A Journey Through the Dark: Exploration of Long Non-Coding RNAs and tRNAs in Chinese Hamster Ovary Cells for Recombinant Protein Production

2019

Davide Vito

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

University of Kent

Faculty of Sciences

Declaration

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

Davide Vito 7th August, 2019

Acknowledgements

My PhD has been a fantastic experience, which radically changed me. I found a second family here, always ready to support me through this challenging journey, making it extraordinary. In particular, I'd like to thank:

My parents for the unconditional support and freedom they've always given me, despite the unconventional routes I chose. It's not easy to be my parents, but they've always given me the energy and never let me feel alone.

Mark for being the most encouraging, understanding and dedicated supervisor anyone could desire, coupled with the extraordinary person he is. It's shocking to see how poorly this translates into his football choices.

Linas for the honest feedbacks and for constantly challenging my points of view. Sharing ideas made me grow, and it was generously rewarded with high quality liquorice.

Theo for being (literally) on my side during this entire time. I can't imagine this journey without airport Wagamama and questionable sambuca.

Budge for the energy he always brings to the people around him. Although the passion he puts in everything is inspiring, his dedication to lose bets and avoid tasty food is unreachable.

Tanya for her solidity in every situation. On top of making all our work possible, she adds priceless reactions while eating novel foods and drinking black coffee.

Dean for the delicate taps on my elbows in my moments of weakness and the constant supply of Complimints.

Jo for introducing me to the wonderful world of protein bars, turkey dinosaurs, women's barbells and problematic "would you rather".

Tulshi for being the hurricane of energy she always is, for the RNA brainstormings, the afternoon coffee breaks, the international aperitifs.

The whole Smales group for generating an exceptionally enjoyable and collaborative environment.

eCHO people for the invaluable interactions and unrepeatable moments. You made this network unique and I felt privileged to be part of it.

Søren and Symphogen for the support during my time in Copenhagen. It allowed me to really grow both as a professional and as a pétanque player.

Minions for the yellow wave of happiness they brought daily to our office, and thanks to everyone who contributed to it.

Myprotein for baked protein cookies.

You all contributed to who I am today and you'll always be part of me. I wish everyone was as fortunate as I am.

Abstract

The importance of biotherapeutic proteins for the treatment of various diseases has grown exponentially over the last few decades, and this growth is predicted to continue in the coming years. Among these, monoclonal antibodies (mAbs) represent the largest class for both revenues and new approvals. Chinese hamster ovary cells (CHO) are the leading platform in industry for the production of these complex molecules requiring human-like post-translational modifications, in particular mAbs. Although CHO cells are capable of producing and secreting mAbs at acceptable yields, numerous attempts to increase the maximum viable cell concentration and productivity have been described in literature exploiting modification of culture conditions, changing the genetic makeup of the vector(s) used to drive expression of the gene(s) of interest, and by engineering host cells. The application of non-coding RNAs, such as miRNAs and siRNAs, to reprogram CHO cells has been explored and allows specific targeting of detrimental genes without loading an additional translational burden on the cell. However, Long non-coding RNAs (LncRNAs), non-coding transcripts >200 nucleotides in length have only recently emerged as key regulators of epigenetics, splicing, microRNAs and translation. Despite the potential for applications in cell engineering, these molecules remain largely unexplored in mammalian expression systems. Further, whilst mRNA translation of coding transcripts is a central regulatory step for cell growth, and thus the yield and quality of recombinant proteins, non-coding tRNAs are an important regulatory molecule in the decoding process. Although recombinant gene sequences are often codon optimized, we do not currently have all the information required around tRNA abundance, modifications and tRNA charging to fully harness codon usage in recombinant sequences. The work reported here presents the first lncRNA and tRNA expression landscape in CHO cells under a variety of conditions and discusses the implications of these on recombinant protein production. To investigate lncRNA and tRNAs in CHO cells, a CHO-S cell line grown under batch and fed-batch culture was sampled at day 4 and 7 of culture while six IgG1-producing CHO cell lines cultivated in an ambr®15 system with different fed strategies were sampled before inoculation and at day 4, 7 and 12 of culture. The whole transcriptomes were investigated using a mouse microarray providing the surveillance of 24,881 mRNAs and 35,923 lncRNAs for CHO-S samples and RNA-Seq for the IgG-producing cell lines. tRNA abundances were quantified using a previously optimized ARM-Seq protocol. Thousands of differentially expressed lncRNAs were filtered by counting the occurrences of each transcript, assessing sequence conservation, secondary structure and RT-qPCR validation. The behaviour of a group of lncRNAs is described for the first time in CHO cells and the applications for cell engineering discussed. In particular, the CHO cell long non-coding RNA (lncRNA) transcriptome from cells grown in controlled miniature bioreactors is defined under fed-batch conditions using RNA-Seq to identify lncRNAs and how the expression of these changes throughout growth and between IgG producers. lncRNAs associated with productivity and growth characteristics are identified, in particular finding that *Adapt15*, linked to ER stress, *GAS5*, linked to mTOR signalling/growth arrest, and *PVT1,* linked to Myc expression, are differentially regulated during fed-batch culture and whose expression relates to productivity (*Adapt15*) or growth (*GAS5*, *PVT1*). Changes in (non)-coding RNA expression between the seed train and the equivalent day of fed-batch culture are also reported, showing large differences in gene expression between these. The ARM-Seq protocol allowed the identification of 4-5 fold more tRNAs compared to standard sequencing, and was applied to yeast and HEK293 cells to allow comparisons with CHO. Ultimately, tRNA quantifications were used in a translation elongation model to calculate the decoding speed of model recombinant proteins and to generate codon optimized versions based on tRNA abundances.

Table of Contents

Chapter 1 Introduction 12

1.1 The emergence and growth of recombinant biotherapeutic proteins for the treatment of disease 12

1.1.1 Comparison of different recombinant protein expression systems 12

1.1.1.1 Bacteria 14

1.1.1.2 Yeast expression systems 15

1.1.1.3 Baculovirus and insect cells 16

1.2 Chinese hamster ovary cells are the leading system for the expression of complex recombinant proteins 16

1.3 Engineering of the Cellular Translational Machinery and Non-Coding RNAs to Enhance CHO Cell Growth, Recombinant Product Yields and Quality 19

1.3.1 Introduction 19

1.3.2 The translation machinery, mRNA analysis and manipulation 23

1.3.3 microRNAs and siRNAs, short non-coding RNAs that repress mRNA translation 28

1.3.4 Long non-coding RNAs (lncRNAs) and their manipulation 31

1.3.5 tRNAs and mRNA translation 37

1.3.6 Challenges in modulation of translation for enhanced CHO cell biotherapeutic protein production 41

1.4 Long non-coding RNAs are emerging as key regulators of cell biology 42

1.5 Translation is influenced by the codon sequence of mRNAs 43

1.5.1 Translation can be computationally modelled to predict decoding speed 44

1.6 Aims of the work described in this thesis 47

Chapter 2 Materials and Methods 48

2.1 Cell culture and RNA extraction 48

2.1.1 CHO-S cell line 48

2.1.2 Symphogen cell lines 48

2.2 Whole transcriptome RNA-Seq on Symphogen samples 50

2.2.1 RNA-Seq and Data analysis 50

2.2.2 RT-qPCR Validation of Differentially Expressed genes 50

2.3 AlkB de-methylase production in *E. coli* 51

2.3.1 Protein expression and purification 51

2.3.2 SDS-PAGE protein characterisation 51

2.3.2.1 Sample preparation procedure 51

2.3.2.2 Gel preparation and electrophoresis 51

2.3.2.3 Bradford assay 52

2.3.3 MS-MS protein characterisation 52

2.4 tRNA sequencing 52

2.4.1 Small RNA de-methylation 52

2.4.2 Library generation and sequencing 53

2.4.3 Alignment and differential expression 53

2.5 Codon Adaptation Index (CAI) and tRNA gene copy numbers 54

2.6 HEK293 cell line 54

2.6.1 Cell culture and RNA extraction 54

2.6.2 Library generation and sequencing 54

2.7 Computational modelling of elongation 55

Chapter 3 The Long Non-Coding RNA Transcriptome Landscape in CHO Cells Under Batch and Fed-Batch Conditions 56

3.1 Introduction 57

3.2 Materials and Methods 59
3.2.1 Model Cell Line and Cell Culture Conditions 59
3.2.2 Sampling from Cell Cultures and Subsequent RNA Extraction 60
3.2.3 lncRNA and Coding RNA Microarray and Data Analysis 60
3.2.3.1 Microarray Details 60
3.2.3.2 Microarray Data Analysis 61
3.2.4 RT-qPCR Validation of lncRNAs Identified as Differentially Expressed by Microarray 61

3.3 RESULTS 62
3.3.1 Growth Characteristics and Sampling of CHO-S Cells Throughout Batch and Fed-Batch Culture for (Non)-Coding RNA Analysis 62
3.3.2 Microarray Analysis of mRNA and lncRNA Transcripts in Batch and Fed-Batch Culture 65
3.3.3 Identification of lncRNAs Differentially Expressed as Potential Engineering Targets for Modulation of Cell Growth 66
3.3.4 Comparison of mRNAs with Existing Datasets 68
3.3.5 GO analysis and Pathway Enrichment 69

3.4 DISCUSSION 72

3.5 CONCLUSIONS 74

Chapter 4 Defining lncRNAs that Underpin CHO Cell Growth and IgG Productivity by RNA-Seq 75

4.1 Introduction 76

4.2 Results and Discussion 78
4.2.1 Analysis of fed-batch culture samples 78
4.2.2 RNA sequencing of ambr15™ generated samples and subsequent analysis of the data: The DAVI experiment 81
4.2.3 RNA sequencing of ambr15™ generated samples and subsequent analysis of the data: The JCE Experiment 85
4.2.4 Investigating Pathway Enrichment in DE Genes 87
4.2.5 Mapping of Long non-coding RNA expression during fed-batch culture 90
4.2.6 Identification of differentially expressed lncRNAs as potential cell engineering targets 90
4.2.7 Comparison of data presented here with existing datasets 92

4.3 Summary 93

Chapter 5 tRNA Expression in CHO and HEK293 Cells and the Impact on Transcript Specific and General mRNA Translation 96

5.1 Introduction 96

5.2 Results 98
5.2.1 Culture of CHO-S cells for tRNA analysis; Fed-batch culture shows a sustained higher proliferation and culture viability compared to batch culture 98
5.2.2 Culture of Symphogen cells for tRNA analysis; Comparison of growth and productivity characteristics of different IgG1 producing cell lines cultivated in an automated ambr15^{TM} system with Cell Boost 6 and Cell Boost 7 feeding 100
5.2.3 AlkB de-methylation treatment on extracted small RNAs increases the detection of full length tRNAs up to 5-fold 101
5.2.4 The CHO-S and the Symphogen datasets both cluster based on culture time point sampling as opposed to feeding regime 105
5.2.5 The de-methylation treatment does not impair the detection of other small RNAs 106

5.2.6 The use of different Chinese hamster genomes does not significantly alter the tRNA gene copy number predictions 107
5.2.7 IgG1 light chain, EPO and Etanercept codon sequences show a strong agreement with the CHO codon bias even when a subset of sequences is considered 109
5.2.8 The CHO-S RNA-Seq tRNA dataset shows a strong correlation with tRNA gene copy numbers and codon usage 110
5.2.9 Differential expression of tRNAs among the CHO-S dataset shows variability at Day 7 regardless of the type of culture 114
5.2.10 Differential expression analysis among the Symphogen dataset shows the tRNA pool between cell lines at the same time point is consistent whilst there is a higher variability in the tRNA pool across culture 116
5.2.11 The use of tRNA abundancies to predict elongation speed shows longer decoding times than using tRNA gene copy numbers 119
5.2.12 HEK293 cold shock dataset 124

5.3 Discussion 126

Chapter 6 Discussion and Future Work 129

6.1 Mapping and Investigation of Long non-coding RNA Expression in CHO Cells .130
6.1.1 CHO-S dataset 130
6.1.2 Symphogen datasets 131
6.1.3 Future work on lncRNAs 132

6.2 tRNA abundance and mRNA translation 132
6.2.1 Quantification of tRNA species through an optimized RNA-Seq protocol 133
6.2.2 mRNA translation decoding speed simulation using an elongation model 134
6.2.3 Future work on tRNAs and translation 135

6.3 Overall Conclusions 136

Bibliography 137

Appendix 162
Supplementary figures 161
RT-qPCR primers 162
Etanercept sequence 171
EPO sequence 171
Published works 172

List of Figures

Figure 1.1 - Eukaryotic mRNA translation initiation and elongation..... 24
Figure 1.2 - mTOR regulated signalling pathway to control translation initiation 27
Figure 1.3 - Schematic depicting the main mechanisms of action for microRNA-mediated mRNA repression in animals. 29
Figure 1.4 - Schematic depicting LncRNA transcription, localisation and functions in the cell..... 32
Figure 1.5 - Graphical representation of the main known lncRNAs functions in the cell 33
Figure 1.6 - Summary of ENSEMBL number of genes for hamster, human and mouse..... 34
Figure 1.7 - Correlations between tRNA gene copy numbers and codon usage..... 37
Figure 1.8 - Summary of chemical modifications to nucleotides 40
Figure 3.1 - Experimental workflow for transcriptomics analysis of CHO-S cell line..... 63
Figure 3.2 - Results of CHO-S batch and fed-batch cultures 64
Figure 3.3 - Graphical summary of the enriched KEGG pathways based on differentially expressed genes 70
Figure 4.1 - Cell culture parameters for the DAVI and JCE datasets 80
Figure 4.2 - Clustering and PCA analysis based on gene expression 83
Figure 4.3 - Differentially expressed genes in the DAVI and JCE datasets 85
Figure 4.4 - Enriched KEGG pathways based on differentially expressed genes in the DAVI dataset . 89
Figure 5.1 - CHO-S and Symphogen cell culture results 99
Figure 5.2 - Small RNA-Seq reads alignments and yeast correlation 103
Figure 5.3 - Clustering and PCA based on tRNA sequencing 104
Figure 5.4 - mt-tRNA and mt-rRNA sequencing 107
Figure 5.5 - Correlations between sequenced tRNAs, gene copy number and codon usage in the Symphogen dataset 112
Figure 5.6 - Correlations between sequenced CHO-S tRNAs, gene copy numbers and codon usage 113
Figure 5.7 - Decoding speed simulations for 3068 IgG1..... 121
Figure 5.8 - Decoding speed simulations for 3080 IgG1..... 122
Figure 5.9 - Decoding speed simulations for Etanercept 123
Figure 5.10 - Correlations between sequenced tRNAs, gene copy number and codon usage in HEK293 126

List of Tables

Table 1.1 - Comparison of different recombinant protein expression systems 13
Table 1.2 - Main classes of non-coding RNAs 21
Table 1.3 - Cell engineering approaches in CHO cells for recombinant protein production 22
Table 3.1 - Selected differentially expressed lncRNAs 68
Table 3.2 - Enriched KEGG pathways based on differentially expressed genes 71
Table 4.1 - Symphogen cell lines used in DAVI and JCE experiments 79
Table 4.2 - Identified potential lncRNA genes for cell engineering 84
Table 5.1 - Summary of the cell lines used in the Symphogen dataset 101
Table 5.2 - Mapping efficiencies for Symphogen dataset 103
Table 5.3 - tRNA gene copy numbers in hamster 108
Table 5.4 - Codon adaptation index calculations 110
Table 5.5 - Correlation coefficients for the sequenced tRNAs 114
Table 5.6 - Differentially expressed tRNAs in CHO-S dataset 115
Table 5.7 - Differentially expressed tRNAs in Symphogen dataset 118
Table 5.8 - Decoding speeds for Symphogen antibodies and Etanercept sequences 124

List of Abbreviations

3'UTRs	3'untranslated regions
4E-BP1	Eukaryotic initiation factor 4E binding protein
arcRNAs	Architectural RNAs
ARM-Seq	AlkB-facilitated RNA methylation sequencing
BAK	BCL2-antagonist/killer
BAX	BCL2-associated X protein
BEVS	Baculovirus expression vector systems
CAGE	Cap analysis of gene expression
CAI	Codon Adaptation Index
ceRNAs	Competing endogenous RNAs
CerS2	Ceramide synthase 2
CH	Chinese hamster
CHO	Chinese hamster ovary cells
CPC2	Coding Potential Calculator 2
DE	Differentially expressed
DHFR	Dihydrofolate reductase
eEF2	Translation elongation factor 2
eEF2	Translation elongator factor 2
Eme1	Essential Meiotic Structure-Specific Endonuclease 1
ER	Endoplasmic reticulum
eRNAs	Enhancer RNA
Exo1	Exonuclease 1
Fancb	FA Complementation Group B
FC	Fold change
FDR	False discovery rate
GO	Gene Ontology
GS	Glutamine synthetase
HC	Heavy chain
HEK	Human embryonic kidney cells
HNRNPs	Heterogeneous nuclear ribonucleoproteins
IVC	Integral of viable cell concentration
KIFC1	Kinesin family member C1
LC	Light chain
LDH-A	Lactate dehydrogenase-A
lncRNAs	Long non-coding RNAs
LoNA	Long nucleolus-specific lncRNA
m^1A	1-methyladenosine
mAbs	Monoclonal antibodies
m^3C	3-methylcytidine

m^1G	1-methylguanosine
miRs/miRNAs	microRNAs
mRNA	Messenger RNA
MSX	Methionine sulfoximine
MTX	Methotrexate
NMD	Nonsense-mediated decay
ORF	Open reading frame
PCA	Principal component analysis
PDHKs	Pyruvate dehydrogenase kinases
PTM	Post-translational modification
PVT1	Plasmacytoma Variant Translocation 1
Qp	Cell specific productivity
rRNA	Ribosomal RNA
SDS	Sodium dodecyl sulphate
siRNAs	Small interfering RNAs
SRP	Signal recognition particle
ST	Seed train
ST6GAL	Alpha-2,6-sialyltransferase
tRFs	tRNA-related fragments
tRNA	transfer RNA
VCD	Viable cell density
VLP	Virus like particle
XIAP	X-linked inhibitor of apoptosis

Chapter 1 Introduction

1.1 The emergence and growth of recombinant biotherapeutic proteins for the treatment of disease

Since the approval by the FDA in 1982 of the first biopharmaceutical Humulin, a recombinant human insulin produced in *E. coli*, the importance of biotherapeutic proteins for the treatment of disease has grown exponentially (Walsh, 2014). Indeed, 47% of the new drug approvals in the US between 2014 and 2018 were biopharmaceuticals, as opposed to just 21% in 2006-2010 and 26% in 2010-2014 (Walsh, 2018). Based on current late-stage clinical trials, where protein-based therapeutics represent the majority of products, this trend is predicted to continue in the coming years (Walsh, 2018). Among these, monoclonal antibodies (mAbs) currently dominate the scene, representing 53% of new approvals between 2015 and 2018 and reaching $123 billion in global sales in 2017, and the sales are still predicted to grow in the next few years due to the flux of recent new approvals and current pipelines (Walsh, 2018). The mAb Humira (adalimumab) alone had global sales just short of $19 billion in 2017. As a comparison, insulins are collectively the second most lucrative product class at $22 billion in 2017 (Walsh, 2018).

1.1.1 Comparison of different recombinant protein expression systems

After human tissue plasminogen activator became the first therapeutic protein to be produced in cultured mammalian cells in 1986 by Genentech, this expression system has now become the most common for recombinant protein production of biopharmaceuticals (F. M. Wurm, 2004). Nonetheless, a variety of alternative organisms are used in industry and academic research for the expression of recombinant proteins. Each one of these platforms present intrinsic advantages and disadvantages which are summarised in Table 1.1 and discussed in the following paragraphs of this chapter.

Table 1.1 - Comparison of different recombinant protein expression systems

Host	Advantages	Disadvantages	Examples on the market	References
E. coli	Fast doubling time High cell density Inexpensive media Extensive biology knowledge Disulphide bonds formation in the periplasm	Limited PTMs Limited glycosylation Presence of endotoxins	Admelog, rapid-acting human insulin analogue Increlex, rh IGF-1 Pegasys, PEGylated IFN-α-2b	(Daegelen, Studier, Lenski, Cure, & Kim, 2009; Marisch et al., 2013)
S. cerevisiae	Fast doubling time High cell density Inexpensive media Extensive biology knowledge Cellular machinery to undertake complex PTMs	Hyper-mannosylation N-linked and O-linked fungal glycans	Ryzodeg, combination of two engineered insulins Victoza, GLP-1 analogue with attached fatty acid Fasturtec, urate oxidase	(Baghban et al., 2019; Fernández, López-Estepa, Querol-García, & Vega, 2016)
P. pastoris	Fast doubling time High cell density Inexpensive media Cellular machinery to undertake complex PTMs Reduced hyper-glycosylation	Presence of N-linked and O-linked fungal glycans	Kalbitor, plasma kallikrein inhibitor Jetrea, truncated form of human plasmin	(Baghban et al., 2019; Macauley-Patrick, Fazenda, McNeil, & Harvey, 2005)
Insect cells	Growth in suspension at 28 °C in serum-free media in the absence of CO_2 Cellular machinery to undertake complex PTMs More complex glycan structures	Costly downstream purification steps Compatibility of the glycosylation patterns with human Low yields	Cervarix, VLP vaccine for the prevention of human papilloma virus infection Provenge, prostatic acid phosphatase Flublok, subunit for the influenza vaccine	(Contreras-Gómez, Sánchez-Mirón, García-Camacho, Molina-Grima, & Chisti, 2014; Jarvis, 2009; van Oers, Pijlman, & Vlak, 2015)
CHO cells	Growth in suspension in chemically defined serum-free media Human-like PTMs Effective folding and secretion of large and complex molecules Yields in excess of 5 g/L	Expensive media and reagents Slower doubling times Examples of immunogenic N-glycosylation patterns	Luveris, rh luteinizing hormone Biopoin, rh EPO Avastin, humanized mAb against VEGF Humira, anti-TNF human mAb Vimizim, rh N-acetylgalactosamine-6-sulfatase	(Fischer, Handrick, & Otte, 2015; J. Y. Kim, Kim, & Lee, 2012)

Summary of the advantages and disadvantages of different cell factories utilised for biopharmaceuticals production with examples of products on the market produced in the respective host and references to relevant literature on the expression system.

1.1.1.1 Bacteria

Bacterial systems are one of the most common expression systems used for recombinant proteins, both at laboratory and manufacturing scale, especially *Escherichia coli* BL21(DE3) strains and derivatives of the K-12 lineage (Daegelen et al., 2009; Marisch et al., 2013; Rosano & Ceccarelli, 2014). After the first human polypeptide was expressed in *E. coli* more than 40 years ago (Itakura et al., 1977), this host gained popularity for recombinant protein production due to various advantages, including extremely fast growth kinetics with high cell densities (Sezonov, Joseleau-Petit, & D'Ari, 2007), significantly lower culture media and reagents costs compared to mammalian systems (Fernández & Vega, 2016) and relatively straightforward genetic manipulation (Pope & Kent, 1996). Different forms of insulin, human hormones, human IgG fragments, fusion proteins and recombinant enzymes are all examples of biopharmaceuticals on the market produced in *E. coli* (Walsh, 2018). Nonetheless, microbial cell factories have intrinsic limitations which can prevent the expression of more complex proteins with acceptable quality attributes.

Human-compatible glycosylation patterns, correct folding and the formation of disulphide bonds can be difficult to obtain in *E. coli*, leading to protein aggregation with the formation of inclusion bodies that require supplementary downstream processing and loss of functional product (Carrió & Villaverde, 2002; Derman, Prinz, Belin, & Beckwith, 1993; Hartley & Kane, 1988). While the reducing environment of the cytoplasm in *E. coli* prevents the formation of disulphide bonds, a common strategy is to direct proteins to the periplasm where these can be formed, although specific steps are then required for downstream purification (Mergulhão, Summers, & Monteiro, 2005; Messens & Collet, 2006). Bacterial endotoxins like lipopolysaccharides (LPSs) are required components in the outer membrane of most Gram-negative bacteria, including *E. coli*, and their presence can affect safety of a biopharmaceutical for use in humans (Raetz & Whitfield, 2002). The recombinant protein itself can be toxic to the host causing detrimental functions in the cell and ultimately slow growth rate, low cell density and insufficient protein yield (Doherty, Connolly, & Worrall, 1993; Dong, Nilsson, & Kurland, 1995). Targeting the protein to the periplasm can solve the problem of protein toxicity as well as promoting disulphide bond formation (de Marco, 2009; Mergulhão et al., 2005).

1.1.1.2 Yeast expression systems

Unicellular yeast organisms retain many of the advantages offered by bacterial hosts, notably fast doubling times, high cell densities and inexpensive media, coupled with an ability to undertake a wider range of post-translational modifications and folding capacity and extracellular secretion into the medium (Fernández et al., 2016). The two main hosts among yeast cell expression systems are the non-methylotrophic *Saccharomyces cerevisiae* and the methylotrophic *Pichia pastoris* (Fernández et al., 2016). While this phenotypical distinction is based on the capacity to utilize methanol as a carbon and energy source, requiring distinct cell culture protocols, these two hosts present specific advantages and disadvantages. The deep knowledge on the microbiology (Barnett & Barnett, 2011), genetics (Duina, Miller, & Keeney, 2014), molecular and cellular biology (Resnick & Cox, 2000) and stress response during cell culture (Bawa et al., 2011) available for *S. cerevisiae* allows extensive genetic engineering and targeted process optimization to increase yield and quality of the product (Bonander & Bill, 2012), making this yeast one of the most common hosts for recombinant protein production. *S. cerevisiae* is currently used for the production of several commercially available biotherapeutics, including rh factor XIII A-subunit, various forms of insulin, glucagon, Hepatitis B antigens, and HPV capsid proteins (Walsh, 2018).

The main limitation for complex heterologous protein expression is frequently the glycan structure of the final product, as *S. cerevisiae* tends to form hyper-mannosylated highly immunogenic structures (Baghban et al., 2019). Genetic engineering approaches have been shown to successfully alleviate this problem through the removal of key enzymes involved in glycosylation pathways (Tang et al., 2016) but the use of *P. pastoris* is often preferred when hyper-glycosylation is a concern (Macauley-Patrick et al., 2005). While maintaining most of the advantages outlined for *S. cerevisiae*, *P. pastoris* adds shorter glycan chains due to the lack of the α-1,3-mannosyltransferase enzyme (Wildt & Gerngross, 2005). Examples of commercially available biopharmaceuticals produced in *P. pastoris* are Kalbitor® (ecallantide), a plasma kallikrein inhibitor introduced onto the market in 2009, and Jetrea®, a truncated form of human plasmin approved for use in 2012 (Walsh, 2018). However, although engineering strategies have been tried to closely mimic mammalian-like glycan structures (Bobrowicz et al., 2004), the presence of N-linked and O-linked fungal glycans can still reduce efficacy or even elicit an immunogenic response in humans (Nett et al., 2013).

1.1.1.3 Baculovirus and insect cells

Back in 1983, Max Summers and co-workers reported the production of a functional human IFN-β in insect cells infected by a recombinant baculovirus for the first time using the polh promoter (G. E. Smith, Summers, & Fraser, 1983). One year later, Lois Miller and colleagues expressed high levels of *Escherichia coli* beta-galactosidase under the p10 promoter (Pennock, Shoemaker, & Miller, 1984). More than three decades since these fundamental contributions, the baculovirus expression vector systems (BEVS) has become a widespread technology in academic research as well as industrial manufacturing of a wide range of proteins (van Oers et al., 2015). Insect cell cultures provide some advantages over other systems, including the ability to grow in suspension in serum free medium at 28°C in the absence of CO_2 with scale-up potential using bioreactors, shake flasks, stirred tanks and single use wavebag systems (Contreras-Gómez et al., 2014; Jarvis, 2009; van Oers et al., 2015). Moreover, being a higher eukaryote, insect cells express protein with more complex glycan structures than yeast or bacteria and can be effectively engineered to partly resemble human-like glycosylation profiles (Mabashi-Asazuma et al., 2013). A particularly successful application for BEVS is the production of virus-like particles (VLPs) (Fernandes, Teixeira, Carinhas, Carrondo, & Alves, 2013; Yamaji, 2014). An example of BEVS produced VLP is the commercially available vaccine for the prevention of human papilloma virus infection, Cervarix (Deschuyteneer et al., 2010). Protein therapeutics expressed in insect cells now account for approximately 1% of all approved biotherapeutics for human use (Yee, Zak, Hill, & Wen, 2018), including the prostatic acid phosphatase Provenge (Burch et al., 2000; Cheever & Higano, 2011; Finer-Moore, Czudnochowski, O'Connell, Wang, & Stroud, 2015), and a subunit for the influenza vaccine Flublok (Cox & Hollister, 2009). Major limitations to a more widespread use of BEVS in manufacturing of biotherapeutics are the presence of baculovirus progeny contamination in the culture along with cellular proteins and debris from lysed cells, requiring costly downstream purification steps (Fernandes et al., 2013), altered less-complex glycosylation patterns than the human equivalents (Harrison & Jarvis, 2006; Hollister, Grabenhorst, Nimtz, Conradt, & Jarvis, 2002) with much lower yields compared to cultured mammalian cells.

1.2 Chinese hamster ovary cells are the leading system for the expression of complex recombinant proteins

The majority of approved recombinant biopharmaceuticals, including monoclonal antibodies (mAbs), are produced in mammalian cell expression systems, predominantly in Chinese

hamster ovary (CHO) cells under fed-batch culture conditions (Butler & Spearman, 2014; Dumont, Euwart, Mei, Estes, & Kshirsagar, 2016; Mauro, 2018). CHO cells are the current leading expression system for the production of biopharmaceuticals due to their demonstrated ability to produce complex and correctly folded proteins with 'human like' post-translational modification (PTMs) coupled with an effective secretion of the mature produced protein (J. Y. Kim et al., 2012). In addition, CHO cells can grow in suspension in chemically defined serum-free media, on a large scale and produce secretory yields of monoclonal antibody in fed-batch culture in excess of 5 g/L (Povey et al., 2014). The secretory yield from such an expression system results from the combination of the number of viable cells across culture (the integral of viable cell concentration or IVC) and the average amount of material expressed by each cell (cell specific productivity, Qp) (Kunert & Reinhart, 2016). These attributes can be positively influenced by selecting improved CHO cell hosts, modification of the culture conditions, changing the genetic makeup of the vector(s) used to drive expression of the gene(s) of interest, and by engineering host cells to enhance their growth and/or recombinant protein capacity. The work described in this thesis is focussed upon providing a better understanding of the non-coding transcriptome in CHO cells, the impact of this on cell growth and recombinant protein production and the potential for engineering the non-coding transcriptome to enhance CHO cell growth recombinant protein production and quality from CHO cells.

A short review of relevant engineering approaches taken to improve CHO cell phenotype under industrially relevant conditions has previously been described (Fischer et al., 2015; J. Y. Kim et al., 2012; J. S. Lee, Grav, Lewis, & Faustrup Kildegaard, 2015). Briefly, approaches such as process and media optimisation have been used extensively to accelerate growth and increase maximum viable cell density (VCD) (F. Li, Vijayasankaran, Shen, Kiss, & Amanullah, 2010). Genetic engineering can further improve cell lines characteristics by introducing external genes, overexpressing transcripts beneficial to the cell or preventing the expression of transcripts with a detrimental effect. Although most PTMs performed in CHO are human-compatible, stably introducing enzymes to decrease immunogenic reaction to certain N-glycosylation patterns has been successfully reported (Davies et al., 2001; Ferrara et al., 2006; Jeong et al., 2008). An example is the introduction of the alpha-2,6-sialyltransferase (ST6GAL) enzyme, allowing the expression of proteins with α2,6-sialylated glycans (E. U. Lee, Roth, & Paulson, 1989). In addition, cell lines with a complete absence of FUT8 activity and subsequent lack of fucosylation in mAbs have been reported (Yamane-Ohnuki et al., 2004).

Altering the metabolism of a CHO cell can reduce nutrient consumption and accumulation of toxic by-products, increasing at the same time growth rate and final yield (Chong et al., 2010; Fogolín, Wagner, Etcheverrigaray, & Kratje, 2004; Tabuchi & Sugiyama, 2013). Further, metabolic manipulations have generated dihydrofolate reductase (DHFR) and glutamine synthetase (GS) defective cell lines, two of the most commonly used CHO cell lines among biopharmaceutical manufacturers (Fan et al., 2012; Urlaub & Chasin, 1980; Urlaub, Käs, Carothers, & Chasin, 1983). Transfecting these cell lines with vectors containing both a transgene and a functional DHFR or GS gene sequence allows metabolic selection of stable transfectants through media containing the dihydrofolate analogue methotrexate (MTX) or methionine sulfoximine (MSX) (F. M. Wurm, 2004).

A substantial increase in specific productivity is often a target of cell engineering. Cells overexpressing the transcription factors ZFP-TF, ATF4 or GADD34 expressed up to 10-fold more recombinant proteins compared to parental cells (Kwon et al., 2006; Ohya et al., 2008; Omasa et al., 2008). The elimination of BCL2-associated X protein (BAX) and BCL2-antagonist/killer (BAK) expression using zinc-finger nucleases causes resistance to apoptosis, leading to a final mAb yield up to 5-fold higher (Cost et al., 2010). Integrating these cell engineering approaches using the precise gene editing capabilities offered by CRISPR/Cas9 can be extremely beneficial for cell line development (Fischer et al., 2015). The disruption of COSMC and FUT8 represented the first example of generating multiplexed knockout (Ronda et al., 2014), followed by simultaneous disruption of FUT8, BAX and BAK and knockout of 10 genes combined with the insertion of the human ST6GAL1 enzyme to generate fully humanized N-glycosylation profiles (Amann et al., 2019).

