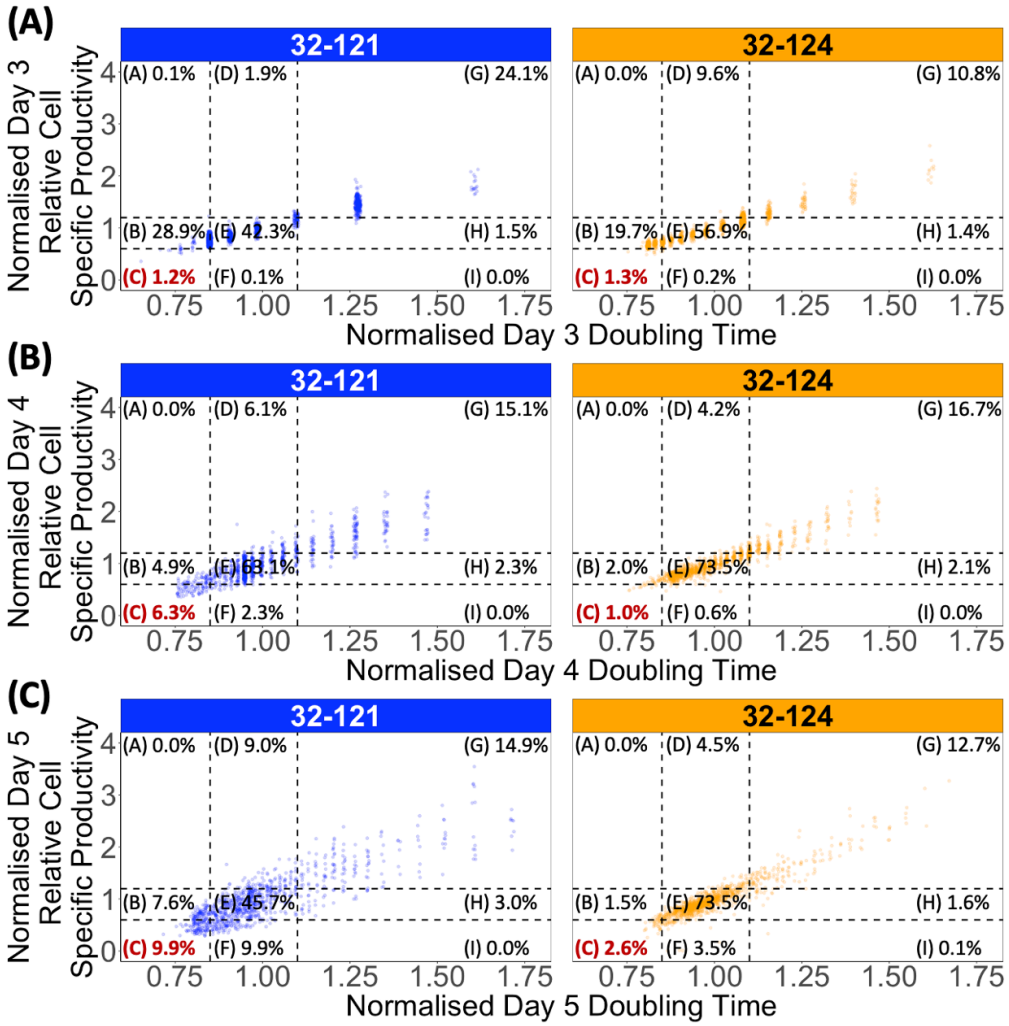


Figure 4.10 - Normalisation Increases the Proportion of Cells in Grid Square E

The association between normalised doubling time and normalised specific productivity for the **(A)** day 3 and **(B)** day 4 and **(C)** day 5 data. The doubling time and cell specific productivity data was normalised for each day by dividing the relative cell-specific productivity or doubling time for each nanopen by the mean relative cell-specific productivity or doubling time for the entire chip. Each data point corresponds to an individual pen. The dashed black lines delineate the grid squares employed for segregating the cell population, labelled A-I. The percentages indicate the proportion of cells in each grid square, labelled A-I. Grid square C represents the low-producing fast-growing subpopulation is highlighted in **red**.

was divided into a 9-square grid with adjusted parameters to ensure that grid square E contained the majority (approximately 74%) of the 32-124 day 4 cell population. The percentage of cells occupying grid square C was used to compare the proportion of low-producing fast-growing cells between the two cell lines.

For the day 3 data, despite a higher proportion of cells falling into grid square E compared with the previous analysis (42.3% for the 32-121 cell line and 56.9% for the 32-124 cell line), a relatively small percentage of cells were encompassed by grid square C for both cell lines (1.2% for 32-121 and 1.3% for 32-124). This was attributed to the shorter culture period, which resulted in nanopens exhibiting similar cell counts and productivities.

On day 4, 6.3% of the 32-121 cells occupied grid square C, while only 1.0% of the 32-124 cells fell into the same grid square. This finding was consistent with the earlier analysis, which demonstrated a 5.7-fold increase in the percentage of cells in grid square C in the 32-121 cell line when compared with the 32-124 cell line.

For the day 5 data, there was a 3.8-fold increase in the percentage of cells occupying grid square C. This increase was lower compared to the previous analysis (section 4.5.6), which showed a 5.6-fold increase in the percentage of 32-121 cells in grid square C compared with the 32-124 cell line. Both the 32-121 and 32-124 cell lines exhibited a significantly lower mean specific productivity on day 5 compared with day 4 (Figure 4.7). Therefore, in the previous analysis without normalisation, the cell population shifted downwards on the graph, resulting in a higher percentage of cells falling into grids C and F (Figure 4.9B). Normalization prevented such a downward shift, leading to a larger proportion of cells in grid E (45.7% for the 32-121 cell line and 73.5% for the 32-124 cell line) compared with the previous analysis. Normalisation is thus likely to help with future comparison studies using the Beacon®.

4.5.9 Increased Variation Around the Regression Line in the 32-121 Cell Line

An additional visible distinction between the two cell lines was the notable variation seen in the 32-121 line around the regression line. On average, the data points appeared to deviate more substantially from the regression line. To confirm this observation, the residual standard error, which measures the standard deviation of the residuals, was calculated (Figure 4.11). Consistent with the initial observations, the residual standard error was higher on days 3, 4, and 5 for the 32-121 cell line, compared with the 32-124 cell line. This observation was intriguing as it suggested a greater degree of clonal variation within the 32-121 cell line. This increased clonal variation could elevate the likelihood of a subpopulation of cells emerging that have lost productivity.

4.5.10 Increased Variation in The 32-121 Cell Line Stems From Increased Variation in Both Growth and Specific Productivity

Given the increased variation observed around the regression line in the 32-121 cell line, the next step in the analysis aimed to identify whether this variation stemmed from differences in doubling time, cell-specific productivity, or both. To test this, the coefficient of variation for doubling time and cell specific productivity was calculated (Figure 4.12). Interestingly, the coefficient of variation for both doubling time and cell-specific productivity was significantly higher on days 3, 4, and 5 in the 32-121 cell line when compared with the 32-124 cell line. This observation is of particular significance because it suggests that the monitoring of both growth and productivity metrics are required for predicting potential cell line instability.

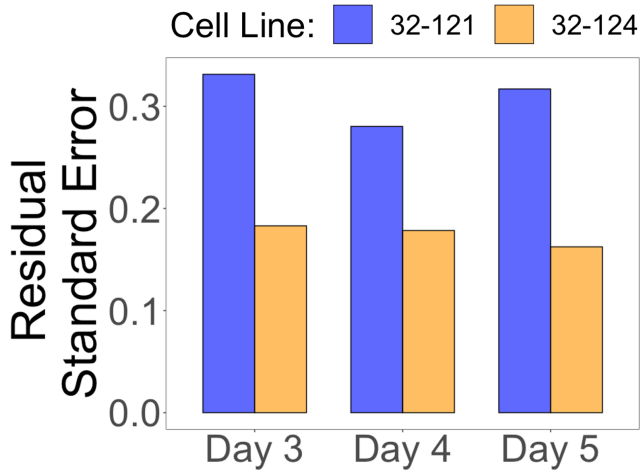


Figure 4.11 - Increased Variation Around the Regression Line in the 32-121 Cell Line.

The residual standard error for the regression lines depicting the relationship between doubling time and relative cell specific productivity (refer to Figure 4.8 and Figure 4.9) was computed for each day.

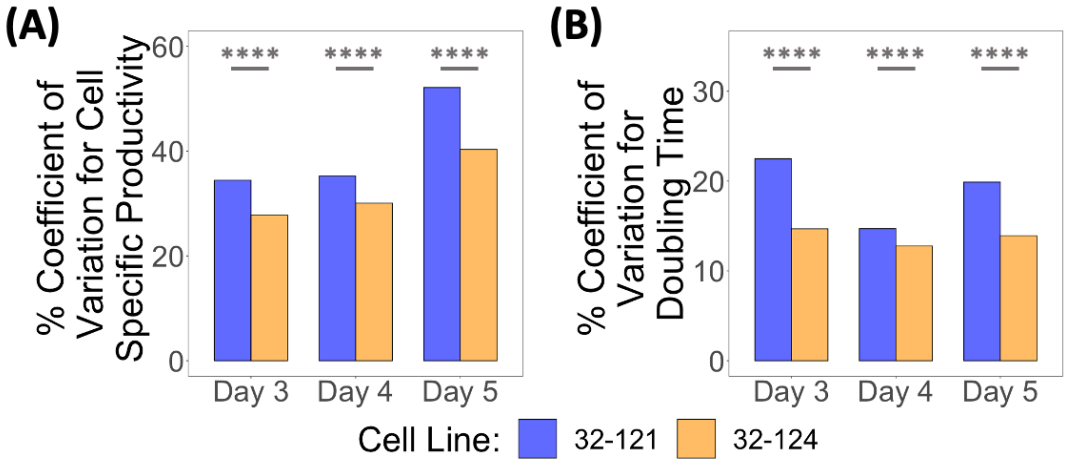


Figure 4.12 - Increase in Clonal Variation is due to Increased Variability in Cell Specific Productivity and Doubling Time

The coefficient of variation for **(A)** Cell Specific Productivity and **(B)** Doubling time was calculated for both cell lines on each day. $^{ns}P > 0.05$, *P ≤ 0.05, **P ≤ 0.01, ***P ≤ 0.001, ****P ≤ 0.0001 using a method to compare coefficients of variation from

4.5.11 Using the day 4 Beacon® Results to Develop an Initial Model of Instability

Our original hypothesis was based on the idea that there was a low-producing, fast-growing subpopulation of cells in the 32-121 cell line, not present in the 32-124 cell line, that gradually outcompeted the rest of the population, leading to a drop in productivity. Using arbitrary parameters for doubling time and cell specific productivity, this point was illustrated by developing a simple model of a population of cells consisting of 2 subpopulations: 99% low-producing fast-growing cells and 1% fast-growing low-producing cells. Originally, it was anticipated that these numbers could be adapted based on the results from the Beacon® runs and used to adapt the initial model for cell line production stability. However, as previously discussed, the Beacon® data did not show distinct subpopulations of cells. Instead, the cells existed on a continuous spectrum of specific productivities and doubling times and thus could not easily be separated into 2 distinct subpopulations.

To interrogate the data further, the graph looking at the association between cell specific productivity and doubling time was divided into a grid, separating the cell population into 9 subpopulations (Figure 4.8). The percentage of cells which fell into each grid square was subsequently monitored and used for comparison between cell lines. While it is important to note that the available data was insufficient at this stage to construct a robust stability prediction model, we were intrigued to explore whether feeding the Beacon® data into an adapted nine-subpopulation model, as an extension of the initial two-subpopulation model, would reveal any discernible differences in production stability when growth over 60 generations was simulated.

To achieve this, a similar methodology as previously detailed in section 4.5.1 was employed using nine subpopulations instead of two. The specific productivities and doubling times of each subpopulation were determined by extracting the centre-point values from each grid square in Figure 4.8. The number of cells allocated to each subpopulation were also derived from the

percentage of cells present in each respective grid square in Figure 4.8. For reference, a summary of the initial cell population for each cell line is provided in Table 1. As previously described, cell growth in routine subculture was simulated over a span of 60 generations. At each generation timepoint, 3-day batch cultures were simulated to gain insights into the changes in population productivity, following the methods outlined in section 4.4.4.

Interestingly, when the parameters from the Beacon® run of the 32-124 cell line (a cell line characterized by stable productivity) were incorporated into the model, no substantial change in productivity was observed over 60 generations (Figure 4.13A). Conversely, when the parameters from the 32-121 cell line (a cell line which exhibited productivity loss) were used, a 10% decrease in productivity over 60 generations was observed. As expected, in both cell lines, subpopulation E was gradually outcompeted by subpopulations B and C (Figure 4.13B). At an early generation, subpopulation E contributed most of the antibody productivity, but after 60 generations, most of the productivity came from subpopulations B and C (Figure 4.13C). Interestingly, the simulated cell population reached a steady-state population after approximately 30 generations, exhibiting minimal changes in both overall productivity and the proportions of each subpopulation beyond this point. However, in contrast, data from the fed-batch cultures in the previous chapter indicated that the decline in productivity of the 32-121 cell line continued between generations 30 and 60. This discrepancy may partially elucidate why, in our experimental findings, the 32-121 cell line exhibited a 27% decrease in productivity over 60 generations (as presented in the previous chapter) compared with the model's prediction of a 10% drop, derived from the Beacon® data. A key assumption made by the model was that each cell, along with its progeny, would consistently maintain the same growth and productivity characteristics throughout the entire 60-generation period. However, it is likely that over this extended period, more 32-121 cells may transition towards a low-producer phenotype, contributing to further decreases in productivity.

Table 4.1 - Parameters used for the 9-Subpopulation Model

Sub-population	Q_p (mAbs/cell/hour)	D_t (hours)	No. of Cells in Starting Population	
			32-121	32-124
A	3.15	16	0	0
B	1.65	16	3	1
C	0.8	16	11	2
D	3.15	19.5	1	2
E	1.65	19.5	61	75
F	0.8	19.5	8	2
G	3.15	26	12	14
H	1.65	26	5	4
I	0.8	26	0	0

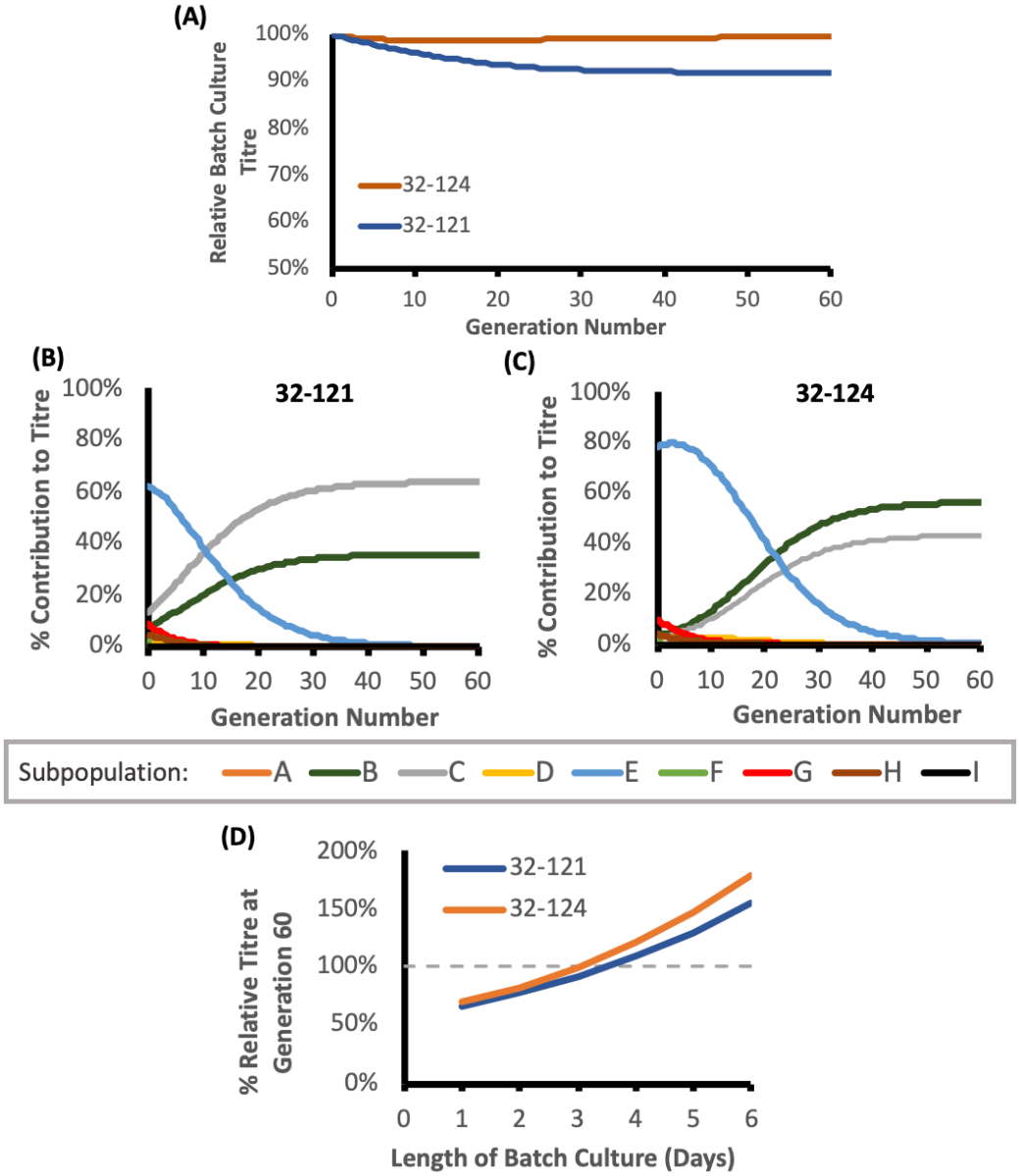


Figure 4.13 - A Simple 9-Cell Model for Production Instability. **(A)** Predicted alterations in the productivity of the 32-121 and 32-124 cell lines using a 9-subpopulation model, with data input from the Beacon® runs at generation 15. Data presented as a relative percentage of the initial expected titre at generation 0. **(B)** Expected change the relative proportions of subpopulations A-I in the 32-121 population. **(C)** Expected change the relative proportions of subpopulations A-I in the 32-124 population. **(D)** Percentage change in productivity between generation 0 and 60 under different batch culture durations.