For the purpose of the work in this thesis, mRNA translation has previously been identified as a potential limitation on recombinant protein production from cultured CHO cells, particularly the production of monoclonal antibodies (mAbs). mRNA translation is a key control point in gene expression and hence in determining CHO cell growth and recombinant protein production capacity. Indeed, mRNA translation is a key biological process determining global and protein specific synthesis, and hence controlling many cellular activities, impacting on IVC, Qp and protein quality (Mead et al., 2015). We recently reviewed the key steps regulating translation, how this can be tuned by small non-coding mRNAs such as microRNAs and siRNAs, and the expanding knowledge around Long Non-Coding RNAs (lncRNAs) and their control of cell phenotype with relevant examples selected from

literature. This review was published in the Journal Current Opinion in Chemical Engineering as follows;

Vito D and Smales CM (2018), Engineering of the cellular translational machinery and non-coding RNAs to enhance CHO cell growth, recombinant product yields and quality, *Current Opinion in Chemical Engineering* 22:199–208, doi: 10.1016/j.coche.2018.11.002.

This review is reported here and expanded to reflect the full background behind the entire PhD project.

1.3 Engineering of the Cellular Translational Machinery and Non-Coding RNAs to Enhance CHO Cell Growth, Recombinant Product Yields and Quality

1.3.1 Introduction

As described above, for the production of biopharmaceuticals, Chinese hamster ovary (CHO) cells are the most widely used mammalian cell expression system, able to produce secretory yields of monoclonal antibody in fed-batch culture in excess of 5 g/L (Povey et al., 2014). The secretory yield from such an expression system is governed by the number of cells in the bioreactor across the culture (the integral of viable cell concentration or IVC) and the average amount of material expressed by each cell, usually referred to as the cell specific productivity (Qp) and expressed as pg of protein/per cell/per day (Kunert & Reinhart, 2016). mRNA translation is a key cellular process that is involve in determining global and protein specific synthesis, and hence control of the abundance of proteins that constitute the cellular machinery, cell growth, division and the IVC of culture. Likewise, mRNA translation plays a key role in determining the Qp of a given cell line and hence is a key regulatory process impacting on the yields and quality of recombinant protein from CHO cells (Mead et al., 2015).

mRNA translation is the process by which the ribosome and associated cellular machinery decodes a target mRNA to yield a polypeptide. Translation is a key step in the gene expression pathway and is the predominant process by which protein cellular abundance is controlled (Schwanhäusser et al., 2011). Over the last few decades it has been established that the control in mammalian cells of mRNA translation, and hence protein synthesis, is not only determined by the translational machinery, modulation of the activity of various translation factors by phosphorylation, and the abundance, availability and makeup of a given mRNA, but also by availability, abundance and activity of non-coding RNAs (H. Janakiraman et al., 2018). Non-coding RNAs are generally described as

either long non-coding RNAs of >200 nucleotides in length or small non-coding RNAs <200 nucleotides and include microRNAs (also referred to as miRs) and tRNAs. A summary of the main classes of non-coding RNAs is reported in Table 1.2.

The discovery of the mechanism(s) by which non-coding RNAs exert an influence on gene expression has opened up new opportunities for the engineering of cells to manipulate cell processes that underpin cell growth and recombinant protein production and quality. Further, manipulation of such non-coding RNAs offers the advantage of not placing an additional translational burden on the cell that over-expression of coding mRNAs does. Here we briefly review our understanding of the control of mRNA translation in CHO cells, describe approaches and outcomes to engineer the translational machinery and non-coding RNAs in CHO cells, and discuss current and future cell engineering opportunities and challenges such approaches present (Summarised in Table 1.3).

Table 1.2 - Main classes of non-coding RNAs

Type	Length	Transcription	Function(s)	Characteristics	References
Long non-coding RNAs	>200 nt	Pol II	Epigenetic modifications Assembly of protein complexes miRNA decoy Translation regulation	With or without polyA Cytoplasmic or nuclear localisation Low sequence conservation between species	(Mattick, 2018)
Transfer RNAs	76-90 nt	Pol III	Translation of mRNAs Generation of short RNAs	Highly modified Compact secondary structure	(Marín, Fernández-Calero, & Ehrlich, 2017)
Ribosomal RNAs	5S - 121 nt 5.8S - 156 nt 28S - 5070 nt	Pol I - 18S, 28S and 5.8S Pol III - 5S	Translation of mRNAs	Form 60S and 40S ribosomal subunits	(Lafontaine & Tollervey, 2001)
small nuclear RNAs	150 nt	Pol II and Pol III	Processing of pre-mRNA in the nucleus Regulation of transcription factors (7SK RNA) or RNA polymerase II (B2 RNA) Maintaining the telomeres	Found within the splicing speckles and Cajal bodies of the cell nucleus	(Matera, Terns, & Terns, 2007)
small nucleolar RNAs	-	Pol II or Pol III	Post-transcriptional modifications to rRNAs, tRNAs and snRNAs	Two large families termed box C/D and H/ACA	(Bachellerie, Cavaillé, & Hüttenhofer, 2002)
microRNAs	21-24 nt	Pol II	Translation attenuation of target mRNAs	Complete or partial sequence complementarity with target	(Ambros, 2004)
short interfering RNAs	20-25 nt	Pol III	Degradation of target mRNAs	Full sequence complementarity with target	(Dana et al., 2017)
Piwi-interacting RNAs	24-32 nt	Pol II	Epigenetic and post-transcriptional silencing of transposons Regulation of other genetic elements in germ line cells	Lack of sequence conservation	(Siomi, Sato, Pezic, & Aravin, 2011)

Summary of all known types of mammalian non-coding RNAs with the respective lengths, functions, specific characteristics and references.

Table 1.3 - Cell engineering approaches in CHO cells for recombinant protein production

Method	Effect	Reference
Cold Shock	Level of luciferase activity increased 6.5-fold at 32 C relative to 37 C due to higher fidelity and accuracy of post-translational events	(Masterton, Roobol, Al-Fageeh, Carden, & Smales, 2010)
	Protein aggregation of TNFR-Fc was partly reduced in a PERK-dependent way at 31°C	(K. Wang et al., 2018)
	Cooling-induced reprogramming of the translatome increases synthesis of RTN3	(Bastide et al., 2017)
mTOR Signalling	Adenosine induced growth arrest improved IFN-γ titer 1.4-fold alongside a 2.5-fold increase in average specific productivity	(Chong, Sim, Wong, & Yap, 2009)
	Expression and phosphorylation profiles of the mTORC1 substrate 4E-binding protein 1 (4E-BP1) fluctuate throughout the course of cell culture, with a positive correlation between the eIF4E/4E-BP1 stoichiometry and cell productivity in an IgG4 producing CHO cell line	(Jossé, Xie, Proud, & Smales, 2016)
	Cultivation of mTOR-transgenic CHO-derived cell lines engineered for secretion of a therapeutic IgG resulted in antibody titers of up to 50 pg/cell/day, which represents a four-fold increase compared to the parental production cell line	(Dreesen & Fussenegger, 2011)
siRNA	LDH-A activities were decreased by 75-89%, while the specific glucose consumption rates were reduced to 54-87% and the specific lactate production rates were reduced to 45-79% of the control cell line level, without impairing either cell proliferation or productivity	(S. H. Kim & Lee, 2006)
	siRNA mediated inhibition of PDHKs and LDH-A in CHO cells expressing a therapeutic monoclonal antibody reduced lactate production, increased specific productivity and volumetric antibody production by 90%, 75% and 68%, respectively, without appreciable impact on cell growth	(Zhou et al., 2011)
	Ribosome profiling identified NeoR as a highly transcribed and translated gene an IgG-producing CHO cell line. Viable cell density was increased by 35% upon siRNA knock-down of NeoR, which was accompanied by an 18% increase in product titer	(Kallehauge et al., 2017)
	Combined transient siRNA-mediated knockdown of the endoplasmic reticulum localized proteins CerS2 and Tbc1D20 resulted in a 50-66% increase in specific productivity of CHO-IgG cells	(Pieper et al., 2017)
miRNA	Co-expression of miR-557 and difficult-to-express antibody resulted in a two-fold increase in product titer	(Fischer et al., 2017)
	miR-143 overexpression resulted in a 20% final increase in mAb productivity	(Schoellhorn, Fischer, Wagner, Handrick, & Otte, 2017)
	Addition of a synthetic 3'UTR to destabilize DHFR expression allowed the generation of stable DG44-derived cell pools expressing a model monoclonal antibody (mAb) with low MTX concentrations	(Jossé, Zhang, & Smales, 2018)
	Inhibition of miR-124-3p and miR-19b-3p in CHO increased X-linked inhibitor of apoptosis protein levels, enhancing CHO cell growth and prolonging culture longevity while additionally boosting productivity	(Griffith et al., 2018)
lncRNA	Engineering of CHO cells with SINEUP long non-coding RNAs resulted in 150% increase in periostin levels in cell supernatant at 72 h post-transfection	(Patrucco et al., 2015)
	The development of a 'universal' protein expression enhancer tool based upon long non-coding RNAs allowed expression enhancement in various mammalian cells of recombinant proteins in the order of 50-1000%, with more than 200% enhancement achieved in most cases	(Yao et al., 2015)
Codon optimization	Modification of human interleukin-2 (IL-2) through codons with high gene copy number and high codon usage bias significantly increased protein productivity in CHO-K1 cells	(Ou et al., 2014)
	The majority of gene sequence optimization studies report an increase in protein expression level but with huge variability, ranging from none to >1000-fold	(Gustafsson et al., 2012)
	Codon de-optimization of a bispecific antibody sequence through the introduction of less frequently occurring codons in CHO showed a doubled final yield	(Magistrelli et al., 2017)
	In vivo expression of various codon context (CC) optimized IFN-γ in CHO cells exhibited at least 13-fold increase in expression level compared to the wild-type IFN-γ while a maximum of 10-fold increase was observed for the individual codon usage (ICU) optimized genes.	(B. K. S. Chung, Yusufi, Mariati, Yang, & Lee, 2013)

Summary of cell engineering studies discussed in this review grouped by type of approach, with the reported experimental effect and the relative reference.

1.3.2 The translation machinery, mRNA analysis and manipulation

As mRNA translation is a key process in defining cell growth, biomass accumulation and recombinant protein yields and quality from cultured CHO cells (McLeod et al., 2011; Mead et al., 2015), the translational machinery and the abundance and availability of global and recombinant mRNAs between cell lines and process conditions has been investigated using a variety of approaches. For example, the phosphorylation of the translation initiation factor eIF2α and attenuation of global protein synthesis during recombinant protein production in CHO cells is known to occur (Figure 1.1A, Steps 1-5) (Underhill, Birch, Smales, & Naylor, 2005). Culture temperature has been shown to impact mRNA translation and the quality of recombinant product produced (Masterton et al., 2010) and the PERK-eIF2α pathway (Figure 1.1A, Steps 1-5) was reported to impact upon the aggregation of a recombinant TNFR-Fc fusion protein (K. Wang et al., 2018). PERK is a type I transmembrane kinase located in the endoplasmic reticulum (ER), that under ER stress conditions oligomerizes and trans-autophosphorylates, inhibiting general protein translation through the phosphorylation of eukaryotic translation initiator factor-2 (eIF2α) at serine 51. This event alleviates the overload of proteins entering the ER but at the same time allows the selective translation of the transcription factor ATF4, thus promoting an antioxidant response, the enhancement of the folding capacity of the ER, and the upregulation of macroautophagy. (Hetz & Papa, 2018). Indeed the activity and availability of translation factors has been shown to change during culture and under different culture conditions, where for example under reduced temperature translation elongation factor 2 (eEF2) becomes phosphorylated (Figure 1.1B) and a reprogramming of translation occurs that means transcripts with particular codon usage can escape the general global attenuation of translation under such conditions and the translation of these transcripts is actually enhanced (Bastide et al., 2017). Further, high producing antibody cell lines have been shown to maintain translation initiation factors at levels that allow such cells to maintain enhanced recombinant protein synthesis above that of lower producing cells (Mead et al., 2015).

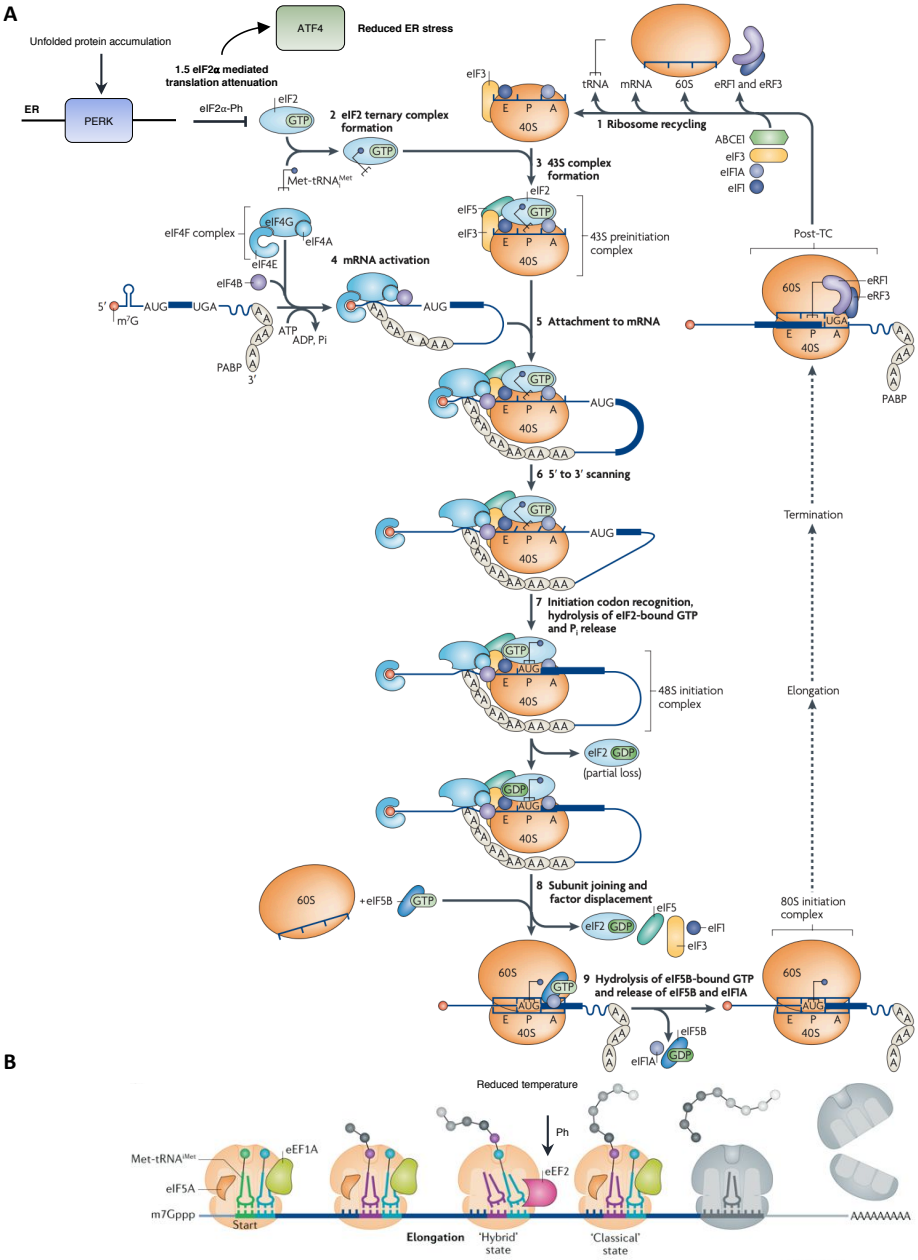


Figure 1.1 - Eukaryotic mRNA translation initiation and elongation

A | Schematic of eukaryotic translation initiation divided into 9 stages. Following translation termination and recycling of the post-termination complex (post-TC)(1), the 40S subunit, eIF1, eIF1A and eIF3 bind to the eIF2 ternary complex composed of eIF2, GTP, Met-tRNAiMet (2) to form the 43S preinitiation complex with eIF5 (3). The phosphorylation of eIF2α, one of the subunits of eIF2, causes global translation attenuation in response to cellular stress. Subsequently to ATP-dependent mRNA activation (4) by eIF4F and eIF4B, the 43S complex attaches on the mRNA (5) and starts scanning of the 5′ UTR in a 5′-3′ direction (6). Once the initiation codon is recognised, the 48S initiation complex is formed, switching the scanning complex to a 'closed' conformation (7). The 60S subunit together with eIF5B-GTP joins the 48S complex while eIF1, eIF3, eIF4B, eIF4F and eIF5 and eIF2–GDP are displaced (8). Finally, the 80S initiation complex ready to start the elongation phase is formed as a result of GTP hydrolysis by eIF5B and release of eIF1A and GDP-bound eIF5B (9). This figure was modified from "The mechanism of eukaryotic translation initiation and principles of its regulation, Jackson RJ, Hellen CU, Pestova TV, Nat Rev Mol Cell Biol. 2010 Feb;11(2):113-27. doi: 10.1038/nrm2838", to show eIF2α phosphorylation control. **B |** Schematic of eukaryotic translation elongation. Starting with a complex formed by the aminoacyl-tRNAs, eIF5A, eEF1A and GTP (not shown) in the ribosome, when the peptide-bond is formed the tRNAs are positioned in a 'hybrid' state with respect to the ribosome subunits. eEF2 drives translocation, causing tRNA repositioning to a 'classical' state, and creating an open A site for the incoming aminoacyl-tRNA. Phosphorylation of eEF2 as a result of reduced temperature reprograms translation promoting the expression of transcripts with a particular codon usage. This figure was modified from "Roadblocks and resolutions in eukaryotic translation. Schuller AP, Green R. Nat Rev Mol Cell Biol. 2018 Aug;19(8):526-541. doi: 10.1038/s41580-018-0011-4", to show eEF2 phosphorylation control.

With regard to monoclonal antibody synthesis in CHO and other cells, investigations have shown that recombinant antibody production is limited by translational efficiency (Mead et al., 2012, 2015; O'Callaghan et al., 2010). Manipulation of the cellular translational machinery is however, not straightforward. One global regulator of ribosome biogenesis and translation is mTORC1, which coordinates cellular responses to signalling pathways involved in sensing growth factors, nutrient availability, intracellular energy status and other perceived cell stresses and modulates translation and ribosome biogenesis in response (Foster & Fingar, 2010). In particular, mTORC1 can influence translation initiation via phosphorylation of eukaryotic initiation factor 4E binding protein (4E-BP1), which when phosphorylated at multiple sites promotes dissociation of 4E-BP1 from the initiation factor eIF4E (Figure 1.2). Increased phosphorylation of 4E-BP1 has been correlated with increased interferon-γ production (Chong et al., 2009) whilst the stoichiometry of 4E-BP1 to eIF4E is reported to relate to recombinant antibody productivity (Jossé et al., 2016) (Figure 1.2). Exogenous mTOR expression has also been shown to enhance recombinant protein expression in CHO cells by improving culture viability, cell growth, proliferation and cell specific productivity (Dreesen & Fussenegger, 2011) (Figure 1.2).

One approach applied to investigate mRNA amounts, and hence determine gene expression profiles of high producing or fast-growing recombinant cell lines is transcriptomics. A general assumption of most such studies is that the amount of mRNA present at a given timepoint reflects the 'state' or 'need' of a cell with regard to the proteins that these mRNAs encode for. As such, transcriptomic profiling has been applied to identify mRNAs whose abundance correlates with cell growth and recombinant protein productivity and quality with a view to using the identification of such targets to engineer the cell for improved performance. Many of these initial studies were hampered by the lack of the CHO genome and appropriate arrays, however the elucidation of the genome (Lewis et al., 2013; Xun Xu et al., 2011) and advent of RNA-Seq has made it possible to undertake such studies with a higher degree of precision. Despite this, there has been little consensus across transcriptomic studies to date with regard to those mRNAs that correlate with cell growth and recombinant protein productivity (Tamošaitis & Smales, 2018). Further, previous correlation analysis has shown that transcript amounts and translation efficiency are uncoupled for around 95% of investigated genes (Courtes et al., 2013), providing strong evidence that global and mRNA specific translational control

needs to be understood and determined to evaluate the impact of mRNAs on phenotype rather than simple mRNA abundance itself.

In order to address this issue, investigators have begun to apply ribosome footprint profiling or RiboSeq analysis to unravel the fine detail of translational control in CHO cells (Tzani et al., 2018). This powerful approach allows genome wide, but also transcript specific, detail on initiation and elongation stages of mRNA translation to be studied and identification of those mRNAs that are being translated at any given time (as opposed to just their abundance), the efficiency of mRNA translation and how this changes during a process or between cell lines to identify targets for cell line engineering (Tzani et al., 2018). Indeed, any given mRNA in the cell may be translated by one or multiple ribosomes (so called polysomes) at any one time (Godfrey et al., 2017). In some cases the number of ribosomes per transcript has been used to estimate translational efficiency of a transcript assuming that more ribosomes on a transcript indicates greater translational efficiency (Botao Liu & Qian, 2016), but this does not account for elongation speed that RiboSeq analysis can. RNA-Seq approaches can also potentially be used to investigate translational activity at the single cell level. The application of such approaches is certain to provide a more detailed understanding of mRNA translation and its control in recombinant protein producing CHO cells, at the population and single cell level and at a global and transcript specific level, revealing new engineering approaches by which translation can be modulated to enhance protein production.

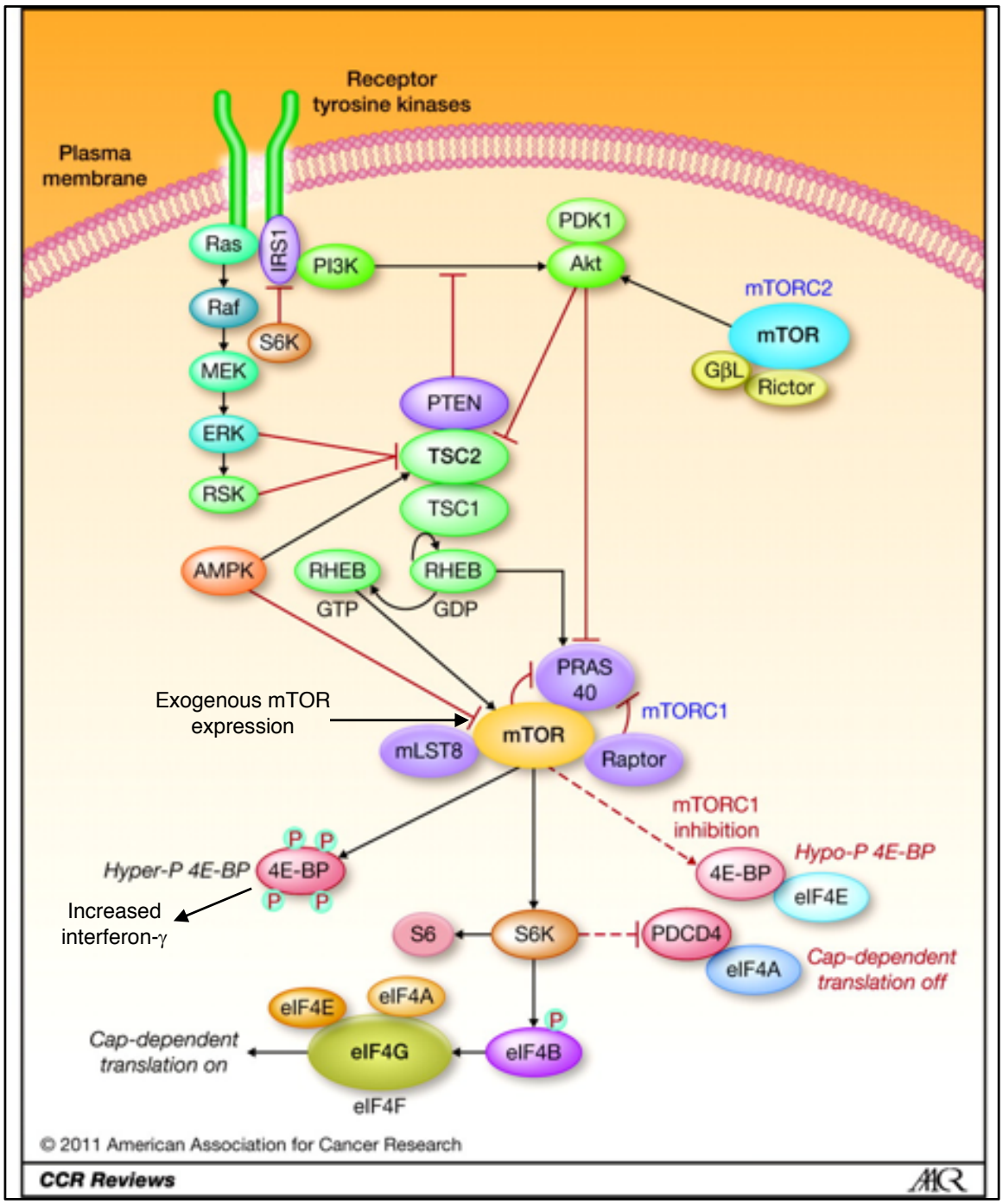


Figure 1.2 - mTOR regulated signalling pathway to control translation initiation

In response to mitogenic and growth-promoting external signals, receptor tyrosine kinases stimulate Ras and PI3K. PI3K causes activation of AKT, which as a result reduces TSC1–TSC2 GAP activity, allowing activation of mTORC1 by Rheb-GTP, and negatively regulates the inhibitor of mTORC1 PRAS40. Alternatively, mTORC1 can be activated as a result of the inhibitory phosphorylation of TSC1/2 complex mediated by extracellular signal-regulated kinase (ERK) or p90RSK. The phosphatase PTEN prevents activation of mTORC1 by PI3K while growth-inhibiting conditions in the environment lead to mTORC1 inhibition through AMP kinase activation.
mTORC1 phosphorylation of the 4E-BPs promotes the dissociation of eIF4E from 4E-BPs and the subsequently interaction with eIF4G and simultaneous stimulation of eIF4A helicase activity caused by S6K-mediated phosphorylation of eIF4B and PDCD4, an inhibitor of eIF4A. This cascade of events culminates in the assembly of the cap-binding complex eIF4F, composed of eIF4E, eIF4G, and eIF4A, stimulating cap-dependent translation. This figure was modified from "Targeting the mTOR/4E-BP pathway in endometrial cancer. Clin Cancer Res. 2011 Dec 15;17(24):7518-28. doi: 10.1158/1078-0432.CCR-11-1664. Epub 2011 Dec 5., Korets SB, Czok S, Blank SV, Curtin JP, Schneider RJ.".

1.3.3 microRNAs and siRNAs, short non-coding RNAs that repress mRNA translation

mRNA translation can also be tuned by non-coding RNAs. One such class of non-coding RNA that has been applied to reprogramming translation in CHO cells is that of microRNAs (also known as miRs or miRNAs). The potential application of microRNAs to CHO cell engineering has recently been reviewed elsewhere (Valdés-Bango Curell & Barron, 2018). These RNAs are transcribed as long primary transcripts but then processed to yield small (20-23 nucleotide) non-coding RNAs and were first described in *C. elegans*. MicroRNAs tend to act as repressors of translation of target mRNAs by interacting with the 3'untranslated regions (3'UTRs) of such mRNAs (Figure 1.3). Functional complexes composed of miRNAs and Argonaute (AGO) proteins bind the target mRNA in the seed region (nucleotides 2 to 8 in the miRNA). These complexes repress the mRNA expression by inhibiting the initiation step of translation, likely through release of eukaryotic initiation factor 4 A-I (eIF4A-I) and eIF4A-II, and subsequently promoting mRNA deadenylation through interaction with protein GW182, which binds polyadenylate-binding protein (PABPC), the deadenylation complexes poly(A)-nuclease deadenylation complex subunit 2 (PAN2)–PAN3 and carbon catabolite repressor protein 4 (CCR4)–NOT. This is then followed by decapping by the complex mRNA-decapping enzyme subunit 1 (DCP1)–DCP2, finally leading to 5'–3' mRNA degradation (Figure 1.3). A given microRNA can in theory target multiple mRNAs via base pairing and hence modulate multiple mRNAs and pathways without placing an additional translational burden on the cell (Valdés-Bango Curell & Barron, 2018).

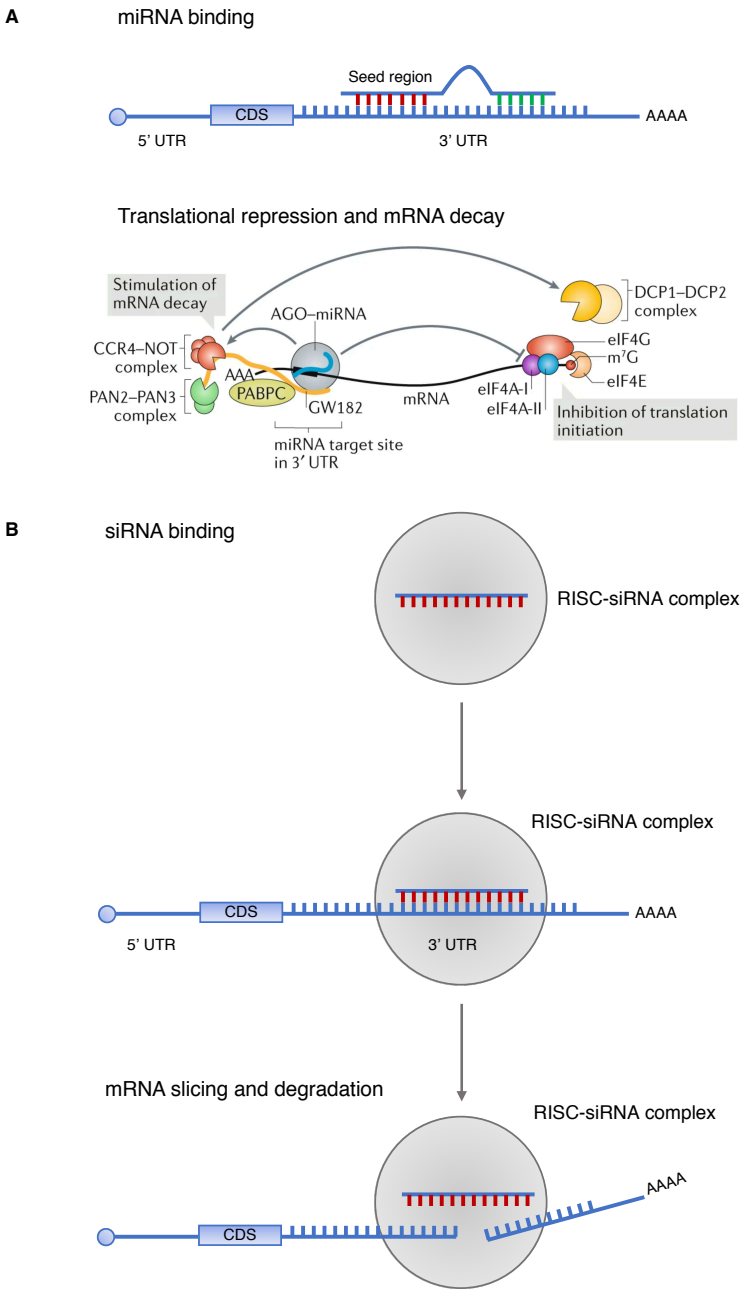


Figure 1.3 - Schematic depicting the main mechanisms of action for microRNA-mediated mRNA repression in animals.

A | Imperfect complementarity between the miRNA and the target mRNA leading to mRNA repression is the most common mechanism in animals. Complementarity in the seed region between nucleotides 2 to 8 of the miRNA (indicated in red) is necessary and sufficient to trigger the inhibition of target mRNA translational initiation and stimulation of RNA decay. Complementarity on the 3' supplementary sites between nucleotides 12 and 16 of the miRNA (indicated in green) can be present to enhance seed region binding. Although not shown in the figure, a single 3' UTR region can be target of multiple miRNAs simultaneously. The second panel of this figure was adapted from: Regulation of microRNA function in animals, LFR. Gebert & IJ. MacRae, Nature Reviews Molecular Cell Biology, 2019. **B |** Extensive complementarity between the siRNA, loaded on the RISC complex, and the mRNA, triggers endonucleolytic cleavage of the target mRNA.

Early microRNA studies in CHO cells were limited by the lack of available Chinese hamster sequence annotation of microRNA primary transcripts, and hence chimeric microRNAs that contained the mature miR sequence but flanking sequences in the primary transcript from other species were used. Subsequent studies showed that endogenous CHO microRNA flanking sequences gave rise to higher expression when over-expressing microRNAs (Klanert et al., 2014). Recent engineering approaches harnessing microRNAs include studies that look to enhance the ability of CHO cells to produce so called 'difficult to express proteins'. For example, one such study showed that a CHO cell line constitutively over-expressing miR-557 and a difficult to express antibody produced twice the antibody yield of cells engineered to express a negative control microRNA (Fischer et al., 2017). A further study reported that both transient and stable miR-143 over-expression resulted in enhanced difficult to express protein production and targeted MAPK7 in CHO cells (Schoellhorn et al., 2017). The natural repertoire of microRNAs has also been harnessed to repress expression of the DHFR selection marker during cell line construction and allow the generation of higher producing cell pools (Jossé et al., 2018). Others have shown that microRNA fingerprints or signatures can be correlated with growth rate across a number of different CHO cell lines (Klanert et al., 2016).

However, although microRNA engineering appears an attractive approach by which to tune translation of multiple mRNAs and translation of specific targets, the potential large number of predicted targets of any given microRNA means that the outcome of such engineering approaches can be difficult to predict as is identifying which targets a given microRNA interacts with. Barron and colleagues have described a system termed 'miR-CATCH' that allows the investigator to identify those microRNAs that interact with a given target and thus validate these for potential cell engineering approaches (Griffith et al., 2018). The authors had identified that the overexpression of the X-linked inhibitor of apoptosis (XIAP) enhanced CHO cell productivity, growth and culture longevity. To avoid overexpressing this gene and placing an additional translational burden on the cell, microRNA regulators of XIAP were identified using a biotin-labelled antisense DNA for XIAP resulting in the capture of interacting microRNAs. Inhibition of two of these microRNAs resulted in increased XIAP protein expression, validating the microRNA catch approach and the utility of this for identifying cell engineering targets.

The use of small interfering RNAs (siRNAs) has also proved to be an effective strategy for CHO cell line engineering to selectively knockdown expression of target genes detrimental to cell growth or productivity. The advantage of knockdown, as opposed to knockout

strategies, is that essential genes can be reduced in their expression and the impact on cell phenotype assessed when knockout proves fatal. siRNAs mediate repression of a target mRNA translation by guiding its sequence-dependent slicing and subsequent degradation. siRNAs begin as long dsRNA molecules which are processed by the endonuclease Dicer into active ~21-25 nt constructs. These are then loaded on the multi-protein RNA-induced silencing complex (RISC), where the guide strand pairs with its fully complementary target mRNAs. This triggers slicing of the target gene by RISC, which is then recycled for another few rounds of slicing, while the target mRNA is degraded (Figure 1.3). One successful application of siRNA engineering has been the inhibition of lactate dehydrogenase-A (LDH-A) on its own (S. H. Kim & Lee, 2006), or in combination with pyruvate dehydrogenase kinases (PDHKs) (Zhou et al., 2011) to significantly reduce lactate accumulation in cultured CHO cells without negatively impacting cell growth and enhancing cell specific productivity. A further group undertook ribosomal profiling in CHO cells and identified the resistance marker NeoR as being highly transcribed and translated, and as expression of this exogenous gene in CHO cells is not required, used siRNA knockdown to reduce its expression with a resultant improvement in production and growth of the host observed (Kallehauge et al., 2017). Finally, a siRNA approach was used to knockdown the expression of the endoplasmic reticulum localized proteins ceramide synthase 2 (CerS2) and Rab1 GAP Tbc domain family member 20 (Tbc1D20) in CHO IgG producing cells with a subsequent observed increase in recombinant protein specific productivity and enhanced cell growth (Pieper et al., 2017).

1.3.4 Long non-coding RNAs (lncRNAs) and their manipulation

Recent genome wide analysis in mammalian cells estimates that 75% of the transcriptome is composed of non-coding sequences (Djebali et al., 2012) and led to the identification of a heterogeneous class of transcripts known as Long non-coding RNAs (lncRNAs) (Kapranov et al., 2007). LncRNAs are defined as transcripts longer than 200 nucleotides that lack a significant open reading frame (ORF) and are usually transcribed by RNA polymerase II and spliced, with or without, 3' polyadenylation (Figure 1.4) (Wilusz, 2016). These molecules are emerging as key regulators in various biological processes both in the nucleus and the cytoplasm (Geisler & Coller, 2013), including epigenetic regulation, transcriptional control, splicing events, and mRNA translation. While most of our current understanding into lncRNAs and the underlying mechanism(s) by which they elicit their responses has come from

studies relating to disease and developmental studies, their potential as targets for cell engineering in mammalian cell factories remains largely unexplored.

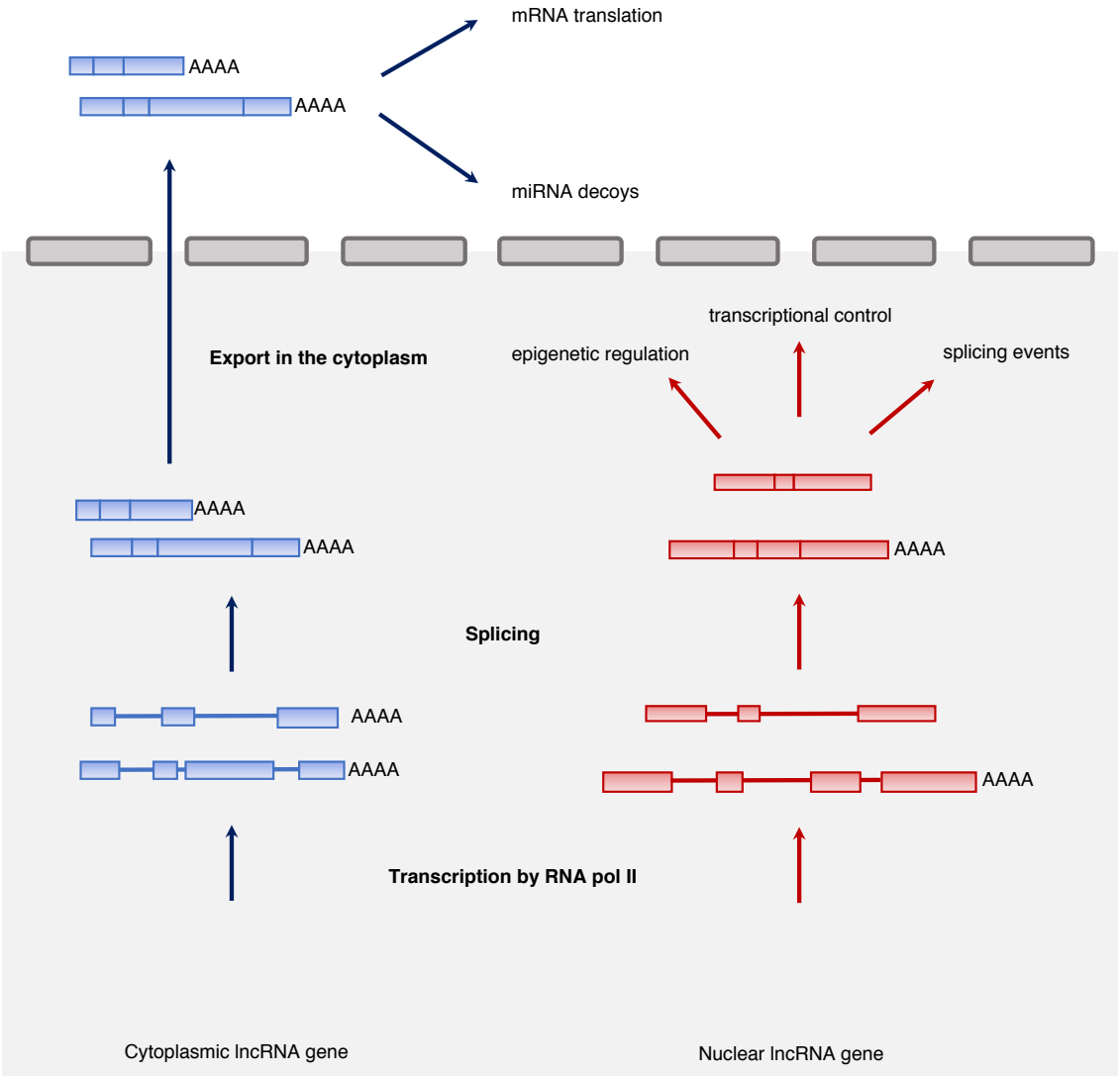


Figure 1.4 - Schematic depicting LncRNA transcription, localisation and functions in the cell

While all lncRNA genes are transcribed by RNA polymerase II, only a subset undergoes co-translational polyadenylation. The transcripts are then spliced and remain in the nucleus, where they act at an epigenetic and transcriptional level to silence or to enhance the expression of single genes or even entire genomic regions and to regulate splice variants formation. Alternatively, transcripts are actively exported in the cytoplasm, where they selectively promote translation of specific transcripts or bind miRNAs working as a decoy to alleviate targeted mRNA repression.

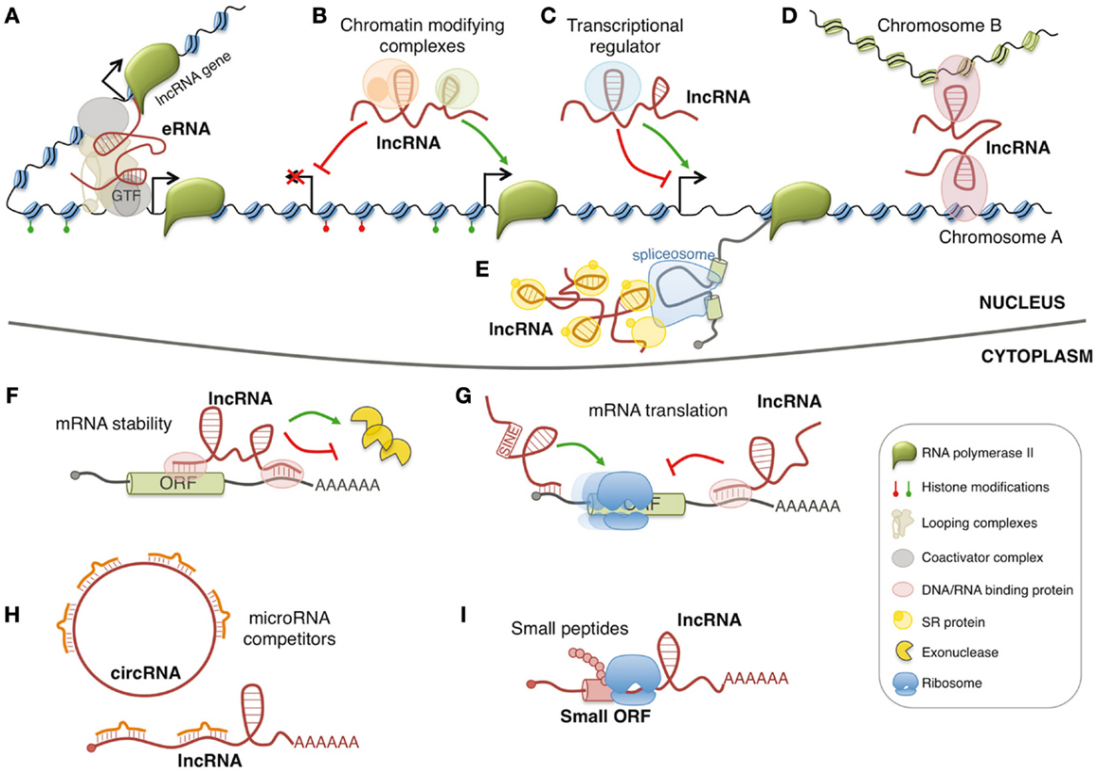


Figure 1.5 - Graphical representation of the main known lncRNAs functions in the cell

LncRNAs in the cell can act in the nucleus at an epigenetic level to silence or to enhance the transcription of entire genomic regions (A to D) and influence alternative splicing (E). When transported in the cytoplasm, they can bind to microRNAs and compete with their natural targets (H) or even interact directly with mRNAs to regulate their stability (F) and selectively promote translation (G,I). The figure is taken from: Long Non-Coding RNAs: New Players in Hematopoiesis and Leukemia, M. Morlando, M. Ballarino, A. Fatica, 2015, Front Med, PMID: 25927065.

The first analysis of the non-coding transcriptome in CHO cells under batch and fed-batch conditions has recently been published, unveiling a number of differentially regulated lncRNAs depending on feed and culture time which could be targets for cell engineering (Vito & Smales, 2018). One of the main challenges in identifying lncRNAs is the low sequence conservation between species. This, coupled with incomplete genome sequences and partial annotations of coding and non-coding genes of most vertebrates including Chinese hamster, have impaired an effective lncRNAs annotation outside from model organisms (Figure 1.6).

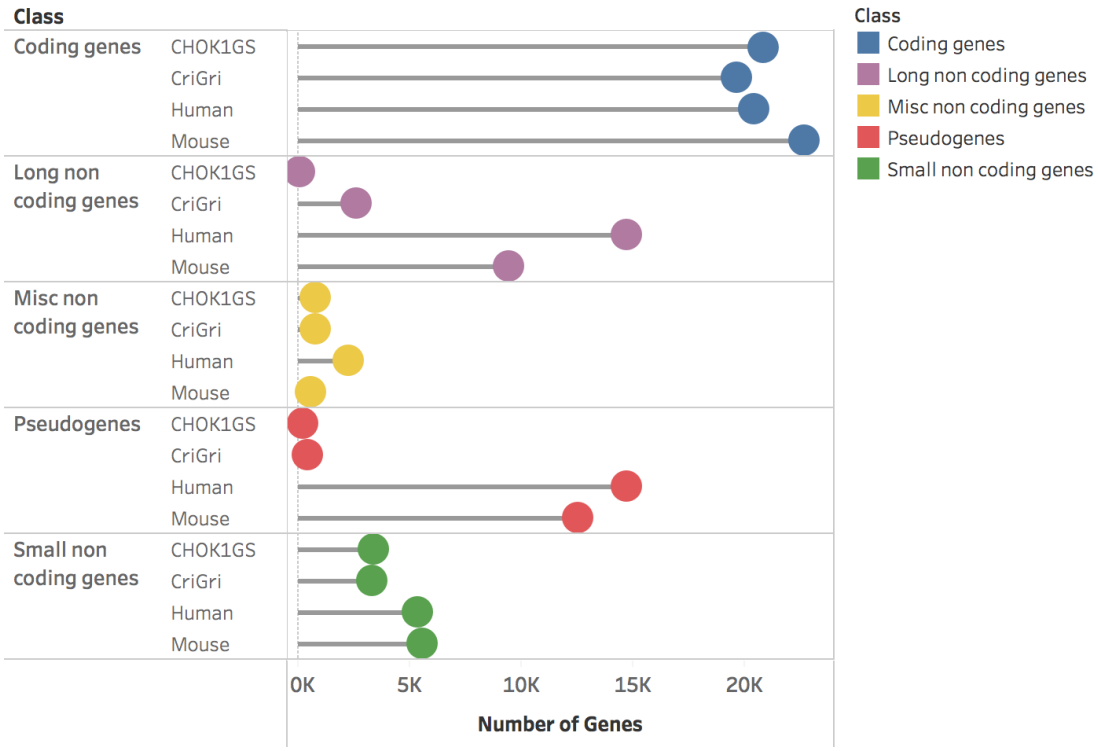


Figure 1.6 - Summary of ENSEMBL number of genes for hamster, human and mouse

The figure shows the number of genes for each annotated class in the ENSEMBL database for Chinese hamster (*Cricetulus griseus*, CriGri_1.0, GCA_000223135.1 and CHOK1GS_HDv1, GCA_900186095.1), mouse (*Mus musculus*, GRCm38.p6, GCA_000001635.8) and human (*Homo sapiens*, GRCh38.p12, GCA_000001405.27). While the number of coding genes is comparable between the four organisms, human contains 14,720 and mouse 9,443 annotated long non-coding genes as compared to 2,563 for CriGri.

A recently published study compared lncRNAs among 16 vertebrates and the echinoid sea urchin finding thousands of human lncRNA homologs with conserved genomic position sharing 5'-biased patches of sequence nested in rewired exonic architectures (Hezroni et al., 2015). The FANTOM consortium applied a cap analysis of gene expression (CAGE) to data obtaining more than 27,919 human lncRNA genes with high-confidence 5' ends and expression profiles across 1,829 samples from the major human primary cell types and tissues (Hon et al., 2017). Through the incorporation of conservation and expression data, the consortium was able to identify 19,175 potentially functional lncRNAs in the human genome. Due to the tissue-specificity of lncRNAs, comparing the expression among several cell types has led to a more robust identification of functional targets. By modelling their effects on the activity of transcription factors, RNA-binding proteins, and microRNAs in 5,185 TCGA tumours and 1,019 ENCODE assays, it was possible to identify potential lncRNAs involved in dysregulated cancer pathways. This approach indicated OIP5-AS1, TUG1, NEAT1, MEG3, and TSIX, as synergic lncRNAs leading to dysregulated cancer pathways in multiple tumour contexts (Chiu et al., 2018). A similar effort using nascent RNA capture sequencing

identified 1145 temporally expressed S-phase-enriched lncRNAs across TCGA data sets in several cancer models showing effects on pathways including FGF/FGFR and its downstream PI3K/AKT and MAPK pathways (Ali et al., 2018).

The NEAT1 lncRNA is a central component of paraspeckles, nuclear bodies that regulate multiple aspects of gene expression, promoting their formation through ATR signalling in response to replication stress and p53 activation (Adriaens et al., 2016). The RNA-binding NONO–PSF heterodimer binds a large number of expressed pri-miRNAs in the paraspeckles to promote processing by the Drosha–DGCR8 microprocessor. NEAT1 thus regulates efficient processing of potentially an entire class of small non-coding RNAs in the nucleus by interaction with the NONO–PSF heterodimer as well as other ribosome binding proteins (RBPs) (Jiang et al., 2017).

The relationship between lncRNAs and the translational machinery was further elucidated with the discovery of a long nucleolus-specific lncRNA (LoNA) (D. Li et al., 2018). LoNA is expressed at high levels at resting state, suppressing rRNAs transcription in the nucleoli through the combined effects of its 5' portion, which binds and sequesters nucleolin, and its snoRNA like 3' end, which recruits and diminishes fibrillarin activity to reduce rRNA methylation. When the cell needs to sustain an elevated translational load, LoNA expression decreases leading to elevated rRNA and ribosome levels, an increased proportion of polysomes, mRNA polysome loading, and eventually protein synthesis.

The first successful engineering of lncRNAs for enhanced recombinant protein production involved manipulation of SINEUPs, natural and synthetic antisense lncRNAs that can activate translation in a gene-specific manner using an inverted SINEB2 sequence (Carrieri et al., 2012). A Binding Domain (BD) located towards the 5' region of the SINEUP overlaps a target mRNA of choice conferring specificity, while an inverted SINEB2 element defined as the Effector Domain (ED) provides the translation activation function (Podbevšek et al., 2018). Synthetic SINEUPs have been used to increase translation and secretion of recombinant proteins in a range of mammalian cell lines, including CHO (Patrucco et al., 2015) and HEK293 (Yao et al., 2015) cells. Translation of secreted proteins usually occurs on the endoplasmic reticulum after recognition of the nascent signal peptide by the signal recognition particle (SRP) (Reid & Nicchitta, 2015). Transiently transfected natural and synthetic SINEUPs targeting the secreting leader sequence of endogenous and exogenous transcripts have been shown to increase the secreted levels up to two-fold for a single chain variable fragment (scFV) directed against the human protein DPP6 (Patrucco et al., 2015).

SINEUPs can also be used for high-throughput gene function screenings with comparable results to RNAi in most higher eukaryotic expression systems, including HEK293T, HEK293A, HeLa and CHO cells (Yao et al., 2015). SINEUPs represent a unique tool to specifically enhance the expression of secreted recombinant proteins in CHO cells and other mammalian systems and to investigate potential targets for cell engineering. As further studies define those lncRNAs present in CHO cells and how these influence cell growth, fate and recombinant protein production, engineering of these non-coding RNAs is sure to offer potential to further tune and enhance mRNA specific and global mRNA translational efficiency.

1.3.5 tRNAs and mRNA translation

The use of specific codons with high gene copy number and high codon bias coupled with the modulation of intracellular tRNA concentration has been shown to improve protein production in CHO cells (Ou et al., 2014). However, despite translational efficiency often being considered the mere result of codon optimization based on the correlation between codon bias and tRNA gene copy numbers (Figure 1.7), recent evidence suggests a considerably more intricate picture where ribosome collisions, co-translational folding, mRNA stability, composition, charge status and post-transcriptional modifications of the tRNA pool all contribute to finely tune protein production in response to the environment (Hanson & Coller, 2018).

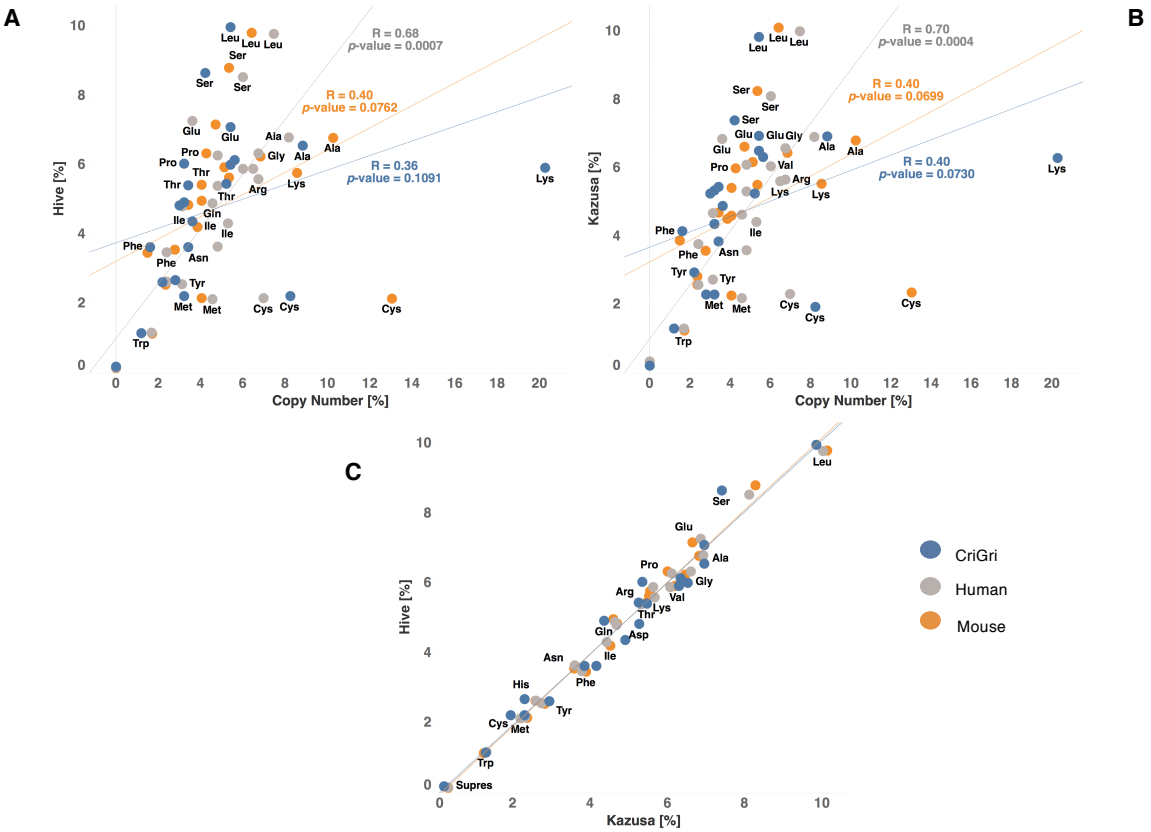


Figure 1.7 - Correlations between tRNA gene copy numbers and codon usage

The gene copy numbers for each codon were retrieved from the genomic tRNA database (GtRNAdb) while the codon usage was sourced from either the Codon Usage Database (kazusa.or.jp/codon/), indicated as "Kazusa", or the Codon Usage Table Database (hive.biochemistry.gwu.edu/review/codon), indicated as 'Hive'. For each aminoacid, the summary of every single codon coding for the same aminoacid is shown as a % of the total. *Cricetulus griseus* (CriGri) is shown in blue, Human in grey, and Mouse in orange. The figure shows in panel **A |** the correlation between the Hive and gene copy number, panel **B |** shows the correlation between the Kazusa and gene copy number and panel **C |** shows the correlation between Kazusa and Hive. The two databases perform similarly, giving a significantly higher correlation (R = 0.68 - 0.70) with a lower p-value (0.0007 and 0.0004) for Human, while Mouse and CriGri have lower correlations and higher p-values.

Controlling the translational capacity of an expression system through the use of alternative codon combinations modulates ribosome decoding speed, impacting protein quality as well as final yield (Zhao, Yu, & Liu, 2017). The use of suboptimal codons has been reported to slow translation at key structural motifs in order to facilitate correct co-translational polypeptide folding and signal recognition particle (SRP) recognition, which assists in protein translocation across membranes (Magistrelli et al., 2017). Thus, although recombinant genes are often 'codon optimized', we do not currently have all the information required around codon usage, context, tRNA abundance, modifications and charging to fully harness codon usage in recombinant sequences or to engineer tRNA abundance.

Codon bias has been referred to as a secondary genetic code that impacts on the fidelity of translation, efficiency of translation, polypeptide/protein folding and mRNA stability/half-life (Hanson & Coller, 2017). The cell utilizes such codon effects to tailor the proteome and allow reprogramming, such as under cold stress whereby reprogramming and synthesis of specific proteins is enhanced through codon bias (Bastide et al., 2017). Codon bias or optimization is also linked to tuning mRNA stability and stable mRNAs are found to be enriched in codons that are considered optimal whilst also impacting on ribosome translocation (Presnyak et al., 2015). Specific combinations of adjacent codons in yeast and mammalian cells can have an effect on translation efficiency resulting in reduced expression, proving how the focus must be on global translation efficiency and codon context as opposed to single codons optimality (Ang, Kyriakopoulos, Li, & Lee, 2016; B. K.-S. Chung, Yusufi, Mariati, Yang, & Lee, 2013; Gamble et al., 2016). On top of this, mRNA secondary structure combined with tRNA abundance modulate translational elongation speed among different regions of the same transcript to avoid excessively slow or fast ribosome movement (Gorochowski, Ignatova, Bovenberg, & Roubos, 2015). As such, there remains enormous potential to enhance recombinant protein yields from further manipulation of codon usage.

In order to further enhance recombinant protein yields by manipulation of codon usage it is necessary to further understand the abundance and modifications of tRNAs and the role these play in their activity. Determination of tRNA copy numbers can now be undertaken using RNA-Seq approaches. tRNA secondary structure and nucleotide modifications, mainly methylations, impair the efficiency of standard sequencing. Dedicated protocols based on an initial de-methylation step were recently developed to overcome this limitation, allowing for direct measurement of each tRNA abundance and detailed mapping of modifications (Cozen et al., 2015; Zheng et al., 2015). While some methods focus exclusively on mature tRNAs

(Shigematsu et al., 2017), partial alkaline RNA hydrolysis complemented with tRNA precursors enrichment identified tRNA leaders, trailers, and introns and showed that around half of all predicted tRNA genes are transcribed in human cells (Gogakos et al., 2017). While tRNA abundance is a major modulator of translational elongation, the aminoacylation state also has to be considered. The addition of chemical steps that specifically remove the 3'A residue in uncharged tRNA coupled with the aforementioned de-methylation RNA-Seq protocols showed most cytosolic tRNAs in HEK293T cells are charged at >80% levels, whereas tRNASer and tRNAThr are charged at lower levels (Evans, Clark, Zheng, & Pan, 2017).

An additional layer of regulation during elongation is chemical modification of nucleotides among tRNAs (Figure 1.8) (T. Pan, 2018). One of the key enzymes to regulate the methylation state of tRNAs is the demethylase ALKBH1, which acts dynamically in response to specific conditions such as variations in glucose availability to impact translation at both the initiation and the elongation phases (F. Liu et al., 2016). These modifications can have different effects depending on the target tRNA and the position in the transcript, as it was shown ALKBH1 is required for the formation of essential methylations at position 34 of anticodon in cytoplasmic tRNALeu and mitochondrial tRNAMet (Kawarada et al., 2017). Advances in high-throughput sequencing and data analysis have also allowed the identification of new classes of small non-coding RNAs derived from tRNAs: stress-induced tRNA halves (tiRs) and tRNA-related fragments (tRFs). These RNAs act on cell proliferation, priming of viral reverse transcriptase, regulation of gene expression, RNA processing, modulation of the DNA damage response, tumour suppression, and stress response (Kumar, Kuscu, & Dutta, 2016). The application of such approaches to study tRNAs in CHO cells will further elucidate the mechanism(s) by which tRNAs and their modifications modulate translation.

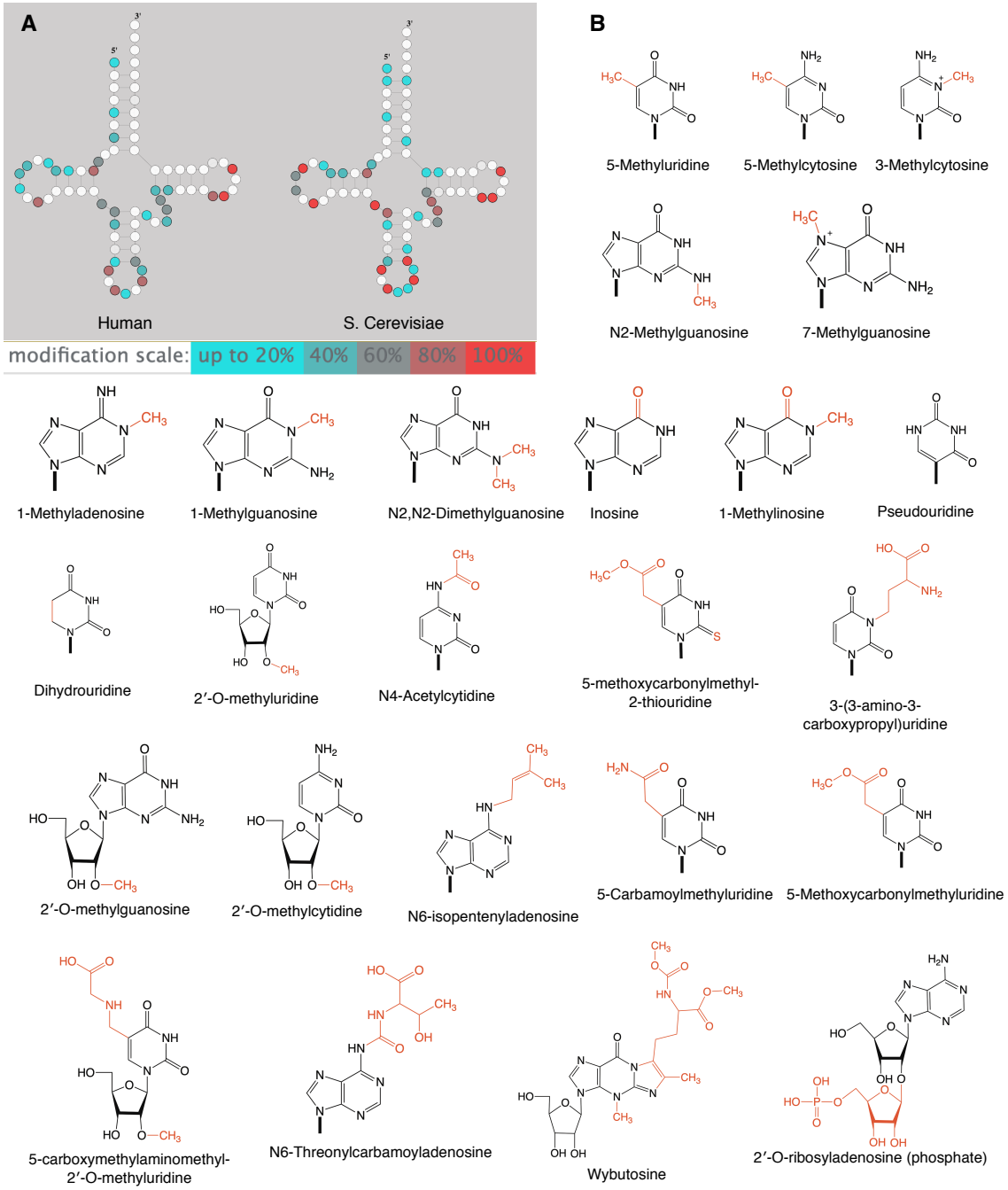


Figure 1.8 - Summary of chemical modifications to nucleotides

A | Known tRNA chemical modification to nucleotides in Human (left) and *S. cerevisiae* (right). The color indicates the percentage of modification frequency for each nucleotide on average considering all tRNAs in the Modomics database. **B |** Chemical structure of each modified nucleotide found in *S. cerevisiae* based on the Modomics database.

1.3.6 Challenges in modulation of translation for enhanced CHO cell biotherapeutic protein production

Our understanding and ability to manipulate the translational machinery and harness non-coding RNAs to enhance global and recombinant protein synthesis in CHO cells has advanced rapidly in the last decade. Further, the advent of the Chinese hamster and CHO cell line genomes has helped in the identification of non-coding RNAs such that these can be studied and manipulated. The ability of non-coding RNAs, in particular microRNAs, siRNAs, lncRNAs and tRNAs to tune both global and transcript specific translation, and hence protein synthesis, offers enormous opportunities to use these to enhance cell growth and proliferation, extend culture lifetimes, and increase recombinant protein yields and quality. However, our ability to harness these non-coding RNAs by engineering of CHO cells is currently limited by our knowledge of the mechanisms and targets by which many of these non-coding RNAs elicit their responses.

The manipulation of microRNAs that can, in theory, tune multiple target transcripts appears an appealing approach, however this approach alone is unlikely to deliver new commercially viable host cells with dramatically enhanced phenotypes due to the fact these are 'tuning' molecules and tend to be negative regulators and off target approaches can be difficult to control. Where these might be more applicable is for the tuning of transcript targets with a specific role, such as enzymes involved in glycosylation or to harness modulation of the cells own endogenous microRNA pool as inducible controllers of exogenous gene circuits.

The potential of lncRNA engineering is very much in its infancy and would appear to offer the potential to act as negative and positive regulators of gene expression. The limitation here is that many of these are, as the name suggests, long RNAs and thus the manipulation is more challenging and we do not yet understand what, if any, role many of these play in the cell. The control of gene expression via the elongation step of mRNA translation and tRNA availability, charging and modification, linked with improved predictive models for how such changes in abundance or modification change elongation rates of target mRNAs is likely to offer advances that can be directly applied industrially to engineering of the target recombinant gene(s) and of pathways in the cell to deliver new engineered host cell lines with improved growth, productivity and post-translational modification abilities. However, the major challenge will be to unravel the mechanisms by which the control on gene expression that these different non-coding RNAs provide are coordinated together, in order to reprogram the translational efficiency of current CHO cell chassis, under appropriate

bioprocessing conditions (including continuous processes) to generate new chassis with enhance bioprocessing properties.

1.4 Long non-coding RNAs are emerging as key regulators of cell biology

While the majority of CHO cell line engineering development studies initially focussed upon manipulation of coding genes, the growing awareness of non-coding RNA as central regulators of cell biology has exponentially increased the interest towards these molecules. Although the manipulation of microRNAs in particular has been successfully used in the research laboratory to enhance CHO cells productivity of recombinant proteins, to the authors knowledge none of these approaches have yet to be translated into manufacturing host cell lines.

The entire class of non-coding RNAs known as long non-coding RNAs (lncRNAs) remains poorly studied in CHO cells (Barron et al., 2011; Bort et al., 2012; Fischer et al., 2017; Stiefel, Fischer, Sczyrba, Otte, & Hesse, 2016). As described earlier, these transcripts are defined as RNAs longer than 200 nucleotides that lack a significant open reading frame (ORF), are usually transcribed by RNA polymerase II and spliced, with or without 3' polyadenylation (Kashi, Henderson, Bonetti, & Carninci, 2016; Kung, Colognori, & Lee, 2013; Wilusz, 2016). They are merging as key regulators in various biological processes including, but not limited to, epigenetic regulation (Betancur, 2016), transcriptional control (Trimarchi et al., 2014), splicing events (Gonzalez et al., 2015), and mRNA translation (Carrieri et al., 2012), both in the nucleus and the cytoplasm (Geisler & Coller, 2013). Our understanding into lncRNAs and their mechanisms is constantly growing, but the majority of information around these comes from studies relating to disease (Schmitt & Chang, 2016) and developmental studies (Perry & Ulitsky, 2016). Therefore, in the work described in this thesis it was decided to obtain the first landscape of long non-coding RNA expression in multiple CHO cell lines under batch and fed-batch conditions with different feeding approaches to investigate the behaviour of these transcripts and identify new targets for manipulation (Chapter 3). We further expanded this study in collaboration with the Danish biotech company Symphogen, sequencing the whole transcriptome of six IgG productive CHO cell lines cultivated in an automated ambr®15 system under controlled conditions (Chapter 4).