Another limitation of the model became evident when the batch culture period was extended beyond 3 days, resulting in the relative mAb productivity at generation 60 surpassing 100% compared to generation 0 for both cell lines (Figure 4.13D). This phenomenon occurred because the growth rate of the simulated populations increased over 60 generations, causing cells to reach a higher cell density during the batch culture at generation 60 than at generation 0. As the batch culture duration was increased, the difference in maximum viable cell density of the cell populations between generation 60 and generation 0 increased concurrently. As a result, the impact of the model cell populations experiencing a drop in mean cell-specific productivity was obscured by the parallel increase in cell density, leading to an increase in productivity over time. It is worth noting that real-world batch cultures are considerably more complex than the model's simplified representation. To account for these complexities fully and gain a comprehensive understanding of the relationship between Beacon® data and cell line performance during FBC, a more extensive dataset encompassing a broader range of cell lines and generation timepoints would be essential, as discussed in section 4.6.1. Consequently, further refinements of the model were not pursued at this point.

4.6 Discussion

The objective of this chapter was to initiate the development of a novel method for predicting production instability in recombinant CHO cells, using the Beacon® system. In the previous chapter, two cell lines (32-121 and 32-124) were shown to exhibit different production stabilities over 60 generations. In this chapter, both cell lines were grown under routine subculture to generation 15, in the absence of MTX, before being compared using the Beacon® system. While it is important to acknowledge that this methodology is still under development (limitations and avenues for future Beacon® studies are discussed in Section 5.6.1) some intriguing findings were obtained.

Our initial hypothesis was that the Beacon® system would help identify discrete subpopulations of cells, characterized by distinct cell-specific productivities and doubling times, particularly in the 32-121 cell line which had previously demonstrated productivity loss over 60 generations. However, what emerged was a continuous spectrum of cell-specific productivities and doubling times within both the 32-121 and 32-124 populations. Nevertheless, upon dividing the graphs associating specific productivity and doubling time into a 9-square grid, it became evident that the 32-121 cell line contained a higher proportion of fast-growing, low-producing cells within its population compared with the 32-124 cell line. Normalisation of the data proved to be a useful tool for comparative analyses between the two cell lines, particularly when trying to identify subpopulations of low-producing, fast-growing cells at timepoints when the two cell lines exhibited different mean specific productivities and/or doubling times. While the proportion of low-producing, fast-growing cells in the 32-121 cell line was not higher than the 32-124 cell line from the day 3 data, likely due to the culture time not being sufficiently long for cells to exhibit divergent cell-specific productivities and doubling times, the day 4 and 5 data revealed a ~6-fold and ~4-fold increase respectively in the proportion of such cells in the 32-121 population compared with the 32-124 population.

Furthermore, the 32-121 cell line displayed a higher degree of variability within the population, as indicated by an increase in the residual standard error, in contrast with the 32-124 cell line. This variability was attributed to increased variability in both doubling time and cell-specific productivity.

4.6.1 Experiment Limitations and Future Avenues for exploration using the Beacon® System

Many of the experimental limitations of this chapter stemmed from the substantial financial costs associated with using the Beacon® system, posing a practical constraint on the number of runs that could be conducted. This problem was further exacerbated by the considerable challenges that were faced when optimising the load and bulk media for on-chip cell growth, which required 5 Beacon® runs, limiting the number of additional runs that could be

performed within the constraints of the available resources at the time. Therefore, many of the experimental limitations of this chapter also provide avenues for further study.

Due to practical limitations, only one Beacon® run could be conducted per cell line at generation 15. The inclusion of replicate runs for each cell line would thus strengthen the statistical robustness of the presented findings. One method that could help to reduce the number of replicates required for each cell line would be to export low producing and high producing subclones from the chip and test them in culture, potentially providing additional evidence for their growth and productivity. Downstream transcriptomic and genomic analysis of a range of low producing and high producing clones during routine subculture is subsequently ongoing in the Rosser lab.

It is important to note that the two cell lines were compared based on data obtained from two distinct Beacon® runs. While efforts were made to maintain consistent conditions across both runs, it is important to acknowledge that comparisons were constrained by the potential for variability between these separate Beacon® runs. For future studies, it is worth emphasising that the Beacon® system allows for the loading of two cell lines on a single chip, a feature that would improve comparison studies. However, it also should be acknowledged that the Beacon® equipment was relatively new to the university, and the full functionality of the system was still being explored at the time of the presented experiments.

While this study provided a comparative analysis of the two cell lines at generation 15, a more comprehensive understanding of the changing population dynamics over 60 generations could be obtained through Beacon® runs conducted at various generation timepoints. Such data could inform a more reliable model of CHO cell population dynamics. For example, a key assumption from the model presented in section 4.5.11 was that each cell, along with its progeny, would consistently maintain the same growth and

productivity characteristics throughout the entire 60-generation period. However, it was reasoned that over this extended period, more cells may transition towards a low-producer phenotype, leading to discrepancies between the predictions from the model and the experimental results presented in the previous chapter. Such complexities underscore the need for further investigation into the dynamic changes occurring within cell populations over extended periods to refine predictive models and gain deeper insights into the intricacies of production instability.

An additional limitation in the current dataset is the limited number of cell lines analysed and compared. To effectively develop a predictive method for cell line production instability using the Beacon® system and to gain a more comprehensive understanding of the dynamic nature of cell populations during LTC, it is imperative to expand the analysis to include a more diverse range of cell lines. This broader scope would facilitate a better grasp of how common factors, such as the increased presence of low-producing, fast-growing subpopulations, are within cell populations that experience productivity decline over time. It would also shed light on whether increased clonal variation is a prevalent characteristic in cell lines that exhibit productivity loss. Several studies have observed that cell lines demonstrating unstable productivity during LTC also tend to display an increase in growth rate, suggesting that the occurrence of low-producing, fast-growing cells within such cell lines is relatively widespread (Baik and Lee, 2017; Dorai et al., 2012). Similarly, there is ample research which shows the substantial increases in clonal variation that recombinant CHO cell lines undergo during routine subculture following cloning (Lee et al., 2019). However, there is limited prior evidence indicating that unstable cell lines exhibit a greater degree of clonal variation compared with stable cell lines. Nevertheless, it is reasonable to anticipate that cell lines which exhibit unstable productivity may have a heightened propensity for the genetic and transcriptomic changes which lead to phenotypic diversity. As a result, such cell lines and are more likely to contain cells that have experienced a loss in productivity while gaining a competitive growth advantage within the population. Further

research encompassing a wider array of cell lines will be instrumental in substantiating these hypotheses and refining our understanding of the underlying mechanisms of production instability.

It is also worth acknowledging the differences in conditions between Beacon® runs and FBCs. Such differences mean that when isolated at the single-cell level during a Beacon® run, the behaviour of an individual cell may not accurately reflect its performance within a bulk population during an FBC. During a Beacon® run, the cell population is physically separated into individual cells which are deposited into individual pens and cultured in nanolitre quantities, with media continuously exchanged to support cell growth. This contrasts with FBCs, during which cells are initially seeded at significantly higher concentrations in millilitre to litre volumes. For example, in the FBC experiments outlined in the previous chapter, cells were seeded at a concentration of 0.5×10^6 cells/ml in a 35 mL culture volume. After seeding, cells are allowed to grow with minimal intervention in terms of media addition or removal, aside from periodic feeds. Furthermore, when dealing with low cell concentrations during Beacon® runs and other single-cell cloning methods, it is often necessary to introduce additional components, such as autocrine factors found in CHO supplements or conditioned media, to promote growth. Although our capability to measure the growth and productivity attributes of recombinant CHO cell clones at the single-cell level is improving, the differences in conditions mean that predictions from such methods for how the clone will behave at a higher cell density have not achieved 100% accuracy. Therefore, it is unlikely that the individual cells that were cultivated within the bulk population (before subcloning by the Beacon®) exhibited precisely the same doubling times and cell-specific productivities during the Beacon® run, due to the disparities in experimental conditions. This presents a significant challenge for any models striving to predict production stability based on information gleaned from the Beacon®. To overcome this challenge, further Beacon® runs, with a wider range of cell lines, will provide insights into whether specific phenotypic traits consistently emerge in cell lines with unstable productivity during early-generation

Beacon® analyses. These traits may include greater variability in productivity and doubling time and an increase in the proportion of low-producing, fast-growing cells, compared with stably producing cell lines.

For the questions that remain from this study to be answered, many more Beacon® runs will need to be conducted, thus requiring substantial investment. While a less expensive alternative method such as FACS could be employed to analyse the variation in productivity of cell populations, our data suggests that both doubling time and cell specific productivity could be important indicators of instability in CHO cell lines. FACS, although less costly, is significantly less high-throughput and poses challenges in assessing growth profiles and cell-specific productivity, especially when compared to the real-time capabilities of the Beacon® system.

4.6.2 Future Avenues for Developing Predictive Methods for Production Instability in Recombinant CHO Cell Lines

Predicting cell line stability, is a complex challenge, given the wide variety of, often product-specific, factors that can contribute to productivity loss during LTC. Continuing efforts to understand the underlying complexity of productivity loss, likely via further omics studies, will thus be incredibly valuable in the development of any predictive strategy. However, the need for predictive strategies which function effectively under short timelines must also be considered, especially given that existing omics strategies may require extensive downstream analyses before predictions can be made. In this chapter, we have introduced a method that holds promise for predicting cell line stability by offering real-time measurements of growth and productivity of single cells within the population. However, as discussed in the previous section, further extensive research involving a broader spectrum of cell lines and additional generation timepoints is essential to comprehensively elucidate the relationship between population dynamics and production instability. Nevertheless, our initial findings have revealed two intriguing early generation phenotypes associated with a cell line exhibiting productivity loss during prolonged culture, as compared to one with minimal productivity loss over 60 generations: a higher proportion of low-producing

fast-growing cells in the population and increased variability stemming from both doubling time and cell-specific productivity. Overall, these findings contribute to our understanding of the dynamics of cell populations with varying production stabilities and lay solid foundations for further refinement of our method for predicting production instability.

Chapter 5 - Final Discussion

Since the first monoclonal antibodies (mAbs) were manufactured using Chinese hamster ovary (CHO) cells in 1986, our capacity to produce mAbs at a large scale has seen remarkable progress, with mAb titres increasing by more than 100-fold (Kelley, 2020; Kelley et al., 2018; Wurm, 2004). Additionally, the lengthy 6-12 month timeline required for developing cell lines for production has been significantly optimised through innovations such as high-throughput single-cell cloning and screening methods (Priola et al., 2016; Tejwani et al., 2021). However, the biopharmaceutical industry continues to evolve rapidly, placing ever-increasing demands on manufacturing to solve challenges such as CHO cell production instability and the production of novel protein therapeutics in sufficient quantities (Bielser et al., 2018)

Given the persistent challenges in mAb production stability and the emerging difficulties associated with novel modalities during CLD, this thesis hypothesised that further optimisation of CLD processes could improve the efficiency and reliability of CHO cell-based recombinant protein production. This hypothesis was to be tested via two main avenues: (1) improving expression vector comparison systems and (2) implementing early predictive methods for assessing cell line production stability. The results chapters have addressed the specific discoveries and methodology limitations of the individual projects, while also offering discussions into potential directions for future research in those specific areas. Therefore, the purpose of this chapter is to provide a more holistic overview of the contributions of this research to testing this initial hypothesis and explore the potential alternative applications and synergies of the methods developed in this thesis for further advancements in the field.

5.1 Thesis Summary and Key Findings

5.1.1 A Landing Pad System for Expression Vector Component Comparison

It was originally hypothesised that, given the vast array of expression vector optimisation work that had been conducted, and the potential value of product specific expression vector optimisation strategies, CLD strategies could benefit from a system for expression vector component comparison. It was also hypothesised that the most efficient and effective method for rapidly comparing components in the same genetic context, thus minimising the impact of independent factors such as transgene copy number and integration site, would be to use a targeted integration strategy.

To test these hypotheses, an expression vector component comparison strategy was developed via the design and integration of a landing pad (LP) into the Apollo X host cell line (Chapter 2). Single copy LP integration, using CRISPR/ Cas9, was previously proven successful in other studies (Gaidukov et al., 2018; Inniss et al., 2017). Despite the cell culture challenges and lack of sequencing data for the cell line, likely leading to unsuccessful integration at the novel loci (Mrpl4, Cdk2ap2 and LemD2), a LP was successfully integrated into the Fer1l4 locus, generating the Fer1l4_c8 clone. While some silencing of the mNeonGreen transgene in the LP of Fer1l4_c8 occurred over 60 generations, the PCR studies led to hypotheses that this silencing was most likely linked to the silencing of mNeonGreen expression in a non-functioning LP that had integrated at an additional genetic locus, rather than silencing at the Fer1l4 locus. The functionality of the Bxb1-based LP design was demonstrated via the successful recombinase-mediated cassette exchange of mCherry cassettes into the LP, successfully displacing the mNeonGreen transgene in the LP. This LP architecture thus offers a promising platform for future studies in the Apollo X system. Furthermore, a proof-of-concept experiment comparing different promoters driving mCherry expression was conducted, demonstrating the utility of the system as an expression vector component comparison tool and helping prove the hypothesis that a TI system can be used for expression vector comparison.

It was subsequently proposed that combining this TI strategy with methods that enable high-throughput assembly of expression vector components, such as golden gate assembly methods, could be used to develop the product-specific strategy for expression vector component development that was initially envisaged. For example, adaptations of the extensible mammalian modular assembly (EMMA) toolkit, which was used in (Chapter 2) to build the LP construct, could provide a useful method for developing such a system. Each component of the expression vector (e.g. promoter, untranslated regions, light chain and heavy chain sequences) would be assigned to a designated EMMA part with the appropriate DNA overhangs. Multiple versions of each component could subsequently be created. The one-pot assembly method of EMMA could then be used to build the range of vectors for comparison in the LP system.

5.1.2 Developing a System for Early Prediction of Production Instability

To ensure that a cell line with unstable productivity does not advance to manufacturing stages, a stability study is usually conducted, which evaluates mAb expression across 60 generations (Dahodwala and Lee, 2019; Dorai et al., 2012; Wurm and Wurm, 2017). This duration roughly aligns with the number of times the cell population needs to double to reach the necessary volume for seeding a manufacture-scale bioreactor (ICH Q5B, 1996). However, such studies typically take 3 months to complete, thus posing a significant bottleneck in CLD timelines. It was thus proposed that by enhancing confidence in a cell line's stability, either by increasing the likelihood of obtaining a stable cell line or by devising a predictive method for assessing its likely stability, the stability study could be uncoupled from the critical path to manufacturing. This would allow cell lines to progress toward manufacturing before the stability study is completed.

To pursue this, the production stabilities of two cell lines, namely 32-121 and 32-124, both expressing the same mAb and derived from the same cell line development (CLD) process, were examined over a period of 60 generations

(Chapter 4). Notably, the 32-124 cell line demonstrated minimal loss of productivity when cultured in the presence (1%) or absence (13%) of methotrexate (MTX) over 60 generations. Conversely, the 32-121 cell line exhibited a more substantial drop in day 14 harvest productivity, with a 24% decrease when cultured in the presence of MTX and a 27% decrease without MTX. It is important to highlight that according to a commonly used definition of instability, which defines cell lines as unstable if they fail to retain >70% productivity over 60 generations, the 32-121 cell line would not be defined as unstable, in either the presence or absence of MTX (Dahodwala and Lee, 2019). However, it falls close to the 30% threshold that separates cell lines from being classified as stable or unstable when cultured without MTX (Dahodwala and Lee, 2019; Dorai et al., 2012). A thorough analysis of growth characteristics, specific productivity, metabolite profiles, gene copy number, and transgene mRNA expression was subsequently conducted for each cell line to discern the root causes of the observed productivity decline in the 32-121 cell line. While culturing cell lines in the presence of MTX improved the production stability of both cell lines, the presence of MTX during long-term culture (LTC) had a negative impact on the growth performance of the 32-121 cell line in the late generation fed-batch cultures. Notably, loss of productivity in the 32-121 cell line could not be correlated with a drop in heavy chain or light chain transgene copy number or mRNA expression. This suggests that other mechanisms, as discussed in section 3.6.2, may be responsible for the observed decrease in productivity. Identifying the root causes of productivity decline in the 32-121 cell line, as well as investigating whether common patterns exist among cell lines that exhibit production instability during LTC at FUJIFILM Diosynth Biotechnologies, represents an important avenue for further exploration to develop strategies for mitigating productivity loss.

To begin to develop a predictive strategy for identifying whether a recombinant cell line is likely to exhibit stable or unstable productivity, it was hypothesised that the Berkley Lights Beacon® System (Beacon®) could be used to identify population differences between stable and unstable cell lines at an early generation timepoint. While the Beacon® system has been mainly

used for single cell sorting and early clonal screening in CLD, it has not been applied for the prediction of production instability in CHO cells. Firstly, the 32-121 and 32-124 cell lines were grown to generation 15, in the absence of MTX, before being subjected to analysis using the Beacon®. This analysis revealed two significant differences in the population dynamics of these cell lines. Firstly, the 32-121 cell lines contained a higher proportion of low-producing fast-growing cells than the 32-124 cell line. Secondly, the 32-121 cell line exhibited increased variability in both growth rate and specific productivity, leading to greater clonal variation. Given the differences in both the growth and productivity characteristics of the two cell lines, it is anticipated that the real-time monitoring capabilities of the Beacon® system will be instrumental in the development of methods for predicting instability. Combining this phenotypic assay with cell line analysis methods, which assess factors such as the transgene integration site and transcriptional analysis of the cell line, could further improve the efficiency of prediction methods.

5.2 Alternative Applications for the Fer1l4_c8 Cell Line

5.2.1 Self-Regulating CHO Cells

While the primary goal of the LP system was to create a tool for comparing expression vector components, this research has opened up several other potential applications in the Apollo X cell line. In section 1.9.2 of Chapter 1, and in the review paper in Appendix 2, dynamic engineering strategies were explored as an alternative to the conventional static approaches in CHO cell engineering. Recent advancements in synthetic biology have provided tools such as receptors, inducible promoters, untranslated regions, and protein degradation systems that can control transcription, translation, or recombinant protein stability in response to external cues (Donaldson et al., 2022). These tools thus offer the possibility of constructing sense-and-respond systems within CHO cells, essentially creating self-regulating CHO cells. Such systems could be highly beneficial for optimising traits such as growth rate and lactate metabolism. For example, systems could incorporate

a quorum sensing-like mechanism where cells grow rapidly and express growth-promoting genes during the first phase of a FBC until a certain cell density is achieved (Weber et al., 2007). At that point, growth would be halted, and genes promoting prolonged culture and high productivity would be activated. To effectively implement these sense-and-respond programs in biopharmaceutical production, genetic circuit components would need stable integration into stable genomic loci (e.g., Fer1l4) to prevent transgene silencing and ensure consistent performance during LTC. Flexibility in the manufacturing process would also be a key feature of such self-regulating systems (Bielser et al., 2018). As discussed earlier, for novel therapeutics, the choice of CLD strategy may need to vary based on the product type. LP systems are thus likely to play a vital role by enabling precise and repeatable targeting of genetic circuits into specific loci, offering a plug-and-play approach for production optimisation.