1.5 Translation is influenced by the codon sequence of mRNAs

As mRNA translation is a complex process composed of sequential phases and requiring the assemble of cellular machinery containing multiple subunits, protein abundance control in the cell can be tuned at multiple levels depending on internal or external factors (Kelen, Beyaert, Inzé, & Veylder, 2009). Due to this central role of translation in controlling gene expression, numerous works in CHO cells have investigated this process, showing a dependence on the phosphorylation of key translation factors like eIF2α causing attenuation of global protein synthesis (Underhill et al., 2005), on global translation efficiency (Mead et al., 2012) and on culture temperature, which impacts on the quantity and quality of recombinant protein produced (Masterton et al., 2010; Underhill et al., 2005; K. Wang et al., 2018). The interplay between these parameters and their variations throughout culture can have an impact with opposed effects for different transcripts.

Under reduced temperature, translation elongator factor 2 (eEF2) becomes phosphorylated leading to a reprogramming of translation and causing global attenuation, which can be escaped by some transcripts by specific codon usage (Bastide et al., 2017). The degeneracy of the genetic code allows the use of different synonymous codon sequences to achieve the same polypeptide, providing the option to tailor an exogenous recombinant coding sequence to the codon usage of the expression system of choice (Welch, Villalobos, Gustafsson, & Minshull, 2009). Indeed, codon usage is routinely considered when designing a recombinant sequence for maximizing protein expression, leading to an increase of up to 1000-fold in protein yield, but much more modest effects in general (Gustafsson et al., 2012; Kimchi-Sarfaty et al., 2013). While this approach assumes that rare codons are rate limiting for mRNA translation and thus protein production (Mauro & Chappell, 2014), recent works suggest a much more complex picture, where mRNA stability, ribosome speed and folding are influenced by specific codons and codon combinations, especially in mammalian cells (Brule & Grayhack, 2017).

A recent study used codon de-optimization of a bispecific antibody sequence through the introduction of less frequently occurring codons in CHO cells, showing an overall increase in final yield (Magistrelli et al., 2017). One of the reasons underpinning this different outcome could be the existence of distinct codon usage patterns depending on the cell type and phase, as reported in the case of proliferation and differentiation (Gingold et al., 2014; Plotkin, Robins, & Levine, 2004), or even among specific Gene Ontology (GO) sets of genes (Rudolph et al., 2016). On top of the intrinsic codon usage variability, mammalian cell factories carry

an additional burden on translation represented by the recombinant gene of interest and the selection marker used to generate the stable cell lines, which can account for up to 15% of the total ribosome capacity in the cell (Kallehauge et al., 2017).

An additional layer of complexity is the determination of the tRNA pool, a key player of translation elongation. Whereas an accurate quantification of single tRNA species by high throughput sequencing techniques has been limited due to secondary structure and nucleotide modifications, new protocols have recently been developed to overcome most of the limitations these place upon using RNA-Seq to determine tRNA abundance (Gogakos et al., 2017; Shigematsu et al., 2017; Zheng et al., 2015). AlkB-facilitated RNA methylation sequencing (ARM-Seq) in particular takes advantage of the bacterial dealkylating enzyme AlkB to remove methyl groups from m^1A, m^3C and m^1G residues in tRNAs allowing a much more accurate determination of tRNAs species present and their abundance (Cozen et al., 2015; Hrabeta-Robinson, Marcus, Cozen, Phizicky, & Lowe, 2017). In Chapter 5 of this work is presented the first description of tRNA quantification in a host CHO-S cell line under batch and fed-batch conditions and both null and IgG1 producing CHO DG44 cell lines cultivated in an automated mini-bioreactor setting using two different feeding strategies. In addition, as cooling is known to globally repress translation and the elongation phase in particular (published work reported in Appendix), the ARM-Seq protocol was applied to Human Embryonic Kidney (HEK293) cells to investigate the effects of temperature on the tRNAs pool.

1.5.1 Translation can be computationally modelled to predict decoding speed

Progress in our knowledge of the concentrations and activities of biological components and the improvement of quantitative methods in high-throughput biology has enabled development of more accurate calculations of kinetic rate constants and other parameters to model mRNA translation. Many attempts to model mRNA translation are reported in literature, but the majority of these do not give a systems-wide and dynamic representation of the process (Dominique Chu, Barnes, & von der Haar, 2011; Gromadski & Rodnina, 2004; Reuveni, Meilijson, Kupiec, Ruppin, & Tuller, 2011; Romano, Thiel, Stansfield, & Grebogi, 2009; Zouridis & Hatzimanikatis, 2007). Therefore, a tool was developed (D. Chu, Zabet, & von der Haar, 2012), and subsequently updated (Tarrant & Von Der Haar, 2014), using mixing agent-based techniques with event-driven stochastic simulation algorithms to address this (Gillespie, 1977). This tool is able to calculate the decoding speed of a given coding sequence and to predict the slowest and the fastest possible combinations of codons that code for the same amino acid sequence. Each transcript, represented as a string of codons with an initial

binding site summarizing a 5'-UTR, can be occupied by multiple ribosomes at the same time, encapsulating information about the particular mRNA they are bound to, the position as well as information about the tRNA they are interacting with.

The model assumes total numbers of tRNA, mRNA and ribosomes remain unchanged over the course of a simulation run. The following events drive the simulation:

1. Binding of free ribosomes to an initial binding site on a transcript
2. Binding of ribosomes located on an initial binding site to the first codon (AUG site)
3. Binding of near/non-cognate aa-tRNAs to unoccupied bound ribosomes
4. Unbinding of near/non-cognate tRNAs from a bound ribosome
5. Translocation of ribosomes to the next codon
6. Aminoacylation of uncharged tRNA
7. Re-initiation/dissociation of ribosome once the ribosome has reached the end of the ORF

In biological systems, translation initiation is composed of multiple steps, partially determined by structural properties of the 5'-UTR, assembly of ribosomal subunits and scanning of the 5'-UTR (von der Haar & McCarthy, 2002). The single parameters of these steps are poorly characterized, therefore this sequence of events was collapsed into a single initiation step (Event 1) following second-order kinetics with a mRNA-type specific rate constant. ORF specific translation rate starts with when the ribosome binds to the first AUG site (Event 2), which follows 0-th order dynamics. Events 3 to 5 summarize the movement of a ribosome from one codon to the next, requiring an aa-tRNA to bind to the ribosomal binding site, to pass a proof-reading step and to deliver the amino acid (Gromadski & Rodnina, 2004). Both cognate and near-cognate aa-tRNAs can bind to the ribosome in this model, then the tRNA is accepted or rejected depending on whether the aa-tRNA is cognate or near-cognate. The model loops into Event 4 if the tRNA is rejected, returning the aa-tRNA to the pool, while it proceeds to Event 5 if the tRNA is accepted, resulting in a de-aminoacylation of the tRNA at the ribosomal P-site. When the next elongation cycle is completed, the uncharged tRNA unbinds and Event 6 summarizes an aminoacylation event using Gillespie's algorithm assuming first-order kinetics and a fixed rate constant that is equal for all tRNA species (Dominique Chu et al., 2011). This model of elongation was applied to mRNAs with roles in neuronal processes identified by polysome profiling to analyse

elongation over the initial 20 codons, showing that a subset of mRNAs, including RTN3 and Noggin, contained codons that require less abundant tRNAs in the 5' end of the transcripts, allowing them to escape the repression of elongation occurring during temperature decrease. This work was published in Current Biology as follows, and is reported in the Appendix:

Bastide A, Peretti D, Knight JR, Grosso S, Spriggs RV, Pichon X, Sbarrato T, Roobol A, Roobol J, Vito D, Bushell M, von der Haar T, Smales CM, Mallucci GR, Willis AE, (2017), RTN3 Is a Novel Cold-Induced Protein and Mediates Neuroprotective Effects of RBM3, *Curr Biol.,* doi: 10.1016/j.cub.2017.01.047, PMID:28238655

Chapter 5 describes the application of this model to predict the decoding speed of model recombinant proteins based on tRNA gene copy numbers and the tRNA abundances measured with ARM-Seq to assess differences in codon combinations of the coding sequence in CHO under different conditions.

1.6 Aims of the work described in this thesis

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

1. Obtain the first landscape of long non-coding RNA (lncRNA) expression in CHO cells in response to different culture conditions.
2. Improve the annotation of lncRNAs in CHO cell databases and their connection to coding genes.
3. Identify candidate non-coding genes to use for genetic manipulation to enhance desirable CHO cell properties, specifically growth and recombinant protein productivity.
4. Quantify tRNAs in CHO cells for the first time using AlkB-facilitated RNA methylation sequencing.
5. Assess the impact of recombinant protein expression, feed and time point sampling on the tRNA pool.
6. Use the quantified tRNAs in a translational elongation phase computational model to calculate the predicted decoding speed of various recombinant transcripts and to predict the fastest and slowest possible codon combinations with the same primary amino sequence.

The outcomes from this work, and arising publications, are detailed in the following Chapters.

Chapter 2 Materials and Methods

2.1 Cell culture and RNA extraction

2.1.1 CHO-S cell line

A CHO-S host cell line (ThermoFisher Scientific, MA, USA) was cultured in a Lab-Therm LT-X (Kühner AG, Basel, Switzerland) shaking incubator at 37°C, 5% CO_2, 125 rpm and 70% humidity in chemically defined serum-free growth medium (CD CHO, ThermoFisher Scientific). Fed-batch cultures were supplemented with CHO CD Efficient Feed B Liquid Nutrient Supplement (ThermoFisher Scientific) immediately on Day 0 with 15% (v/v) of the initial starting volume, followed by 10% (v/v) supplementation on Day 3 and Day 6. Four biological replicates of each culture process (batch and fed-batch) were seeded at $3x10^5$ viable cells/mL from 20 mL cultures with a culture viability ≥ 98.5% in 120 mL CD CHO starting working volume in 500 mL polycarbonate Erlenmeyer flasks with vented caps (Corning). Viable cell concentrations and culture viability were determined daily using a Vi-CELL XR Cell Viability Analyzer (Beckman Coulter, Inc., Brea, CA, USA) on 1 mL culture samples. Cell samples of 1×10^7 viable cells were taken from each biological replicate flask after 96 hours (day 4) and 144 hours (day 7) of culture and small RNAs under 200 nt in length were immediately extracted using the commercially available mirVana miRNA Isolation Kit (ThermoFisher Scientific) and treated with RapidOut DNA Removal Kit (ThermoFisher Scientific) according to the manufacturers' instructions. The recovered RNA was quantified using a NanoDrop ND-1000 (ThermoFisher Scientific) instrument and the small RNAs integrity and enrichment were assessed by standard denaturing agarose gel electrophoresis. *S. cerevisiae* was cultured in standard YPD media and the small RNAs extracted using the same described method.

2.1.2 Symphogen cell lines

Four Symphogen in-house mAb CHO cell clones (designated 3068, 3077, 3080, 4384 producing the same IgG1 subtype) and two cell pools, of which one (3936) was IgG1-producing and the other (3478) was a null-producer, were used in this study. All were generated from a modified dihydrofolate reductase-deficient (DHFR-) CHO DG44 host cell line (Urlaub et al., 1983) through transfection with a vector containing the DHFR gene and the genes for the antibody heavy (HC) and light chains (LC) and methotrexate (MTX) mediated stable selection . Clones were isolated by fluorescence activated cell sorting (FACS). Cells were routinely maintained and expanded in PowerCHO-2 CD (Lonza) as basal

media in shake flasks with shaking at190- 200 rpm in a 37°C humidified culture incubator supplied with 5% CO_2. For upstream experiments to generate samples for analysis, cells were inoculated at a starting concentration of either 0.4x 10^6 viable cells/ml (DAVI experiments) or 0.6 x 10^6 viable cells/ml (JCE experiments, cell culture and RNA extractions executed by Dr Eriksen J.C. at Symphogen), in a total culture volume of 13 ml. The ambr15™ micro bioreactor by Sartorius (Goettingen, Germany) was used to run fed-batch culture experiments for either 12 (DAVI) or 14 days (JCE). For feeding of cultures, HyClone Cell Boost 6 Supplement (GE Healthcare Bio-Sciences AB, Uppsala, Sweden) + 8% FunctionMAX™ TiterEnhancer (ThermoFisher Scientific, Massachusetts, United States) was used with a feeding regime of every 2nd or 3rd day for JCE experiments or daily from the 2nd day of culture for DAVI experiments. Culture viability and viable cell concentration (VCD) was measured using a Vi-CELL XR (Beckman Coulter, Brea, CA) instrument. For JCE experiments, glucose, glutamine, lactate, ammonium, glutamate, pH, and osmolality were measured using a Bioprofile 100plus (Nova BioMedical, Waltham, WA) instrument while IgG titer was determined by biolayer interferometry using an Octet QK384 instrument equipped with Protein G biosensors (ForteBio, Menlo Park, CA). For DAVI experiments, a Cedex Bio HT analyser (Roche, Basel, Switzerland) was used for all measurements. For seed train samples, cells were harvested 48 hours after adjusting their VCD to 0.3x10^6 viable cells/ml, when a steady doubling time and high culture viability were maintained. Cell suspensions from each seed-train in the JCE dataset were added to 15 ml Falcon tubes and centrifuged at 1000 rpm for 4 minutes, followed by Direct-zol RNA extraction (Zymo Research, California, USA) following the manufacturer's instructions. For samples taken from the fed-batch cultures, 1 x 10^7 viable cells were taken before feeding from each biological replicate ambr15™ bioreactor after 96 h of culture time (day 4, both DAVI and JCE experiments) and after either 168 h (day 7, JCE experiment) or 288 h (day 12, DAVI experiment) of culture. Cells were lysed using TRI-Reagent (Sigma-Aldrich, Missouri, USA) followed by extraction of total RNA using a Direct-zol RNA Kit (Zymo Research) and in-column DNAse treated. Samples for tRNAs sequencing were extracted using the mirVana miRNA Isolation Kit (ThermoFisher Scientific) and eluted in RNase free water. The RNA quantity/quality was determined using a NanoDrop instrument (ThermoFisher Scientific) and RNA 6000 Pico Kit (Agilent, California, USA).

2.2 Whole transcriptome RNA-Seq on Symphogen samples

2.2.1 RNA-Seq and Data analysis

For DAVI experiments: the total RNA was sent to the Oxford Genomics Centre (Oxford, UK) where the Ribo-Zero ribosomal RNA (rRNA) removal kit (Illumina, California, USA) was used to remove rRNAs followed by multiplex barcoded sequencing on an HiSeq4000 by Illumina to generate 150 bp long paired ends reads, with coverage over 40M fragments per sample. The fastq files were initially checked for quality using the FastQC software on default settings, sorted using samtools, deduplicated using Picard MarkDuplicates then aligned with HISAT2 to the CriGri_1.0 reference genome. Aligned files are publicly available on GEO, accession GSE140671. The gene counts were calculated using featureCounts, then the differential expression analysis was undertaken with the R/Bioconductor package DESeq2 (Love, Huber, & Anders, 2014).

For JCE experiments: samples were sequenced at Genotypic (Bengaluru, India) to generate 75 nt long paired end reads, with coverage over 20M fragments per sample. The fastq files were initially checked for quality using the FastQC software on default settings, then trimmed using Trim_Galore! and aligned to GCF_000419365.1 with STAR. The gene counts were calculated using featureCounts, then the differential expression analysis was conducted with the R package DESeq2 (Love et al., 2014) as described for the DAVI experiments. Genes with a fold change (FC) > 1.50 and Benjamini-Hochberg adjusted *p*-value <0.10 in both datasets were selected for GO and KEGG enrichment using the Bioconductor package gProfileR (Supplementary Material 1).

2.2.2 RT-qPCR Validation of Differentially Expressed genes

Primers for RT-qPCR experiments were designed using Primer-BLAST and synthetized by Integrated DNA Technologies (Illinois, USA). Primers used during this study are described in Supplementary Material 2. RT-qPCR reactions were conducted using a Mastercycler EP Realplex instrument (Eppendorf, Hamburg, Germany) using the PrecisionPLUS Onestep qRT-PCR Master Mix kit by Primer Design (Southampton, UK). The specificity of amplification was checked by the generation of Tm curves and by analysis of the reaction products using 2% agarose gel electrophoresis to confirm the presence of a single amplicon of the expected size. The results were analysed applying the standard ΔCt method and normalized to B-actin and B2M housekeeping genes expression.

2.3 AlkB de-methylase production in *E. coli*

2.3.1 Protein expression and purification

The original AlkB-Facilitated RNA de-Methylation (ARM-Seq) workflow for tRNAs sequencing is described in detail here (Hrabeta-Robinson et al., 2017). A plasmid containing the wild-type *E. coli* AlkB sequence, with the addition of a His-Tag, was kindly provided by the Phizicky group, based at the University of Rochester, NY, USA. The plasmid was amplified in DH5-Alpha cells and expressed in BL21*(DE3) pLysS *E. coli* cells in the presence of chloramphenicol and ampicillin and induced with IPTG. Cells were harvested, resuspended in Binding Buffer (20 mM Tris-HCl pH 8, 500 mM NaCl, 5 mM Imidazole pH 8) and sonicated. A nickel affinity chromatography column was prepared by loading 5 mL of Chelating Sepharose to an empty column, washed with 20 mL of water, charged with 15 mL of Charge Buffer (100 mM $NiSO_4$) and equilibrated with 25 mL of Binding Buffer. The cell lysate was loaded onto the column, then washed in succession with 20 mL of Binding Buffer, 10 mL of Wash Buffer I (20 mM Tris-HCl pH 8, 500 mM NaCl, 50 mM Imidazole pH 8), 10 mL of Wash Buffer II (20 mM Tris-HCl pH 8, 500 mM NaCl, 100 mM Imidazole pH 8) and eluted in 15 mL of Elution Buffer (20 mM Tris-HCl pH 8, 500 mM NaCl, 400 mM Imidazole) divided in single 1 mL fractions. The purified AlkB was stored in a buffer composed of 20 mM Tris-HCl, 200 mM NaCl, 2 mM DTT, 50% Glycerol and RNase free water.

2.3.2 SDS-PAGE protein characterisation

2.3.2.1 Sample preparation procedure

Samples were prepared adding to the eluted fractions a 5X sample buffer composed of 10% (v/v) glycerol, 120 mM Tris-HCl pH 6.8, 2% (w/v) SDS, a few grains of bromophenol bleu, a few grains of pyronin G diluted in distilled water to a 1X final buffer concentration, then vortexed thoroughly, boiled for 5 min at 95˚C and loaded into the polyacrylamide gel.

2.3.2.2 Gel preparation and electrophoresis

Sodium dodecyl sulphate (SDS) polyacrylamide gels were composed of a stack phase (5% acrylamide; 5% (v/v) acrylamide/Bis solution 37.5:1 (BioRad, USA), 125 mM Tris-HCl pH 6.8, 1% (w/v) SDS, 1% (w/v) Ammonium persulfate, 0.1% TEMED) and a separating phase between 8 and 12% acrylamide (8-12% (v/v) acrylamide/Bis solution 37.5:1 (BioRad, USA), 375 mM Tris-HCl pH 8.8, 1% (w/v) SDS, 1% (w/v) Ammonium persulfate, 0.06% TEMED) set in a Novex cassette (Life Technologies, USA). Samples were loaded on the gel and electrophoresis was performed in running buffer (200 mM glycine, 25 mM Tris-HCl, 0.1% (w/v) SDS, pH 8.8) at 100 V for 30 min and then at 150 V until desired migration.

2.3.2.3 Bradford assay

To determine AlkB protein concentration, 1 mL of Bradford reagent (120 mM Coomassie Blue G250, 15% (v/v) ethanol, 8.5% (v/v) phosphoric acid) was added to 50 µL of sample, incubated 10 min at room temperature before using a calibrated spectrophotometer for the Bradford reagent used.

2.3.3 MS-MS protein characterisation

The SDS-PAGE gel was washed twice with water for 10 minutes, then the band corresponding to AlkB was excised using a clean scalpel, cut into 1 x1 mm squares and transferred to a 0.5 mL centrifuge tube. The gel particles were washed with 100 µL of freshly prepared 50 mM NH_4HCO_3/acetonitrile (1:1) for 15 minutes, then the liquid was removed and 100 µL acetonitrile added for 15 minutes until the gel pieces had shrunk. After removing the liquid, the gel pieces were covered with 10 mM DTT in 50 mM NH_4HCO_3, incubated for 30 minutes at 56°C and briefly shrunk again in acetonitrile. The acetonitrile was then removed and 55 mM Iodoacetamide in NH_4HCO_3 was added for 20 minutes in the dark at room temperature. The iodoacetamide solution was removed and the gel pieces washed two times with 100 µL of NH_4HCO_3 for 15 minutes, shrunk again with acetonitrile and dried in a vacuum centrifuge. The dried gel pieces were rehydrated in 20 µL of digestion buffer (10 mM NH_4HCO_3, 10% Acetonitrile) containing 10 ng/µL of trypsin at 4°C for 30 minutes, then the remaining supernatant was removed and 10 µL of digestion buffer without trypsin added to cover the gel pieces and keep wet during enzyme cleavage at 37°C overnight. 5 µL acetonitrile was then added to the gel pieces, followed by 15 minutes sonication, centrifugation, collection of the supernatant and addition of 50% acetonitrile with 5% formic acid. After additional 15 minutes of sonication, the gel pieces were span down and the supernatant collected for MALDI TOF-TOF MS/MS analysis (Supplementary Figure 5.1).

2.4 tRNA sequencing

2.4.1 Small RNA de-methylation

Between 4.0 and 7.2 ug of RNA were de-methylated at 37°C for 100 minutes using 10 µg of AlkB enzyme in a 100 µL reaction composed of a freshly prepared reaction buffer (50 mM HEPES KOH pH 8, 75 µM ferrous ammonium sulfate pH 5, 1 mM alpha-ketoglutarate, 2 mM ascorbate Na+, 50 µg/mL BSA). For each different condition, a single non-demethylated sample ($AlkB^-$) was generated by substituting the volume of AlkB with the protein storage

buffer only. The RNA was then recovered using the Zymo Research RNA CleanUp kit (Zymo Research, California, USA), eluted RNase free water and quantified by Nanodrop. The samples collected at Symphogen were on-column DNase treated during RNA recovery. After the de-methylation treatment and before sequencing, an integrity control on the RNA samples was performed by the sequencing facility using RNA Screentape (Supplementary Figure 5.2).

2.4.2 Library generation and sequencing

For the CHO-S dataset and the yeast samples, the library was generated using a NEXTFLEX Small RNA-Seq Kit (Bioo Scientific, TX, USA) and sequenced on a HiSeq 2500 (Illumina), generating single end 50 bp long reads. For the Symphogen dataset, 150 ng of small RNAs where used to generate a library with the TruSeq Small RNA Library Prep kit (Illumina, California, USA). The standard protocol was followed with the exception of the reverse transcription step, changed to a 1-hour incubation at 60^{o}C with SuperScript IV Reverse Transcriptase (ThermoFisher Scientific) and the purification of cDNA constructs, where the size window for the fragment was 140 to 250 bp to include all tRNAs fragments. cDNA fragments were checked with a High Sensitivity D1000 ScreenTape assay (Agilent, California, USA) and sequenced in biological duplicates on a NextSeq 500 (Illumina), generating single end 75 bp long reads.

2.4.3 Alignment and differential expression

Quality control on the fastq files was run using FastQC, adapters were trimmed using cutadapt 1.16 (Martin, 2011), discarding reads that were shorter than 16 bases after trimming. A reference set was created by downloading tRNA sequences from GtRNAdb (Chan & Lowe, 2016), removing any predicted intron sequences and adding a 3' terminal CCA to each tRNA. The reads were mapped to this reference set of tRNA sequences using Bowtie2 2.3.1 (Langmead & Salzberg, 2012), using sensitive options and primary mappings. Reads per tRNA codon were counted using an in-house Perl script. Reads were also mapped, using Bowtie2 with the same options, to a reference set of all small RNA sequences. This reference set was created by using Ensembl (Zerbino et al., 2018) BioMart to retrieve all transcripts not labelled as protein coding, and adding in the sequences from the tRNA reference set described above. Reads were again counted using an in-house Perl script. Differential expression analysis was performed using DESeq2 1.18.1 (Love et al., 2014), using raw counts at the individual codon level.

2.5 Codon Adaptation Index (CAI) and tRNA gene copy numbers

The Codon Adaptation Index (CAI) of the selected coding sequences was calculated using the publicly available domain https://www.biologicscorp.com/tools/CAICalculator/ where the complete Chinese hamster genome was taken as a reference. The CAI for top and bottom 500 most abundant coding sequences as a custom set of references was calculated using the CAI python package (B. D. Lee, 2018). Gene copy numbers for each tRNA were calculated in Chinese hamster genomes (CriGri_1.0, C_griseus_v1.0, CHOK1GS_HDv1) and *Mesocricetus auratus* (MesAur1.0), using the publicly available tool tRNAscan-SE 2.0 (Lowe & Chan, 2016).

2.6 HEK293 cell line

2.6.1 Cell culture and RNA extraction

Four T175 flasks of HEK293 cells were grown at 37°C for 48 h then two of the flasks were transferred to 32°C for 24 h and two flasks were left at 37°C. After the 24 h, medium was removed from the cells, the flasks were washed with 10 mL PBS and then cells were detached with 2.5 mL trypsin-EDTA/flask (either at 37 or 32°C as appropriate) and the trypsin then quenched by adding 4.5 mL growth medium per flask. The two 37°C flask contents were pooled as were the 32°C flask contents into 15 mL falcon tubes. 20 uL samples were removed for cell counts and the remainder of the cell suspensions were harvested at 1000 rpm x 5min. The medium was removed from the cell pellets and the falcon tubes containing the cell pellets were immediately frozen in dry ice and stored at -80°C. The small RNAs samples were extracted following the mirVana protocol for small RNAs purification and subsequently treated using the bacterial demethylase AlkB as described in Section 2.4.1. After the de-methylation treatment and before sequencing, an integrity control on the RNA samples was performed by the sequencing facility using RNA Screentape (Supplementary Figure 5.2).

2.6.2 Library generation and sequencing

150 ng of each RNA sample was processed using the Illumina TruSeq Small RNA library prep kit (Illumina). The standard protocol was followed with the exception of the reverse transcription step, changed to a 1-hour incubation at 55°C with SuperScript III Reverse Transcriptase (ThermoFisher Scientific), a 20 cycle PCR amplification step, and the purification of cDNA constructs between 146 bp and 200 bp. The libraries were then run on a NextSeq instrument, using the 75 bp single end read metric on one lane with the high-output option, and analysed as described in Section 2.4.3.

2.7 Computational modelling of elongation

Translation elongation rates on all indicated coding sequences were estimated using a published computational model (D. Chu et al., 2012; Dominique Chu et al., 2014), described in Section 1.5.1. The model was adapted for the Chinese hamster decoding system using relative sequenced tRNA abundances obtained as described in Section 2.4.

Chapter 3 The Long Non-Coding RNA Transcriptome Landscape in CHO Cells Under Batch and Fed-Batch Conditions

While the biological functions and the use of coding genes has been investigated extensively for CHO cell engineering, the focus on non-coding RNAs has almost exclusively been on siRNAs and miRNAs. An emerging class of non-coding RNAs, defined as Long Non-Coding RNAs (lncRNAs), has in recent years shown a broad range of actions in cell biology, but the majority of our knowledge on these molecules is limited to diseases and model organisms. We evaluated lncRNA expression in a mammalian expression system (CHO cells) under industrially relevant conditions and assessed their potential as cell engineering targets. Through the use of a mouse microarray providing the surveillance of 24,881 mRNAs and 35,923 lncRNAs, this work delivered to the community the first landscape of expression for both coding genes and lncRNAs in CHO cells under batch and fed-batch conditions at Day 4 and Day 7 of culture. When comparing fed-batch against batch time points, I found thousands of differentially expressed mRNAs and lncRNAs. Single gene functions and their potential application as targets for cell engineering, as well as entire pathway variations, were identified and discussed in detail. This work was published in the Biotechnology journal as follows:

Vito D, Smales CM *(*2018), The Long Non-Coding RNA Transcriptome Landscape in CHO Cells Under Batch and Fed-Batch Conditions, *Biotechnol J.,* doi: 10.1002/biot.201800122, PMID:29781203

This work is reported in the following pages and integrated with the rest of the PhD project described in the thesis.

3.1 Introduction

Mammalian expression systems are widely used for the production of recombinant protein biopharmaceuticals, largely due to their ability to correctly fold, assemble, post-translationally modify and secrete complex human like proteins (Walsh, 2010). Among these, the Chinese hamster ovary (CHO) cell is the most widely utilised expression platform used in industry, especially for the production of monoclonal antibodies (mAbs), and are used for the expression of more than 60% of biotherapeutic proteins made in mammalian cells (Dumont et al., 2016; Jayapal, Wlaschin, Hu, & Yap, 2007; J. Y. Kim et al., 2012; Kunert & Reinhart, 2016; Walsh, 2014). The most widely used bioprocess for the production of biopharmaceuticals from CHO cells involves fed-batch culture, this offering an advantage over batch culture in terms of cell growth, culture viability and product yields due to the supplementation of nutrients resulting in higher biomass accumulation, less build-up of toxic metabolic by-products and enhanced productivity (Durocher & Butler, 2009; X. Pan, Streefland, Dalm, Wijffels, & Martens, 2017; Wong et al., 2006).

The prominence of CHO cell expression systems has driven innovation in the industry such that CHO expression systems and associated bioprocesses have been developed that can deliver yields of mAb in excess of 5-10 g/L in stably expressing, fed-batch systems (Povey et al., 2014; Reinhart, Damjanovic, Kaisermayer, & Kunert, 2015). Despite this, some biotherapeutic recombinant proteins, and particularly some of the non-mAb novel biotherapeutics in development, remain difficult to express in CHO cells or other mammalian cell expression systems (Alves & Dobrowsky, 2017; Jossé et al., 2016). Using *a prior* knowledge about cellular systems, various approaches have been taken to engineer cells to deliver enhanced product yields and quality including the engineering of chaperones (Josse, Smales, & Tuite, 2010; G. M. Lee, 2008), glycosylation machinery (Ferrara et al., 2006; Malphettes et al., 2010; Yamane-Ohnuki et al., 2004) and proliferation control strategies, including manipulation of the cell cycle (Bi, Shuttleworth, & Al-Rubeai, 2004; Fussenegger, Mazur, & Bailey, 1997), apoptosis (Choi, Rhee, Kim, & Park, 2006; N. S. Kim & Lee, 2002) and autophagy (Hwang & Lee, 2009; J. S. Lee, Ha, Park, & Lee, 2013).

Alongside this approach, there have been studies to further our understanding of the potential cellular constraints on the production of mAbs (Mason, Sweeney, Cain, Stephens, & Sharfstein, 2012; Pybus et al., 2014), other recombinant biotherapeutics (Johari, Estes, Alves, Sinacore, & James, 2015) and difficult to express recombinant proteins (Thoring et al., 2016) in order to identify bottlenecks in the recombinant gene expression pathway and to

develop new bioprocesses or adapt/engineer novel hosts for enhanced production and/or quality of such molecules (Mead et al., 2012, 2015). The majority of these studies to date have focussed upon manipulation of coding genes, however with the discovery and improved understanding of non-coding RNA in the control of cellular processes, there has been much interest in the last decade or so in these non-coding RNAs in CHO cells. In particular, the manipulation of microRNAs to enhance the ability of CHO cells to produce biotherapeutic proteins has attracted much attention (Barron et al., 2011; Bort et al., 2012; Fischer et al., 2017; Stiefel et al., 2016). An advantage of manipulating such non-coding RNAs for modulating CHO cell phenotypes is that such engineering does not place an addition protein synthetic burden upon the host cell (Hackl, Borth, & Grillari, 2012), unlike engineering of coding genes, and often such non-coding RNAs can modulate whole pathways rather than individual steps or processes as when manipulating many coding genes.

It has been estimated that at least 75% of transcripts originate from non-coding sequences (Djebali et al., 2012). Investigations into these transcripts has resulted in the identification of a class of transcript collectively known as Long non-coding RNAs (lncRNAs) (Kapranov et al., 2007). LncRNAs are defined as transcripts longer than 200 nucleotides that lack a significant open reading frame (ORF), are usually transcribed by RNA polymerase II and spliced, with or without 3' polyadenylation (Kashi et al., 2016; Kung et al., 2013; Wilusz, 2016). A number of lncRNA molecules have been shown to play key regulatory roles in various biological processes including epigenetic regulation (Betancur, 2016), transcriptional control (Trimarchi et al., 2014), splicing events (Gonzalez et al., 2015), and mRNA translation (Carrieri et al., 2012). Indeed, lncRNAs are capable of modulating a wide range of cellular processes and mechanisms both in the nucleus and the cytoplasm (Geisler & Coller, 2013). The majority of our understanding into lncRNAs and the mechanism(s) by which they elicit their responses has come from studies relating to disease (Schmitt & Chang, 2016) and developmental studies (Perry & Ulitsky, 2016). LncRNAs elicit their effects by acting as competing endogenous RNAs (ceRNAs) by binding to and sequestering microRNAs (Tay, Rinn, & Pandolfi, 2014), by acting as architectural RNAs (arcRNAs) whereby they form functional structures (Chujo, Yamazaki, & Hirose, 2016), act as *cis* molecules to enhance (eRNAs) coding gene expression (Melo et al., 2013) or as *trans* protein binding RNA molecules that can recruit chromatin modifying (Gomez et al., 2013; Koziol & Rinn, 2010), as microRNA precursors (Keniry et al., 2012), modulators of mRNA stability (Rashid, Shah, & Shan, 2016) and to impact upon post-translational modifications (Bodu Liu et al., 2015; P. Wang et al., 2014).

Despite the importance of lncRNAs in controlling cellular processes, and unlike small non-coding RNAs (e.g. siRNAs, microRNAs), the impact(s) of lncRNA expression on CHO cell bioprocessing with regard to growth/proliferation and recombinant protein yields and quality has barely been explored, with only a small number of studies reported (Patrucco et al., 2015; Tabuchi, 2013). However, these two studies show that manipulation of lncRNAs can impact upon recombinant protein production from CHO cells. One of the reasons for the small number of studies on lncRNAs in CHO cells is the lack of a comprehensive annotation of non-coding transcripts in CHO, hampering their identification, genome wide assessment of their expression and modulation during culture, functional studies and hence identification of target lncRNAs for cell engineering. One way to address this problem is to take advantage of the reported similarities between the Chinese hamster, CHO cell and mouse genomes (Becker et al., 2011), where the number of well annotated non-coding transcripts is much higher. The similarity between CHO cells and mouse has already been utilised for the identification of 416 ncRNAs based on sequenced transcripts from a pooled CHO sample compared to the fRNAdb database of non-coding RNAs using BLAST, with most hits coming from mouse (X Xu et al., 2011). Here, I report the first analysis in CHO cells of both the coding and the non-coding transcriptome (specifically the lncRNAs) during batch and fed-batch culture at two different time points. I report on the identification of differentially expressed lncRNAs and mRNA, their interconnections and their potential impact on cellular pathways. This has allowed the mapping of the lncRNA landscape in CHO cells and identification of lncRNAs targets in CHO for further manipulation with a view to increase proliferation and to sustain viability throughout batch and fed-batch culture.

3.2 Materials and Methods

3.2.1 Model Cell Line and Cell Culture Conditions

The CHO-S Freestyle host cell line (ThermoFisher Scientific, MA, USA) was used as a model CHO cell line for this study. Cells were routinely cultured in a Lab-Therm LT-X (Kühner AG, Basel, Switzerland) shaking incubator at 37°C, 5% CO_2, 125 rpm and 70% humidity in chemically defined serum-free growth medium (CD CHO, ThermoFisher Scientific, MA, USA). Fed-batch cultures were supplemented with CHO CD Efficient Feed B Liquid Nutrient Supplement (ThermoFisher Scientific, MA, USA). Initial supplementation testing was conducted in duplicate following Conditions 3 and 9 as described in the CHO CD Efficient Feed manual. Cultures with a viability >98% were used to seed initial fed experiments at $3x10^5$ viable cells/mL in a 50 mL working volume in 250 mL polycarbonate Erlenmeyer flasks

with vented caps (Corning, Wiesbaden, Germany). From the initial feeding experiments, Condition 3 was used for all future fed-batch cultures and for generating samples analysed during this study. This adopted feeding strategy consisted of a 15% (v/v) supplementation of CHO CD Efficient Feed B to the CD CHO starting volume immediately on Day 0, followed by 10% (v/v) supplementation on Day 3 and Day 6. Four biological replicates of each culture process (batch and fed-batch) were seeded at $3x10^5$ viable cells/mL from 20 mL cultures with a culture viability ≥ 98.5% in 120 mL CD CHO starting working volume in 500 mL polycarbonate Erlenmeyer flasks with vented caps (Corning). Viable cell concentrations and culture viability were determined daily using a Vi-CELL XR Cell Viability Analyzer (Beckman Coulter, Inc., Brea, CA, USA) on 1 mL culture samples.

3.2.2 Sampling from Cell Cultures and Subsequent RNA Extraction

Samples of 1 x 10^7 viable cells were taken from each flask after 96 hours (day 4) and 144 hours (day 7) of culture and total RNA immediately extracted using the commercially available mirVana miRNA Isolation Kit (ThermoFisher Scientific, MA, USA) and treated with RapidOut DNA Removal Kit (ThermoFisher Scientific, MA, USA). The RNA quantity and quality were determined using a NanoDrop ND-1000 (ThermoFisher Scientific) instrument and the integrity of RNA assessed by standard denaturing agarose gel electrophoresis.

3.2.3 lncRNA and Coding RNA Microarray and Data Analysis

3.2.3.1 Microarray Details

Three of the 4 cultures were selected for analysis. Analysis of extracted RNA for coding and lncRNAs was undertaken using the commercially available ArrayStar Mouse lncRNA Microarray V3.0 (Rockville, MD, USA). RNA labelling and array hybridization were performed according to the Agilent One-Color Microarray-Based Gene Expression Analysis protocol (Agilent Technologies, Santa Clara, CA, USA) with minor modifications. Ribosomal RNAs were removed from total RNA using the mRNA-ONLY Eukaryotic mRNA Isolation Kit, (Epicentre, Madison, WI, USA). Each sample was then amplified and transcribed into fluorescent cRNA along the entire length of the transcripts without 3' bias utilizing a mixture of oligo(dT) and random primers using the Arraystar Flash RNA Labelling Kit (Arraystar). The labelled cRNAs were purified by RNeasy Mini Kit (Qiagen, Hilden, Germany). The concentration and specific activity of the labelled cRNAs were determined by NanoDrop ND-1000 (ThermoFisher Scientific). A total of 1 μg of each labelled cRNA was then fragmented by adding 5 μL of 10 × Blocking Agent and 1 μL of 25 × Fragmentation Buffer before heating at 60°C for 30 min.

Finally, 25 µL of 2 × GE Hybridization buffer was added to dilute the labelled cRNA. A sample of 50 µL of the hybridization solution was then dispensed into the gasket slide and assembled to the LncRNA expression microarray slide. The slides were incubated for 17 hours at 65°C in a Microarray Hybridization Oven (Agilent Technologies). The hybridized arrays were then washed, fixed and scanned using the G2505C DNA Microarray Scanner (Agilent Technologies).

3.2.3.2 Microarray Data Analysis

Agilent Feature Extraction software (version 11.0.1.1) was used to analyse acquired array images. Quantile normalization and subsequent data processing were performed using the GeneSpring GX v12.1 software package (Agilent Technologies). After quantile normalization of the raw data, lncRNAs and mRNAs that were present in ≥3 of 12 samples were selected for further data analysis. Raw *p*-values were calculated by unpaired t-test, then the differentially expressed lncRNAs and mRNAs with statistical significance between compared groups were filtered for a FC ≥ 2 and a false discovery rate (FDR) ≤ 0.10 calculated by the Benjamini–Hochberg procedure. GO analysis was performed using the Bioconductor package topGO (Alexa & Rahnenfuhrer, 2016) and with a raw *p*-value cut-off of 0.05 calculated by Fisher's exact test, subsequently filtered for an FDR ≤ 0.10. Pathway analysis for differentially expressed mRNAs was performed based on the KEGG database (http://www.genome.jp/kegg) with a raw *p*-value cut-off of 0.05 calculated by Fisher's exact test, subsequently filtered for an FDR ≤ 0.10.

3.2.4 RT-qPCR Validation of lncRNAs Identified as Differentially Expressed by Microarray

Primers for RT-qPCR were designed using OligoPerfect Designer (ThermoFisher Scientific) and synthetized by Eurofins Scientific (Luxembourg) (listed in Appendix). RT-qPCR reactions were conducted using a Mastercycler EP Realplex instrument (Eppendorf) following the QuantiFast SYBR Green RT-PCR Kit protocol (Qiagen). Specificity of amplification was checked by the generation of T_m curves and by analysis of the reaction products by 2% agarose gel electrophoresis to confirm the presence of a single amplicon of the expected size. The results were analysed applying the standard ΔCt method and normalized to GAPDH, B-actin and B2M housekeeping gene expression.

3.3 RESULTS

3.3.1 Growth Characteristics and Sampling of CHO-S Cells Throughout Batch and Fed-Batch Culture for (Non)-Coding RNA Analysis

The steps followed to generate the results presented in this work are summarized in Figure 3.1. Initially, growth comparisons and supplementation testing were undertaken using CD CHO and the associated commercially available feds (Efficient Feds A and B). From these preliminary experiments, I selected Efficient Feed B Liquid Nutrient Supplement for experimental Fed-batch cultures as this gave the highest viable cell concentrations across a 10-day process from the feed strategies investigated (data not shown). Cells under Fed or Batch culture grew almost identically up until day 4 of culture whereupon their growth characteristics differed (Figure 3.2A). Whilst growth of the Batch cultures slowed dramatically after day 4, with a mean peak viable cell concentration on day 6 of approximately 1 x 10^7 cells/mL, in the Fed-batch cultures growth and proliferation continued until day 6 where a mean peak viable cell concentration of almost 2 x 10^7 cells/mL was obtained (Figure 3.2A). After day 6 the viable cell number and viability of both cultures decreased with time, however while culture viability rapidly declined in the Batch culture such that by day 9 culture viability was close to 0%, in the Fed-batch mode cultures had a mean viability of 83.6% at day 10 (Figure 3.2A). The first-time point at which samples were harvested was Day 4, when both types of culture had grown in a similar way and were still in the exponential growth phase (late exponential for batch), whilst the second time point at which cells were harvested for RNA analysis was on Day 7, corresponding to the end of the stationary phase/first day of decline for both cultures. In the case of the Fed-batch cultures, these two sampling points were also 24 hours after addition of the Efficient Feed B.

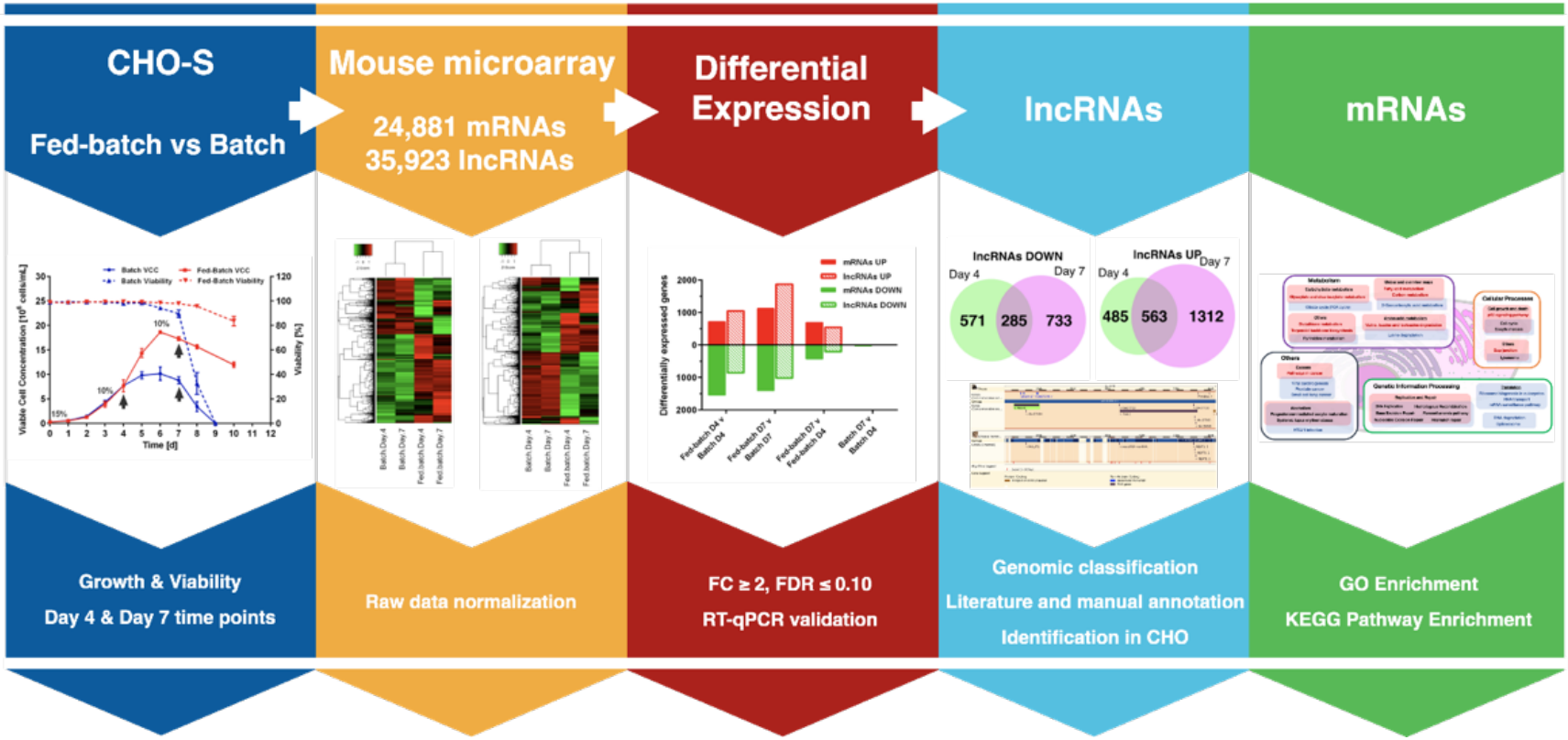


Figure 3.1 - Experimental workflow for transcriptomics analysis of CHO-S cell line

Summary of the experimental workflow. Growth and culture viability of a CHO-S cell line in Batch and Fed-batch cultures were measured for 10 days and samples for RNA extraction taken at Day 4 and at Day 7. The samples were analysed on a mouse array containing all the coding and non-coding transcripts stored in the main public databases. The measured intensities were log-normalized and differentially expressed transcripts/genes were filtered for a fold change (FC) ≥ 2 and an FDR ≤ 0.10. A selected group of genes was validated through RT-qPCR (Supplementary Table 3.1). Due to the poor annotation of lncRNAs in CHO, the identification of potential targets with a described biological role required the comparison of human and mouse literature and databases, followed by alignment against the Chinese hamster genome, leading to predicted lncRNAs transcripts and previously un-annotated genomic regions (Table 3.1). At the same time, GO and pathway enrichment was implemented on mRNAs.

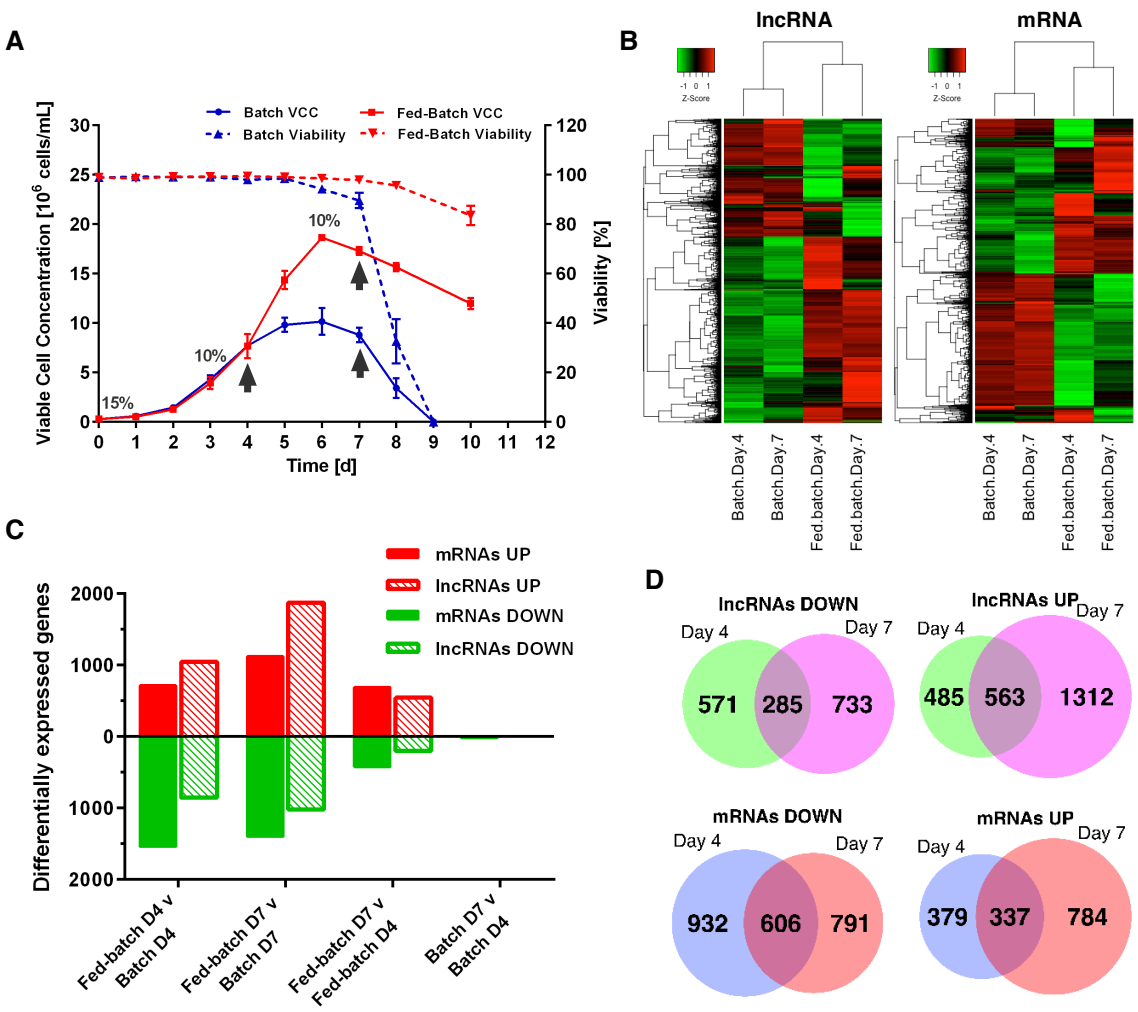


Figure 3.2 - Results of CHO-S batch and fed-batch cultures

A | Growth profiles for the model CHO-S host cell line throughout Batch and Fed-batch culture. Viable cell concentration (VCC) and culture viability are shown (red for Batch and blue for Fed-batch). The percentage of CHO CD Efficient Feed B Liquid Nutrient Supplement added to the existing working volume are shown for each feeding day (days 0, 3 and 6). The arrows indicate when samples for microarray analysis were harvested (Day 4 and Day 7). **B |** Hierarchical clustering heatmaps arranging samples into groups based on their averaged log-normalized expression levels. Only transcripts/genes with an expression variance between each group above the 80th percentile are shown. The dendrogram shows the relationships for lncRNAs (left panel) and mRNAs (right panel). **C |** Down-regulated (green bars) and up-regulated (red bars) mRNAs (full bars) and lncRNAs (textured bars) in CHO-S cells during batch and fed-batch culture for each of the compared pairs showing a fold change ≥ 2 and an FDR ≤ 0.10. D) Venn diagrams showing the number of lncRNAs and mRNAs differentially expressed (DE) in Fed-batch vs Batch (FC ≥ 2, FDR ≤ 0.10). Genes DE at both Day 4 and Day 7 are represented in the overlaps while genes DE only at one time point are represented inside their corresponding circle.

3.3.2 Microarray Analysis of mRNA and lncRNA Transcripts in Batch and Fed-Batch Culture

The Arraystar Mouse LncRNA Microarray V3.0 used in this study is based on publicly available databases and publications, allowing for the potential simultaneous surveillance of 35923 lncRNAs and 24881 coding transcripts. The lncRNAs collection is based on *Mus musculus* gene expression in all tissues as stored in the NCBI Refseq, UCSC Known Gene 6.0, Ensembl 38.71, Fantom3, RNAdb 2.0, NRED databases, a number of literature publications, T-UCRs, and evolutionary constrained lncRNAs (Amaral, Clark, Gascoigne, Dinger, & Mattick, 2011; Bejerano et al., 2004; Benson, Karsch-Mizrachi, Lipman, Ostell, & Wheeler, 2004; Cunningham et al., 2015; Guttman et al., 2009; Mercer, Dinger, Sunkin, Mehler, & Mattick, 2008; Pruitt, Tatusova, & Maglott, 2007; Rinn et al., 2007). Positive probes for housekeeping genes and negative probes are printed onto the array for hybridization quality control. After quantile normalization of the raw data, 24603 unique lncRNAs and 19617 mRNAs were selected for analysis. Firstly, I implemented Principal Component Analysis (PCA) on the log2 transformed intensities and plotted the first 3 components against each other, where lncRNAs and mRNAs showed very comparable results, especially where for the first two components samples from Batch cultures at both days grouped closely compared to Fed-batch samples, which showed a wider separation even between the same condition (Supplementary Figure 3.1). Then I applied hierarchical clustering, arranging samples into groups based on their averaged log-normalized expression levels to show the relationships among gene expression patterns of samples for both lncRNAs and mRNAs (Figure 3.2B). We continued performing differential expression analysis (DE) where I compared Fed-batch against Batch at Day 4, Fed-batch against Batch at Day 7, Fed-batch at Day 7 against Fed-batch at Day 4 and Batch at Day 7 against Batch at Day 4. At this point, the expression of a selected number of the identified DE transcripts was confirmed by RT-qPCR, that showed these transcripts were present but that lower fold-changes were observed between samples than from the array analysis (Supplementary Table 3.1). The genes identified as DE were then filtered based on a threshold fold change (FC) ≥2 and a false discovery rate (FDR) ≤ 0.10 (Supplementary Tables 3.2, 3.3). When I compared Batch Day 7 to Batch Day 4 samples, I found surprisingly low numbers of mRNAs or lncRNAs where the expression changed beyond the set thresholds, with 0 mRNAs being up and 19 down regulated between the two days whilst for the lncRNAs there were no transcripts that changed about the thresholds set (Figure 3.2C). To check whether this lack of identified transcripts changing was due to the FC ≥ 2 threshold, we repeated the analysis for FC ≥ 1.5 and FC ≥ 1.25, FDR ≤ 0.10 (Supplementary

Tables 3.4-3.11). For a FC ≥ 1.25, 99 mRNAs were down and 30 were up-regulated while none met the threshold for the lncRNAs. I then carried on with the FC ≥ 2, FDR ≤ 0.10 thresholds for the remaining comparisons. Fed-batch Day 7 vs Day 4 revealed 693 mRNAs up-regulated and 421 down-regulated whilst for the lncRNAs I found 545 up-regulated and 200 down-regulated genes (Figure 3.2C). When I compared Fed-batch against Batch at Day 4 and Day 7, 1048 and 1875 lncRNAs were up-regulated respectively in addition to 856 down-regulated at day 4 and 1018 down-regulated at day 7. For the same comparisons I saw more down-regulated mRNAs at day 4 (1538) and day 7 (1397) while there were 716 up-regulated mRNAs at day 4 and 1121 at day 7 (Figure 3.2C).

3.3.3 Identification of lncRNAs Differentially Expressed as Potential Engineering Targets for Modulation of Cell Growth

As one of the aims of this work was to identify new transcripts for manipulation in CHO cells, I reduced the list of potential targets to a manageable group for further experimental validation. Firstly, I aligned all the 60 nucleotide mouse probes corresponding to the differentially expressed genes identified for each comparison against the Chinese hamster (CH) genome using the discontiguous megablast algorithm to check how many transcripts had an annotation in CHO. Only 16-28% of the probes corresponding to differentially expressed lncRNAs had a matching transcript in CH, as opposed to 58-80% of the mRNAs (Supplementary Table 3.12). This is likely due to the poor annotation of the CH transcriptome compared to mouse, where the number of described lncRNAs is significantly higher. If we consider the ENSEMBL 91 database release alone, 2563 lncRNAs and 446 pseudogenes are listed for Chinese hamster as opposed to 9308 lncRNAs and 12363 pseudogenes in mouse. To address this issue, I complemented this approach with literature and databases mining (NCBI, ENSEMBL, lncRNAdb, LNCipedia) to identify differentially expressed lncRNAs already described in mouse but not in Chinese hamster, revealing genomic positions that align with potential lncRNAs (Supplementary Figures 3.2-3.4). The identified sequences were then examined using the Rfam database (Kalvari et al., 2018) to assess their resemblance to existing non-coding RNA families. A list of potential lncRNAs found using the described approaches is reported in Table 1, along with literature references describing their biological function.

The role of lncRNAs in diseases is well established, especially linked to cell proliferation in cancer (W. Sun, Yang, Xu, & Guo, 2017) and several of those identified here have been investigated in such systems. NEAT1 (nuclear-enriched abundant transcript 1) and MALAT1

(also known as NEAT2), found to be up-regulated at both days in our Fed-batch cultures (Table 3.1), are among the most well characterized lncRNAs and have been reported to promote cell proliferation through regulation of gene expression at the nuclear level. MALAT1 is a mostly un-spliced transcript around 6.7 kb in mouse, with a long half-life due to the tRNA-like structure adopted at the 3' end (Wilusz, 2016). This structure is cleaved to generate a 61 nt mascRNA (MALAT1-associated small cytoplasmic RNA) exported to the cytoplasm where it is subjected to canonical CCA nucleotides addition and accumulates in the cytoplasm (Wilusz, 2015; Wilusz, Freier, & Spector, 2008). I was able to identify the JH002628.1 genomic region in Chinese hamster using the abovementioned complementary approaches and we could recognize all of its main conserved domains (Supplementary Figure 3.2). The structure of this genomic region is conserved in the most recent CHOK1GS_HDv1 genome and resembles the pattern found in human and mouse, where close to MALAT1 is the NEAT1 locus. The NEAT1 locus is regulated by alternative 3' end processing, where the primary transcript can be cleaved and polyadenylated to generate a 3.2 kb long MENε isoform or cleaved and non-polyadenylated to generate the 20.8 kb long MENβ isoform (Wilusz, 2016). We were able to identify a predicted 3.2 kb long non-coding transcript (NCBI Ref XR_478750.2) arising from a locus (LOC103159497, C_griseus_v1.0) with similar primary structure to both the 3.2 kb and the 20.8 kb isoforms (E-value 0.0, Identity 80%, algorithm megablast). MALAT1 and NEAT1 both localize to the paraspeckles stress-responsive nuclear bodies in the cell (Nakagawa & Hirose, 2012), where they are reported to influence the splicing machinery (W. Sun et al., 2017; Tripathi et al., 2010) and the DNA repair machinery (Adriaens et al., 2016). The Plasmacytoma Variant Translocation 1 (PVT1) lncRNA is considered a biomarker for various cancers due to its ability to promote cell proliferation with a range of proposed mechanisms (Colombo, Farina, Macino, & Paci, 2015a; Tseng & Bagchi, 2015; F. Wang et al., 2014; Shikai Zhu et al., 2017). I identified a predicted non-coding RNA in Chinese hamster (NCBI Ref XR_478426.2) which contained the PVT1_3 RFAM domain and was up-regulated at both Day 4 (FDR = 1.2E-01) and Day 7 in the Fed-batch compared to Batch (Table 3.1).

Table 3.1 - Selected differentially expressed lncRNAs

Gene symbol	NCBI Ref	FC Day 4	FC Day 7	FDR Day 4	FDR Day 7	Rfam Family	Rfam ID	E-value	References
MALAT1	JH002628.1	3.5	4.5	2.4E-02	1.0E-02	MALAT1	RF01871	7.5E-13	*Sun* 2017, *Wilusz* 2016
						mascRNA-menRNA	RF01684	7.0E-16	
MEG3	JH001208.1	2.9	2.8	5.9E-02	4.7E-02	MEG3_2	RF01872	3.5E-14	*He* 2017
MIAT	XM_007625231.2	2.6	-	2.6E+00	1.0E-02	MIAT_exon1	RF01874	7.0E-12	*Liao* 2016
NEAT1	XR_478750.2	6.1	3.3	3.9E-02	6.3E-02	NEAT1_1	RF01955	5.4E-18	*Adriaens* 2016, *Hirose* 2014
						NEAT1_2	RF01956	1.2E-17	
						NEAT1_3	RF01957	2.2E-19	
PVT1	XR_478426.2	2.5	3.5	1.2E-01	2.7E-02	PVT1_3	RF02166	1.7E-16	*Colombo* 2015, *Zhu* 2017
TERC	AF221928.1	-	6.3	-	4.0E-02	Telomerase-vert	RF00024	5.7E-58	*Engreitz* 2016
TUG1	XR_483407.1	-	2.1	-	8.0E-02	TUG1_3	RF01891	1.6E-46	*Li* 2016
						TUG1_1	RF01882	1.3E-21	
						TUG1_4	RF01892	2.1E-20	
						TUG1_2	RF01883	3.2E-17	

Summary of the lncRNAs identified using the described approaches, consisting in a direct alignment of the differentially expressed gene probe against the Chinese hamster genome to literature search and RNA families conservation in RFAM. From left to right, the columns show the gene symbol, the NCBI accession reference, the fold-change measured for Fed-batch vs Batch samples respectively at Day 4 and Day 7, with the corresponding FDRs, the ID of the non-coding RNA family found for the sequence in Rfam database with the corresponding ID and E-value and the references describing the biological function of the gene.

3.3.4 Comparison of mRNAs with Existing Datasets

I compared our Fed-batch against Batch dataset with previous work to identify coding genes related to growth in CHO. Although all these works used different approaches and cell lines, I saw the opportunity to find common patterns of expression in CHO across various conditions. The first dataset compared was Clarke 2011 (Clarke et al., 2011), where I observed 10 of their reported down-regulated genes at Day 4 and 7 at Day 7 of our dataset (Supplementary Table 3.13). On comparing our data to Clarke 2012 (Clarke et al., 2012), only the kinesin family member C1 (KIFC1) gene, up-regulated at both Day 4 and Day 7, was observed to behave in the same way. Lastly, when I considered the translatome analysis in Courtes 2013 I identified 3 genes at Day 4 and 8 genes at Day 7 among our down-regulated transcripts, together with 1 gene at Day 4 and 2 genes at Day 7 among the up-regulated transcripts (Courtes et al., 2013). Interestingly, a group of the common genes between the selected works and our dataset (CDC20, MAD2L1, MCM7, MCM4, GTF2H4) are involved in cell cycle and DNA replication, supporting the findings in the pathway enrichment analysis (described in Section 3.3.5) (Buch et al., 2012; Yu, 2002; Zhai et al., 2017). In addition, I found single genes in consistently enriched pathways in our pathway analysis such as HNRNPC, involved in RNA molecule binding or LGMN, participating in protein degradation in the

lysosome (Dall & Brandstetter, 2016). Interestingly, HNRNPs proteins are well known to bind and mediate the functions of lncRNAs, and HNRNPC in particular has been reported to interact with MALAT1 in a tightly regulated N^6-methyladenosine-dependent manner (N. Liu et al., 2015; X. Sun, Haider Ali, & Moran, 2017).

3.3.5 GO analysis and Pathway Enrichment

I performed GO term analysis on the differentially expressed mRNAs with an FDR cut-off of 0.10 calculated by the BH method, followed by pathway enrichment analysis based on the KEGG database (Supplementary Tables 3.14-3.29), allowing the determination of the significantly enriched biological pathways filtered for an FDR ≤ 0.10 and grouped based on KEGG class annotation (Table 3.2). The pathway enrichment identified 22 pathways containing down-regulated genes at Fed-batch vs Batch Day 4 and 24 at Fed-batch vs Batch Day 7 while only 1 contained up-regulated genes (Lysosome, ID: 04142) at either day. This is most likely a reflection of the lower number of up-regulated genes compared to down-regulated (Figure 3.3). The Metabolism domain includes 7 enriched pathways at Day 4 while only 1 of these was still enriched at Day 7, suggesting a central role of metabolism together with p53 signalling predominantly during the exponential growth phase. On the contrary, towards Day 7 I see the prevalent enrichment of pathways related to translation regulation and RNA interaction at different levels, from transport to splicing. The most evident pattern of enrichment between Day 4 and Day 7 is represented by the Replication and Repair class, where the majority of the pathways involved in genome maintenance and diverse repair mechanisms are consistently enriched, indicating an early and sustained regulation of these genes throughout culture. DNA damage is reported to stimulate the expression of NEAT1 and, together with MALAT1, to promote the formation of paraspeckles, which regulate alternative splicing and promote proliferation (Adriaens & Marine, 2017; Adriaens et al., 2016; Nakagawa & Hirose, 2012). I then compared our results with relevant KEGG pathway enrichment datasets available for CHO (Harreither et al., 2015) and found 4 common pathways, which were enriched exclusively at Day 7 in our dataset: RNA Transport (ID: 03013), mRNA surveillance (ID: 03015), RNA degradation (ID: 03018), Spliceosome (ID: 03040). Taken together, these results suggest the importance of cell cycle and genome repair mechanism control likely due to the high proliferation of the fed-batch system. In addition, the active regulation of RNA transport, RNA maintenance and splicing seem to be particularly important towards the later stages of our Fed-batch cultures.