5.2.2 Solving Product-Specific Production Challenges

Optimising the product quality of therapeutic mAbs is crucial during CLD. The Fc region of a mAb contains two N-glycans, with sugar molecules such as fucose, galactose, and sialic acid attaching to the core structure in specific patterns known as N-glycosylation profiles. These profiles have a direct impact on the efficacy, safety, and pharmacokinetics of the drug. For example, excessive fucose can hinder antibody-dependent cellular cytotoxicity, while increased galactose can enhance complement-dependent cytotoxicity. Sialic acid on N-glycans may reduce complement-dependent cytotoxicity and impair antibody-dependent cellular cytotoxicity in some cases but removing it can lead to faster drug clearance and a shorter half-life, potentially affecting efficacy. Therefore, the ability to tailor the product quality attributes of a mAb in a product-specific manner to its intended use is important. Past research has employed LPs to overexpress glycosyltransferase genes, allowing fine-tuning of N-glycosylation profiles (Nguyen et al., 2021). This approach, when applied to the Fer1l4_c8 cell line, could facilitate the overexpression of various glycotransferases and other beneficial proteins, such as chaperones that enhance Fc protein secretion

(Pybus et al., 2014), thereby enabling the customisation of cell line attributes to meet specific therapeutic requirements.

5.3 Beacon® Analysis of LP Cell Lines

As previously mentioned, an important finding from the Beacon® study (presented in Chapter 4) was the difference in clonal variation between the 32-124 cell line, which exhibited minimal productivity loss, and the 32-121 cell line. Previous research has emphasised the potential of TI as a means of reducing clonal variation in CHO cell lines. For example, in a previous study, an LP was integrated into the CHO cell genome using the CRISPR/Cas9 system. Subsequently, various transgenes were integrated into the LP through recombinase-mediated cassette exchange. Subclones of the resulting cell populations displayed low clonal variation, consistent growth, stable transgene transcript levels, and minimal differences in the global transcriptome in response to recombinant protein production. Investigating whether the Beacon® can detect reductions in clonal variation in TI cell lines compared with random integration (RI) cell lines would be intriguing. In addition, exploring how clonal variation changes over time in RI cell lines, as proposed in section 4.6.1 of Chapter 4, could provide valuable insights. A meaningful comparison could involve assessing clonal variation during LTC in RI cell lines versus TI cell lines, with the Fer1l4_c8 clone serving as a potential model for this investigation. Such a study would create an interesting synergy between the two research directions presented in this thesis. To advance these experiments, an interesting approach would be to use the Beacon® to compare the population dynamics of a RI cell line with a TI cell line that produces a similar quantity of mAb. Currently, the single LP copy in Fer1l4_c8 makes it unlikely that comparable mAb production levels to a RI cell line would be achieved. Previous studies have increased titres from TI systems by targeting LPs to multiple locations in the CHO genome and increasing the number of transgene copies on the expression vector (Gaidukov et al., 2018; Sergeeva et al., 2020b). Although targeting LPs to multiple loci was considered as a potential future direction for the TI system

(section 2.5.2 in Chapter 2), the challenges associated with targeting the LemD2, Mrpl4, and Cdk2ap2 loci prevented its implementation.

5.4 Combining Predictive Measures with Stability Enhancers

While high-throughput screening has accelerated the identification of high producers during CLD, the extended nature of LTC stability studies remains a bottleneck in CLD timelines. In addition to methods which predict production stability, methods which improve the likelihood of producing a stable cell line during CLD are required. The screening of expression vector components, such as promoters which are resistant to silencing, will play a pivotal role for this purpose (Brown et al., 2017, 2014; Johari et al., 2019). However, high-throughput comparison of expression vector components that enhance stability presents challenges. LPs may offer a promising solution for expression vector component comparison when integrated into a region of the genome prone to silencing (Harraghy et al., 2015). Subsequent integration of expression vectors into the LP which rescue transgene expression could be used to identify components which prevent transgene silencing. Additionally, cell lines, such Fer1l4_c8, with LPs in stable, high-expressing regions of the genome will remain invaluable for such studies. Overall productivity will remain a priority and novel expression vector components, that potentially improve production stability, must still yield comparable titres to existing components.

5.5 Final Conclusions

In conclusion, this thesis has provided an investigation into the production of mAbs using CHO cells. It has explored the challenges and opportunities within the biopharmaceutical industry, highlighting the increasing demand for innovative solutions to accelerate CLD timelines, improve the productivity of novel difficult-to-express therapeutics, and enhance the stability of mAb-producing cell lines. Through a combination of TI strategies, which will facilitate the comparison of future novel expression vector component developments, and predictive methods for production stability assessment,

this research has contributed significantly towards addressing these challenges. The development of predictive strategies for identifying stable cell lines, such as the utilisation of the Beacon® system, represents an important advancement in the field. By better understanding the population dynamics and clonal variation within cell lines, it could become possible to foresee potential LTC productivity issues, allowing for timely interventions and reduced timelines. Moreover, the LP system offers a versatile platform for the comparison of expression vector components. This system is likely to play an important role in the development of product-specific CLD strategies for novel therapeutics, such as bi-specific antibodies and toxic proteins, which each present unique manufacturing challenges. Ultimately, this work contributes to the continued innovation in CLD which is required to meet the ever-growing demands of the biopharmaceutical market and ensure the timely delivery of life-saving therapeutics to patients worldwide.

Appendix

Appendix 1

Copy of “Decoupling Growth and Protein Production in CHO Cells: A Targeted Approach” by James S. Donaldson, Matthew P. Dale and Susan J. Rosser. Published on 2nd June 2021 in Frontiers in Bioengineering and Biotechnology

Appendix 2

Copy of “Synthetic biology approaches for dynamic CHO cell engineering” by James S. Donaldson, Dirk-Jan Kleinjan and Susan J. Rosser. Published on 1st October 2022 in Current Opinion in Biotechnology

Appendix 3

Homology Arms

Target Gene	5' Homology Arm Sequence	3' Homology Arm Sequence
Fer1l4	aaaCGTCTCGTAGGCATCCCCCCAGAGCACACAGGGGGA CGTCAGGACCATCCACCCCACAAAACGAACTGTTCTAAG CCAGGCTCTATTGAGATGATCTGgtgtctgtaaggctaagatatctttat ttagagaaacaccagccaaacgcaagccagagcgtgccaggggcagcaaggc agggaggccagagagaaggggctcccatgtgccttgcagccacttaaaaccctat cacgtgtcatcctgacctcaccttggccacgccctgacaggcatggtcaggcacacc tgtagccagcttctagcaggcgtggcttacactgtccctacaaggCTCTGTCT TCCAACCatgacccccaccccaccccaccccactgttCCAGTCAAGCC TTGCTGCAAAATGACCTCTCTGAAGTCAGTGGTTTCAGTA TCCACACCCTCTCCGCCAATTACCACACCAGTACTTCTAC CCCACTCTCAGCTCCCCTCTGAAGTCTAGAGAGGTCCTC TGTCTGCTACTGTTTGTCACAGACCTGTTCCCTTCCAGAA CTcagccttctctttccctcccccaacacctGGAGGATACAAAGGTGTA CAGTCCAGAGCAGTCTCCATTTACCCTGTGCTCTCCTCCA AACCCTGTTCTTGTCCCCCATAGCGGGGTTTAACCCCACT TTTGGTCCAGCCTGGGTGCCCCTCTATGGCTCGCTCCCC AGTGGCAGGTTCCGGGATGGTCTTCAAAGTCTCAATGAA GGTGTTGGAGAAGGCATTTGGTTCCGGGGCCGAATTCTG GTAGCTGTGTCCATGGAGGTGTTGGAAGGCAGAGTTGAA CCTAATGGcGAGACGaaa	aaaCGTCTCgTCACTCCAGATTGTCCCGACtcactgggaagaaaaa gaagaaaaccaagcaGCGTCAAACCACGGCTGGCGTTCCGAAGC CTGTGGATGCCAGTGCCAGCGGGGACGCACCTGAGATTCC TAGCGCaatggaggtggaggtggaggagttGCTGCCTCTGCCGGAGG TGGGGCTGGTGGTGGTGACCCTCGGCTCTAGGGCTTGTAA ATCCCTTTCCAAGCTTTAGAAATCATGATGCTGTTATCAATTT GGGGACTTGGGTTGAGAATTCAGTGTATGTGCACACGTGTT CTGATGTACACAGACATGGTATATATAGAGGGCGTGGACCA GAGAGAATCTGTCCTCTGGACCAATGTATAGGGAAGCTACC AGGGGCCAAGAGGGAGGTAAGACACTCACAGAGCCCCCTG GACCTTAGCTGATATCTTTCTGTGGGCATTGTGACCAAGCAT AGAGCAAACTGCTCCTGGGGAGGATGGGAGCTGACCCCAG CTGACCTTGCTCTCTTTCCTAGAATGCCCTGGCACCCTGCG AAGATTTCCTGCTTTTTGCAGTGGTCTTTGAGGCTACCATGA TTGATCCTTCGCTGGCCAAGCAGCCTATCAGCTTGGAGATT TCCATTGGTGTGTGGCCTAAATAAGCACTGAAGGTCATGGG TTTGCCTTCTTGTTTCAGCTTTAGGCACTGGAGGGAGGTTCA CCTTCAGGGAGAGTGGTGGGCTCTGCTCCTACTGGGGGAG AGGCTGCAGAGGGATGCATAGGACCCTCAGCCTGAGGCTC TGATCCTTGATGTTCTGGGGGCAGGTCACACAGGCCGCCG AGAGGAGCAATTGGGCCAAAGGTCAAGGGCTGATGAGGGA GCCGAAGGCAGCACACTAGAGAGCGAATCTCGGCCTCTCC TGGGATCTGAGGACAGAGGGGcagggcaggaagaggaagagttgCT GGGGACACCGGATCAGTGGCCCGAGCCTGTAGATGGCAAT GGGTAAGTGATCGGGCGGACGTCCATCCCCCCAGAGCACA CAGGGGCTACcGAGACGaaa
LemD2	aaaCGTCTCGTAGGCCACCCTGTTATCTGCCACATGCGAC GTCCCGAAGGGGAGCTGGGACTGAGCCGGGCAAAAAAC AAATGACCGTTATCATTGTCACCTTGTGATAGTAGTTTTCA GCGTCCACAGCAAAAACCACATAGCCAACCCACTTATTTC TGCCTCCTAAACAAAACTGGATGGGAAGTTTCATTCTAAC TTAATTTTGTATTAacaaatgttcatttatttataaactgAAGCCACGTA GGAATCGCAGCTCCTATGATCAGCATAGGTTCCTGGCGA TcactgtgccccaccccactgccaGCTGCCTTCCTGATGACAGAG AGATCATGGTTTTGGAGAATGCTACCAGGACAAGAACCC CCCCACCCTGCCACCCTTGTTTGGGACGCTGAATTCCCT GGGACTTCTGAGTGTACAGACCCTGCACTGCTGCAAGAC TCCGTGCATTCCCAAAACCACAAGCCTCAGACCGGCTGT GAGTGTCTGGGAACTGATGCCAGGATTTGACCTCCAGAT GCAGAGGCCTGCCACCCCGACTCCACCTCACCCATCCTG GCTCCCACAGCCAGTATGGcGAGACGaaa	aaaCGTCTCgTCACTTCCAAAAGCAGGAAAAACAAGAGGCTC CAGCGTTCTTCCAGGTAAGAAAAATCATTTGTATTTTTCGTC ACCCTCTGCCCTACGGTGAGAACCCCTTCCCCTGTCACCAG GAGGGAACTTCCATTTCTAGGGTGTTGGGTGGGGGACTGGT ACCCTCGGGAGTTTTTCCTAAGTTCCCGAGGTTTGGAGAGC AGAGCAGATCTGGGAACTCCCTGGGTGATTCCTGCCACGTG CAATCTTTttagaattctttctttattggtaTTTTTCAGCTGGAAATTGGTT GGCGGAGCATGGTCCCTGGGGATTGTGGTGCTGTAGGCTC TGCAGAAGAGCAGGGGTTCACTCAGCAGGAGTAGCGCTGA CCCCACCCTGCAGCCCTTCCTTGTGTCTGGAAGTAACTCAG GAGGCCATGGCAGCAGtttgcaccccaccccacccccgccccggGTG AGGCCTGTCAgcccctctcctgccctcctcctgcccctcctctgttgctcagTC GGATCAAATAAGtggatttttggtttgtttgtttttatgaacaAGCTCCTCTA GTTTCCTGGGCTGTGGATTCCACTGCCTGGAGCATGCTCAC AAACCCTGGAAGTAATGAAATCGgaggggggtggaggtggggaggatc caTGAGCTGAACGGAGATgaaatgaaatgatcatgttttcagaaaacaagaa aagtgaGACATAGTCTCgcttcaaaaagcaaaacaaacccttcaGAATCAG GACTGCagaaatggctcatcagttaaaatatttattgttctAGAAGGACCTGAG GACGTCCCACCCTGTTATCTGCCACATGCCTACcGAGACGaa a
Mrpl4	aaaCGTCTCGTAGGTTACAGGCCTCTGAGCCCAGTGGGA CGTCgtgtgtgtattatatatagaGCAACTGCCCAGCCGCTGAGGC TCTAGGCTCATTACCCTGAACACAGCACTAAGTACCTCTC TCTGAGAGATGATGTGTGCCCAGTCTACCTACTGATTTTT GTTCACTTCTCCCACTAGATTGAGTCCTCAGGACAGGGAT CTGCTGACTTGTAAATCCTCTAACATAATACCCACCATGA TGAGTGCTGAGATAGGATCTGAGAAGTGACCCTGCCTTT GCATGTACAGGTATACAAGGCCTGAACTGGGCTCTTTCC AGAGCACACATGGcGAGACGaaa	aaaCGTCTCgTCACGAAGCTCTGCCATGCTACTCAGGACACT CACAATTCTCTTGAAGTTCTTCTGCCAGGTGGCAACCTGGT GCAGAATGTCCAGcctggaaagagggaagagaaacctAAGAGGGACA CCCACCCAACTTAAACTCTCCCTCACCTGCATCCCAGTAAAG GAGGCCTGGCTCTCCCATGGATATTCAGGGCCCTATGTCCT CTCTTCCTTGACTACAGCTGACACCTCAACTCAAGCTAGTCT ATGCCTGACCTCCTGGTACTCAAGAGGTTAAGGCAAGAGGA TGGAAATTTAAAGGCCTGGCTGTGCTTCACGGTAAGAATCT GTCTCTGAAAACCAGAGTGGAGTCGAACATCAGAACCTCTG TACCTGCTCTCTAACTGAAATATTCTTGTCTTGACTGGATCC AGTAGGTTCcagaggtgtgtgtatgtgaggggagggaggttgcagagatggctca gtggttaagagtacttctgCTTTGGTGGAGGGgccaggtccaattcccagcacc

Target Gene	5’ Homology Arm Sequence	3’ Homology Arm Sequence
		caaaataGCTCACAGCTAGTTTTGCTTTGGACGTCTTACAGGC CTCTGAGCCCAGTGGCTACcGAGACGaaa
Cdk2ap 2	aaaCGTCTCGTAGGGTCACCTTGCCTCATCTCTGGGGGAC GTCAGACCACTGTTTAACGACTTTGGACCGCCCTCCATG GGGTATGTACAGGTGAGTGTGGCCAGGAGCTTGCATATG GATGAGGACAGGAGGTCCCCTCGGGGTCTGAGGTCATG TCTCTATTTTCTTTTGACATAGGCGATGAAGCCACCTGGT TCCCAGGGCTCTCAGAGCACCTACACGGACTTGCTGTCT GTCAtagaggagatgggaaaagagaTCCGGCCCACCTATGCTGG TAGCAAGAGTGCTATGGAGCGCCTGAAGAGAGGTGAGTG AAGCCGCAGACCCACCCTTGCACTGCCCAGAACACAAAC ACGGAACGCTAAGACATTTCATTAAGGAGGGAGGCTATC TGGGCCAGCCTAGATTGAACCAGAACCTCTATCCAGACA GGTGTTTCAAAATGAACTTGAGTCAGTTTCCAAAAAAGTG CCTGGGAACTCAGGGTGATTGTGAGGGTTCACTAAACTC TTAATCTCATCCGACCGCATTAGATTAGAGGTGCTCTTGC TGAACCACCAAGTGGGTGGGGGAGTATCTGCTCTAAAGG GTGGGAATGGTGGCTGTGGAAGGTTGCTGAGACCCTGTT TGTTACTTCCCCGTCCCCATCCCAcctcaggcatcatccatgcACG GGCACTGGTCAGAGAGTGCTTGGCAGAGACAGAACGCA ACGCCCGCACGTAACAGGAAGCACCTTGGCCTTTTCATC GGGACCTTTCCGGCCACTGCAGAGCACCTGCTTCTCCTT GGCCTTCACCCCAAGTTGCTTCCTATCCTGGGCTTCCTGT CCTGTGTCCCTTGGGGGGCATCCTCCAGGAACCAGGTAG CAGATCTTCCTCCAGTTGGGCTCCTCTCTCTGCAGGAGT GAGCAGAGAGGCCACCGCGAGTTCCACCCATGACCTGC TCACAAATCCGATCCAACTGGAGCTTGCTCCTCTCTGTGG CTTCTTCTCAACTCTTAGTCCAGGCCCCCTCCCCCTCCAA CGGGAGTGCAGAAGATGGcGAGACGaaa	aaaCGTCTCgTCACCCACAGCCGACTCTTTTTTTGGTGTGTCT TTTACTAAATACgccctttttatataaataaaagatgATTTGGAGTCATTC TCTCAGCTCTGagcctttctgtgtctgttccTCTGGGGAAGCCTGAGTG GGCCAGGCAGAGGGGAAAGGGCAGAGCAGGAATCAGGGA GGGTCTTGAACCTAGTTTTGCCTTCTCACCCTCTCTGTGCCT GCGAGCTGGGGGTAGCTAAAGGCCTGAACACCCGTGGTGT GAAACCCTGCATCGTGTGGTGCCACTAGGGGGCAGCGCAG GGCCTTGATGCCTTCACAGCTGGAGGCACTGTTGGACAGCC ACAGCCTGGAGGGCCCTTCCCCTGGTTAGGCAGGAGTGCC TGGCAGGGTGGAAATACCAGTTCACCTGAGTCTCTCCTTTG AATGCCCTAATGCCCCTACACTACTCTCTGCATTGCCCTTCG AAGACTCTAAAAACACCAGATCTCTTAACTTCTCGGGGAGAC AGTGGGATGTCCCCTTTCCTCAGCTCCTTGGCTTGATGCTG CTATCCACTTCCACCAGGCTGGGTGTGGCTCCATGAGCCAC CCTCGGTGTCTCAGCCTTTGTACAGAGTCGAGGCAACTCAG ATGGGTGGCCAAATTTGAGAACCTAGCAGCCATAAGTAGGG AAGGACTCGGACTCTATTAAGGGGATGGCCTAGAGAATTGT GGTAACAGAGACCTCTACCTGGGTCTGTAGCTGAGTGACCC TGAAATTCCAACCCTAGGCAGAGACAAGTGAGAGCTACCAG CTTCTGGCTAACTTTGGGGTACAGCGTTTCCTTTACCCTGG GACCCAGCCATGTCAAATTCCACCTCTTCAACATACACAGGA GGGCTGCCCTGGCTGGGCCAGGGTGGGAAATGAAACCGAA AATCCTTGGGGTGTCCATGCAGCACAGAAGTAAAGAACTGT GGCATATGGGTACCGACCTCCTCCCCATCAGCCATGAAGAA GGACGTCGTCACCTTGCCTCATCTCTGGGGCTACcGAGACG aaa

Sequence Results from Diagnostic PCRs of the Fer1l4_c8 Cell Line

This section provides the sequencing data that is referred to in section 2.5.8.