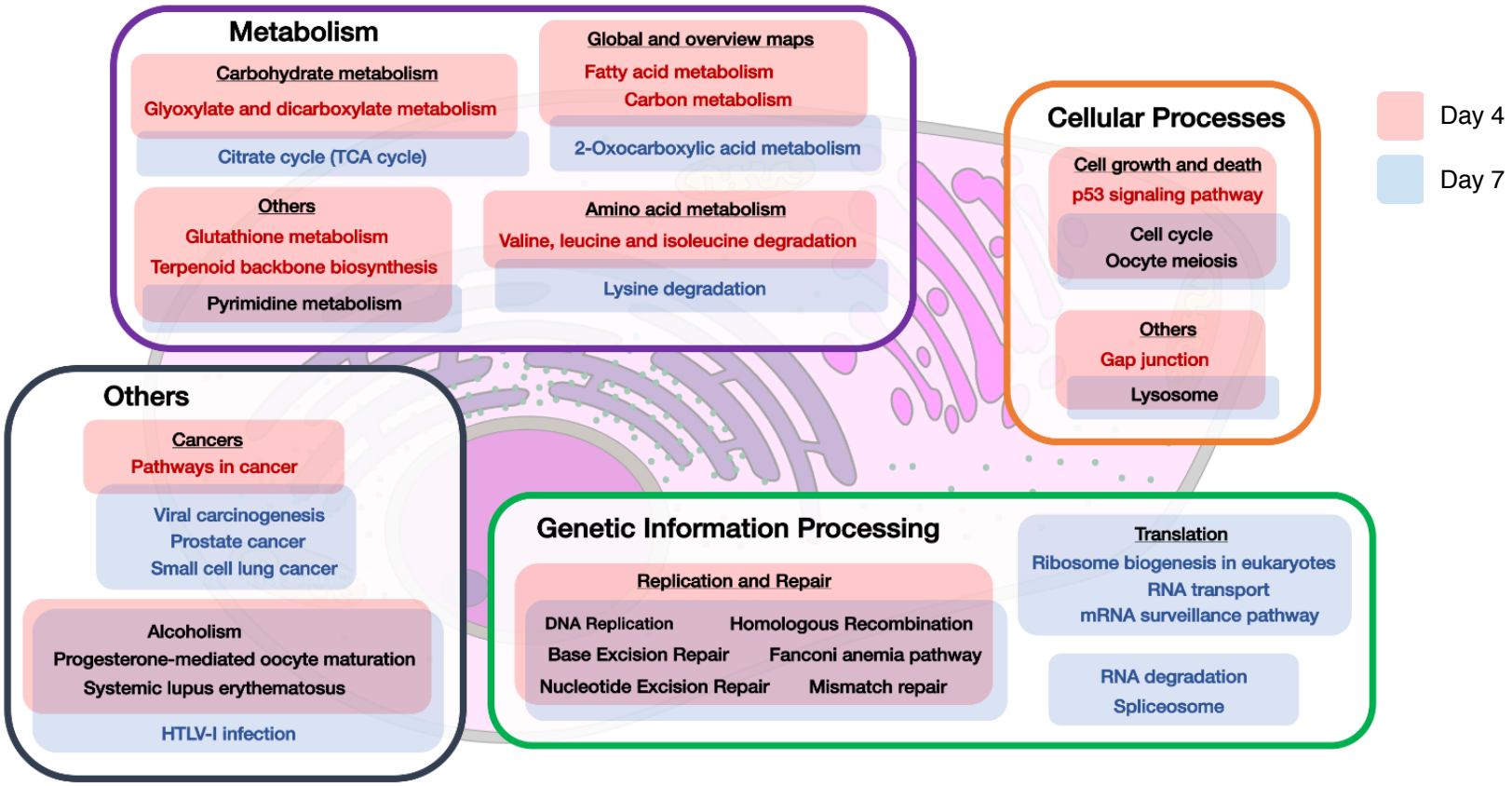


Figure 3.3 - Graphical summary of the enriched KEGG pathways based on differentially expressed genes

Graphical representation of the enriched KEGG pathways listed in (Table 3.2), hierarchically grouped based on KEGG Pathway Maps. Each filled rectangle (in red for Day 4, in blue for Day 7) contains the corresponding enriched pathways, with pathways enriched at both Day 4 and Day 7 enclosed in the overlaps between the filled rectangles. The enrichment is based on genes differentially expressed in Fed-batch vs Batch at Day 4 and Day 7 with an FDR ≤ 0.10.

Table 3.2 - Enriched KEGG pathways based on differentially expressed genes

Enriched pathways

KEGG ID	Pathway	FDR	Diff. Expressed (DE) genes	Total Genes (TG)	DE:TG ratio
04142	**Lysosome**	6.3E-02	14	124	0.11

KEGG ID	Pathway	FDR	Diff. Expressed (DE) genes	Total Genes (TG)	DE:TG ratio
05034	**Alcoholism**	8.5E-11	48	203	0.24
04110	**Cell cycle**	2.3E-09	34	125	0.27
05322	**Systemic lupus erythematosus**	6.8E-07	33	147	0.22
03030	**DNA replication**	3.5E-04	12	35	0.34
03460	**Fanconi anemia pathway**	8.5E-04	14	51	0.27
05203	**Viral carcinogenesis**	1.1E-03	37	243	0.15
03410	**Base excision repair**	1.3E-03	11	35	0.31
03430	**Mismatch repair**	4.4E-03	8	22	0.36
01200	**Carbon metabolism**	4.5E-03	21	117	0.18
00480	**Glutathione metabolism**	5.7E-03	13	56	0.23
00900	**Terpenoid backbone biosynthesis**	6.4E-03	8	24	0.33
04114	**Oocyte meiosis**	8.5E-03	20	116	0.17
05200	**Pathways in cancer**	1.3E-02	48	397	0.12
03440	**Homologous recombination**	1.6E-02	8	28	0.29
04914	**Progesterone-mediated oocyte maturation**	1.8E-02	16	90	0.18
01212	**Fatty acid metabolism**	2.1E-02	11	51	0.22
04540	**Gap junction**	2.8E-02	15	86	0.17
00280	**Valine, leucine and isoleucine degradation**	4.2E-02	11	56	0.20
04115	**p53 signaling pathway**	6.4E-02	12	68	0.18
00240	**Pyrimidine metabolism**	6.4E-02	16	104	0.15
03420	**Nucleotide excision repair**	6.5E-02	9	44	0.20
00630	**Glyoxylate and dicarboxylate metabolism**	7.6E-02	7	30	0.23

Enriched pathways

KEGG ID	Pathway	FDR	Diff. Expressed (DE) genes	Total Genes (TG)	DE:TG ratio
04142	**Lysosome**	7.1E-04	21	124	0.17

KEGG ID	Pathway	FDR	Diff. Expressed (DE) genes	Total Genes (TG)	DE:TG ratio
04110	**Cell cycle**	7.9E-14	38	125	0.30
03030	**DNA replication**	1.2E-09	17	35	0.49
03013	**RNA transport**	1.2E-09	38	170	0.22
03040	**Spliceosome**	1.3E-07	30	134	0.22
05203	**Viral carcinogenesis**	2.8E-06	40	243	0.16
03430	**Mismatch repair**	2.2E-05	10	22	0.45
03460	**Fanconi anemia pathway**	2.3E-05	15	51	0.29
03410	**Base excision repair**	3.3E-04	11	35	0.31
05322	**Systemic lupus erythematosus**	9.3E-04	24	147	0.16
03440	**Homologous recombination**	1.4E-03	9	28	0.32
03420	**Nucleotide excision repair**	2.5E-03	11	44	0.25
05034	**Alcoholism**	3.5E-03	28	203	0.14
04114	**Oocyte meiosis**	4.2E-03	19	116	0.16
03008	**Ribosome biogenesis in eukaryotes**	2.0E-02	14	83	0.17
00240	**Pyrimidine metabolism**	2.3E-02	16	104	0.15
03015	**mRNA surveillance pathway**	2.6E-02	15	96	0.16
00310	**Lysine degradation**	3.1E-02	10	52	0.19
04914	**Progesterone-mediated oocyte maturation**	3.4E-02	14	90	0.16
03018	**RNA degradation**	4.4E-02	13	83	0.16
05166	**HTLV-I infection**	4.8E-02	32	294	0.11
05222	**Small cell lung cancer**	4.9E-02	13	85	0.15
00020	**Citrate cycle (TCA cycle)**	5.5E-02	7	32	0.22
05215	**Prostate cancer**	6.7E-02	13	89	0.15
01210	**2-Oxocarboxylic acid metabolism**	1.0E-01	5	20	0.25

Summary of the enriched pathways based on the list of up-regulated (lysosome) and down-regulated (lower tables) transcripts in Fed-batch vs Batch comparison at Day 4 (left panel) and Day 7 (right panel). The columns show from left to right the ID from KEGG, the pathway name, the FDR associated with the enrichment (FDR ≤ 0.10 threshold), the number of differentially expressed genes in the pathway, the number of total genes listed in the pathway, the differentially expressed genes over total genes ratio.

3.4 DISCUSSION

By taking advantage of a comprehensive commercially available mouse microarray containing 35923 lncRNAs and 24881 mRNAs, I have provided the first mapping of the CHO lncRNA landscape, together with the coding transcriptome. Previous reports have shown that more than 70% of the assembled CHO transcriptome is similar to mouse (*Mus musculus*) and closely related to rat (*Rattus norvegicus*) transcriptomes (Becker et al., 2011) suggesting that this approach was likely to be valid for lncRNAs as well. Due to the species and tissue-specificity of lncRNAs compared to mRNAs, the number of detectable lncRNAs in CHO is likely to be lower than the 35923 probes included in the array. Nevertheless, using this approach I was able to detect 24603 lncRNAs (68.5% of the total probes) and 19617 mRNAs (78.8% of the total probes), and found that several hundreds of lncRNAs exhibit changing expression profiles on different days of culture and between Batch and Fed-batch culture in a model CHO-S system. This was especially true for the Fed system, where comparing Day 4 and Day 7 we observed 1114 differentially expressed mRNAs and 745 lncRNAs, as opposed to the Batch, were we saw only 19 differentially expressed mRNAs and no lncRNA for the same comparison, suggesting a prevalent variability induced by the Fed supplement as compared to time only.

Among the differentially expressed genes, between 16-28% of the lncRNAs probes had a matching transcript in CH, as opposed to 58-80% of the mRNAs. This required a specific approach where the Chinese hamster genome, literature search and databases mining were combined to detect lncRNAs differentially expressed in my system with an established biological function. Within these lncRNAs, we focused on MALAT1, NEAT1 and PVT1 to provide a comparison between the mouse gene and the Chinese hamster putative homologues. The number of lncRNAs with a fully understood role in the cell remains small, however these three non-coding genes are among the most well characterized (Kopp & Mendell, 2018). MALAT1 and NEAT1 in particular are associated with increased proliferation and participate in the regulation of alternative splicing and DNA repair, which I found to be strongly regulated in our pathway enrichment analysis. This suggests a potential role for these lncRNAs in CHO, although further experimental studies on the single genes are now required to assess the actual effects on the cell under different conditions.

Several approaches to investigate and confirm the functional annotation of lncRNAs in other organisms have been described, including the perturbation of lncRNA expression by overexpression, knockout or knockdown (S. J. Liu et al., 2016; Shiyou Zhu et al., 2016), in

addition to complementary strategies (Kashi et al., 2016). Future developments in CHO will have to proceed with a mix of functional prediction tools to assess the properties of the transcriptome and evaluate the degree of conservation with other species (Iwakiri, Hamada, & Asai, 2015; Signal, Gloss, & Dinger, 2016; Ulitsky, 2016) and of targeted approaches to ameliorate the annotation and propose mechanisms of action for the single transcripts (Cabili et al., 2015; Clark et al., 2015; Lai, Blumenthal, & Shiekhattar, 2016; Mercer et al., 2014). Further, the data reported here is for both lncRNAs and mRNAs and hence will allow investigators to further probe the relationships between the expression and regulation of these two classes of RNA. As the majority of the lncRNAs reported in the literature are discussed and related to human or model organism systems, our work aimed at unveiling the role of lncRNAs in CHO under industrially relevant conditions to identify new targets for manipulation to sustain proliferation. Examples of successful cell engineering of lncRNAs to selectively enhance translation (Zucchelli, Patrucco, Persichetti, Gustincich, & Cotella, 2016) and product yield (Tabuchi, 2013) have already been reported in CHO, demonstrating the potential of manipulation of lncRNAs for enhancing industrial processes. Moreover, since it was reported up to 15% of the total ribosome occupancy can be occupied by a single recombinant mRNA, the intrinsic characteristics of lncRNAs place them as ideal candidates for cell line engineering of protein production cell factories, as they do not add any translational burden on top of the coding gene of interest (Kallehauge et al., 2017).

My work has identified potential lncRNA targets differentially expressed in Fed-batch compared to Batch culture from which we selected a group of molecules to be experimentally studied (Table 3.1). In addition to the expression of lncRNAs I also looked at expression of mRNAs (coding transcripts) and found a consistent change in differentially expressed mRNAs when comparing batch and fed-batch cultures. Pathway enrichment analysis (Figure 3.3) underlined the importance of genes involved in cell cycle and genome maintenance pathways along with the regulation of lysosome formation as potential targets for cell engineering to enhance proliferation. My approach allowed the identification of previously undescribed lncRNAs in CHO along with mRNAs to identify the connections between them and compared these with existent literature. This network of reciprocal interactions is beginning to be unveiled in other organisms (Han & Chang, 2015; Kornienko, Guenzl, Barlow, & Pauler, 2013; Mallory & Shkumatava, 2015) and my work will help pave the way for the definition of new layers of regulation involving single transcripts or even entire pathways in CHO.

3.5 CONCLUSIONS

He I report on the lncRNA landscape and how this changes in CHO cells, presenting a full dataset of those lncRNAs present as determined from an array study and how these change through a Batch and Fed-batch culture and between the two culture systems. From analysis of the data, I have determined those lncRNAs whose expression changes the most between 2 days in culture and between fed and batch culture that are attractive targets for cell engineering. This resource will now provide the community with the opportunity to undertake functional validation studies by undertaking single or multiple knock downs/outs, or by the up-regulation of target lncRNAs, and determine the impact on growth, and productivity, characteristics of CHO cells. Ultimately, I anticipate such a resource will be incorporated into wider genome analysis datasets including coding mRNAs and other non-coding RNAs to develop a wider appreciation of the role of RNAs in controlling recombinant CHO cell line growth and productivity characteristics.

Chapter 4 Defining lncRNAs that Underpin CHO Cell Growth and IgG Productivity by RNA-Seq

In Chapter 3 is described the use of a mouse microarray to obtain a transcriptomic landscape of expression in CHO cells and the identification of candidate lncRNAs for cell engineering. Based on this work, we further explored lncRNAs expression in six CHO cell lines in collaboration with an industrial partner, Symphogen. Small-scale bioreactors used in Chinese hamster ovary (CHO) cell line development allow transcriptomic studies on multiple cell lines. Here I define the CHO cell long non-coding RNA (lncRNA) transcriptome from cells grown in controlled miniature bioreactors under fed-batch conditions using RNA-Seq to identify lncRNAs and how the expression of these changes throughout growth and between IgG producers. I identify lncRNAs associated with productivity and growth characteristics, finding that *Adapt15*, linked to ER stress, *GAS5*, linked to mTOR signalling/growth arrest, and *PVT1,* linked to Myc expression, are differentially regulated during fed-batch culture and whose expression relates to productivity (Adapt15) or growth (*GAS5*, *PVT1*). Changes in (non)-coding RNA expression between the seed train and the equivalent day of fed-batch culture are also reported, showing large differences in gene expression between these. Collectively, I present a comprehensive lncRNA CHO cell profiling and identify targets for engineering growth and productivity characteristics of CHO cells.

This work has been submitted to the scientific journal iScience as a research manuscript titled:

"Defining lncRNAs that Underpin CHO Cell Growth and IgG Productivity by RNAseq", Vito D, Eriksen JC, Skjødt C, Weilguny D, Rasmussen SK, Smales CM;

and is currently under revision after initial review at the time of submitting this thesis.

4.1 Introduction

Many recombinant protein biopharmaceuticals are expressed in mammalian expression systems due to the ability of such systems to correctly fold, assemble, and undertake 'human-like' post-translational modifications and secrete the target protein out of the cell (Walsh, 2010). Of mammalian cell expression systems one predominates, with more than 60% of mammalian made biotherapeutic proteins produced from cultured Chinese hamster ovary (CHO) cells (Kunert & Reinhart, 2016; Leu, Dumont, Hafey, Murphy, & George, 2004; Mead et al., 2015; Povey et al., 2014; Walsh, 2010). Fed-batch culture is currently the most common bioprocess used for the industrial production of proteins in CHO cells, generating increased cell concentrations (and hence biomass) and sustained culture viability compared to batch culture, ultimately resulting in higher productivity and final product yields (Durocher & Butler, 2009; X. Pan et al., 2017; Wong et al., 2006). The introduction of small-scale parallel bioreactors allowing automated sampling and continuous control of fundamental culture parameters, including pH, stirring and temperature has enhanced the ability to screen a wider range of culture parameters and cell lines leading to improved upstream development timelines and experimental throughput (Bareither & Pollard, 2011).

The ambr15™ cell culture system (Sartorius Stedim Biotech) has been shown to give similar cell growth and productivity data to those achieved in larger scale stirred bioreactors, enabling more accurate predictions compared to shake flasks on the behaviour of a cell line at larger scale (Alsayyari et al., 2018; V. Janakiraman, Kwiatkowski, Kshirsagar, Ryll, & Huang, 2015; Nienow et al., 2013; Rouiller et al., 2016). This capacity to conduct small scale experiments under controlled conditions, of a highly predictive nature at larger scale, allows the investigation of the behaviour of different cell lines under alternative feeding regimes to determine how each respond. Indeed, recent reports state that the ambr15™ small-scale automated and controlled bioreactor system provides an excellent scale down model to facilitate studies on multiple cell lines under controlled industrially relevant conditions to identify robust targets linked to productivity for cell engineering and material and data for future regulatory submissions (Sandner et al., 2018).

Despite advancements in the ability of CHO cells to reach higher cell concentrations and generate increasing amounts of target biotherapeutic proteins, particularly monoclonal antibodies (mAbs), there remains a desire to further understand the limitations upon CHO cell phenotypes and to engineer cells for the production of more difficult to express products (Godfrey et al., 2017; Jossé et al., 2018; Mead, Chiverton, Smales, & Von Haar, 2009). One

approach that has been applied towards improving our understanding of the limitations on CHO cell growth and recombinant protein production is the field of transcriptomics (Tamošaitis & Smales, 2018). Transcriptomic studies in particular could benefit from generating material from controlled miniature bioreactors that predict behaviour at larger scale, as an issue of such previous studies is reproducibility and robustness across different transcriptomic datasets, given the high heterogeneity of CHO cell lines and their intrinsic genetic instability (Chen, Le, & Goudar, 2017; F. M. Wurm, 2013; F. Wurm & Wurm, 2017). The availability of CHO cell and Chinese hamster genome sequences (Lewis et al., 2013; X Xu et al., 2011) has greatly enabled omics-based studies (Faustrup Kildegaard, Baycin-Hizal, Lewis, & Betenbaugh, 2013), and since there has been an increasing number of publicly available databases for different CHO cell lines. However, the focus of these studies has been on either coding genes or microRNAs (miRNAs) with few studies investigating other classes of RNAs and their impact on CHO cell behaviour (Singh, Kildegaard, & Andersen, 2018; Tamošaitis & Smales, 2018).

Here I investigate the long non-coding transcriptome in CHO cells during fed-batch culture under controlled bioreactor conditions. Since the unravelling of multiple organisms genomes, particularly eukaryotic genomes, associated with the development of high throughput sequencing technologies, new classes of non-coding RNA have been identified (Djebali et al., 2012). Among these, a class of transcripts known as long non-coding RNAs (lncRNAs) was identified. LncRNAs are defined as transcripts longer than 200 nucleotides that lack a significant open reading frame (ORF), are usually transcribed by RNA polymerase II, and are spliced with, or without, 3' polyadenylation (Kashi et al., 2016; Kung et al., 2013; Wilusz, 2016). In the nucleus, *cis*-acting lncRNAs regulate the chromatin state and transcription of nearby genes while *trans*-acting lncRNAs can recruit RNA binding proteins to form chromatin modifying complexes, modulate splicing or organise functional nuclear domains (Kopp & Mendell, 2018). When transported to the cytoplasm, lncRNAs act at a post-transcriptional level by promoting specific mRNA translation or turnover and by competitively binding microRNAs (miRNAs), attenuating the repression of target genes (Geisler & Coller, 2013). The wide range of processes involving lncRNAs suggests that some of these may be potential cell engineering targets to rewire CHO cell phenotypes for enhanced cell growth and/or recombinant protein production and quality without placing a translational burden on the cell compared to overexpression of coding genes. However, the majority of our knowledge around lncRNAs comes from studies in model organisms related to disease and development (Perry & Ulitsky, 2016; Schmitt & Chang, 2016) with lncRNAs

poorly annotated in the CHO cell genome and little known about their role in defining CHO cell phenotypes.

A recent report described the first study into the lncRNA landscape in CHO cells, showing regulated expression of thousands of lncRNAs under batch and fed-batch conditions over time (Vito et al 2018). Others have demonstrated the potential power of lncRNA cell engineering to manipulate the cells ability to produce target recombinant proteins with the first reports of engineering of a class of lncRNAs named SINEUPs to enhance the translation of specific target mRNAs in various mammalian cell factories (Patrucco et al., 2015; Zucchelli, Patrucco, Persichetti, Gustincich, & Cotella, 2016). However, the limited number of studies and poor annotation of non-coding regions in the Chinese hamster genome means that transcriptomics across multiple cell lines associated with phenotypes of interest under industrially relevant and controlled conditions is required to identify lncRNAs whose manipulation may enhance mammalian cell factories ability to generate secreted target products (Vishwanathan, Le, Le, & Hu, 2014). Here I present a comprehensive coding and non-coding, particularly lncRNAs, transcriptome analysis using RNA-Seq of 5 IgG1 producing CHO cell lines and 1 stable pool harbouring the plasmid cassette without genes encoding for the IgG1, cultivated under fed-batch conditions in an ambr15™ system to unveil regulated lncRNA targets for cell engineering. The RNA-Seq datasets and analyses are made openly available to the community to promote further studies and comparisons and provide a first CHO cell lncRNA transcriptomic resource.

4.2 Results and Discussion

I set out to provide the first analysis of the long noncoding RNA (lncRNA) transcriptome using RNA-Seq in CHO cells producing three different model monoclonal antibodies during fed-batch culture, defining those lncRNAs expressed in CHO cells and the flux of these during culture and between cell lines. To do this, I undertook RNA-Seq analysis on a panel of IgG expressing CHO DHFR cell lines, sampling throughout fed-batch cultures in an ambr15™ microbioreactor system generating profiles of the flux of coding RNAs and lncRNAs.

4.2.1 Analysis of fed-batch culture samples

The DAVI dataset included the 3068, 3080 and 3077 IgG1 producing cell lines and the null pool 3478 (Summarised in Table 4.1). These cell lines showed comparable culture viability, maintaining a culture viability >80% throughout the 12 days of fed-batch culture. However,

there was a diverse range of viable cell concentrations achieved across the cell lines with cell line 3077 reaching the highest maximum viable cell concentration of 14.48 x 10^6 viable cells/mL with the other cell lines reaching lower concentrations of 10.66 x 10^6 (for 3478), 9.38 x 10^6 (for 3068) and 8.56 x 10^6 (for 3080, Figure 4.1A). The overall product titres and Qp for the producer cell lines at Day 12 were 1.96 g/L and 23.05 pg/cell/day (3077), 1.48 g/L and 26.53 pg/cell/day (3068), and 1.80 g/L and 31.71 pg/cell/day (3080) respectively (Figure 4.1B,D). The glutamate, glutamine and ammonia profiles were similar across all cell lines (Supplementary Figure 4.1), while lactate accumulated during the first few days of culture, with a peak concentration of 2.2 g/L for 3080 at Day 5 and 3478 at Day 12 and <1.5 g/L for 3068 and 3077 (Figure 4.1C).

Table 4.1 - Symphogen cell lines used in DAVI and JCE experiments

DAVI Experiment				
Cell line	Clonality	Peak VCD [viable cells/day]	Yield [g/l]	Qp [pg/cell/day]
3478	Pool	10.66 x 10^6	-	-
3068	Clone	9.38 x 10^6	1.48	26.53
3077	Clone	14.48 x 10^6	1.96	23.05
3080	Clone	8.56 x 10^6	1.80	31.71
JCE Experiment				
Cell line	Clonality	Peak VCD [viable cells/day]	Yield [g/l]	Qp [pg/cell/day]
3068	Clone	8.52 x 10^6	2.64	35.20
3080	Clone	6.94 x 10^6	2.25	35.81
4384	Clone	11.01 x 10^6	2.43	23.39
3936	Pool	10.67 x 10^6	1.23	13.25

Cell line detail/name, clonality, peak viable cell number (VCD), the IgG yield at Day 12 for DAVI and Day 14 for JCE and the cell specific productivity for the model IgG1 expressing cells used in the experiments described in the study.

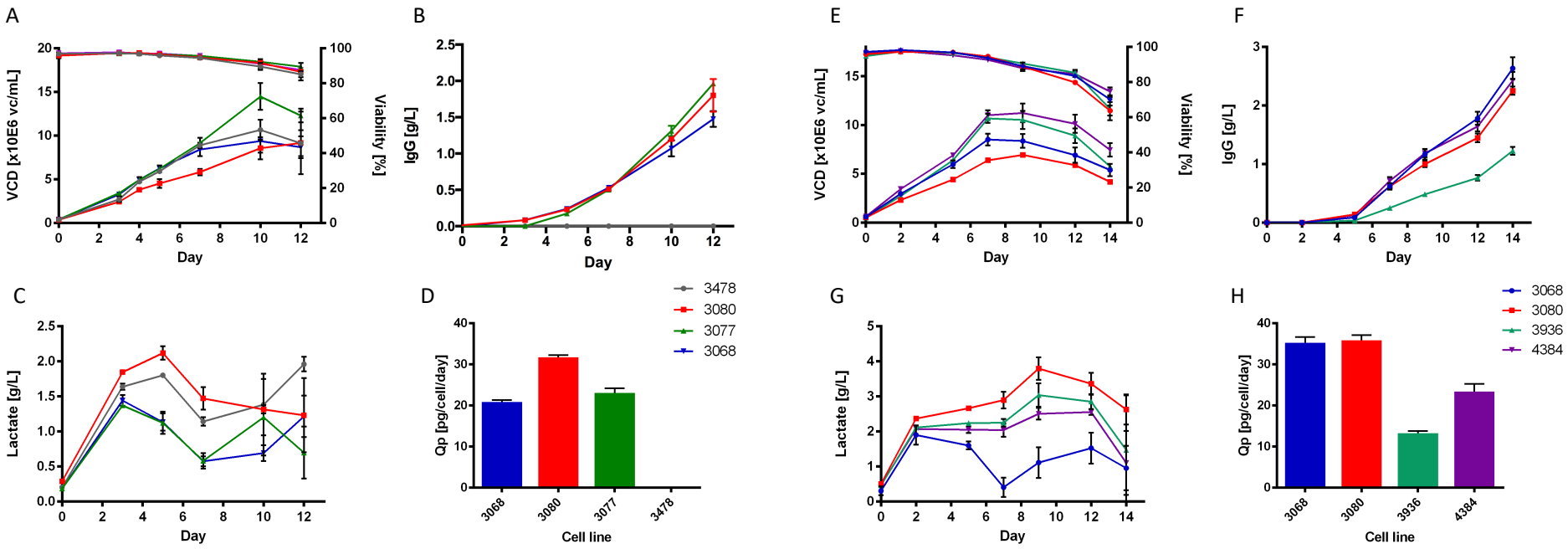


Figure 4.1 - Cell culture parameters for the DAVI and JCE datasets

The parameters measured and monitored during fed-batch culture of model IgG1 expressing CHO cell clones and a null pool for the DAVI dataset (**A, B, C, D**) and the JCE dataset (**E,F,G,H**). **A |E |** viable cell density (VCD) and viability over time. **B |F |** yield of IgG1 antibody over time. **C |G |** Lactate concentration over time. **D |H |** productivity (Qp) for each cell line.

The JCE dataset included the 3068, 3080 and 4384 IgG1 producing cell lines and the 3936 IgG1 producing pool (Table 4.1). The 3068 cell line reached a maximum viable cell concentration of $8.52x10^6$ viable cells/mL, while the 3080 cell line obtained a maximum viable cell concentration of $6.94x10^6$ viable cells/mL. The 4384 cell line reached the highest maximum viable cell concentration of the cells used in this experiment of $11.01x10^6$ viable cells/mL followed closely by the 3936 cell line at $10.67x10^6$ viable cells/mL (Figure 4.1E). The culture viability profiles were broadly comparable across the cell lines, being within the range of 60-80% on the last day of culture (Figure 4.1E). The highest IgG titre reached at the end of culture on Day 14 was achieved by the 3068 cell line of 2.64 g/L, followed closely by 2.43 g/L and 2.25 g/L achieved by the 4384 and 3080 cell lines respectively. The 3936 cell line generated the lowest titre of only 1.23 g/L, as would be expected from a pool of clones (Figure 4.1F). When the Qp was calculated, the 3068 and 3080 cell lines had cell specific productivities of 35.20 and 35.81 pg/cell/day respectively, while the 4384 cell line had a Qp of 23.39 pg/cell/day and 3936 a Qp of 13.25 pg/cell/day (Figure 4.1H). The glutamate, glutamine and ammonia profiles were broadly similar across the cell lines while lactate never accumulated over 2 g/L for 3068 (Figure 4.1), while the lactate in the culture supernatant for the other cell lines fluctuated above this value, with a peak at approximately 3.5 g/L for the 3080 cell line (Figure 4.1G).

4.2.2 RNA sequencing of ambr15™ generated samples and subsequent analysis of the data: The DAVI experiment

As described in the methods section, cell pellet samples in biological duplicates for RNA sequencing were collected in duplicate at Day 4 and Day 12 of fed-batch culture from the ambr15™ reactors. Clustering analysis of the RNA-Seq data revealed that the samples showed a consistent hierarchical clustering for the biological replicates and an evident separation between the two time-points based on gene expression (Figure 4.2A, B). This separation was confirmed by Principal Component Analysis (PCA), however an additional layer of clustering emerged among the producers upon PCA with a difference in Qp, clearly showing clusters formed of 3077 and 3068 (23.05 and 26.53 pg/cell/day), that were distanced from cell line 3080 (31.71 pg/cell/day) and the null-pool 3478 (Figure 4.2C). Overall, this preliminary cluster analysis based on total gene expression suggests a grouping of cell lines directly related to differences in Qp more than maximum viable cell concentrations for this dataset. Differential transcript expression (DE) analysis was then conducted using the DESeq2 R software package, setting a fold change (FC) threshold of 1.50. DE genes were considered significant if the adjusted *p*-value for this FC threshold calculated

in DESeq2 using the Benjamini-Hochberg method was below 0.10. Comparing DE within the same cell line between Days 12 and Day 4 identified the highest numbers of DE genes, with the producing clones 3068, 3077 and 3080 having 2295, 2538 and 2255 DE genes respectively, while there was a much higher number of DE genes identified between these two days in the null-pool 3478 (4464 DE genes) (Figure 4.3). When comparing the gene expression between the different producers on the same culture day, a comparison of cell line 3077 against 3068 revealed 224 (Day 4) and 281 (Day 12) DE genes, while a comparison of cell line 3077 against 3080 revealed 883 (Day 12) and 604 (Day 4) and for 3080 against 3068 revealed 813 (Day 12) and 570 (Day 4) (Figure 4.3). The number of identified DE genes using the criteria outlined when comparing the producer cell lines to the null-pool 3478 was much larger than when comparing the producers to each other. As such, when comparing the null to cell line 3068 there were 371 (Day 4) and 1653 (Day 12) identified DE genes, whilst for cell line 3077 there were 509 (Day 4) and 2255 (Day 12) and for cell line 3080 there were 491 (Day 4) and 813 (Day 12) (Figure 4.3). Within the identified DE genes, lncRNAs made up 10-30% of the total number identified (Figure 4.3). A representative group of differentially expressed coding and non-coding RNAs were then selected (see Section 4.2.6) for RT-qPCR validation, resulting in a positive correlation between the fold changes measured by RNA-Seq and by RT-qPCR (results summarised in Supplementary Material 4.2, primers listed in Appendix).

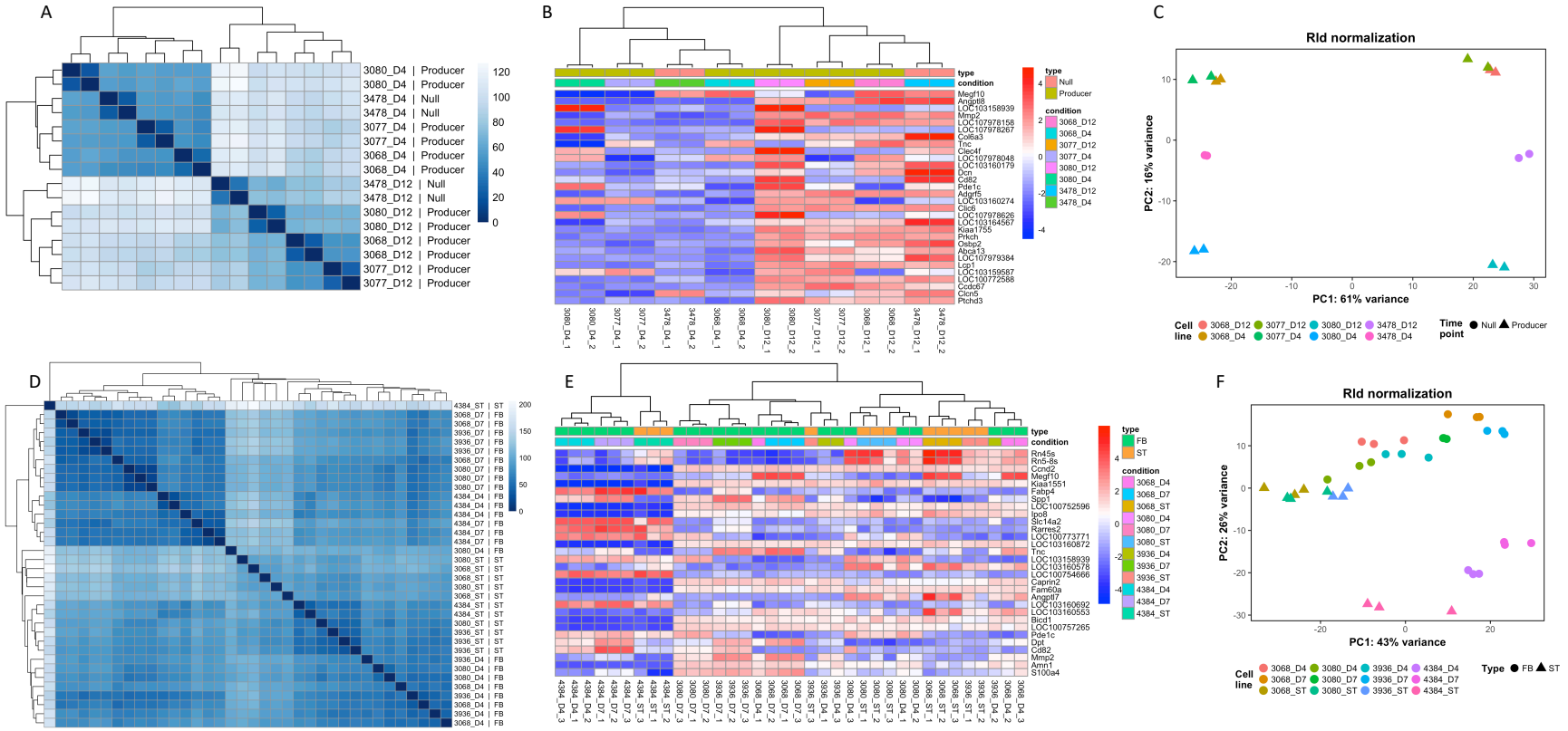


Figure 4.2 - Clustering and PCA analysis based on gene expression

Coding and non-coding gene expression measured with RNA-Seq. From left to right, clustering based on the distance between each sample, clustering based on the top 30 most differentially expressed genes expression and PCA of normalized gene expression for the first two principal components. Figures A to C show the DAVI dataset. The type indicates whether the samples are from the non-producing pool (Null) or from producing clones (Producer) while the condition groups replicates of the same cell line and time point. Figures D to F show the JCE dataset. The type indicates whether the samples are from fed-batch culture (FB) or from the seed-train (ST) while the condition groups replicates of same cell line and time point.