Figure 5.1 - Sequencing Results from Diagnostic PCRs of the Fer1l4_c8 Cell Line.

PCR products from section 3.5.8 were ran on a gel and sequenced. **(A)** Alignment of the sequencing results from B3 (from Figure 2.8) with the expected resultant sequence from a LP integrating into the Fer1l4 locus. **(B)** PCR reactions were designed to amplify the attB sites of the LP (see Figure 2.8C). Figure shows alignment of the 5'-attB PCR product with the expected 5'-attB sequence. **(C)** Alignment of the 3'-attB PCR product with the expected 5'-attB sequence.

Sanger Sequencing of the PCR products show recombination at the AttB sites has occurred

This section provides the sequencing data from section 2.5.13 indicates successful integration into the LP.

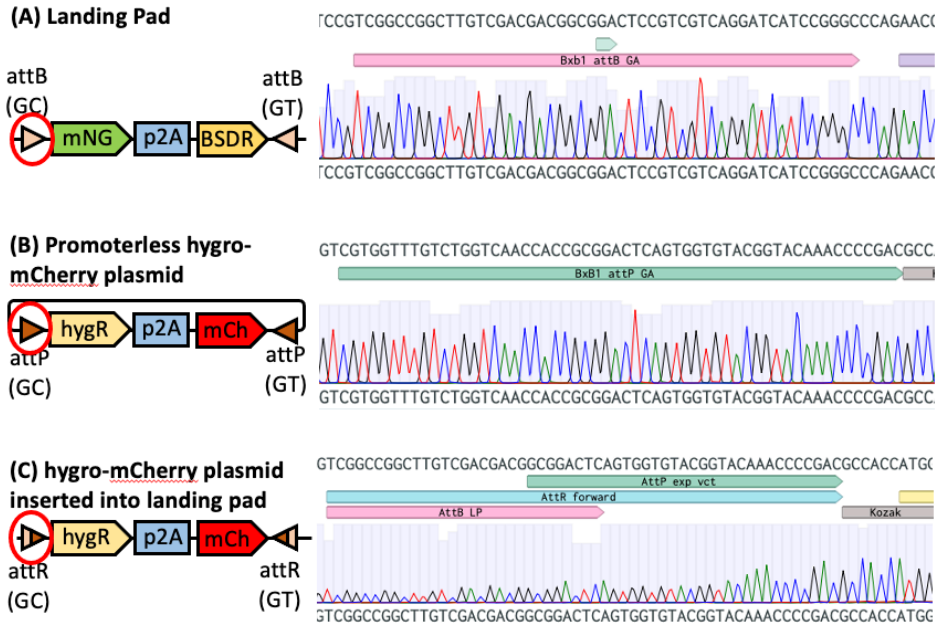


Figure 5.2 - Sanger Sequencing of the PCR products show recombination at the AttB sites has occurred

(A) Sequence of the 5' *attB* site of the LP plasmid. **(B)** Sequence of the corresponding 5' *attP* on the donor vector. **(C)** Following gel extraction and purification of Band 3 (with insert) from Figure 2.12, the DNA was sequenced using Sanger sequencing. Image shows sequence of the recombined 5' *att* site.

References

Aithal, M.G.S., Rajeswari, N., 2015. Validation of Housekeeping Genes for Gene Expression Analysis in Glioblastoma Using Quantitative Real-Time Polymerase Chain Reaction. Brain Tumor Res Treat 3, 24–29. https://doi.org/10.14791/btrt.2015.3.1.24

Akama-Garren, E.H., Joshi, N.S., Tammela, T., Chang, G.P., Wagner, B.L., Lee, D.-Y., Rideout III, W.M., Papagiannakopoulos, T., Xue, W., Jacks, T., 2016. A Modular Assembly Platform for Rapid Generation of DNA Constructs. Sci Rep 6, 16836. https://doi.org/10.1038/srep16836

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., Walter, P., 2002. B Cells and Antibodies, in: Molecular Biology of the Cell. 4th Edition. Garland Science.

Alejandra, W.-P., Miriam Irene, J.-P., Fabio Antonio, G.-S., Patricia, R.-G.R., Elizabeth, T.-A., Aleman-Aguilar, J.P., Rebeca, G.-V., 2023. Production of monoclonal antibodies for therapeutic purposes: A review. International Immunopharmacology 120, 110376. https://doi.org/10.1016/j.intimp.2023.110376

Allen, M.L., Antoniou, M., 2007. Correlation of DNA Methylation with Histone Modifications Across the HNRPA2B1-CBX3 Ubiquitously-Acting Chromatin Open Element (UCOE). Epigenetics 2, 227–236. https://doi.org/10.4161/epi.2.4.5231

Alqahtani, A.A., Jansen, R.K., 2021. The evolutionary fate of rpl32 and rps16 losses in the Euphorbia schimperi (Euphorbiaceae) plastome. Sci Rep 11, 7466. https://doi.org/10.1038/s41598-021-86820-z

Alves, C.S., Dobrowsky, T.M., 2017. Strategies and Considerations for Improving Expression of “Difficult to Express” Proteins in CHO Cells, in: Meleady, P. (Ed.), Heterologous Protein Production in CHO Cells: Methods and Protocols, Methods in Molecular Biology. Springer, New York, NY, pp. 1–23. https://doi.org/10.1007/978-1-4939-6972-2_1

Antoniou, M., Harland, L., Mustoe, T., Williams, S., Holdstock, J., Yague, E., Mulcahy, T., Griffiths, M., Edwards, S., Ioannou, P.A., Mountain, A., Crombie, R., 2003. Transgenes encompassing dual-promoter CpG islands from the human TBP and HNRPA2B1 loci are resistant to heterochromatin-mediated silencing. Genomics 82, 269–279. https://doi.org/10.1016/S0888-7543(03)00107-1

Badding, M.A., Vaughan, E.E., Dean, D.A., 2012. Transcription factor plasmid binding modulates microtubule interactions and intracellular trafficking during gene transfer. Gene Ther 19, 338–346. https://doi.org/10.1038/gt.2011.96

Baghini, S.S., Razeghian, E., Malayer, S.K., Pecho, R.D.C., Obaid, M., Awfi, Z.S., Zainab, H.A., Shamsara, M., 2023. Recent advances in the application of genetic and epigenetic modalities in the improvement of antibody-producing cell lines. International Immunopharmacology 123, 110724. https://doi.org/10.1016/j.intimp.2023.110724

Bahr, S.M., Borgschulte, T., Kayser, K.J., Lin, N., 2009. Using microarray technology to select housekeeping genes in Chinese hamster ovary cells. Biotechnology and Bioengineering 104, 1041–1046. https://doi.org/10.1002/bit.22452

Bai, H., Lester, G.M.S., Petishnok, L.C., Dean, D.A., 2017. Cytoplasmic transport and nuclear import of plasmid DNA. Bioscience Reports 37, BSR20160616. https://doi.org/10.1042/BSR20160616

Baik, J.Y., Han, H.-J., Lee, K.H., 2021. DNA Double-Strand Breaks Affect Chromosomal Rearrangements during Methotrexate-Mediated Gene Amplification in Chinese Hamster Ovary Cells. Pharmaceutics 13, 376. https://doi.org/10.3390/pharmaceutics13030376

Baik, J.Y., Lee, K.H., 2018. Growth Rate Changes in CHO Host Cells Are Associated with Karyotypic Heterogeneity. Biotechnology Journal 13, 1700230. https://doi.org/10.1002/biot.201700230

Baik, J.Y., Lee, K.H., 2017. A framework to quantify karyotype variation associated with CHO cell line instability at a single-cell level. Biotechnology and Bioengineering 114, 1045–1053. https://doi.org/10.1002/bit.26231

Bailey, L.A., Hatton, D., Field, R., Dickson, A.J., 2012. Determination of Chinese hamster ovary cell line stability and recombinant antibody expression during long-term culture. Biotechnology and Bioengineering 109, 2093–2103. https://doi.org/10.1002/bit.24485

Bandyopadhyay, A.A., O'Brien, S.A., Zhao, L., Fu, H.-Y., Vishwanathan, N., Hu, W.-S., 2019. RECURRING GENOMIC STRUCTURAL VARIATION LEADS TO CLONAL INSTABILITY AND LOSS OF PRODUCTIVITY. Biotechnol Bioeng 116, 41–53. https://doi.org/10.1002/bit.26823

Banerjee, R., Gherasim, C., Padovani, D., 2009. The tinker, tailor, soldier in intracellular B12 trafficking. Current Opinion in Chemical Biology, Analytical Techniques/Mechanisms 13, 484–491. https://doi.org/10.1016/j.cbpa.2009.07.007

Barnes, L.M., Dickson, A.J., 2006. Mammalian cell factories for efficient and stable protein expression. Current Opinion in Biotechnology, Protein technologies 17, 381–386. https://doi.org/10.1016/j.copbio.2006.06.005

Beckmann, T.F., Krämer, O., Klausing, S., Heinrich, C., Thüte, T., Büntemeyer, H., Hoffrogge, R., Noll, T., 2012. Effects of high passage cultivation on CHO cells: a global analysis. Appl Microbiol Biotechnol 94, 659–671. https://doi.org/10.1007/s00253-011-3806-1

Berkley Lights, 2023. Beacon® Optofluidic System [WWW Document]. Berkeley Lights. URL https://www.berkeleylights.com/systems/beacon/ (accessed 10.16.23).

Betts, Z., Croxford, A.S., Dickson, A.J., 2015. Evaluating the interaction between UCOE and DHFR-linked amplification and stability of recombinant protein expression. Biotechnology Progress 31, 1014–1025. https://doi.org/10.1002/btpr.2083

Bi, J.-X., Shuttleworth, J., Al-Rubeai, M., 2004. Uncoupling of cell growth and proliferation results in enhancement of productivity in p21CIP1-arrested CHO cells. Biotechnology and Bioengineering 85, 741–749. https://doi.org/10.1002/bit.20025

Bielser, J.-M., Wolf, M., Souquet, J., Broly, H., Morbidelli, M., 2018. Perfusion mammalian cell culture for recombinant protein manufacturing – A critical review. Biotechnology Advances 36, 1328–1340. https://doi.org/10.1016/j.biotechadv.2018.04.011

Blanco, N., Williams, A.J., Tang, D., Zhan, D., Misaghi, S., Kelley, R.F., Simmons, L.C., 2020. Tailoring translational strength using Kozak sequence variants improves bispecific antibody assembly and reduces product-related impurities in CHO cells. Biotechnology and Bioengineering 117, 1946–1960. https://doi.org/10.1002/bit.27347

Bojar, D., Scheller, L., Hamri, G.C.-E., Xie, M., Fussenegger, M., 2018. Caffeine-inducible gene switches controlling experimental diabetes. Nat Commun 9, 2318. https://doi.org/10.1038/s41467-018-04744-1

Borth, N., Mattanovich, D., Kunert, R., Katinger, H., 2005. Effect of Increased Expression of Protein Disulfide Isomerase and Heavy Chain Binding Protein on Antibody Secretion in a Recombinant CHO Cell Line. Biotechnology Progress 21, 106–111. https://doi.org/10.1021/bp0498241

Borth, N., Zeyda, M., Katinger, H., 2000. Efficient selection of high-producing subclones during gene amplification of recombinant Chinese hamster ovary cells by flow cytometry and cell sorting. Biotechnology and Bioengineering 71, 266–273. https://doi.org/10.1002/1097-0290(2000)71:4<266::AID-BIT1016>3.0.CO;2-2

Braselmann, S., Graninger, P., Busslinger, M., 1993. A selective transcriptional induction system for mammalian cells based on Gal4-estrogen receptor fusion proteins. PNAS 90, 1657–1661. https://doi.org/10.1073/pnas.90.5.1657

Brezinsky, S.C.G., Chiang, G.G., Szilvasi, A., Mohan, S., Shapiro, R.I., MacLean, A., Sisk, W., Thill, G., 2003. A simple method for enriching populations of transfected CHO cells for cells of higher specific productivity. Journal of Immunological Methods 277, 141–155. https://doi.org/10.1016/S0022-1759(03)00108-X

Brinkmann, U., Kontermann, R.E., 2017. The making of bispecific antibodies. MAbs 9, 182–212. https://doi.org/10.1080/19420862.2016.1268307

Brown, A.J., Gibson, S., Hatton, D., James, D.C., 2018. Transcriptome-Based Identification of the Optimal Reference CHO Genes for Normalisation of qPCR Data. Biotechnology Journal 13, 1700259. https://doi.org/10.1002/biot.201700259

Brown, A.J., Gibson, S.J., Hatton, D., James, D.C., 2017. In silico design of context-responsive mammalian promoters with user-defined functionality. Nucleic Acids Research 45, 10906–10919. https://doi.org/10.1093/nar/gkx768

Brown, A.J., James, D.C., 2017. Constructing Strong Cell Type-Specific Promoters Through Informed Design, in: Gould, D. (Ed.), Mammalian Synthetic Promoters, Methods in Molecular Biology. Springer, New York, NY, pp. 131–145. https://doi.org/10.1007/978-1-4939-7223-4_10

Brown, A.J., Sweeney, B., Mainwaring, D.O., James, D.C., 2014. Synthetic promoters for CHO cell engineering. Biotechnology and Bioengineering 111, 1638–1647. https://doi.org/10.1002/bit.25227

Budge, J.D., Roobol, J., Singh, G., Mozzanino, T., Knight, T.J., Povey, J., Dean, A., Turner, S.J., Jaques, C.M., Young, R.J., Racher, A.J., Smales, C.M., 2021. A proline metabolism selection system and its application to the engineering of lipid biosynthesis in Chinese hamster ovary cells. Metabolic Engineering Communications 13, e00179. https://doi.org/10.1016/j.mec.2021.e00179

Burioni, R., Plaisant, P., Delli Carri, V., Vannini, A., Spanu, T., Clementi, M., Fadda, G., Varaldo, P.E., 1997. An improved phage display vector for antibody repertoire cloning by construction of combinatorial libraries. Res Virol 148, 161–164. https://doi.org/10.1016/s0923-2516(97)89903-7

Cabrera, A., Edelstein, H.I., Glykofrydis, F., Love, K.S., Palacios, S., Tycko, J., Zhang, M., Lensch, S., Shields, C.E., Livingston, M., Weiss, R., Zhao, H., Haynes, K.A., Morsut, L., Chen, Y.Y., Khalil, A.S., Wong, W.W., Collins, J.J., Rosser, S.J., Polizzi, K., Elowitz, M.B., Fussenegger, M., Hilton, I.B., Leonard, J.N., Bintu, L., Galloway, K.E., Deans, T.L., 2022. The sound of silence: Transgene silencing in mammalian cell engineering. Cell Systems 13, 950–973. https://doi.org/10.1016/j.cels.2022.11.005

Cao, J., Novoa, E.M., Zhang, Z., Chen, W.C.W., Liu, D., Choi, G.C.G., Wong, A.S.L., Wehrspaun, C., Kellis, M., Lu, T.K., 2021. High-throughput 5′ UTR engineering for enhanced protein production in non-viral gene therapies. Nat Commun 12, 4138. https://doi.org/10.1038/s41467-021-24436-7

Cao, Y., Kimura, S., Itoi, T., Honda, K., Ohtake, H., Omasa, T., 2012. Construction of BAC-based physical map and analysis of chromosome rearrangement in chinese hamster ovary cell lines. Biotechnology and Bioengineering 109, 1357–1367. https://doi.org/10.1002/bit.24347

Carvalhal, A.V., Marcelino, I., Carrondo, M.J.T., 2003. Metabolic changes during cell growth inhibition by p27 overexpression. Appl Microbiol Biotechnol 63, 164–173. https://doi.org/10.1007/s00253-003-1385-5

Carver, J., Ng, D., Zhou, M., Ko, P., Zhan, D., Yim, M., Shaw, D., Snedecor, B., Laird, M.W., Lang, S., Shen, A., Hu, Z., 2020. Maximizing antibody production in a targeted integration host by optimization of subunit gene dosage and position. Biotechnology Progress 36, e2967. https://doi.org/10.1002/btpr.2967

Cha, M., Yu, F., 2014. Pharma's first-to-market advantage [WWW Document]. McKinsey & Company. URL https://www.mckinsey.com/industries/life-sciences/our-insights/pharmas-first-to-market-advantage (accessed 10.3.23).

Chen, Z.Y., He, C.Y., Meuse, L., Kay, M.A., 2004. Silencing of episomal transgene expression by plasmid bacterial DNA elements in vivo. Gene Ther 11, 856–864. https://doi.org/10.1038/sj.gt.3302231

Chen, Z.-Y., Riu, E., He, C.-Y., Xu, H., Kay, M.A., 2008. Silencing of Episomal Transgene Expression in Liver by Plasmid Bacterial Backbone DNA Is Independent of CpG Methylation. Molecular Therapy 16, 548–556. https://doi.org/10.1038/sj.mt.6300399

Cheng, J.K., Alper, H.S., 2016. Transcriptomics-Guided Design of Synthetic Promoters for a Mammalian System. ACS Synth. Biol. 5, 1455–1465. https://doi.org/10.1021/acssynbio.6b00075

Cheng, J.K., Morse, N.J., Wagner, J.M., Tucker, S.K., Alper, H.S., 2019. Design and Evaluation of Synthetic Terminators for Regulating Mammalian Cell Transgene Expression. ACS Synth. Biol. 8, 1263–1275. https://doi.org/10.1021/acssynbio.8b00285

Chiu, M.L., Goulet, D.R., Teplyakov, A., Gilliland, G.L., 2019. Antibody Structure and Function: The Basis for Engineering Therapeutics. Antibodies 8, 55. https://doi.org/10.3390/antib8040055
Chung, J.H., Whiteley, M., Felsenfeld, G., 1993. A 5' element of the chicken beta-globin domain serves as an insulator in human erythroid cells and protects against position effect in Drosophila. Cell 74, 505–514. https://doi.org/10.1016/0092-8674(93)80052-g
Chung, J.Y., Lim, S.W., Hong, Y.J., Hwang, S.O., Lee, G.M., 2004. Effect of doxycycline-regulated calnexin and calreticulin expression on specific thrombopoietin productivity of recombinant chinese hamster ovary cells. Biotechnology and Bioengineering 85, 539–546. https://doi.org/10.1002/bit.10919
Chusainow, J., Yang, Y.S., Yeo, J.H.M., Toh, P.C., Asvadi, P., Wong, N.S.C., Yap, M.G.S., 2009. A study of monoclonal antibody-producing CHO cell lines: What makes a stable high producer? Biotechnology and Bioengineering 102, 1182–1196. https://doi.org/10.1002/bit.22158
Clappier, C., Böttner, D., Heinzelmann, D., Stadermann, A., Schulz, P., Schmidt, M., Lindner, B., 2023. Deciphering integration loci of CHO manufacturing cell lines using long read nanopore sequencing. New Biotechnology 75, 31–39. https://doi.org/10.1016/j.nbt.2023.03.003
Cockett, M.I., Bebbington, C.R., Yarranton, G.T., 1990. High Level Expression of Tissue Inhibitor of Metalloproteinases in Chinese Hamster Ovary Cells Using Glutamine Synthetase Gene Amplification. Nat Biotechnol 8, 662–667. https://doi.org/10.1038/nbt0790-662
Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., Zhang, F., 2013. Multiplex genome engineering using CRISPR/Cas systems. Science 339, 819–823. https://doi.org/10.1126/science.1231143
Cooper, G.M., 2000. The Endoplasmic Reticulum. The Cell: A Molecular Approach. 2nd edition.
Da Silva, J.D., Costa, M.D., Almeida, B., Lopes, F., Maciel, P., Teixeira-Castro, A., 2021. Case Report: A Novel GNB1 Mutation Causes Global Developmental Delay With Intellectual Disability and Behavioral Disorders. Frontiers in Neurology 12.
Dahodwala, H., Lee, K.H., 2019. The fickle CHO: a review of the causes, implications, and potential alleviation of the CHO cell line instability problem. Current Opinion in Biotechnology, Pharmaceutical Biotechnology ● Chemical Biotechnology 60, 128–137. https://doi.org/10.1016/j.copbio.2019.01.011
Davis, 2020. The History and Future Prospects of Antibodies [WWW Document]. St John's Laboratory Ltd. URL https://stjohnslabs.com/blog/the-history-and-future-prospects-of-antibodies (accessed 9.28.23).
Derouazi, M., Martinet, D., Besuchet Schmutz, N., Flaction, R., Wicht, M., Bertschinger, M., Hacker, D.L., Beckmann, J.S., Wurm, F.M., 2006. Genetic characterization of CHO production host DG44 and derivative recombinant cell lines. Biochemical and Biophysical Research Communications 340, 1069–1077. https://doi.org/10.1016/j.bbrc.2005.12.111

Dhiman, H., Campbell, M., Melcher, M., Smith, K.D., Borth, N., 2020. Predicting favorable landing pads for targeted integrations in Chinese hamster ovary cell lines by learning stability characteristics from random transgene integrations. Computational and Structural Biotechnology Journal 18, 3632–3648. https://doi.org/10.1016/j.csbj.2020.11.008

Ding, W., Cheng, J., Guo, D., Mao, L., Li, J., Lu, L., Zhang, Y., Yang, J., Jiang, H., 2018. Engineering the 5′ UTR-Mediated Regulation of Protein Abundance in Yeast Using Nucleotide Sequence Activity Relationships. ACS Synth. Biol. 7, 2709–2714. https://doi.org/10.1021/acssynbio.8b00127

Donaldson, J., Kleinjan, D.-J., Rosser, S., 2022. Synthetic biology approaches for dynamic CHO cell engineering. Current Opinion in Biotechnology 78, 102806. https://doi.org/10.1016/j.copbio.2022.102806

Donaldson, J.S., Dale, M.P., Rosser, S.J., 2021. Decoupling Growth and Protein Production in CHO Cells: A Targeted Approach. Front. Bioeng. Biotechnol. 9. https://doi.org/10.3389/fbioe.2021.658325

Doolan, P., Meleady, P., Barron, N., Henry, M., Gallagher, R., Gammell, P., Melville, M., Sinacore, M., McCarthy, K., Leonard, M., Charlebois, T., Clynes, M., 2010. Microarray and proteomics expression profiling identifies several candidates, including the valosin-containing protein (VCP), involved in regulating high cellular growth rate in production CHO cell lines. Biotechnology and Bioengineering 106, 42–56. https://doi.org/10.1002/bit.22670

Dorai, H., Corisdeo, S., Ellis, D., Kinney, C., Chomo, M., Hawley-Nelson, P., Moore, G., Betenbaugh, M.J., Ganguly, S., 2012. Early prediction of instability of chinese hamster ovary cell lines expressing recombinant antibodies and antibody-fusion proteins. Biotechnology and Bioengineering 109, 1016–1030. https://doi.org/10.1002/bit.24367

Dou, Y., Lin, Y., Wang, T., Wang, X.-Y., Jia, Y., Zhao, C., 2021. The CAG promoter maintains high-level transgene expression in HEK293 cells. FEBS Open Bio 11, 95–104. https://doi.org/10.1002/2211-5463.13029

Dreesen, I.A.J., Fussenegger, M., 2011. Ectopic expression of human mTOR increases viability, robustness, cell size, proliferation, and antibody production of chinese hamster ovary cells. Biotechnology and Bioengineering 108, 853–866. https://doi.org/10.1002/bit.22990

Du, Z., 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 Bioproc E 18, 419–429. https://doi.org/10.1007/s12257-012-0577-1

Dumont, J., Euwart, D., Mei, B., Estes, S., Kshirsagar, R., 2016. Human cell lines for biopharmaceutical manufacturing: history, status, and future perspectives. Crit Rev Biotechnol 36, 1110–1122. https://doi.org/10.3109/07388551.2015.1084266

Dvir, S., Velten, L., Sharon, E., Zeevi, D., Carey, L.B., Weinberger, A., Segal, E., 2013. Deciphering the rules by which 5′-UTR sequences affect protein expression in yeast. PNAS 110, E2792–E2801. https://doi.org/10.1073/pnas.1222534110

Eisenhut, P., Mebrahtu, A., Moradi Barzadd, M., Thalén, N., Klanert, G., Weinguny, M., Sandegren, A., Su, C., Hatton, D., Borth, N., Rockberg, J., 2020. Systematic use of synthetic 5′-UTR RNA structures to tune protein translation improves yield and quality of complex proteins in mammalian cell factories. Nucleic Acids Research 48, e119–e119. https://doi.org/10.1093/nar/gkaa847

Fagraeus, A., 1948. The plasma cellular reaction and its relation to the formation of antibodies in vitro. J Immunol 58, 1–13.

Fan, J., Li, Y., Jia, R., Fan, X., 2018. An inherited FGFR2 mutation increased osteogenesis gene expression and result in Crouzon syndrome. BMC Medical Genetics 19, 91. https://doi.org/10.1186/s12881-018-0607-8

Fan, L., Kadura, I., Krebs, L.E., Hatfield, C.C., Shaw, M.M., Frye, C.C., 2012. Improving the efficiency of CHO cell line generation using glutamine synthetase gene knockout cells. Biotechnology and Bioengineering 109, 1007–1015. https://doi.org/10.1002/bit.24365

Fann, C.H., Guirgis, F., Chen, G., Lao, M.S., Piret, J.M., 2000. Limitations to the amplification and stability of human tissue-type plasminogen activator expression by Chinese hamster ovary cells. Biotechnology and Bioengineering 69, 204–212. https://doi.org/10.1002/(SICI)1097-0290(20000720)69:2<204::AID-BIT9>3.0.CO;2-Z

Feary, M., Moffat, M.A., Casperson, G.F., Allen, M.J., Young, R.J., 2021. CHOK1SV GS-KO SSI expression system: A combination of the Fer1L4 locus and glutamine synthetase selection. Biotechnology Progress 37, e3137. https://doi.org/10.1002/btpr.3137

Figueroa Jr., B., Ailor, E., Osborne, D., Hardwick, J.M., Reff, M., Betenbaugh, M.J., 2007. Enhanced cell culture performance using inducible anti-apoptotic genes E1B-19K and Aven in the production of a monoclonal antibody with Chinese hamster ovary cells. Biotechnology and Bioengineering 97, 877–892. https://doi.org/10.1002/bit.21222

Fischer, S., Handrick, R., Otte, K., 2015. The art of CHO cell engineering: A comprehensive retrospect and future perspectives. Biotechnology Advances 33, 1878–1896. https://doi.org/10.1016/j.biotechadv.2015.10.015

Foquet, B., Song, H., 2020. There is no magic bullet: the importance of testing reference gene stability in RT-qPCR experiments across multiple closely related species. PeerJ 8, e9618. https://doi.org/10.7717/peerj.9618

Fu, T., Zhang, C., Jing, Y., Jiang, C., Li, Z., Wang, S., Ma, K., Zhang, D., Hou, S., Dai, J., Kou, G., Wang, H., 2016. Regulation of cell growth and apoptosis through lactate dehydrogenase C over-expression in Chinese hamster ovary cells. Appl Microbiol Biotechnol 100, 5007–5016. https://doi.org/10.1007/s00253-016-7348-4

Fus-Kujawa, A., Prus, P., Bajdak-Rusinek, K., Teper, P., Gawron, K., Kowalczuk, A., Sieron, A.L., 2021. An Overview of Methods and Tools for Transfection of Eukaryotic Cells in vitro. Frontiers in Bioengineering and Biotechnology 9.

Fussenegger, M., Mazur, X., Bailey, J.E., 1997. A novel cytostatic process enhances the productivity of Chinese hamster ovary cells. Biotechnology and Bioengineering 55, 927–939.

https://doi.org/10.1002/(SICI)1097-0290(19970920)55:6<927::AID-BIT10>3.0.CO;2-4
Gagnon, M., Hiller, G., Luan, Y.-T., Kittredge, A., DeFelice, J., Drapeau, D., 2011. High-End pH-controlled delivery of glucose effectively suppresses lactate accumulation in CHO Fed-batch cultures. Biotechnology and Bioengineering 108, 1328–1337. https://doi.org/10.1002/bit.23072
Gaidukov, L., Wroblewska, L., Teague, B., Nelson, T., Zhang, X., Liu, Y., Jagtap, K., Mamo, S., Tseng, W.A., Lowe, A., Das, J., Bandara, K., Baijuraj, S., Summers, N.M., Lu, T.K., Zhang, L., Weiss, R., 2018. A multi-landing pad DNA integration platform for mammalian cell engineering. Nucleic Acids Res 46, 4072–4086. https://doi.org/10.1093/nar/gky216
Gaillet, B., Gilbert, R., Broussau, S., Pilotte, A., Malenfant, F., Mullick, A., Garnier, A., Massie, B., 2010. High-level recombinant protein production in CHO cells using lentiviral vectors and the cumate gene-switch. Biotechnology and Bioengineering 106, 203–215. https://doi.org/10.1002/bit.22698
Gallagher, C., Kelly, P.S., 2017. Selection of High-Producing Clones Using FACS for CHO Cell Line Development, in: Meleady, P. (Ed.), Heterologous Protein Production in CHO Cells: Methods and Protocols, Methods in Molecular Biology. Springer, New York, NY, pp. 143–152. https://doi.org/10.1007/978-1-4939-6972-2_9
Ghorbaniaghdam, A., Chen, J., Henry, O., Jolicoeur, M., 2014. Analyzing Clonal Variation of Monoclonal Antibody-Producing CHO Cell Lines Using an In Silico Metabolomic Platform. PLOS ONE 9, e90832. https://doi.org/10.1371/journal.pone.0090832
Girod, P.-A., Zahn-Zabal, M., Mermod, N., 2005. Use of the chicken lysozyme 5′ matrix attachment region to generate high producer CHO cell lines. Biotechnology and Bioengineering 91, 1–11. https://doi.org/10.1002/bit.20563
Gossen, M., Freundlieb, S., Bender, G., Muller, G., Hillen, W., Bujard, H., 1995. Transcriptional activation by tetracyclines in mammalian cells. Science 268, 1766–1769. https://doi.org/10.1126/science.7792603
Granot-Attas, S., Elson, A., 2004. Protein tyrosine phosphatase epsilon activates Yes and Fyn in Neu-induced mammary tumor cells. Experimental Cell Research 294, 236–243. https://doi.org/10.1016/j.yexcr.2003.11.003
Grav, L.M., la Cour Karottki, K.J., Lee, J.S., Kildegaard, H.F., 2017. Application of CRISPR/Cas9 Genome Editing to Improve Recombinant Protein Production in CHO Cells, in: Meleady, P. (Ed.), Heterologous Protein Production in CHO Cells: Methods and Protocols, Methods in Molecular Biology. Springer, New York, NY, pp. 101–118. https://doi.org/10.1007/978-1-4939-6972-2_7
Grav, L.M., Lee, J.S., Gerling, S., Kallehauge, T.B., Hansen, A.H., Kol, S., Lee, G.M., Pedersen, L.E., Kildegaard, H.F., 2015. One-step generation of triple knockout CHO cell lines using CRISPR/Cas9 and fluorescent enrichment. Biotechnol J 10, 1446–1456. https://doi.org/10.1002/biot.201500027

Grav, L.M., Sergeeva, D., Lee, J.S., Marin de Mas, I., Lewis, N.E., Andersen, M.R., Nielsen, L.K., Lee, G.M., Kildegaard, H.F., 2018. Minimizing Clonal Variation during Mammalian Cell Line Engineering for Improved Systems Biology Data Generation. ACS Synth. Biol. 7, 2148–2159. https://doi.org/10.1021/acssynbio.8b00140

Gray, F., Kenney, J.S., Dunne, J.F., 1995. Secretion capture and report web: use of affinity derivatized agarose microdroplets for the selection of hybridoma cells. Journal of Immunological Methods 182, 155–163. https://doi.org/10.1016/0022-1759(94)00319-R

Grens, A., Scheffler, I.E., 1990. The 5'- and 3'-untranslated regions of ornithine decarboxylase mRNA affect the translational efficiency. J Biol Chem 265, 11810–11816.

Grosveld, F., van Assendelft, G.B., Greaves, D.R., Kollias, G., 1987. Position-independent, high-level expression of the human β-globin gene in transgenic mice. Cell 51, 975–985. https://doi.org/10.1016/0092-8674(87)90584-8

Gu, M.B., Todd, P., Kompala, D.S., 1995. Metabolic burden in recombinant CHO cells: effect ofdhfr gene amplification andlacZ expression. Cytotechnology 18, 159–166. https://doi.org/10.1007/BF00767763

Gu, M.B., Todd, P., Kompala, D.S., 1994. Analysis of Foreign Protein Overproduction in Recombinant CHO Cells. Annals of the New York Academy of Sciences 721, 194–207. https://doi.org/10.1111/j.1749-6632.1994.tb47392.x

Guo, D., Gao, A., Michels, D.A., Feeney, L., Eng, M., Chan, B., Laird, M.W., Zhang, B., Yu, X.C., Joly, J., Snedecor, B., Shen, A., 2010. Mechanisms of unintended amino acid sequence changes in recombinant monoclonal antibodies expressed in Chinese Hamster Ovary (CHO) cells. Biotechnology and Bioengineering 107, 163–171. https://doi.org/10.1002/bit.22780

Hamaker, N.K., Lee, K.H., 2018. Site-specific Integration Ushers in a New Era of Precise CHO Cell Line Engineering. Curr Opin Chem Eng 22, 152–160. https://doi.org/10.1016/j.coche.2018.09.011

Hamilton, 2021. The Integral of Viable Cell Density in Biopharma.