Table 4.2 - Identified potential lncRNA genes for cell engineering

Gene ID	RefSeq	lncRNA	Expression in DAVI	Expression in JCE	RFAM	RFAM ID	Coding probability	Mouse homologue	BLAST E-value	Function
100689050	NR_045124.1	Adapt15	3068 D12 v D4, -1.80 3080 D12 v D4, 3.36 3478 D12 v D4, -5.21 D12 3068 v 3478, 2.86 D12 3077 v 3478, 3.36 D12 3080 v 3478, 1.88	-	-	-	0.023	NR_040384.1	5e-11	linked to oxidative lipotoxicity, resulting in ER stress and cell death
103158913	XR_478428.1	GAS5	3077 D12 v D4, -1.98 3478 D12 v D4, -2.47 D12 3068 v 3478, 2.17	-	SNORD44 SNORD78	RF00287 RF00592	0.023	NR_002840.2	1e-65	snoRNA host gene tumour suppressor
103158906	XR_478426.2	PVT1	3068 D12 v D4, 1.91 3080 D12 v D4, 3.59 D4 3077 v 3478, 2.70 D12 3068 v 3478, 2.71 D12 3077 v 3478, 3.59 D12 3080 v 3478, 2.13	D4 3068 v 3936, 2.53 D4 3936 v 3080, -2.63 D4 4384 v 3080, -2.04 D7 3068 v 3936, 2.53	PVT1_3	RF02166	0.029	NR_003368.2	2e-53	Oncogene, interacts with miR200 family and Myc

Three identified potential lncRNAs targets for cell engineering with an established function in the literature and a homologue in mouse. From left to right, the table indicates the GeneID and NCBI RefSeq accessions, the gene name, the statistically significant differential expression in DAVI and JCE datasets, the RFAM secondary structure family and accession numbers, the coding probability measured in the Coding Potential Calculator 2 (CPC2) (Kang et al., 2017), the mouse homologue transcript with the corresponding E-value obtained using the disc megablast algorithm and a summary of the biological function.

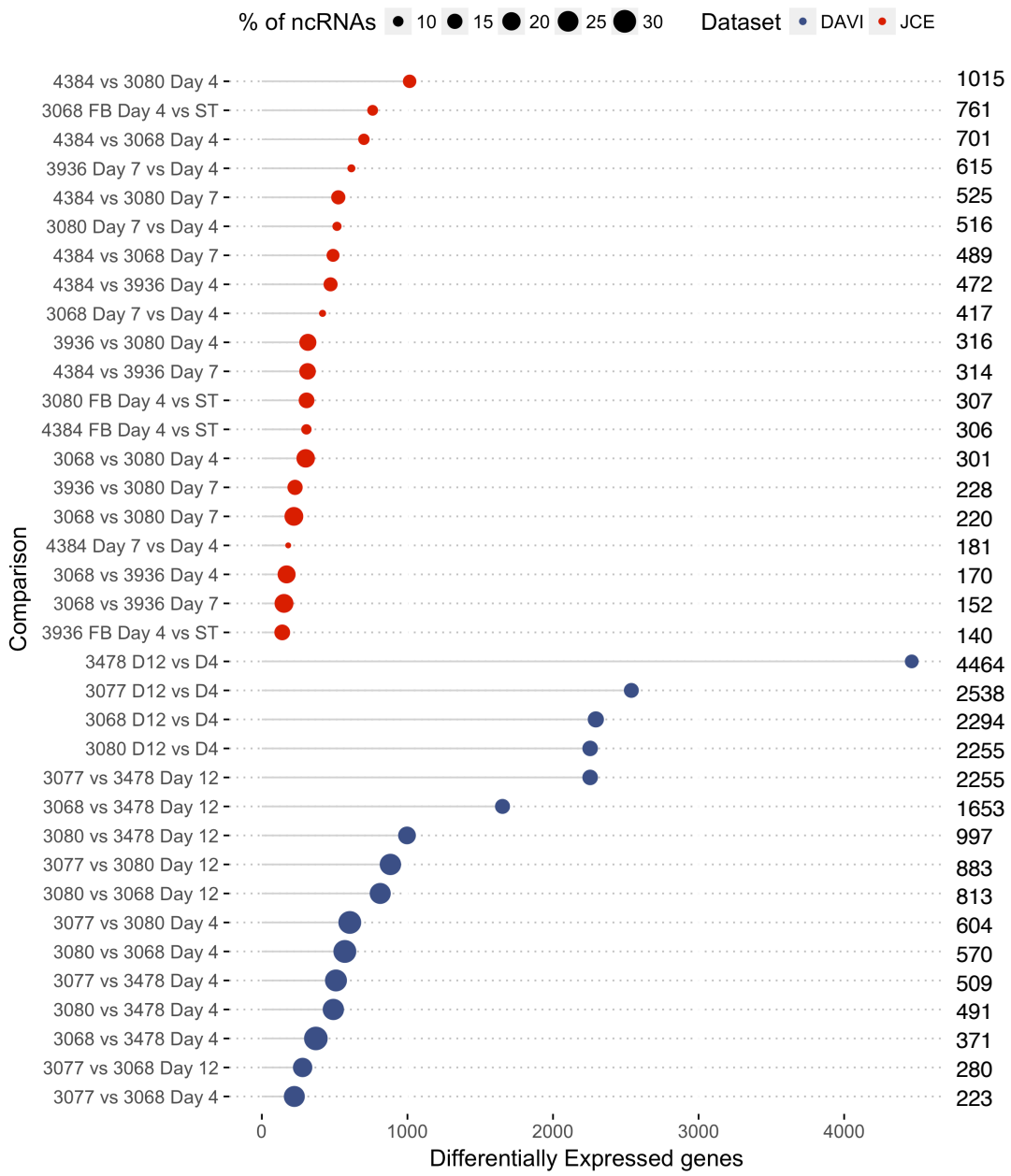


Figure 4.3 - Differentially expressed genes in the DAVI and JCE datasets

The number of differentially expressed genes with a FC ≥ 1.5, *adj* p-value < 0.1. In blue is shown the DAVI dataset and in red is shown the JCE dataset while the size of the dot indicates the % of lncRNAs for each comparison. For each comparison, the exact number of genes is indicated on the right.

4.2.3 RNA sequencing of ambr15™ generated samples and subsequent analysis of the data: The JCE Experiment

In the JCE experiment, samples for RNA sequencing were collected in biological triplicate from the seed train (ST) flasks and at Day 4 and Day 7 of fed-batch culture. The samples showed a hierarchical clustering for each biological triplicate but the separation into groups

as observed in the DAVI experiment was not as evident (Figure 4.2D, Figure 4.2E). PCA revealed a similar pattern where only a cluster composed of the 4384 cell line samples was distinguishable from the other samples with the other clones not grouping separately, with the exception of the seed-train samples, based upon analysis of the complete RNA-Seq datasets (Figure 4.2F). The clustering of the 4384 cell line samples as different from the others (as observed by PCA) was confirmed by the high number of DE genes identified when comparing this cell line (4384) against the others on Day 4 (vs 3068, 701 DE genes; vs 3080, 1015 DE genes; vs 3936, 472 DE genes). When the other cell lines were compared to each other on Day 4 the number of identified DE genes was much lower than when comparing to the 4384 cell line, indicating the gene expression profiles of these cell lines was closer than to that of 4384 (3068 vs 3080, 301 DE genes; 3068 vs 3936,170 DE genes; 3936 vs 3080, 316 DE genes). The number of DE genes identified between the cell lines was much lower when the profiles at Day 7 were compared (220 DE genes for 3068 vs 3080; 152 DE genes for 3068 vs 3936; 228 DE genes for 3936 vs 3080; 489 DE genes for 4384 vs 3068; 525 DE genes for 4384 vs 3080, 314 DE genes for 4384 vs 3936; see Figure 3).

When comparing changes in gene expression within cell lines between days 4 and 7 of fed-batch culture there was a large difference in the total number of DE genes. When comparing Day 7 against Day 4 of culture there were 181 DE genes identified in cell line 4384 whereas in cell line 3068 there were 417 DE genes identified, for cell line 3080 there were 516 DE genes and for cell line 3936 there were 615 DE genes. I then compared the gene expression profiles of the seed-train cultures that were used to start the fed-batch process, obtained from cells during logarithmic growth phase, to the Day 4 gene expression profiles of the fed-batch culture experiments. Although cells from the seed train and fed-batch day 4 ambr15™ bioreactor might be expected to be in a similar growth and metabolic state, there were changes in gene expression identified between the samples. The seed train samples gene expression when compared to the Day 4 bioreactor sample comparison for the 3936 cell line revealed the fewest DE genes with 140 genes changing in expression between the samples. The same comparison for the 4384 and 3080 cell lines revealed 306 and 307 genes DE respectively between the seed train and fed-batch reactor samples whilst for the 3068 cell line there was more than double the number of DE genes identified (761). Overall, the hierarchical clustering, PCA and DE analysis suggest the 4384 cell line has a different gene expression profile to the other cell lines whilst the 3068, 3936 and 3080 cell lines have a much closer gene expression profile. Further, the seed train samples of each cell line show,

to varying degrees, different gene expression profiles than that of cells taken from the fed-batch cultures in an equivalent growth phase.

4.2.4 Investigating Pathway Enrichment in DE Genes

KEGG pathway functional enrichment of the RNA-Seq datasets based on statistically significant differentially expressed genes showed two distinct patterns across the datasets. Firstly, a major theme of enrichment in the DAVI dataset was in the Replication and Repair area, where DE genes were found to be enriched in DNA replication, base excision repair, nucleotide excision repair, mismatch repair, homologous recombination and Fanconi anemia pathways among the 3077, 3068 and 3478 cell lines when comparing Day 12 v Day 4 expression profiles within the same cell line (Figure 4.4). Interestingly, the only cell line in the DAVI dataset in which none of these pathways was enriched was the 3080 cell line. In the DAVI dataset this was the cell line with the highest Qp and also the cell line that was distinct from the others when the RNA-Seq data was analysed by PCA. On the other hand, comparing the different cell lines gene expression profiles between Day 4 or Day 12 to each other did not reveal any enriched pathways related to genome maintenance (Supplementary Figure 4.2). Thus, within a given cell line changes in genome maintenance pathways were observed between days 4 and 12 of culture, however when Day 4 or Day 12 of different cell lines were compared this was not observed.

I then applied the same enrichment analysis to the JCE dataset and surprisingly none of the pathways involved in genome maintenance were enriched within the DE genes between Day 4 and Day 7 (Supplementary Figure 4.3). An analysis of the seed train v fed-batch DE genes for the 3080 and 3068 cell lines revealed enrichment of the DNA replication pathway and for the seed train v 4384 cell line enrichment of the nucleotide excision repair, mismatch repair and DNA replication pathways (Supplementary Figure 4.4). The pathway enrichment analysis also consistently revealed the enrichment of the PI3K-Akt signalling pathway, focal adhesion and ECM-receptor interaction pathways in the DE genes across both the DAVI and JCE datasets. Overall, the enrichment suggests a prominent regulation of genome maintenance mechanisms is conserved across different cell lines at the passage from seed train to fed-batch culture and towards the end of culture, while the most differentially regulated pathways at the same stage in culture for the same cell lines are the PI3K-Akt signalling, focal adhesion and ECM-receptor pathways.

When considering the differential expression of individual genes involved in the Replication and Repair domain, I identified Exonuclease 1 (*Exo1*), *Rad51*, Essential Meiotic Structure-

Specific Endonuclease 1 (*Eme1*) and FA Complementation Group B (*Fancb*) in common among the top 30 most differentially expressed genes in both the DAVI and JCE datasets (Supplementary Figure 4.5). These genes are involved in a wide range of genome repair mechanisms from mismatch repair to homologous recombination and DNA double-strand break repair, suggesting a co-regulation of multiple facets of genome maintenance and the importance of high fidelity in these pathways to maintain cell integrity, viability and growth as culture progresses.

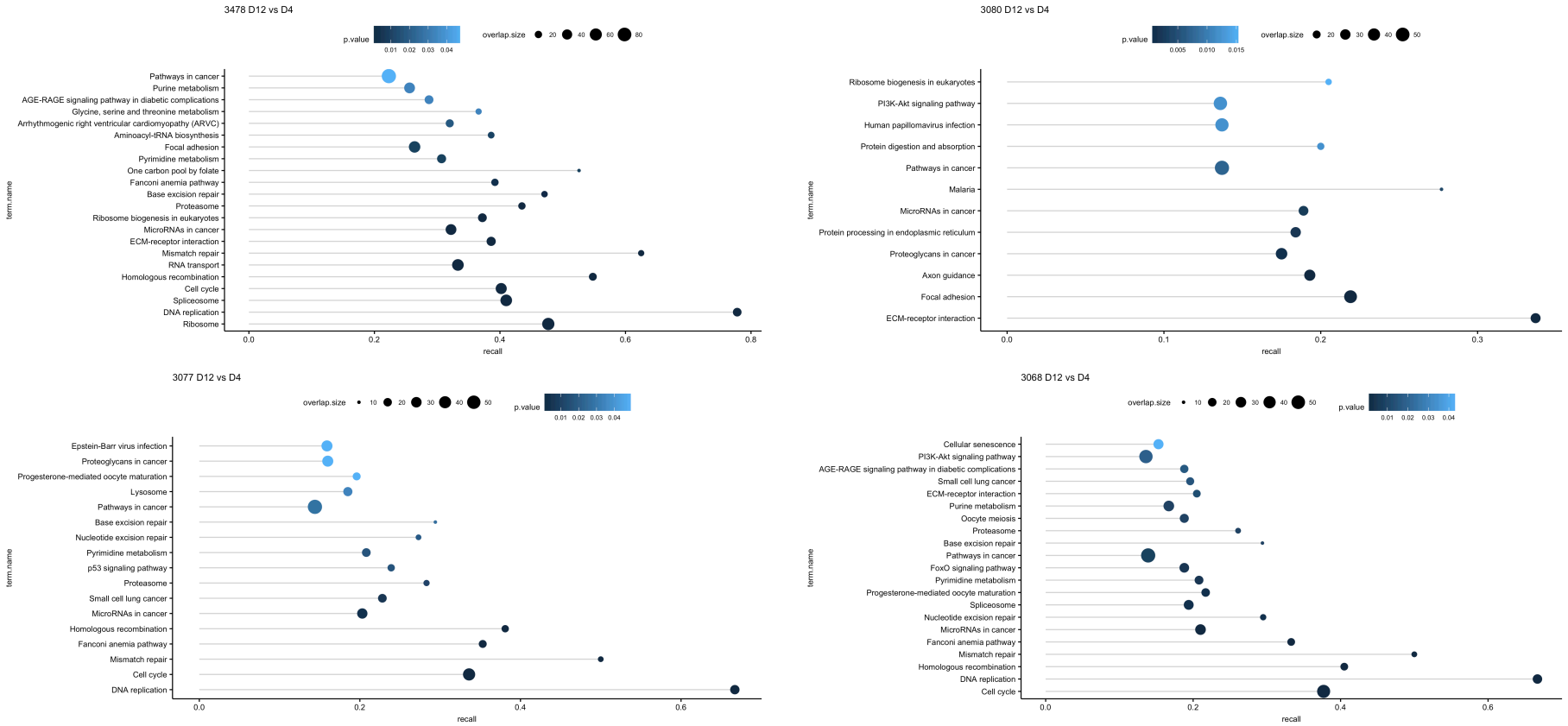


Figure 4.4 - Enriched KEGG pathways based on differentially expressed genes in the DAVI dataset

Enriched KEGG pathways based on differentially expressed genes for each comparison among the same cell line at day 12 against day 4 in the DAVI dataset. Each dot represents a pathway, with colour shade representing the p-value, size proportional to the overlap size (differentially expressed genes in the pathway) and x-coordinate recall (overlap size divided by the total number of genes in the pathway).

4.2.5 Mapping of Long non-coding RNA expression during fed-batch culture

As outlined in the introduction section, a key aspect of this study was to generate a detailed description of lncRNA expression under controlled bioreactor conditions in different IgG producing CHO cells and to identify lncRNAs whose manipulation may enhance the CHO cell factory ability to generated secreted target products. In order to investigate and identify non-coding RNAs, all the significant differentially expressed non-coding genes were filtered based on NCBI annotation and are shown as a percentage of the total number of differentially expressed genes for each comparison in
Figure 4.3 and in Supplementary Table 4.1. The complete RNA-Seq dataset is provided reporting those lncRNAs identified as being expressed in CHO cells and hence providing a reference for the community to investigate individual lncRNAs in CHO cells (Supplementary Material 4.1).

The DAVI dataset showed a higher percentage of ncRNAs on average (22.2%) compared to JCE (14.0%) as a percentage of the total DE RNAs identified, most likely due to the higher coverage in the sequencing data for the DAVI dataset. Among the DAVI dataset, comparisons listing the highest number of DE genes, in particular Day 12 v Day 4 for each cell line, contained the lowest percentage of lncRNAs on average at 14.8% while the opposite was true for comparisons at Day 4 and partly at Day 12, where the lower number of DE genes resulted in a higher percentage of ncRNAs respectively at 28.7% and 20.6% on average. I then calculated how many ncRNAs were differentially expressed at both time points in the two datasets to assess the variability of lncRNAs expression as culture progresses (Supplementary Material 4.1). When comparing producers against non-producers in the DAVI dataset, between 33.4% and 44.4% ncRNAs were differentially expressed at both day 4 and day 12, while the percentages raised to 56.6-71.1% when comparing producers. The JCE dataset showed a narrower range, between 38.6% and 53.1% ncRNAs differentially expressed at both day 4 and day 7. Overall, these results confirm the higher coverage for the DAVI dataset resulting in a higher number of identified ncRNAs.

4.2.6 Identification of differentially expressed lncRNAs as potential cell engineering targets

I then looked to filter and refine the list of DE lncRNAs by counting the occurrence of each transcript in all the DE comparisons, assessing sequence conservation across mammalian species through the discontiguous megablast algorithm and secondary structure prediction based on the RFAM database (Kalvari et al., 2018). Each lncRNA identified via this strategy

was then experimentally validated by RT-qPCR to confirm the differential expression between conditions in CHO cells (Supplementary Material 4.2). The first lncRNA identified via this approach was *Adapt15*. LncRNA *Adapt15* (also known as growth arrested DNA-damage inducible gene 7, *Gadd7*) was discovered in hamster cells, with homologs identified in the closely related Long-tailed Dwarf Hamster (*C. longicaudatus*) with sequence conservation across rodents (Crawford, Schools, Salmon, & Davies, 1996). More recently *Adapt15* has been linked to oxidative lipotoxicity, with knockdown alleviating ER stress and cell death (Brookheart, Michel, Listenberger, Ory, & Schaffer, 2009). In the DAVI dataset at Day 12, *Adapt15* was upregulated compared to day 4 in the 3077, 3068 and 3080 cell lines with a fold change in expression of 3.36, 2.86 and 1.88 respectively compared to the null pool 3478. Conversely, within the JCE dataset, which is focused only on producers and comparison of earlier stages of culture, there was no significant DE of *Adapt15* identified, suggesting DE expression of this transcript is observed in producer cell lines compared to a non-producing control later in fed-batch culture (Table 4.2). This observation of increased *Adapt15* transcript expression at day 12 relative to day 4 of fed-batch culture in the DAVI dataset was confirmed by qPCR experiments.

A second lncRNA that showed significant DE was growth arrest specific transcript 5 (*GAS5*), a non-protein-coding multiple small nucleolar RNA (snoRNA) host gene (C. M. Smith & Steitz, 1998) with a short ORF and a well-known tumour suppressor lncRNA in human cancer biology (Ma et al., 2016). As a result of cell growth arrest and mTOR pathway activity repression, translation of the *GAS5* short ORF is blocked and the transcript accumulates, escaping the nonsense-mediated decay (NMD) pathway which depends on active translation (Ma et al., 2016; Tani, Torimura, & Akimitsu, 2013). I identified the uncharacterized gene *LOC103158913*, upregulated in its expression at Day 12 in 3068 compared to 3478 null pool with a fold change in expression of 2.17 and an adj *p*-value = 7.23×10^{-5}, as a homologue of mouse *GAS5* in CHO using BLAST. I suggest this lncRNA is upregulated in CHO cell lines whereby growth has been arrested and mTOR signalling attenuated and may be a characteristic of such cell lines.

A 3rd well characterised lncRNA in other systems that was identified in our system as differentially regulated was plasmacytoma variant translocation 1 (*PVT1*). *PVT1* is a non-coding oncogene related to poor prognosis in different cancer types (Shikai Zhu et al., 2017) with reports of competing endogenous RNA (ceRNA) activity on the miR200 microRNA family and direct interaction with the MYC oncogene (Colombo, Farina, Macino, & Paci, 2015b). In

a previous work (Vito & Smales, 2018), we identified the uncharacterized gene *LOC103158906* as a homologue of mouse *PVT1* in CHO using BLAST and secondary structure prediction in RFAM. *PVT1* is annotated in C_griseus_v1.0 as *LOC103162981*, giving the same BLAST alignment scores and RFAM. Myc proto-oncogene protein is predicted to be encoded in CHO as *LOC100758352*, containing Myc amino-terminal region (Myc_N), Helix-loop-helix DNA-binding domain (HLH) and a Myc leucine zipper domain (Myc-LZ). The expression of *Myc* and *PVT1* in DAVI was inversely correlated, with a Pearson coefficient of -0.73 and a *p*-value of 0.0015. In particular, *PVT1* was expressed at lower levels at Day 4 for all cell lines and at Day 12 for the null pool 3478 and 3080, when *Myc* expression was at its highest levels. Interestingly, when *PVT1* expression increased at Day 12 for 3077, *Myc* levels decreased to a similar extent. By contrast, the correlation between *PVT1* and *Myc* in the JCE dataset was positive, with a Pearson coefficient of 0.60 and a *p*-value = 0.0001. When only the seed train samples were considered, the coefficient was 0.93, indicating a very strong correlation at the early stage of fed-batch culture, when *PVT1* expression is lower (Supplementary Figure 4.6). This suggests a change in *PVT1* expression as culture progresses that relates to cell growth and *Myc* expression.

PVT1 expression was upregulated with a two-fold change at Day 12 in every producer compared to the null pool in the DAVI dataset and it was upregulated a fold-change of 2 at both Day 4 and Day 7 in the 3068 cell line compared to 3936, the clone with the highest yield and the pool with the lowest yield respectively in the JCE dataset. When assessing the direct interaction probability of CHO *PVT1* (XR_478426.2) and *Myc* (XP_003516054.2) using lncPro and RPISeq prediction tools, the output suggested a strong likelihood of interaction, scoring respectively 85.1 and 0.9 (Lu et al., 2013; Muppirala, Honavar, & Dobbs, 2011). Collectively these data suggest that *PVT1* is likely to be related to *Myc* expression and hence cell growth and proliferation in CHO cells during fed-batch culture.

4.2.7 Comparison of data presented here with existing datasets

A recent meta-analysis publication compared transcriptomics studies in CHO cells, commenting on the difficulty in comparing these datasets but also identifying the most recurrent genes identified as related to Qp and growth (Tamošaitis & Smales, 2018). Here, I find that of the coding genes in this study a number were regulated in agreement with this meta-analysis study. The gene ranking first in the list of the meta-analysis was *Cd36*, a multifunctional glycoprotein acting as receptor for a broad range of ligands of proteinaceous or lipidic nature (Yang et al., 2017), and was consistently downregulated in producers in this

study compared to the 3478 null pool in the DAVI dataset, with a fold-change between 2.53 - 3.41 at Day 4 and 4.14 - 6.64 at Day 12. *Cd36* was also downregulated in 3068 and 3080, the cell lines with the highest Qp in the JCE dataset, compared to 3936 and 4384 at both Day 4 and Day 7, with a fold change down in expression > 4.29. Heat Shock Protein Family A Member 8 (*Hspa8*), a molecular chaperone implicated in the protein quality control system and protection of the proteome from stress (Stricher, Macri, Ruff, & Muller, 2013), was downregulated at Day 12 against Day 4 in the DAVI dataset among every producer cell line and when compared to the null 3478 cell line at Day 12. The same pattern of differential expression was shown by both *Serpinh1*, a collagen-specific molecular chaperone localized to the endoplasmic reticulum (Ito & Nagata, 2017), and vimentin (*Vim*) a type III intermediate filament protein responsible for maintaining cell shape and stabilizing cytoskeletal interactions (Musaelyan et al., 2018). A summary of all the common genes between the meta study and the current one is provided in Supplementary Material 4.3. Overall, *Cd36, Hspa8, Serpinh1* and *Vim* were consistently downregulated in cell lines with higher Qp in my datasets in agreement with previous transcriptomics studies summarised in the meta-analysis (Tamošaitis & Smales, 2018).

4.3 Summary

In this work I present and make available to the community two RNA-Seq derived transcriptomic datasets that comprehensively detail coding and non-coding transcript expression analysis of 5 IgG1 producing CHO cell lines and 1 null pool at different time points cultivated under fed-batch conditions in an ambr15™ system. In particular, I provide the first CHO RNA-Seq study detailing the lncRNA profile of such cells, confirming the expression of lncRNAs in CHO cells and identifying those whose expression is differentially regulated throughout fed-batch culture and between cell lines with different characteristics. The different time points for sample collection throughout culture and Qp of cell lines was clearly reflected in the PCA clustering of the transcript expression analysis and in the numbers of differentially expressed genes for the DAVI dataset. In addition, feeding was shown to be a significant source of variability even at early stage of culture, as shown by the comparison of seed train flask data used to inoculate the fed-batch process. This data shows for the first time that there is a significant change in gene expression after only 4 days of culture among the same clone when the cells are still rapidly growing and dividing and before the major phase of production of the protein of interest.

KEGG functional enrichment analysis confirmed a tendency of pathways in the Replication and Repair domain to be differentially regulated in response to feeding when the seed train is inoculated in the fed-batch process and towards Day 12 in particular. The only exception was the clone with the highest Qp at Day 12, 3080, which interestingly did not show any differentially regulated pathways in the Replication and Repair domain. These data suggest that those cell lines that can maintain genome integrity and its surveillance may be better suited to prolonged culture and recombinant protein productivity. Although our datasets contained 6 different cell lines, I wanted to identify coding genes related to an increase in Qp across literature to improve robustness across different systems, leading to the identification of *Cd36*, *Hspa8*, *Serpinh1* and *Vim* as genes negatively correlated with Qp in both our datasets and the most recent transcriptomics meta-analysis in CHO cells (Tamošaitis & Smales, 2018). Although the functions of those genes are heterogenous, the conserved pattern of expression among very different experimental settings and cell lines suggests conserved roles with detrimental effects on Qp. Knock-Out (KO) or Knock-Down (KD) strategies on the aforementioned genes with CRISPR or RNA interference could be implemented to investigate these effects (J. S. Lee et al., 2015; Wu, 2009).

In conclusion, while many studies in CHO cells have investigated coding genes, my work aimed to unveil the non-coding transcriptome variation, specifically lncRNAs. I identified *Adapt15*, *GAS5* and *PVT1* among many others as lncRNAs linked to Qp, although there is a lack of their annotation in Chinese hamster genomes there is a well-established effect of these in model organisms and human diseases (Colombo et al., 2015b; Hollander, Alamo, & Fornace, 1996; Ma et al., 2016). While *Adapt15* was initially identified in hamster and later linked to ER stress and cell death, its role in CHO cells has never been further investigated. *Adapt15* upregulation in producing cell lines towards day 12 of culture indicates an increasing stress on the ER as the recombinant protein is produced, suggesting *Adapt15* as a target for knock-down with RNAi or knock-out with CRISPR. *GAS5* transcript accumulation occurs as a result of mTOR signalling repression, suggesting its use as a marker of translation repression and cell growth arrest in specific cell lines. My data also suggests a close link between the expression of the lncRNA *PVT1* and of *Myc*, the relationship between the two depends on the stage of fed-batch culture. When compared to an existent public database of gene expression (Singh et al., 2018), these lncRNAs are constitutively expressed in different CHOS, CHOK1 and DG44 cell lines, suggesting a conserved functional role which could be exploited for cell engineering. Although secondary structure and RNA-protein interaction predictions can suggest functional mechanisms of action, focused experimental

studies on single transcripts will be required to assess their effects in mammalian cell factories.

Chapter 5 tRNA Expression in CHO and HEK293 Cells and the Impact on Transcript Specific and General mRNA Translation

5.1 Introduction

The majority of approved recombinant biopharmaceuticals, including monoclonal antibodies (mAbs), are produced in mammalian cell expression systems, predominantly in Chinese hamster ovary (CHO) cells under fed-batch culture conditions (Butler & Spearman, 2014; Dumont et al., 2016; Mauro, 2018). CHO cells are the current leading system due to their demonstrated ability to produce complex correctly folded proteins with post-translational modification (PTMs) compatible with human coupled with an effective secretion of the protein produced (J. Y. Kim et al., 2012). In addition, CHO cells can grow in suspension in chemically defined serum-free media, on a large scale and produce secretory yields of monoclonal antibody in fed-batch culture in excess of 5 g/L (Povey et al., 2014). The secretory yield from such an expression system results from the combination of the number of cells across culture (the integral of viable cell concentration or IVC) and the average amount of material expressed by each cell (cell specific productivity, Qp) (Kunert & Reinhart, 2016). These attributes can be positively influenced by selecting the most appropriate culture conditions and by engineering cell clones, but mRNA translation remains the main biological processes determining global and protein specific synthesis, and hence controlling most cellular activities, impacting on IVC, Qp and protein quality (McLeod et al., 2011; Mead et al., 2015; Schwanhäusser et al., 2011). As mRNA translation is a complex process composed of sequential phases and requiring the assemble of cellular machineries with multiple subunits, protein abundance control in the cell can be tuned at multiple levels depending on internal or external dynamics, such as the availability of translation factors in the cell or extracellular signalling stimulating growth (Kelen et al., 2009). Due to this central role of translation, numerous works in CHO cells have investigated this process showing a dependence on the phosphorylation of key translation factors such as eIF2α causing attenuation of global protein synthesis (Underhill et al., 2005), on global translation efficiency (Mead et al., 2012; O'Callaghan et al., 2010; Roobol et al., 2015) and on culture temperature, which impacts on quantity and quality of recombinant protein produced (Underhill et al., 2005). The interplay of these parameters and their variations throughout culture can have a huge impact with

opposed effects for different transcripts. Under reduced temperature, translation elongator factor 2 (eEF2) becomes phosphorylated leading to a reprogramming of translation and causing global attenuation which can be escaped by transcripts with a specific codon usage (Bastide et al., 2017). The degeneracy of the genetic code allows the use of different synonymous codon sequences to achieve the same polypeptide, providing the option to tailor an exogenous recombinant coding sequence to the codon usage of the expression system of choice (Welch et al., 2009).

Indeed, codon usage is routinely considered when designing a recombinant sequence for maximizing protein expression, leading to a peak increase up to 1000 fold in protein yield, but much more modest effects in general (Gustafsson et al., 2012; Kimchi-Sarfaty et al., 2013). While this approach assumes that rare codons are rate limiting for protein production (Mauro & Chappell, 2014), recent work suggests a much more complex picture, where mRNA stability, ribosome speed and folding are influenced by specific codons and codon combinations, especially in mammalian cells (Brule & Grayhack, 2017; Hanson & Coller, 2017). A recent report used codon de-optimization of a bispecific antibody sequence through the introduction of less frequently occurring codons in CHO, showing an overall increase in final yield (Magistrelli et al., 2017). One of the reasons underneath this different behaviour could be the existence of distinct codon usage patterns depending on the cell type and growth phase, as reported in the case of proliferation and differentiation (Gingold et al., 2014; Plotkin et al., 2004), or even among specific Gene Ontology (GO) sets of genes (Rudolph et al., 2016). On top of the intrinsic codon usage variability, mammalian cell factories bare an additional burden on translation represented by the recombinant gene of interest, which can account for up to 15% of the total ribosome capacity in the cell (Kallehauge et al., 2017).

An additional layer of complexity is the determination of the tRNAs pool, key players in translation elongation. Whereas an accurate quantification of single tRNAs species by high throughput sequencing techniques has been limited due to secondary structure and nucleotide modifications, new protocols recently emerged to overcome most of the limitations (Gogakos et al., 2017; Shigematsu et al., 2017; Zheng et al., 2015). AlkB-facilitated RNA methylation sequencing (ARM-Seq) in particular takes advantage of the bacterial dealkylating enzyme AlkB to remove methyl groups from m^1A, m^3C and m^1G residues in tRNAs allowing a much more accurate determination of tRNAs species (Cozen et al., 2015; Hrabeta-Robinson et al., 2017). Here I present the first attempt of tRNA quantification in a host CHO-S cell line under batch and fed-batch conditions and in both null and IgG1

producing CHO DG44 cell lines cultivated in an automated mini-bioreactor setting using two different feeding strategies, and I compare these to HEK293 cells cultivated at 37°C and 32°C. In addition, I calculate the speed of elongation of selected IgG1 sequences and Etanercept based on the computational model of elongation described in Section 1.5.1.