Harraghy, N., Calabrese, D., Fisch, I., Girod, P.-A., LeFourn, V., Regamey, A., Mermod, N., 2015. Epigenetic regulatory elements: Recent advances in understanding their mode of action and use for recombinant protein production in mammalian cells. Biotechnology Journal 10, 967–978. https://doi.org/10.1002/biot.201400649

Hartley, F., Walker, T., Chung, V., Morten, K., 2018. Mechanisms driving the lactate switch in Chinese hamster ovary cells. Biotechnology and Bioengineering 115, 1890–1903. https://doi.org/10.1002/bit.26603

Haryadi, R., Ho, S., Kok, Y.J., Pu, H.X., Zheng, L., Pereira, N.A., Li, B., Bi, X., Goh, L.-T., Yang, Y., Song, Z., 2015. Optimization of Heavy Chain and Light Chain Signal Peptides for High Level Expression of Therapeutic Antibodies in CHO Cells. PLOS ONE 10, e0116878. https://doi.org/10.1371/journal.pone.0116878

Hershey, J.W.B., Sonenberg, N., Mathews, M.B., 2012. Principles of Translational Control: An Overview. Cold Spring Harb Perspect Biol 4, a011528. https://doi.org/10.1101/cshperspect.a011528

Hilliard, W., Lee, K.H., 2021. Systematic identification of safe harbor regions in the CHO genome through a comprehensive epigenome analysis. Biotechnology and Bioengineering 118, 659–675. https://doi.org/10.1002/bit.27599

Ho, S.C.L., Koh, E.Y.C., Soo, B.P.C., Mariati, Chao, S.-H., Yang, Y., 2016. Evaluating the use of a CpG free promoter for long-term recombinant protein expression stability in Chinese hamster ovary cells. BMC Biotechnology 16, 71. https://doi.org/10.1186/s12896-016-0300-y

Holmes, P., Al-Rubeai, M., 1999. Improved cell line development by a high throughput affinity capture surface display technique to select for high secretors. Journal of Immunological Methods 230, 141–147. https://doi.org/10.1016/S0022-1759(99)00181-7

Hong, M.S., Severson, K.A., Jiang, M., Lu, A.E., Love, J.C., Braatz, R.D., 2018. Challenges and opportunities in biopharmaceutical manufacturing control. Computers & Chemical Engineering 110, 106–114. https://doi.org/10.1016/j.compchemeng.2017.12.007

Huang, Y., Li, Y., Wang, Y.G., Gu, X., Wang, Y., Shen, B.F., 2007. An efficient and targeted gene integration system for high-level antibody expression. Journal of Immunological Methods 322, 28–39. https://doi.org/10.1016/j.jim.2007.01.022

ICH Q5B, 1996. Q5B Quality of Biotechnological Products: Analysis of the Expression Construct in Cells Used for Production of r-DNA Derived Protein Products.

Ilieva, K.M., Cheung, A., Mele, S., Chiaruttini, G., Crescioli, S., Griffin, M., Nakamura, M., Spicer, J.F., Tsoka, S., Lacy, K.E., Tutt, A.N.J., Karagiannis, S.N., 2018. Chondroitin Sulfate Proteoglycan 4 and Its Potential As an Antibody Immunotherapy Target across Different Tumor Types. Frontiers in Immunology 8.

Inniss, M.C., Bandara, K., Jusiak, B., Lu, T.K., Weiss, R., Wroblewska, L., Zhang, L., 2017. A novel Bxb1 integrase RMCE system for high fidelity site-specific integration of mAb expression cassette in CHO Cells. Biotechnology and Bioengineering 114, 1837–1846. https://doi.org/10.1002/bit.26268

Izumi, M., Gilbert, D.M., 2000. Homogeneous tetracycline-regulatable gene expression in mammalian fibroblasts. Journal of Cellular Biochemistry 76, 280–289. https://doi.org/10.1002/(SICI)1097-4644(20000201)76:2<280::AID-JCB11>3.0.CO;2-0

Jaluria, P., Betenbaugh, M., Konstantopoulos, K., Shiloach, J., 2007. Enhancement of cell proliferation in various mammalian cell lines by gene insertion of a cyclin-dependent kinase homolog. BMC Biotechnol 7, 71. https://doi.org/10.1186/1472-6750-7-71

Jamnikar, U., Nikolic, P., Belic, A., Blas, M., Gaser, D., Francky, A., Laux, H., Blejec, A., Baebler, S., Gruden, K., 2015. Transcriptome study and identification of potential marker genes related to the stable expression of recombinant proteins in CHO clones. BMC Biotechnology 15, 98. https://doi.org/10.1186/s12896-015-0218-9

Jenkins, N., Murphy, L., Tyther, R., 2008. Post-translational Modifications of Recombinant Proteins: Significance for Biopharmaceuticals. Mol Biotechnol 39, 113–118. https://doi.org/10.1007/s12033-008-9049-4

Johari, Y.B., Brown, A.J., Alves, C.S., Zhou, Y., Wright, C.M., Estes, S.D., Kshirsagar, R., James, D.C., 2019. CHO genome mining for synthetic promoter design. Journal of Biotechnology 294, 1–13. https://doi.org/10.1016/j.jbiotec.2019.01.015

Johari, Y.B., Estes, S.D., Alves, C.S., Sinacore, M.S., James, D.C., 2015. Integrated cell and process engineering for improved transient production of a “difficult-to-express“ fusion protein by CHO cells. Biotechnology and Bioengineering 112, 2527–2542. https://doi.org/10.1002/bit.25687

Johari, Y.B., Mercer, A.C., Liu, Y., Brown, A.J., James, D.C., 2021. Design of synthetic promoters for controlled expression of therapeutic genes in retinal pigment epithelial cells. Biotechnology and Bioengineering 118, 2001–2015. https://doi.org/10.1002/bit.27713

Johnson, A.O., Fowler, S.B., Webster, C.I., Brown, A.J., James, D.C., 2022. Bioinformatic Design of Dendritic Cell-Specific Synthetic Promoters. ACS Synth. Biol. 11, 1613–1626. https://doi.org/10.1021/acssynbio.2c00027

Jusiak, B., Jagtap, K., Gaidukov, L., Duportet, X., Bandara, K., Chu, J., Zhang, L., Weiss, R., Lu, T.K., 2019. Comparison of Integrases Identifies Bxb1-GA Mutant as the Most Efficient Site-Specific Integrase System in Mammalian Cells. ACS Synth. Biol. 8, 16–24. https://doi.org/10.1021/acssynbio.8b00089

Kallehauge, T.B., Li, S., Pedersen, L.E., Ha, T.K., Ley, D., Andersen, M.R., Kildegaard, H.F., Lee, G.M., Lewis, N.E., 2017. Ribosome profiling-guided depletion of an mRNA increases cell growth rate and protein secretion. Sci Rep 7, 40388. https://doi.org/10.1038/srep40388

Kaneyoshi, K., Kuroda, K., Uchiyama, K., Onitsuka, M., Yamano-Adachi, N., Koga, Y., Omasa, T., 2019. Secretion analysis of intracellular “difficult-to-express” immunoglobulin G (IgG) in Chinese hamster ovary (CHO) cells. Cytotechnology 71, 305–316. https://doi.org/10.1007/s10616-018-0286-5

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

Karollus, A., Avsec, Ž., Gagneur, J., 2021. Predicting mean ribosome load for 5’UTR of any length using deep learning. PLOS Computational Biology 17, e1008982. https://doi.org/10.1371/journal.pcbi.1008982

Kaufman, R.J., 2000. Overview of vector design for mammalian gene expression. Mol Biotechnol 16, 151–160. https://doi.org/10.1385/MB:16:2:151

Kaufmann, S.H.E., 2017. Remembering Emil von Behring: from Tetanus Treatment to Antibody Cooperation with Phagocytes. mBio 8, e00117-17. https://doi.org/10.1128/mBio.00117-17

Kaunitz, J.D., 2017. Development of Monoclonal Antibodies: The Dawn of mAb Rule. Dig Dis Sci 62, 831–832. https://doi.org/10.1007/s10620-017-4478-1

Kelley, B., 2020. Developing therapeutic monoclonal antibodies at pandemic pace. Nature Biotechnology 38, 540–545. https://doi.org/10.1038/s41587-020-0512-5

Kelley, B., 2009. Industrialization of mAb production technology The bioprocessing industry at a crossroads. MAbs 1, 443–452.
Kelley, B., Kiss, R., Laird, M., 2018. A Different Perspective: How Much Innovation Is Really Needed for Monoclonal Antibody Production Using Mammalian Cell Technology?, in: Kiss, B., Gottschalk, U., Pohlscheidt, M. (Eds.), New Bioprocessing Strategies: Development and Manufacturing of Recombinant Antibodies and Proteins, Advances in Biochemical Engineering/Biotechnology. Springer International Publishing, Cham, pp. 443–462. https://doi.org/10.1007/10_2018_59
Kesik-Brodacka, M., 2018. Progress in biopharmaceutical development. Biotechnol Appl Biochem 65, 306–322. https://doi.org/10.1002/bab.1617
Kim, M., O'Callaghan, P.M., Droms, K.A., James, D.C., 2011. A mechanistic understanding of production instability in CHO cell lines expressing recombinant monoclonal antibodies. Biotechnology and Bioengineering 108, 2434–2446. https://doi.org/10.1002/bit.23189
Kim, M.S., Kim, W.H., Lee, G.M., 2008. Characterization of site-specific recombination mediated by Cre recombinase during the development of erythropoietin producing CHO cell lines. Biotechnol Bioproc E 13, 418–423. https://doi.org/10.1007/s12257-008-0151-z
Kim, S.J., Kim, N.S., Ryu, C.J., Hong, H.J., Lee, G.M., 1998. Characterization of chimeric antibody producing CHO cells in the course of dihydrofolate reductase-mediated gene amplification and their stability in the absence of selective pressure. Biotechnology and Bioengineering 58, 73–84. https://doi.org/10.1002/(SICI)1097-0290(19980405)58:1<73::AID-BIT8>3.0.CO;2-R
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 Biotechnol 60, 442–448. https://doi.org/10.1007/s00253-002-1134-1
Kober, L., Zehe, C., Bode, J., 2013. Optimized signal peptides for the development of high expressing CHO cell lines. Biotechnology and Bioengineering 110, 1164–1173. https://doi.org/10.1002/bit.24776
Köhler, G., Milstein, C., 1975. Continuous cultures of fused cells secreting antibody of predefined specificity. Nature 256, 495–497. https://doi.org/10.1038/256495a0
Kol, S., Ley, D., Wulff, T., Decker, M., Arnsdorf, J., Schoffelen, S., Hansen, A.H., Jensen, T.L., Gutierrez, J.M., Chiang, A.W.T., Masson, H.O., Palsson, B.O., Voldborg, B.G., Pedersen, L.E., Kildegaard, H.F., Lee, G.M., Lewis, N.E., 2020. Multiplex secretome engineering enhances recombinant protein production and purity. Nat Commun 11, 1908. https://doi.org/10.1038/s41467-020-15866-w
Kunert, R., Reinhart, D., 2016. Advances in recombinant antibody manufacturing. Appl Microbiol Biotechnol 100, 3451–3461. https://doi.org/10.1007/s00253-016-7388-9
Kwaks, T.H.J., Barnett, P., Hemrika, W., Siersma, T., Sewalt, R.G.A.B., Satijn, D.P.E., Brons, J.F., van Blokland, R., Kwakman, P., Kruckeberg, A.L., Kelder, A., Otte, A.P., 2003. Identification of anti-repressor elements that confer high and stable protein production in

mammalian cells. Nat Biotechnol 21, 553–558. https://doi.org/10.1038/nbt814
Kwaks, T.H.J., Otte, A.P., 2006. Employing epigenetics to augment the expression of therapeutic proteins in mammalian cells. Trends in Biotechnology 24, 137–142. https://doi.org/10.1016/j.tibtech.2006.01.007
Kwon, R.-J., Kim, S.K., Lee, S.-I., Hwang, S.-J., Lee, G.M., Kim, J.-S., Seol, W., 2006. Artificial Transcription Factors Increase Production of Recombinant Antibodies in Chinese Hamster Ovary Cells. Biotechnol Lett 28, 9–15. https://doi.org/10.1007/s10529-005-4680-7
Lakshmanan, M., Kok, Y.J., Lee, A.P., Kyriakopoulos, S., Lim, H.L., Teo, G., Poh, S.L., Tang, W.Q., Hong, J., Tan, A.H.-M., Bi, X., Ho, Y.S., Zhang, P., Ng, S.K., Lee, D.-Y., 2019. Multi-omics profiling of CHO parental hosts reveals cell line-specific variations in bioprocessing traits. Biotechnology and Bioengineering 116, 2117–2129. https://doi.org/10.1002/bit.27014
Lam, C., Santell, L., Wilson, B., Yim, M., Louie, S., Tang, D., Shaw, D., Chan, P., Lazarus, R.A., Snedecor, B., Misaghi, S., 2017. Taming hyperactive hDNase I: Stable inducible expression of a hyperactive salt- and actin-resistant variant of human deoxyribonuclease I in CHO cells. Biotechnology Progress 33, 523–533. https://doi.org/10.1002/btpr.2439
Laurell, H., Iacovoni, J.S., Abot, A., Svec, D., Maoret, J.-J., Arnal, J.-F., Kubista, M., 2012. Correction of RT–qPCR data for genomic DNA-derived signals with ValidPrime. Nucleic Acids Res 40, e51. https://doi.org/10.1093/nar/gkr1259
Le, H., Kabbur, S., Pollastrini, L., Sun, Z., Mills, K., Johnson, K., Karypis, G., Hu, W.-S., 2012. Multivariate analysis of cell culture bioprocess data—Lactate consumption as process indicator. Journal of Biotechnology 162, 210–223. https://doi.org/10.1016/j.jbiotec.2012.08.021
Le, H., Vishwanathan, N., Kantardjieff, A., Doo, I., Srienc, M., Zheng, X., Somia, N., Hu, W.-S., 2013. Dynamic gene expression for metabolic engineering of mammalian cells in culture. Metabolic Engineering 20, 212–220. https://doi.org/10.1016/j.ymben.2013.09.004
Le, K., Tan, C., Gupta, S., Guhan, T., Barkhordarian, H., Lull, J., Stevens, J., Munro, T., 2018. A novel mammalian cell line development platform utilizing nanofluidics and optoelectro positioning technology. Biotechnology Progress 34, 1438–1446. https://doi.org/10.1002/btpr.2690
Lee, J.S., Ha, T.K., Park, J.H., Lee, G.M., 2013. Anti-cell death engineering of CHO cells: Co-overexpression of Bcl-2 for apoptosis inhibition, Beclin-1 for autophagy induction. Biotechnology and Bioengineering 110, 2195–2207. https://doi.org/10.1002/bit.24879
Lee, J.S., Kildegaard, H.F., Lewis, N.E., Lee, G.M., 2019. Mitigating Clonal Variation in Recombinant Mammalian Cell Lines. Trends in Biotechnology 37, 931–942. https://doi.org/10.1016/j.tibtech.2019.02.007
Leonard, M., Hone, M., Cooley, C., Mccarthy, K., Crowe, K., Heller-Harrison, R., 2009. Managing Cell Line Instability and Its Impact During Cell

Line Development. BioPharm International, BioPharm International-06-02-2009 2009 Supplement.
Li, F., Vijayasankaran, N., Shen, A. (Yijuan), Kiss, R., Amanullah, A., 2010. Cell culture processes for monoclonal antibody production. MAbs 2, 466–477. https://doi.org/10.4161/mabs.2.5.12720
Li, H., Chen, K., Wang, Z., Li, D., Lin, J., Yu, C., Yu, F., Wang, X., Huang, L., Jiang, C., Gu, H., Fang, J., 2015. Genetic analysis of the clonal stability of Chinese hamster ovary cells for recombinant protein production. Mol. BioSyst. 12, 102–109. https://doi.org/10.1039/C5MB00627A
Li, H., Sharp, R., Rutherford, K., Gupta, K., Duyne, G.D.V., 2018. Serine integrase attP binding and specificity. Journal of molecular biology 430, 4401. https://doi.org/10.1016/j.jmb.2018.09.007
Li, S., Gao, X., Peng, R., Zhang, S., Fu, W., Zou, F., 2016. FISH-Based Analysis of Clonally Derived CHO Cell Populations Reveals High Probability for Transgene Integration in a Terminal Region of Chromosome 1 (1q13). PLOS ONE 11, e0163893. https://doi.org/10.1371/journal.pone.0163893
Liu, J.K.H., 2014. The history of monoclonal antibody development – Progress, remaining challenges and future innovations. Ann Med Surg (Lond) 3, 113–116. https://doi.org/10.1016/j.amsu.2014.09.001
López, G.Y., Van Ziffle, J., Onodera, C., Grenert, J.P., Yeh, I., Bastian, B.C., Clarke, J., Oberheim Bush, N.A., Taylor, J., Chang, S., Butowski, N., Banerjee, A., Mueller, S., Kline, C., Torkildson, J., Samuel, D., Siongco, A., Raffel, C., Gupta, N., Kunwar, S., Mummaneni, P., Aghi, M., Theodosopoulos, P., Berger, M., Phillips, J.J., Pekmezci, M., Tihan, T., Bollen, A.W., Perry, A., Solomon, D.A., 2019. The genetic landscape of gliomas arising after therapeutic radiation. Acta Neuropathol 137, 139–150. https://doi.org/10.1007/s00401-018-1906-z
Lu, R.-M., Hwang, Y.-C., Liu, I.-J., Lee, C.-C., Tsai, H.-Z., Li, H.-J., Wu, H.-C., 2020. Development of therapeutic antibodies for the treatment of diseases. Journal of Biomedical Science 27, 1. https://doi.org/10.1186/s12929-019-0592-z
Ma, X., Zhang, Ling, Zhang, Luming, Wang, C., Guo, X., Yang, Y., Wang, L., Li, X., Ma, N., 2020. Validation and identification of reference genes in Chinese hamster ovary cells for Fc-fusion protein production. Exp Biol Med (Maywood) 245, 690–702. https://doi.org/10.1177/1535370220914058
Ma, Y., Budde, M.W., Mayalu, M.N., Zhu, J., Lu, A.C., Murray, R.M., Elowitz, M.B., 2022. Synthetic mammalian signaling circuits for robust cell population control. Cell 185, 967-979.e12. https://doi.org/10.1016/j.cell.2022.01.026
Maehle, A.-H., 2009. A binding question: the evolution of the receptor concept. Endeavour 33, 135–140. https://doi.org/10.1016/j.endeavour.2009.09.001
Mahmuda, A., Bande, F., Al-Zihiry, K.J.K., Abdulhaleem, N., Majid, R.A., Hamat, R.A., Abdullah, W.O., Unyah, Z., 2017. Monoclonal antibodies: A review of therapeutic applications and future prospects.

Tropical Journal of Pharmaceutical Research 16, 713–722. https://doi.org/10.4314/tjpr.v16i3
Martella, A., Matjusaitis, M., Auxillos, J., Pollard, S.M., Cai, Y., 2017. EMMA: An Extensible Mammalian Modular Assembly Toolkit for the Rapid Design and Production of Diverse Expression Vectors. ACS Synth. Biol. 6, 1380–1392. https://doi.org/10.1021/acssynbio.7b00016
Martinelli, R., De Simone, V., 2005. Short and highly efficient synthetic promoters for melanoma-specific gene expression. FEBS Letters 579, 153–156. https://doi.org/10.1016/j.febslet.2004.11.068
Martinet, D., Derouazi, M., Besuchet, N., Wicht, M., Beckmann, J., Wurm, F.M., 2007. Karyotype of CHO DG44 cells, in: Smith, R. (Ed.), Cell Technology for Cell Products. Springer Netherlands, Dordrecht, pp. 363–366. https://doi.org/10.1007/978-1-4020-5476-1_59
Marx, N., Grünwald-Gruber, C., Bydlinski, N., Dhiman, H., Ngoc Nguyen, L., Klanert, G., Borth, N., 2018. CRISPR-Based Targeted Epigenetic Editing Enables Gene Expression Modulation of the Silenced Beta-Galactoside Alpha-2,6-Sialyltransferase 1 in CHO Cells. Biotechnology Journal 13, 1700217. https://doi.org/10.1002/biot.201700217
Mason, M., Sweeney, B., Cain, K., Stephens, P., Sharfstein, S.T., 2012. Identifying bottlenecks in transient and stable production of recombinant monoclonal-antibody sequence variants in chinese hamster ovary cells. Biotechnology Progress 28, 846–855. https://doi.org/10.1002/btpr.1542
Matte, A., 2022. Recent Advances and Future Directions in Downstream Processing of Therapeutic Antibodies. International Journal of Molecular Sciences 23, 8663. https://doi.org/10.3390/ijms23158663
Mazur, X., Fussenegger, M., Renner, W.A., Bailey, J.E., 1998. Higher Productivity of Growth-Arrested Chinese Hamster Ovary Cells Expressing the Cyclin-Dependent Kinase Inhibitor p27. Biotechnology Progress 14, 705–713. https://doi.org/10.1021/bp980062h
Merrick, C.A., Zhao, J., Rosser, S.J., 2018. Serine Integrases: Advancing Synthetic Biology. ACS Synth. Biol. 7, 299–310. https://doi.org/10.1021/acssynbio.7b00308
Misaghi, S., Chang, J., Snedecor, B., 2014. It's time to regulate: Coping with product-induced nongenetic clonal instability in CHO cell lines via regulated protein expression. Biotechnology Progress 30, 1432–1440. https://doi.org/10.1002/btpr.1970
Molecular Devices, 2023. ClonePix 2 Mammalian Colony Picker [WWW Document]. URL https://www.moleculardevices.com/products/clone-screening/mammalian-screening/clonepix-2-mammalian-colony-picker (accessed 10.13.23).
Moritz, B., Becker, P.B., Göpfert, U., 2015. CMV promoter mutants with a reduced propensity to productivity loss in CHO cells. Sci Rep 5, 16952. https://doi.org/10.1038/srep16952
Moritz, B., Woltering, L., Becker, P.B., Göpfert, U., 2016. High levels of histone H3 acetylation at the CMV promoter are predictive of stable expression in Chinese hamster ovary cells. Biotechnology Progress 32, 776–786. https://doi.org/10.1002/btpr.2271

Morrison, S.L., Johnson, M.J., Herzenberg, L.A., Oi, V.T., 1984. Chimeric human antibody molecules: mouse antigen-binding domains with human constant region domains. Proc Natl Acad Sci U S A 81, 6851–6855.

Mullard, A., 2021. FDA approves 100th monoclonal antibody product. Nature Reviews Drug Discovery 20, 491–495. https://doi.org/10.1038/d41573-021-00079-7

Mullick, A., Xu, Y., Warren, R., Koutroumanis, M., Guilbault, C., Broussau, S., Malenfant, F., Bourget, L., Lamoureux, L., Lo, R., Caron, A.W., Pilotte, A., Massie, B., 2006. The cumate gene-switch: a system for regulated expression in mammalian cells. BMC Biotechnol 6, 43. https://doi.org/10.1186/1472-6750-6-43

Mulukutla, B.C., Mitchell, J., Geoffroy, P., Harrington, C., Krishnan, M., Kalomeris, T., Morris, C., Zhang, L., Pegman, P., Hiller, G.W., 2019. Metabolic engineering of Chinese hamster ovary cells towards reduced biosynthesis and accumulation of novel growth inhibitors in fed-batch cultures. Metabolic Engineering 54, 54–68. https://doi.org/10.1016/j.ymben.2019.03.001

Nawaz, A.A., Chen, Y., Nama, N., Nissly, R.H., Ren, L., Ozcelik, A., Wang, L., McCoy, J.P., Levine, S.J., Huang, T.J., 2015. Acoustofluidic Fluorescence Activated Cell Sorter. Anal Chem 87, 12051–12058. https://doi.org/10.1021/acs.analchem.5b02398

Neville, J.J., Orlando, J., Mann, K., McCloskey, B., Antoniou, M.N., 2017. Ubiquitous Chromatin-opening Elements (UCOEs): Applications in biomanufacturing and gene therapy. Biotechnology Advances 35, 557–564. https://doi.org/10.1016/j.biotechadv.2017.05.004

NextMSC, 2022. Biopharmaceutical Market Size and Share | Analysis - 2030 [WWW Document]. Biopharmaceutical Market by Product (Monoclonal Antibody, Interferon, Insulin, Growth and Coagulation Factor, Erythropoietin, Vaccine, Hormone and Others) and by Application (Oncolology, Blood Disorder, Metabolic Disease, Infectious Disease, Cardiovascular Disease, Neurological Disease, Immunology, and Others) - Global Opportunity Analysis and Industry Forecast 2022-2030. URL https://www.nextmsc.com/report/biopharmaceutical-market (accessed 9.26.23).

Nguyen, L.N., Baumann, M., Dhiman, H., Marx, N., Schmieder, V., Hussein, M., Eisenhut, P., Hernandez, I., Koehn, J., Borth, N., 2019. Novel Promoters Derived from Chinese Hamster Ovary Cells via In Silico and In Vitro Analysis. Biotechnology Journal 14, 1900125. https://doi.org/10.1002/biot.201900125

Nguyen, L.N., Novak, N., Baumann, M., Koehn, J., Borth, N., 2020. Bioinformatic Identification of Chinese Hamster Ovary (CHO) Cold-Shock Genes and Biological Evidence of their Cold-Inducible Promoters. Biotechnology Journal 15, 1900359. https://doi.org/10.1002/biot.201900359

Nguyen, N.T.B., Lin, J., Tay, S.J., Mariati, Yeo, J., Nguyen-Khuong, T., Yang, Y., 2021. Multiplexed engineering glycosyltransferase genes in CHO cells via targeted integration for producing antibodies with diverse complex-type N-glycans. Sci Rep 11, 12969. https://doi.org/10.1038/s41598-021-92320-x

Niazi, S., Lokesh, S., 2021. Introduction to biopharmaceuticals, in: Biopharmaceutical Manufacturing, Volume 1: Regulatory Processes. IOP Publishing. https://doi.org/10.1088/978-0-7503-3175-3ch1
Nijhout, H.F., 2013. Stochastic Gene Expression: Dominance, Thresholds and Boundaries, in: Madame Curie Bioscience Database [Internet]. Landes Bioscience.
Nixon, A.E., Sexton, D.J., Ladner, R.C., 2014. Drugs derived from phage display. MAbs 6, 73–85. https://doi.org/10.4161/mabs.27240
Nolan, R.P., Lee, K., 2011. Dynamic model of CHO cell metabolism. Metabolic Engineering 13, 108–124. https://doi.org/10.1016/j.ymben.2010.09.003
O'Callaghan, P.M., Racher, A.J., 2015. Building a Cell Culture Process with Stable Foundations: Searching for Certainty in an Uncertain World, in: Al-Rubeai, M. (Ed.), Animal Cell Culture, Cell Engineering. Springer International Publishing, Cham, pp. 373–406. https://doi.org/10.1007/978-3-319-10320-4_12
Ohya, T., Hayashi, T., Kiyama, E., Nishii, H., Miki, H., Kobayashi, K., Honda, K., Omasa, T., Ohtake, H., 2008. Improved production of recombinant human antithrombin III in Chinese hamster ovary cells by ATF4 overexpression. Biotechnology and Bioengineering 100, 317–324. https://doi.org/10.1002/bit.21758
Osterlehner, A., Simmeth, S., Göpfert, U., 2011. Promoter methylation and transgene copy numbers predict unstable protein production in recombinant chinese hamster ovary cell lines. Biotechnology and Bioengineering 108, 2670–2681. https://doi.org/10.1002/bit.23216
Otte, A.P., Kwaks, T.H.J., Van Blokland, R.J.M., Sewalt, R.G.A.B., Verhees, J., Klaren, V.N.A., Siersma, T.K., Korse, H.W.M., Teunissen, N.C., Botschuijver, S., Van Mer, C., Man, S.Y., 2007. Various Expression-Augmenting DNA Elements Benefit from STAR-Select, a Novel High Stringency Selection System for Protein Expression. Biotechnology Progress 23, 801–807. https://doi.org/10.1021/bp070107r
Ozdemir Kutbay, N., Biray Avci, C., Sarer Yurekli, B., Caliskan Kurt, C., Shademan, B., Gunduz, C., Erdogan, M., 2020. Effects of metformin and pioglitazone combination on apoptosis and AMPK/mTOR signaling pathway in human anaplastic thyroid cancer cells. Journal of Biochemical and Molecular Toxicology 34, e22547. https://doi.org/10.1002/jbt.22547
Pan, X., Dalm, C., Wijffels, R.H., Martens, D.E., 2017. Metabolic characterization of a CHO cell size increase phase in fed-batch cultures. Appl Microbiol Biotechnol 101, 8101–8113. https://doi.org/10.1007/s00253-017-8531-y
Park, H.-S., Kim, I.-H., Kim, I.-Y., Kim, K.-H., Kim, H.-J., 2000. Expression of carbamoyl phosphate synthetase I and ornithine transcarbamoylase genes in Chinese hamster ovary dhfr-cells decreases accumulation of ammonium ion in culture media. Journal of Biotechnology 81, 129–140. https://doi.org/10.1016/S0168-1656(00)00282-0
Park, J.-H., Lee, H.-M., Jin, E.-J., Lee, E.-J., Kang, Y.-J., Kim, S., Yoo, S.-S., Lee, G.M., Kim, Y.-G., 2022. Development of an in vitro screening system for synthetic signal peptide in mammalian cell-based protein

production. Appl Microbiol Biotechnol 106, 3571–3582. https://doi.org/10.1007/s00253-022-11955-6
Patel, Y.D., Brown, A.J., Zhu, J., Rosignoli, G., Gibson, S.J., Hatton, D., James, D.C., 2021. Control of Multigene Expression Stoichiometry in Mammalian Cells Using Synthetic Promoters. ACS Synth. Biol. 10, 1155–1165. https://doi.org/10.1021/acssynbio.0c00643
Pichler, J., Hesse, F., Wieser, M., Kunert, R., Galosy, S.S., Mott, J.E., Borth, N., 2009. A study on the temperature dependency and time course of the cold capture antibody secretion assay. Journal of Biotechnology 141, 80–83. https://doi.org/10.1016/j.jbiotec.2009.03.001
Plass, M., Rasmussen, S.H., Krogh, A., 2017. Highly accessible AU-rich regions in 3' untranslated regions are hotspots for binding of regulatory factors. PLoS Comput Biol 13, e1005460. https://doi.org/10.1371/journal.pcbi.1005460
Posner, J., Barrington, P., Brier, T., Datta-Mannan, A., 2019. Monoclonal Antibodies: Past, Present and Future, in: Barrett, J.E., Page, C.P., Michel, M.C. (Eds.), Concepts and Principles of Pharmacology: 100 Years of the Handbook of Experimental Pharmacology, Handbook of Experimental Pharmacology. Springer International Publishing, Cham, pp. 81–141. https://doi.org/10.1007/164_2019_323
Poulain, A., Mullick, A., Massie, B., Durocher, Y., 2019. Reducing recombinant protein expression during CHO pool selection enhances frequency of high-producing cells. Journal of Biotechnology 296, 32–41. https://doi.org/10.1016/j.jbiotec.2019.03.009
Poulain, A., Perret, S., Malenfant, F., Mullick, A., Massie, B., Durocher, Y., 2017. Rapid protein production from stable CHO cell pools using plasmid vector and the cumate gene-switch. Journal of Biotechnology 255, 16–27. https://doi.org/10.1016/j.jbiotec.2017.06.009
Powell, K.T., Weaver, J.C., 1990. Gel Microdroplets and Flow Cytometry: Rapid Determination of Antibody Secretion by Individual Cells Within a Cell Population. Nat Biotechnol 8, 333–337. https://doi.org/10.1038/nbt0490-333
Priola, J.J., Calzadilla, N., Baumann, M., Borth, N., Tate, C.G., Betenbaugh, M.J., 2016. High-throughput screening and selection of mammalian cells for enhanced protein production. Biotechnology Journal 11, 853–865. https://doi.org/10.1002/biot.201500579
Prioleau, M.N., Nony, P., Simpson, M., Felsenfeld, G., 1999. An insulator element and condensed chromatin region separate the chicken beta-globin locus from an independently regulated erythroid-specific folate receptor gene. EMBO J 18, 4035–4048. https://doi.org/10.1093/emboj/18.14.4035
Puck, T.T., Cieciura, S.J., Robinson, A., 1958. GENETICS OF SOMATIC MAMMALIAN CELLS : III. LONG-TERM CULTIVATION OF EUPLOID CELLS FROM HUMAN AND ANIMAL SUBJECTS. Journal of Experimental Medicine 108, 945–956. https://doi.org/10.1084/jem.108.6.945
Pybus, L.P., Dean, G., West, N.R., Smith, A., Daramola, O., Field, R., Wilkinson, S.J., James, D.C., 2014. Model-directed engineering of "difficult-to-express" monoclonal antibody production by Chinese

hamster ovary cells. Biotechnology and Bioengineering 111, 372–385. https://doi.org/10.1002/bit.25116
Pybus, L.P., Hawke, E., Knowles, Christopher, 2020. Abstracts from the 26th European Society for Animal Cell Technology Meeting - Cell culture technologies: bridging academia and industry to provide solutions for patients. BMC Proc 14, 5. https://doi.org/10.1186/s12919-020-00188-y
Pybus, L.P., Kalsi, D., Matthews, J.T., Hawke, E., Barber, N., Richer, R., Young, A., Saunders, F.L., 2022. Coupling picodroplet microfluidics with plate imaging for the rapid creation of biomanufacturing suitable cell lines with high probability and improved multi-step assurance of monoclonality. Biotechnology Journal 17, 2100357. https://doi.org/10.1002/biot.202100357
Quiroz, J., Tsao, Y.-S., 2016. Statistical analysis of data from limiting dilution cloning to assess monoclonality in generating manufacturing cell lines. Biotechnology Progress 32, 1061–1068. https://doi.org/10.1002/btpr.2290
Raj, A., Oudenaarden, A. van, 2008. Nature, Nurture, or Chance: Stochastic Gene Expression and Its Consequences. Cell 135, 216–226. https://doi.org/10.1016/j.cell.2008.09.050
Ran, F.A., Hsu, P.D., Wright, J., Agarwala, V., Scott, D.A., Zhang, F., 2013. Genome engineering using the CRISPR-Cas9 system. Nat Protoc 8, 2281–2308. https://doi.org/10.1038/nprot.2013.143
Rang, H.P., Hill, R.G., 2013. Chapter 14 - Drug development: Introduction, in: Hill, R., Rang, H. (Eds.), Drug Discovery and Development (Second Edition). Churchill Livingstone, pp. 203–209. https://doi.org/10.1016/B978-0-7020-4299-7.00014-7
Reinhart, D., Damjanovic, L., Kaisermayer, C., Sommeregger, W., Gili, A., Gasselhuber, B., Castan, A., Mayrhofer, P., Grünwald-Gruber, C., Kunert, R., 2019. Bioprocessing of Recombinant CHO-K1, CHO-DG44, and CHO-S: CHO Expression Hosts Favor Either mAb Production or Biomass Synthesis. Biotechnology Journal 14, 1700686. https://doi.org/10.1002/biot.201700686
Reinhart, D., Sommeregger, W., Debreczeny, M., Gludovacz, E., Kunert, R., 2014. In search of expression bottlenecks in recombinant CHO cell lines—a case study. Appl Microbiol Biotechnol 98, 5959–5965. https://doi.org/10.1007/s00253-014-5584-z
Renard, J.M., Spagnoli, R., Mazier, C., Salles, M.F., Mandine, E., 1988. Evidence that monoclonal antibody production kinetics is related to the integral of the viable cells curve in batch systems. Biotechnol Lett 10, 91–96. https://doi.org/10.1007/BF01024632
Roca, B.C., Lao, N., Barron, N., Doolan, P., Clynes, M., 2019. An arginase-based system for selection of transfected CHO cells without the use of toxic chemicals. Journal of Biological Chemistry 294, 18756–18768. https://doi.org/10.1074/jbc.RA119.011162
Romanova, N., Noll, T., 2018. Engineered and Natural Promoters and Chromatin-Modifying Elements for Recombinant Protein Expression in CHO Cells. Biotechnology Journal 13, 1700232. https://doi.org/10.1002/biot.201700232
Rosser, S.J., Kleinjan, D.-J., 2022. Cho cell modification. WO2022123242A1.

Rutherford, K., Van Duyne, G.D., 2014. The ins and outs of serine integrase site-specific recombination. Curr Opin Struct Biol 24, 125–131. https://doi.org/10.1016/j.sbi.2014.01.003
Sample, P.J., Wang, B., Reid, D.W., Presnyak, V., McFadyen, I., Morris, D.R., Seelig, G., 2019. Human 5′ UTR design and variant effect prediction from a massively parallel translation assay. Nat Biotechnol 37, 803–809. https://doi.org/10.1038/s41587-019-0164-5
Samsung Biologics, 2023. Cell Line Development | CDO | Samsung Biologics | The Leading Global CDMO with End-to-End Services [WWW Document]. Cell Line Development | CDO. URL https://samsungbiologics.com/services/cdo/cell-line-development (accessed 10.16.23).
Saunders, F., Sweeney, B., Antoniou, M.N., Stephens, P., Cain, K., 2015. Chromatin Function Modifying Elements in an Industrial Antibody Production Platform - Comparison of UCOE, MAR, STAR and cHS4 Elements. PLoS One 10, e0120096. https://doi.org/10.1371/journal.pone.0120096
Saxena, P., Bojar, D., Fussenegger, M., 2017. Design of Synthetic Promoters for Gene Circuits in Mammalian Cells, in: Gould, D. (Ed.), Mammalian Synthetic Promoters, Methods in Molecular Biology. Springer, New York, NY, pp. 263–273. https://doi.org/10.1007/978-1-4939-7223-4_19
Scherzinger, J., Türk, D., Aprile-Garcia, F., 2022. An optimized and validated workflow for developing stable producer cell lines with >99.99% assurance of clonality and high clone recovery. https://doi.org/10.1101/2022.12.16.520697
Schuster, S.L., Hsieh, A.C., 2019. The untranslated regions of mRNAs in cancer. Trends Cancer 5, 245–262. https://doi.org/10.1016/j.trecan.2019.02.011
Sergeeva, D., Lee, G.M., Nielsen, L.K., Grav, L.M., 2020a. Multicopy Targeted Integration for Accelerated Development of High-Producing Chinese Hamster Ovary Cells. ACS Synth. Biol. 9, 2546–2561. https://doi.org/10.1021/acssynbio.0c00322
Sergeeva, D., Lee, G.M., Nielsen, L.K., Grav, L.M., 2020b. Multicopy Targeted Integration for Accelerated Development of High-Producing Chinese Hamster Ovary Cells. ACS Synth. Biol. 9, 2546–2561. https://doi.org/10.1021/acssynbio.0c00322
Shin, S.W., Lee, J.S., 2020a. CHO Cell Line Development and Engineering via Site-specific Integration: Challenges and Opportunities. Biotechnol Bioproc E 25, 633–645. https://doi.org/10.1007/s12257-020-0093-7
Shin, S.W., Lee, J.S., 2020b. Optimized CRISPR/Cas9 strategy for homology-directed multiple targeted integration of transgenes in CHO cells. Biotechnology and Bioengineering 117, 1895–1903. https://doi.org/10.1002/bit.27315
Silver, N., Best, S., Jiang, J., Thein, S.L., 2006. Selection of housekeeping genes for gene expression studies in human reticulocytes using real-time PCR. BMC Molecular Biology 7, 33. https://doi.org/10.1186/1471-2199-7-33
Sizer, R.E., White, R.J., 2023. Use of ubiquitous chromatin opening elements (UCOE) as tools to maintain transgene expression in biotechnology.

Computational and Structural Biotechnology Journal 21, 275–283. https://doi.org/10.1016/j.csbj.2022.11.059
Smith, G.P., 1985. Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. Science 228, 1315–1317. https://doi.org/10.1126/science.4001944
Sommeregger, W., Prewein, B., Reinhart, D., Mader, A., Kunert, R., 2013. Transgene copy number comparison in recombinant mammalian cell lines: critical reflection of quantitative real-time PCR evaluation. Cytotechnology 65, 811–818. https://doi.org/10.1007/s10616-013-9606-y
Sphere Fluidics, 2019. The Future of Cell Line Development.
Srila, W., Baumann, M., Borth, N., Yamabhai, M., 2022. Codon and signal peptide optimization for therapeutic antibody production from Chinese hamster ovary (CHO) cell. Biochemical and Biophysical Research Communications 622, 157–162. https://doi.org/10.1016/j.bbrc.2022.06.072
Stadermann, A., Gamer, M., Fieder, J., Lindner, B., Fehrmann, S., Schmidt, M., Schulz, P., Gorr, I.H., 2022. Structural analysis of random transgene integration in CHO manufacturing cell lines by targeted sequencing. Biotechnology and Bioengineering 119, 868–880. https://doi.org/10.1002/bit.28012
Steger, K., Brady, J., Wang, W., Duskin, M., Donato, K., Peshwa, M., 2015. CHO-S Antibody Titers >1 Gram/Liter Using Flow Electroporation-Mediated Transient Gene Expression followed by Rapid Migration to High-Yield Stable Cell Lines. J Biomol Screen 20, 545–551. https://doi.org/10.1177/1087057114563494
Sun, T., Kwok, W.C., Chua, K.J., Lo, T.-M., Potter, J., Yew, W.S., Chesnut, J.D., Hwang, I.Y., Chang, M.W., 2020. Development of a Proline-Based Selection System for Reliable Genetic Engineering in Chinese Hamster Ovary Cells. ACS Synth. Biol. 9, 1864–1872. https://doi.org/10.1021/acssynbio.0c00221
Suzuki, M., Kato, C., Kato, A., 2015. Therapeutic antibodies: their mechanisms of action and the pathological findings they induce in toxicity studies. J Toxicol Pathol 28, 133–139. https://doi.org/10.1293/tox.2015-0031
Szaniszlo, P., Rose, W.A., Wang, N., Reece, L.M., Tsulaia, T.V., Hanania, E.G., Elferink, C.J., Leary, J.F., 2006. Scanning cytometry with a LEAP: Laser-enabled analysis and processing of live cells in situ. Cytometry Part A 69A, 641–651. https://doi.org/10.1002/cyto.a.20291
Tanguay, R.L., Gallie, D.R., 1996. Translational efficiency is regulated by the length of the 3' untranslated region. Molecular and Cellular Biology 16, 146–156. https://doi.org/10.1128/MCB.16.1.146
Tejwani, V., Chaudhari, M., Rai, T., Sharfstein, S.T., 2021. High-throughput and automation advances for accelerating single-cell cloning, monoclonality and early phase clone screening steps in mammalian cell line development for biologics production. Biotechnology Progress 37, e3208. https://doi.org/10.1002/btpr.3208
Thermo Fisher, 2020. Freedom™ CHO-S™ Kit User Guide.

Tihanyi, B., Nyitray, L., 2020. Recent advances in CHO cell line development for recombinant protein production. Drug Discovery Today: Technologies 38, 25–34. https://doi.org/10.1016/j.ddtec.2021.02.003
Torres, M., Altamirano, C., Dickson, A.J., 2018. Process and metabolic engineering perspectives of lactate production in mammalian cell cultures. Current Opinion in Chemical Engineering, Biotechnology and bioprocess engineering 22, 184–190. https://doi.org/10.1016/j.coche.2018.10.004
Torres, M., Betts, Z., Scholey, R., Elvin, M., Place, S., Hayes, A., Dickson, A.J., 2023. Long term culture promotes changes to growth, gene expression, and metabolism in CHO cells that are independent of production stability. Biotechnology and Bioengineering n/a. https://doi.org/10.1002/bit.28399
Torres, M., Hussain, H., Dickson, A.J., 2022. The secretory pathway – the key for unlocking the potential of Chinese hamster ovary cell factories for manufacturing therapeutic proteins. Critical Reviews in Biotechnology 0, 1–18. https://doi.org/10.1080/07388551.2022.2047004
Tzani, I., Herrmann, N., Carillo, S., Spargo, C.A., Hagan, R., Barron, N., Bones, J., Dillmore, W.S., Clarke, C., 2021. Tracing production instability in a clonally derived CHO cell line using single-cell transcriptomics. Biotechnology and Bioengineering 118, 2016–2030. https://doi.org/10.1002/bit.27715
Ulianov, S.V., Gavrilov, A.A., Razin, S.V., 2012. Spatial organization of the chicken beta-globin gene domain in erythroid cells of embryonic and adult lineages. Epigenetics Chromatin 5, 16. https://doi.org/10.1186/1756-8935-5-16
Urlaub, G., Chasin, L.A., 1980. Isolation of Chinese hamster cell mutants deficient in dihydrofolate reductase activity. Proceedings of the National Academy of Sciences 77, 4216–4220. https://doi.org/10.1073/pnas.77.7.4216
Urlaub, G., Käs, E., Carothers, A.M., Chasin, L.A., 1983. Deletion of the diploid dihydrofolate reductase locus from cultured mammalian cells. Cell 33, 405–412. https://doi.org/10.1016/0092-8674(83)90422-1
Valaperta, R., Rizzo, V., Lombardi, F., Verdelli, C., Piccoli, M., Ghiroldi, A., Creo, P., Colombo, A., Valisi, M., Margiotta, E., Panella, R., Costa, E., 2014. Adenine phosphoribosyltransferase (APRT) deficiency: identification of a novel nonsense mutation. BMC Nephrol 15, 102. https://doi.org/10.1186/1471-2369-15-102
Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paepe, A., Speleman, F., 2002. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biology 3, research0034.1. https://doi.org/10.1186/gb-2002-3-7-research0034
Vcelar, S., Jadhav, V., Melcher, M., Auer, N., Hrdina, A., Sagmeister, R., Heffner, K., Puklowski, A., Betenbaugh, M., Wenger, T., Leisch, F., Baumann, M., Borth, N., 2018a. Karyotype variation of CHO host cell lines over time in culture characterized by chromosome counting and chromosome painting. Biotechnology and Bioengineering 115, 165–173. https://doi.org/10.1002/bit.26453

Vcelar, S., Melcher, M., Auer, N., Hrdina, A., Puklowski, A., Leisch, F., Jadhav, V., Wenger, T., Baumann, M., Borth, N., 2018b. Changes in Chromosome Counts and Patterns in CHO Cell Lines upon Generation of Recombinant Cell Lines and Subcloning. Biotechnology Journal 13, 1700495. https://doi.org/10.1002/biot.201700495

Veith, N., Ziehr, H., MacLeod, R.A.F., Reamon-Buettner, S.M., 2016. Mechanisms underlying epigenetic and transcriptional heterogeneity in Chinese hamster ovary (CHO) cell lines. BMC Biotechnol 16, 6. https://doi.org/10.1186/s12896-016-0238-0

Wahrheit, J., Niklas, J., Heinzle, E., 2014. Metabolic control at the cytosol–mitochondria interface in different growth phases of CHO cells. Metabolic Engineering 23, 9–21. https://doi.org/10.1016/j.ymben.2014.02.001

Walsh, G., Walsh, E., 2022. Biopharmaceutical benchmarks 2022. Nat Biotechnol 40, 1722–1760. https://doi.org/10.1038/s41587-022-01582-x

Wang, J., Jia, S.T., Jia, S., 2016. New Insights into the Regulation of Heterochromatin. Trends in Genetics 32, 284–294. https://doi.org/10.1016/j.tig.2016.02.005

Wang, X., Du, Q., Zhang, W., Xu, D., Zhang, X., Jia, Y., Wang, T., 2022. Enhanced Transgene Expression by Optimization of Poly A in Transfected CHO Cells. Frontiers in Bioengineering and Biotechnology 10.

Wang, X., Kawabe, Y., Hada, T., Ito, A., Kamihira, M., 2018. Cre-Mediated Transgene Integration in Chinese Hamster Ovary Cells Using Minicircle DNA Vectors. Biotechnology Journal 13, 1800063. https://doi.org/10.1002/biot.201800063

Wang, Y., O'Malley, B.W., Tsai, S.Y., O'Malley, B.W., 1994. A regulatory system for use in gene transfer. PNAS 91, 8180–8184. https://doi.org/10.1073/pnas.91.17.8180

Weber, W., Daoud-El Baba, M., Fussenegger, M., 2007. Synthetic ecosystems based on airborne inter- and intrakingdom communication. Proceedings of the National Academy of Sciences 104, 10435–10440. https://doi.org/10.1073/pnas.0701382104

Weber, W., Fussenegger, M., 2007. Inducible product gene expression technology tailored to bioprocess engineering. Curr Opin Biotechnol 18, 399–410. https://doi.org/10.1016/j.copbio.2007.09.002

Weber, W., Fux, C., Daoud-El Baba, M., Keller, B., Weber, C.C., Kramer, B.P., Heinzen, C., Aubel, D., Bailey, J.E., Fussenegger, M., 2002. Macrolide-based transgene control in mammalian cells and mice. Nat Biotechnol 20, 901–907. https://doi.org/10.1038/nbt731

Weber, W., Lienhart, C., Daoud-El Baba, M., Fussenegger, M., 2009. A biotin-triggered genetic switch in mammalian cells and mice. Metabolic Engineering 11, 117–124. https://doi.org/10.1016/j.ymben.2008.12.001

Weber, W., Rimann, M., Spielmann, M., Keller, B., Baba, M.D.-E., Aubel, D., Weber, C.C., Fussenegger, M., 2004. Gas-inducible transgene expression in mammalian cells and mice. Nat Biotechnol 22, 1440–1444. https://doi.org/10.1038/nbt1021

Weinguny, M., Klanert, G., Eisenhut, P., Jonsson, A., Ivansson, D., Lövgren, A., Borth, N., 2020. Directed evolution approach to enhance efficiency and speed of outgrowth during single cell subcloning of Chinese Hamster Ovary cells. Comput Struct Biotechnol J 18, 1320–1329. https://doi.org/10.1016/j.csbj.2020.05.020

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, 17. https://doi.org/10.1186/1472-6750-5-17

Wong, H.E., Chen, C., Le, H., Goudar, C.T., 2022. From chemostats to high-density perfusion: the progression of continuous mammalian cell cultivation. Journal of Chemical Technology & Biotechnology 97, 2297–2304. https://doi.org/10.1002/jctb.6841

Wu, M.-R., Nissim, L., Stupp, D., Pery, E., Binder-Nissim, A., Weisinger, K., Enghuus, C., Palacios, S.R., Humphrey, M., Zhang, Z., Maria Novoa, E., Kellis, M., Weiss, R., Rabkin, S.D., Tabach, Y., Lu, T.K., 2019. A high-throughput screening and computation platform for identifying synthetic promoters with enhanced cell-state specificity (SPECS). Nat Commun 10, 2880. https://doi.org/10.1038/s41467-019-10912-8

Wurm, F.M., 2004. Production of recombinant protein therapeutics in cultivated mammalian cells. Nat Biotechnol 22, 1393–1398. https://doi.org/10.1038/nbt1026

Wurm, F.M., Hacker, D., 2011. First CHO genome. Nat Biotechnol 29, 718–720. https://doi.org/10.1038/nbt.1943

Wurm, F.M., Wurm, M.J., 2017. Cloning of CHO Cells, Productivity and Genetic Stability—A Discussion. Processes 5, 20. https://doi.org/10.3390/pr5020020

Wurm, M.J., Wurm, F.M., 2021. Naming CHO cells for bio-manufacturing: Genome plasticity and variant phenotypes of cell populations in bioreactors question the relevance of old names. Biotechnology Journal 16, 2100165. https://doi.org/10.1002/biot.202100165

Xu, W.-J., Lin, Y., Mi, C.-L., Pang, J.-Y., Wang, T.-Y., 2023. Progress in fed-batch culture for recombinant protein production in CHO cells. Appl Microbiol Biotechnol 107, 1063–1075. https://doi.org/10.1007/s00253-022-12342-x

Yamane-Ohnuki, N., Kinoshita, S., Inoue-Urakubo, M., Kusunoki, M., Iida, S., Nakano, R., Wakitani, M., Niwa, R., Sakurada, M., Uchida, K., Shitara, K., Satoh, M., 2004. Establishment of FUT8 knockout Chinese hamster ovary cells: An ideal host cell line for producing completely defucosylated antibodies with enhanced antibody-dependent cellular cytotoxicity. Biotechnology and Bioengineering 87, 614–622. https://doi.org/10.1002/bit.20151

Yang, Y., Li, Z., Li, Q., Ma, K., Lin, Y., Feng, H., Wang, T., 2022. Increase recombinant antibody yields through optimizing vector design and production process in CHO cells. Appl Microbiol Biotechnol 106, 4963–4975. https://doi.org/10.1007/s00253-022-12051-5

Yang, Y., Mariati, Chusainow, J., Yap, M.G.S., 2010. DNA methylation contributes to loss in productivity of monoclonal antibody-producing

CHO cell lines. Journal of Biotechnology 147, 180–185. https://doi.org/10.1016/j.jbiotec.2010.04.004
Yang, Z., Wang, S., Halim, A., Schulz, M.A., Frodin, M., Rahman, S.H., Vester-Christensen, M.B., Behrens, C., Kristensen, C., Vakhrushev, S.Y., Bennett, E.P., Wandall, H.H., Clausen, H., 2015. Engineered CHO cells for production of diverse, homogeneous glycoproteins. Nat Biotechnol 33, 842–844. https://doi.org/10.1038/nbt.3280
Yin, J., Yang, L., Mou, L., Dong, K., Jiang, J., Xue, S., Xu, Y., Wang, X., Lu, Y., Ye, H., 2019. A green tea–triggered genetic control system for treating diabetes in mice and monkeys. Science Translational Medicine 11, eaav8826. https://doi.org/10.1126/scitranslmed.aav8826
Young, A., Lovelady, C., Shivare, M., Glassey, J., Coleman, S., Porter, A., 2016. Establishing a robust two-step cloning strategy for the generation of cell lines with a high probability of monoclonality. Cell Culture Engineering XV.
Young, J.D., 2013. Metabolic flux rewiring in mammalian cell cultures. Current Opinion in Biotechnology, Chemical biotechnology • Pharmaceutical biotechnology 24, 1108–1115. https://doi.org/10.1016/j.copbio.2013.04.016
Zboray, K., Sommeregger, W., Bogner, E., Gili, A., Sterovsky, T., Fauland, K., Grabner, B., Stiedl, P., Moll, H.P., Bauer, A., Kunert, R., Casanova, E., 2015. Heterologous protein production using euchromatin-containing expression vectors in mammalian cells. Nucleic Acids Research 43, e102. https://doi.org/10.1093/nar/gkv475
Zhang, L., Inniss, M.C., Han, S., Moffat, M., Jones, H., Zhang, B., Cox, W.L., Rance, J.R., Young, R.J., 2015. Recombinase-mediated cassette exchange (RMCE) for monoclonal antibody expression in the commercially relevant CHOK1SV cell line. Biotechnology Progress 31, 1645–1656. https://doi.org/10.1002/btpr.2175
Zhang, P., Haryadi, R., Chan, K.F., Teo, G., Goh, J., Pereira, N.A., Feng, H., Song, Z., 2012. Identification of functional elements of the GDP-fucose transporter SLC35C1 using a novel Chinese hamster ovary mutant. Glycobiology 22, 897–911. https://doi.org/10.1093/glycob/cws064
Zhang, S., Hughes, J.D., Murgolo, N., Levitan, D., Chen, J., Liu, Z., Shi, S., 2016. Mutation Detection in an Antibody-Producing Chinese Hamster Ovary Cell Line by Targeted RNA Sequencing. Biomed Res Int 2016, 8356435. https://doi.org/10.1155/2016/8356435
Zhao, C.-P., Guo, X., Chen, S.-J., Li, C.-Z., Yang, Y., Zhang, J.-H., Chen, S.-N., Jia, Y.-L., Wang, T.-Y., 2017. Matrix attachment region combinations increase transgene expression in transfected Chinese hamster ovary cells. Sci Rep 7, 42805. https://doi.org/10.1038/srep42805
Zhou, H., Liu, Z., Sun, Z., Huang, Y., Yu, W., 2010. Generation of stable cell lines by site-specific integration of transgenes into engineered Chinese hamster ovary strains using an FLP-FRT system. Journal of Biotechnology 147, 122–129. https://doi.org/10.1016/j.jbiotec.2010.03.020