5.2 Results

5.2.1 Culture of CHO-S cells for tRNA analysis; Fed-batch culture shows a sustained higher proliferation and culture viability compared to batch culture

A CHO-S host cell line was used as a model CHO cell host expression system to accurately measure tRNA species abundance and variation throughout culture at different time points, and to allow a comparison between tRNA abundancies and their variation during batch or fed-batch culture processes. The CHO-S batch and fed-batch culture growth characteristics are described in detail in Chapter 3. Cell growth under fed or batch culture was almost identical up until day 4 of culture ; after this time cell growth of the batch cultures started to slow dramatically (Figure 5.1A).

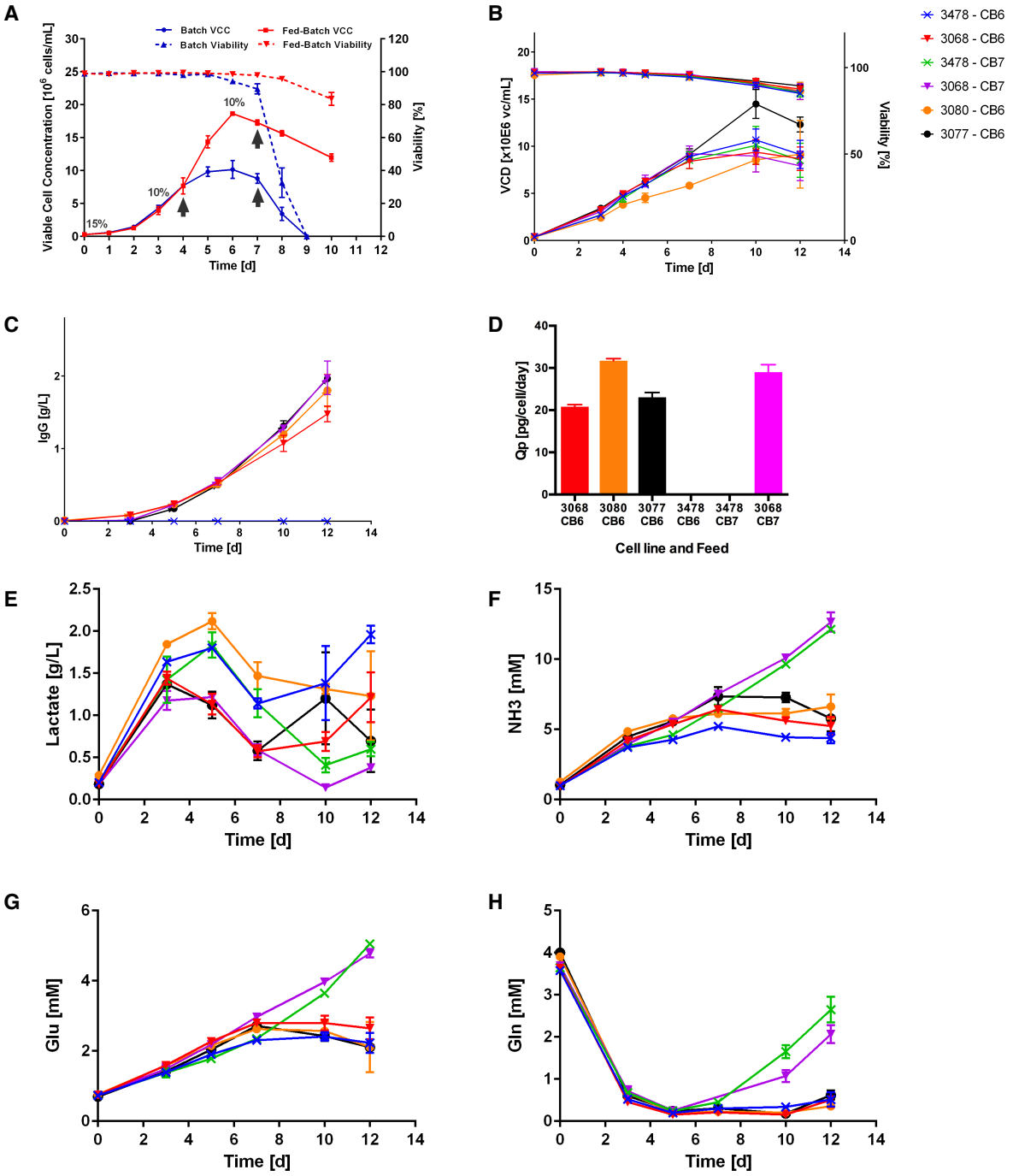


Figure 5.1 - CHO-S and Symphogen cell culture results

A | Growth profiles of the CHO-S host cell line throughout batch and fed-batch culture over time. Viable cell concentration (VCC) and culture viability are shown (red for Batch and blue for Fed-batch). This data was used in Chapter 3 and is reported for the convenience of the reader. The percentages of CHO CD Efficient Feed B Liquid Nutrient Supplement added to the existing working volume are shown for each feeding day (days 0, 3 and 6). The arrows indicate when samples for tRNA sequencing were harvested (Day 4 and Day 7). **B** to **H |** Main parameters of the three Symphogen IgG1 producing cell clones (3068, 3080, 3077) and one null producer cell pool (3478) cultured in an ambr15™ system under fed-batch culture with feeding using either the Cell Boost 6 (CB6) or Cell Boost 7 (CB7) protocol. **B |** Viable cell density (VCD) and culture viability over time. **C |** IgG1 titer over time. **D |** Specific productivity (Qp) for each cell line. **E |** Lactate accumulation over time **F |** Ammonia accumulation over time **G |** Glutamate accumulation over time **H |** Glutamine accumulation over time.

Fed-batch cultures growth continued with an increase in viable cell concentrations observed until day 6 where a mean peak viable cell concentration of approximately 2 x 10^7 cells/mL was obtained. After this time point the viable cell number and culture viability of both cultures started to decrease, albeit at different rates, such that the cultures were terminated with a culture viability of 0% at day 9 for the batch cultures as opposed to a 83.6% mean culture viability at day 10 for the fed-batch cultures (Figure 5.1A). The first-time point at which samples were harvested for tRNA analysis was on day 4, when cell growth and culture viability were high and maintained in both cultures. On-the-other-hand, the second time point at which cells were harvested for RNA analysis was on day 7 of culture, corresponding to the end of the stationary phase/first day of decline for both cultures. In the case of the fed-batch cultures, these two sampling points were also 24 hours after addition of the Efficient Feed B.

5.2.2 Culture of Symphogen cells for tRNA analysis; Comparison of growth and productivity characteristics of different IgG1 producing cell lines cultivated in an automated ambr15^{TM} system with Cell Boost 6 and Cell Boost 7 feeding

As described in Chapter 4, three Symphogen DG44 derived in-house IgG1-producing CHO cell clones (3068, 3080, 3077) and one non-producing cell pool (3478) were used in this study (Table 5.1). This cell lines showed a comparable culture viability, well above 80%, throughout the 12 days of culture, while the VCD (viable cells/mL) peaked at 14.48 x 10^6 for cell line 3077, 10.66 x 10^6 for cell line 3478, 9.38 x 10^6 for cell line 3068 and 8.56 x 10^6 for cell line 3080 (Figure 5.1B). The IgG titres and Qp (cell specific productivity) for the producer cell lines at Day 12 were 1.96 g/L and 23.05 pg/cell/day respectively for cell line 3077, 1.48 g/L and 20.82 pg/cell/day for cell line 3068, and 1.80 g/L and 31.71 pg/cell/day for cell line 3080 (Figure 5.1C). Monitoring of the extracellular glutamate, glutamine and ammonia concentrations throughout culture did not reveal any significant differences in concentrations between the cell lines (Figure 5.1F,G,H). There were some differences observed in lactate concentrations throughout culture between the cell lines with lactate accumulating to a peak concentration slightly over 2 g/L for cell line 3080 at Day 5 and for cell line 3478 at Day 12 whilst the peak lactate concentration observed was below 1.5 g/L for cell lines 3068 and 3077 (Figure 5.1E).

To allow a comparison of any impact between different feeds on tRNAs in particular, the 3478 (non-producing cell line) and 3068 cell lines were cultured with Cell Boost 7a + b (CB7) supplement in parallel to Cell Boost 6 (CB6). In this case, when using the Cell Boost 7 fed, the peak VCDs were 10.09 x 10^6 viable cells/mL for cell line 3478 and 9.21 x 10^6 viable cells/mL

for cell line 3068. When grown in fed-batch culture mode with Cell Boost 7, a Qp of 29.03 pg/cell/day was observed for cell line 3068 with a final IgG titre of 1.98 g/L being achieved (compared to a titre of 1.48 g/L and a Qp of 20.82 pg/cell/day using Cell Boost 6). After day 7 of culture, lactate formation for both cell lines was decreased when using the CB7 feed compared to CB6, ending below 0.70 g/L at day 12 (Figure 5.1E). However, although the lactate concentration was reduced in CB7 fed cultures compared to CB6 fed cultures, the accumulation of the remaining measured metabolites was increased substantially in both the 3478 and 3068 cell line cultures when using CB7 compared to CB6 fed-batch culture, reaching concentrations respectively of 12.14 (cell line 3478) and 12.64 (cell line 3068) mM for ammonia (Figure 5.1F), 5.05 and 4.78 mM for glutamate (Figure 5.1G), and 2.65 and 2.19 mM for glutamine (Figure 5.1H).

Table 5.1 - Summary of the cell lines used in the Symphogen dataset

Cell line	Clonality	Feed	Peak VCD [viable cells/day]	Yield [g/l]	Qp [pg/cell/day]
3478	Pool	Cell Boost 6	10.66×10^6	-	-
3478	Pool	Cell Boost 7	10.09×10^6	-	-
3068	Clone	Cell Boost 6	9.38×10^6	1.48	26.53
3068	Clone	Cell Boost 7	9.21×10^6	1.98	29.03
3077	Clone	Cell Boost 6	14.48×10^6	1.96	23.05
3080	Clone	Cell Boost 6	8.56×10^6	1.80	31.71

The cell line number, clonality (not assessed for Pool, assessed by FACS for Clone), the peak viable cell concentration (VCD), the total IgG yield at Day 12 and the cell specific productivity.

5.2.3 AlkB de-methylation treatment on extracted small RNAs increases the detection of full length tRNAs up to 5-fold

Once cell pellet samples had been taken from the cultures at the appropriate days for analysis, it was necessary to extract the RNA, and then for tRNA RNA-Seq analysis, specifically to remove modification of tRNAs that can interfere with sequencing. Thus, it was necessary to assess the efficiency of tRNA de-methylation treatment (as described in section 2.4.1) we included one negative control sample for each condition, where the AlkB protein that undertakes the demethylation reaction was not added (AlkB). On top of this, as tRNA gene copy number is known to correlate strongly with codon usage and tRNA abundance in yeast (Iben & Maraia, 2012), we de-methylated and sequenced small RNAs from *S. cerevisiae* to use as a reference for the assessment of the technical accuracy of the RNA-Seq analysis.

After adapter and short read (<16 nt) trimming, between 28.9% and 48.9% of the total reads aligned to tRNAs in the different AlkB treated samples, as opposed to only 5.8-12.2%

obtained in the no AlkB added samples, and therefore no demethylation, controls (Figure 5.2A). Therefore, using the demethylase treatment and ARM-Seq allowed the identification of 4- to 5-fold more tRNAs compared to standard small RNA sequencing. Interestingly, optimizing the library generation protocol for the DAVI dataset with the use of a more processive Reverse Transcriptase and a longer reaction time (see Materials and Methods) led to an even higher percentage of reads aligning to tRNAs, ranging from 74.3% to 88.5% (Table 5.2). As expected, the yeast samples showed a strong correlation between the tRNA abundances measured by sequencing and the gene copy numbers (r = 0.75, *p*-value = 5.1e-09, Figure 5.2B). Taken together, these results show a much higher identification of tRNAs following the AlkB treatment protocol to detect tRNAs, with a 4 to 5-fold increase, and a reliable accuracy, as shown by the *S. cerevisiae* samples used as a reference.

Table 5.2 - Mapping efficiencies for Symphogen dataset

Sample	Reads mapping to tRNA [%]	Reads mapping to sRNA [%]	Reads not mapped [%]
3478 D4 CB6	86	8	6
3478 D12 CB6	86	9	5
3068 D4 CB6	84	10	6
3068 D12 CB6	82	13	5
3478 D4 CB7	86	8	6
3478 D12 CB7	81	13	6
3068 D4 CB7	88	7	5
3068 D12 CB7	77	16	7
3077 D4	84	10	6
3077 D12	83	11	6
3080 D4	80	13	7
3080 D12	82	11	7

For each sequenced Symphogen cell line sample, the table shows from left to right the percentage of reads aligning to tRNAs, the percentage of reads aligning to small RNAs and the percentage of reads not mapped.

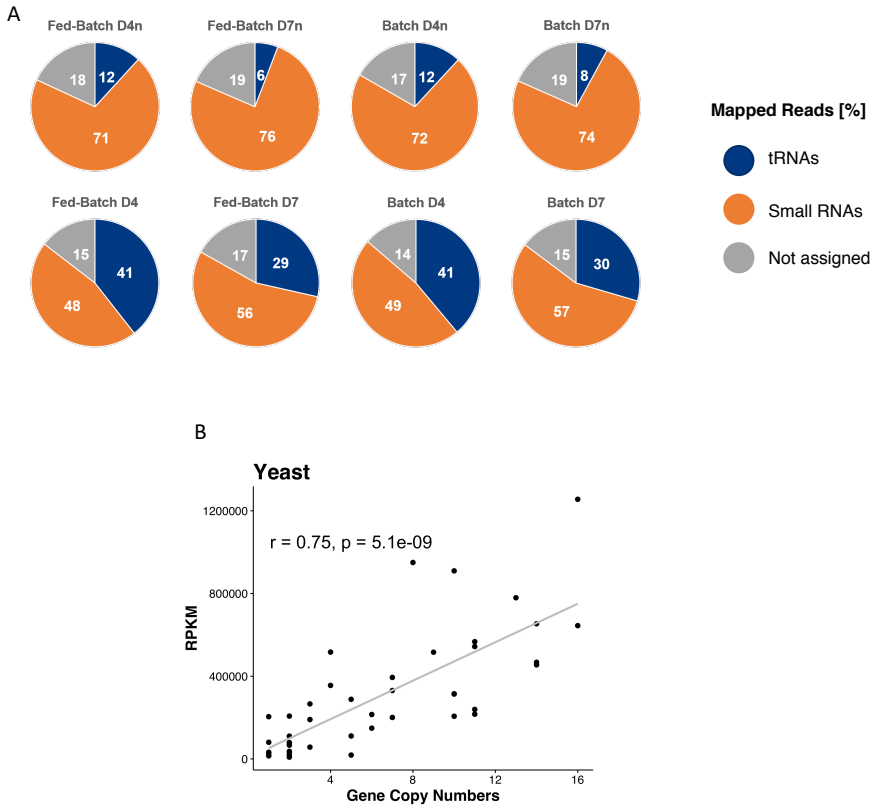


Figure 5.2 - Small RNA-Seq reads alignments and yeast correlation

A | Pie chart summary of the CHO-S cell small RNAs sequenced dataset mapping to GtRNAdb for tRNAs and ENSEMBL for the other small RNAs. In blue is indicated the proportion of reads mapping to tRNAs, in orange the proportion of reads mapping to small RNAs and in grey the proportion of reads unassigned. Samples indicated with the suffix –n (first row of graphs) were not demethylated before sequencing, while the corresponding samples below were treated with the bacterial de-methylase AlkB before library generation. **B |** Pearson correlation between the sequenced *S. cerevisiae* tRNAs and the tRNAs gene copy numbers in the genome. The coefficient (r = 0.75) indicates a strong correlation.

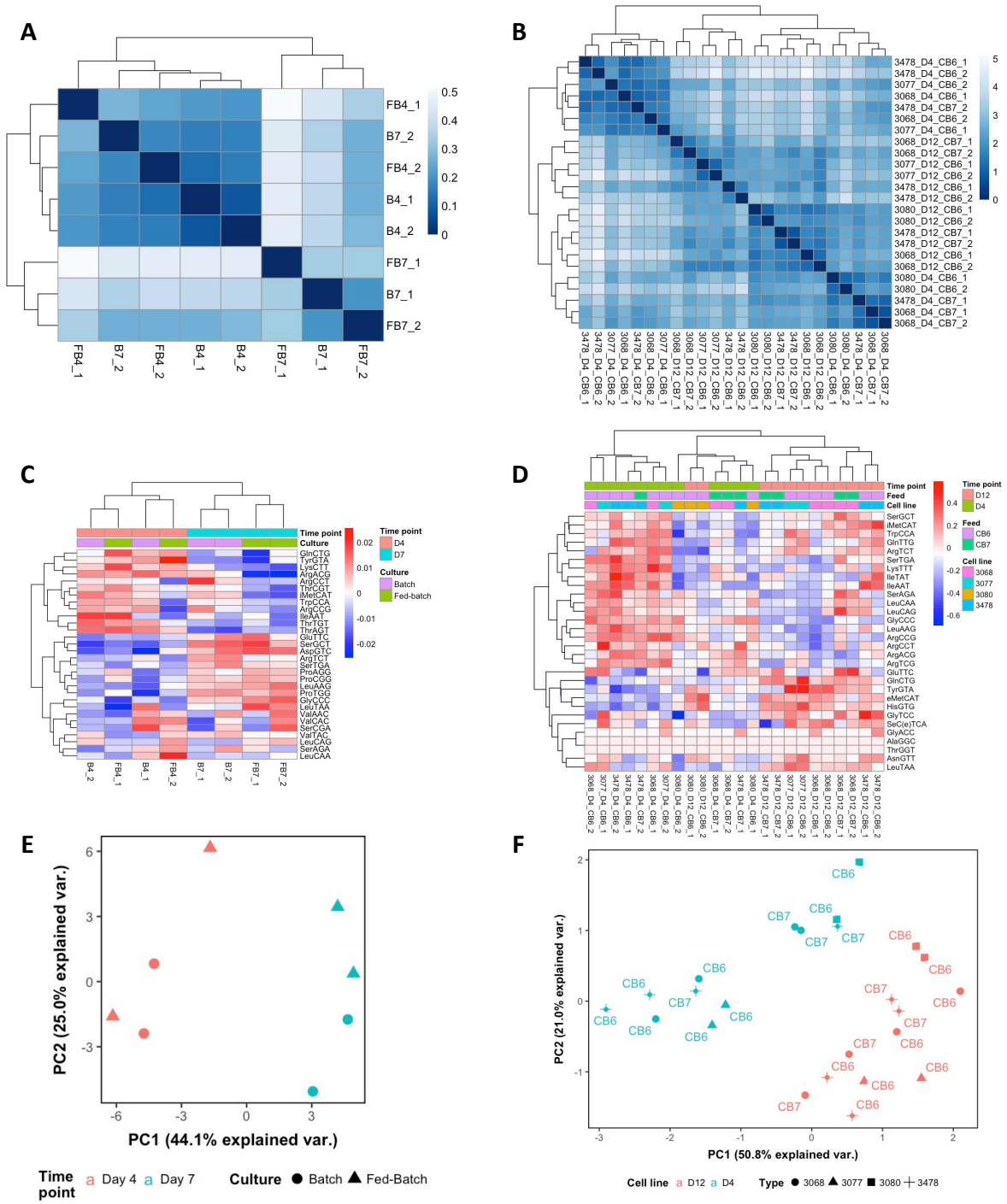


Figure 5.3 - Clustering and PCA based on tRNA sequencing

From top to bottom: clustering of samples based on the Euclidean distance between each sample, clustering based on the top 30 most differentially expressed genes and the PCA of normalized gene expression for the first two principal components. **A, C, E |** CHO-S dataset. **B, D, F |** Symphogen dataset.

5.2.4 The CHO-S and the Symphogen datasets both cluster based on culture time point sampling as opposed to feeding regime

The calculation of the Euclidean distances between samples provides an effective visual representation of a datasets global heterogeneity. When this was calculated for the CHO-S dataset, we observed two main clusters composed respectively of samples collected at Day 4 and those collected on Day 7 of culture. The one exception to this was for one of the replicates for Batch culture Day 7 (Figure 5.3A). The Symphogen dataset showed a similar trend, with a main block composed of samples collected at Day 12 of culture and two blocks formed from samples collected at Day 4 of culture (Figure 5.3B). While for Day 4 I noticed a preferential separation for CB7 fed samples, with the exception of 3478_D4_CB7_2, this was not the case at Day 12 of culture.

Focusing the analysis on the 30 tRNAs with the highest expression variance across samples gives a more accurate overview of the single gene behaviours and global clustering. The CHO-S dataset showed a clear clustering based on time point sampling, while there was no clear separation between batch and fed-batch culture (Figure 5.3E). The three tRNA genes with the highest variance were ThrGGT, ValTAC and LeuCAG, while the most represented amino acids were Arg (ArgCCG, ArgCCT, ArgTCG, ArgTCT), Leu (LeuAAG, LeuCAA, LeuCAG, LeuTAA), Ser (SerAGA, SerCGA, SerGCT, SerTGA), Thr (ThrAGT, ThrCGT, ThrTGT), and Val (ValAAC, ValCAC, ValTAC).

Clustering for the Symphogen dataset followed a similar pattern showing distinct Day 4 and Day 12 culture clusters, with the exception of 3080_D12_CB6_1 and 3080_D12_CB6_2, while feeding and cell line did not group together (Figure 5.3B,D). The three tRNAs with the highest variance in this data set were AlaGGC, ThrGGT and GlyACC, while the most represented amino acids were Arg (ArgACG, ArgCCG, ArgCCT, ArgTCG, ArgTCT), Gly (GlyACC, GlyCCC, GlyTCC), Leu (LeuAAG, LeuCAA, LeuCAG, LeuTAA), and Ser (SerAGA, SerGCT, SerTGA). Thus, Arg, Leu and Ser were commonly observed as the most represented amino acids across both datasets. 19 tRNA genes were observed as being the most differentiated in both datasets: AlaGGC, ArgCCG, ArgCCT, ArgTCT, GlnCTG, GluTTC, GlyACC, GlyCCC, IleAAT, iMetCAT, LeuAAG, LeuCAA, LeuCAG, LeuTAA, SerAGA, SerGCT, SerTGA, ThrGGT, TrpCCA.

A useful method to visualize grouping of different samples is to reduce the dimensionality of the dataset using Principal Component Analysis (PCA). PCA for the CHO-S dataset confirmed a clear separation between the Day 4 and Day 7 culture samples, while there was no evident clustering based on feeding (Figure 5.3E). The same trend was observed for the Symphogen dataset, with two main clusters corresponding to Day 4 and Day 12 culture samples and no

obvious clustering based on cell line or feeding type (Figure 5.3F). For both datasets, more than 69% of the variation was explained by the first two components.

5.2.5 The de-methylation treatment does not impair the detection of other small RNAs

In addition to tRNAs, the sequencing protocol allows the identification and quantification of most small RNAs, including in order of abundance: small nucleolar RNAs (snoRNAs), microRNAs, mitochondrial tRNAs, ribosomal RNAs and small nuclear RNAs. When these were included in the alignment, the percentage of reads mapping to either tRNAs or small RNAs increased to 92.1-95.3% of the total in the DAVI dataset, leaving a very small proportion of non-aligned reads (Table 5.2). Although my analysis focused primarily on cytosolic RNAs, the sequencing protocol allows the identification of mitochondrial small RNA species, including mitochondrial tRNAs (mt-tRNAs) and mitochondrial rRNAs (mt_rRNAs). Both species were consistently overexpressed in the DAVI dataset at Day 12 compared to Day 4, and the same trend was observed in the CHO-S dataset at Day 7 compared to Day 4 (Figure 5.4). Unfortunately, ENSEMBL annotation does not specify the type of mitochondrial tRNA for Chinese hamster except ENSCGRT00000000009 as tRNA methionine, and blasting each identified sequence to mouse and human did not produce any significant analogy.

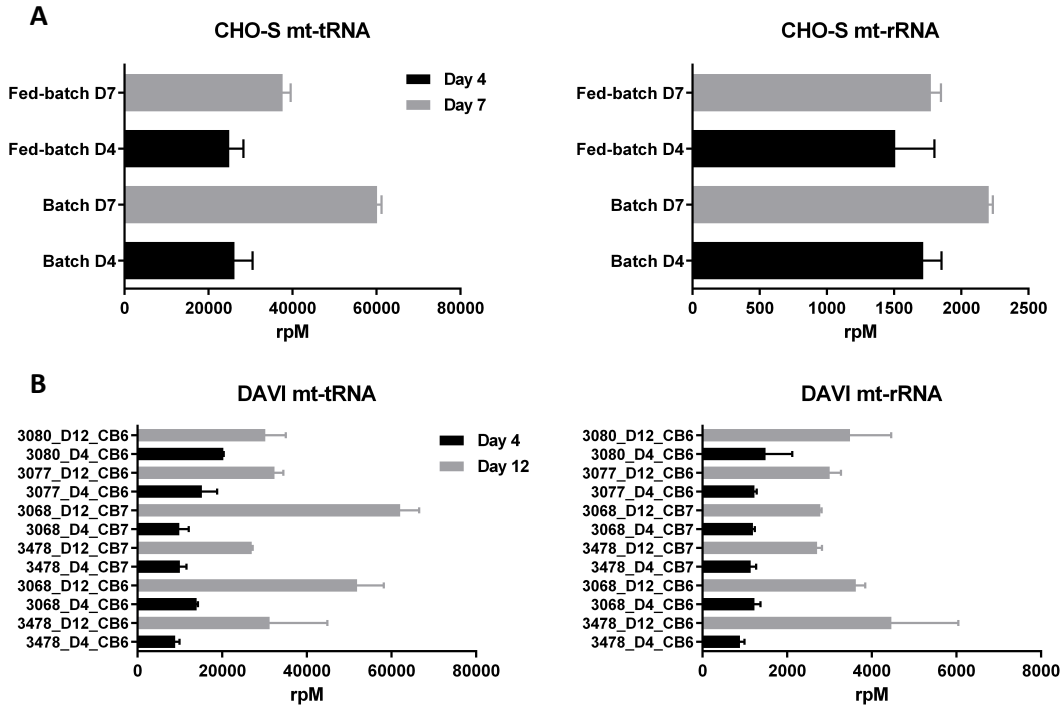


Figure 5.4 - mt-tRNA and mt-rRNA sequencing

Sequenced mitochondrial transfer RNAs (mt-tRNA) and mitochondrial ribosomal RNAs (mt-rRNA) for the CHO-S dataset (Panel **A**) and the Symphogen dataset (Panel **B**). The plots show a noticeable increase in the expression of mt-tRNAs and mt-rRNAs towards Day 7 and Day 12 of culture compared to Day 4 in both datasets.

5.2.6 The use of different Chinese hamster genomes does not significantly alter the tRNA gene copy number predictions

The tRNAscan-SE algorithm allows the identification of tRNA genes in DNA sequences using two previously described tRNA detection programs as fast, first-pass prefilters to identify candidate tRNAs, followed by a highly selective tRNA covariance model (Lowe & Chan, 2016). As tRNA prediction in the GtRNAdb only included those from the CriGri_1.0 genome assembly, I decided to assess the variability of tRNA gene copy numbers across the different ENSEMBL publicly available Chinese hamster genomes and the related species *Mesocricetus auratus* (Table 5.3) (Lewis et al., 2013). The majority of the genes showed comparable numbers across Chinese hamster, with the biggest discrepancies measured for LysCTT and CysGCA, the tRNAs with the highest copy numbers (36-41 and 109-124 respectively). The number of genes for each tRNA shows a higher variability when comparing Chinese hamster and *Mesocricetus auratus*, in particular for AlaAGC, AlaTGC, CysGCA and LysCTT.

Table 5.3 - tRNA gene copy numbers in hamster

Anticodon	Type	CriGri_1.0	C_griseus_v1.0	CHOK1GS_HDv1	MesAur1.0
AGC	Ala	19	21	20	10
CGC	Ala	3	3	3	5
GGC	Ala	1	1	1	2
TGC	Ala	13	14	14	5
ACG	Arg	6	7	6	7
CCG	Arg	3	3	3	3
CCT	Arg	7	6	7	7
TCG	Arg	6	5	6	3
TCT	Arg	4	4	5	6
ATT	Asn	1	0	1	0
GTT	Asn	14	16	15	13
GTC	Asp	14	13	13	13
GCA	Cys	41	36	41	25
CTG	Gln	12	12	12	10
TTG	Gln	5	5	5	4
CTC	Glu	14	13	16	6
TTC	Glu	7	7	8	6
CCC	Gly	7	7	7	6
GCC	Gly	11	12	13	8
TCC	Gly	6	4	8	6
GTG	His	13	11	12	5
AAT	Ile	13	14	14	9
TAT	Ile	3	4	4	4
CAT	iMet	8	8	8	10
AAG	Leu	5	5	5	4
CAA	Leu	3	3	4	3
CAG	Leu	9	8	8	7
TAA	Leu	3	3	3	2
TAG	Leu	4	4	5	4
CTT	Lys	109	111	124	31
TTT	Lys	16	16	17	11
CAT	Met	8	8	8	7
GAA	Phe	8	7	8	8
AGG	Pro	7	6	8	9
CGG	Pro	3	4	3	3
TGG	Pro	6	5	6	5
TCA	SeC	2	2	2	1
AGA	Ser	6	7	7	5
CGA	Ser	2	3	3	3
GCT	Ser	8	9	9	9
TGA	Ser	3	3	3	3

Anticodon	Type	CriGri_1.0	C_griseus_v1.0	CHOK1GS_HDv1	MesAur1.0
AGT	Thr	7	7	8	8
CGT	Thr	4	5	5	3
TGT	Thr	4	5	4	3
CCA	Trp	6	6	7	5
GTA	Tyr	11	11	11	12
AAC	Val	11	12	11	6
CAC	Val	8	8	8	4
TAC	Val	5	5	5	8

Summary of the gene copy numbers for each tRNA and the correspondent aminoacid type in three Chinese hamster genomes (CriGri_1.0, C_griseus_v1.0, CHOK1GS_HDv1) and *Mesocricetus auratus* (MesAur1.0), predicted using tRNAscan-SE 2.0.

5.2.7 IgG1 light chain, EPO and Etanercept codon sequences show a strong agreement with the CHO codon bias even when a subset of sequences is considered

As discussed in the introduction of this thesis, the proportion of specific synonymous codons for the same amino acids can significantly impact translation efficiency. The codon adaptation index (CAI) is a score ranging from 0 to 1 which measures the codon usage bias of a sequence based on a set of reference coding genes (Sharp & Li, 1987). A significant advantage of the Symphogen dataset is the availability of total RNA-Seq results for the 3478, 3080, 3068 and 3077 cell lines at Day 4 and Day 12 of culture, along with the sequenced tRNAs. Therefore, I used the CAI python package (B. D. Lee, 2018) to calculate the index of adaptation for the 3068 and 3080 cell lines with respect to the IgG light and heavy chain sequence/codon usage, along with the publicly available Etanercept and EPO sequences, reported in the Appendix. To assess whether highly expressed genes show a different codon usage bias to genes expressed at a lower level, I decided to split the CAI calculation using the whole CHO genome and the 500 most and least expressed coding genes measured by RNA-Seq (Table 5.4). While the heavy chain sequences showed a strong CAI between 0.84 and 0.87, the IgG light chain CAI for both cell lines was always above 0.91, with a peak around 0.95. The EPO and Etanercept sequences displayed a CAI similar to the IgG light chain, while the use of the reference set of genes sequenced from different cell lines did not produce significant variations in the CAI of any of the coding sequences. Overall, all the coding sequences showed a strong agreement with the CHO codon bias, with little variation when considering only part of the expressed genes as a reference and a higher adaptation for the light chain compared to the heavy chain for both 3068 and 3080 IgGs.

Table 5.4 - Codon adaptation index calculations

Sequence	Chinese hamster	Top 500 3478 D12	Top 500 3080 D12	Bottom 500 3478 D12	Bottom 500 3080 D12
HC 3068	0.86	0.85	0.84	0.84	0.84
HC 3080	0.87	0.86	0.85	0.84	0.85
LC 3068	0.95	0.94	0.92	0.91	0.91
LC 3080	0.95	0.94	0.92	0.91	0.92
EPO	0.96	0.96	0.96	0.91	0.93
Etanercept	0.95	0.95	0.95	0.89	0.91

Summary of the Codon Adaptation Index (CAI) for the 3068 and 3080 cell line IgG heavy (HC) and light (LC) chains, EPO and Etanercept coding sequences. From left to right, CAI was calculated using a reference set of sequences from either the whole Chinese hamster coding genome or the top and bottom 500 coding genes ordered by transcript abundance based on the whole transcriptome RNA-Seq dataset described in Chapter 4.

5.2.8 The CHO-S RNA-Seq tRNA dataset shows a strong correlation with tRNA gene copy numbers and codon usage

The correlation between tRNA gene copy numbers and codon usage in Chinese hamster, human and mouse are discussed in Chapter 1 (see section 1.3.5). The sequenced tRNA datasets add a new layer of data where actual quantified tRNA abundances in the cell were determined and are compared to codon usage and gene copy number (Summarized in Table 5.5). As opposed to data from yeast (described in Section 5.2.3 and Figure 5.2B), the Symphogen CHO cell sequenced tRNAs indicated a weak Pearson correlation with gene copy numbers, showing coefficients (r) ranging between 0.14 and 0.20 and *p*-values above 0.1 (Figure 5.5A for day 12). Thus, for the Symphogen CHO cell samples there appears little correlation between tRNA gene copy number and tRNA abundance which could have implications when designing recombinant gene sequences and codon usage when based upon gene copy number alone.

To the contrary, the CHO-S cell samples showed a much stronger correlation between tRNA abundance and gene copy number than the Symphogen CHO cell samples, with a range between 0.32 and 0.38 with *p*-values all below 0.025 (Figure 5.6), although this is still much lower than observed with the yeast samples. A similar pattern was observed when comparing sequenced tRNAs to codon usage between CHO cell samples, where the correlation coefficient was between 0.13 and 0.19 for the Symphogen dataset (Figure 5.5B for day 12), while it was between 0.41 and 0.48 for CHO-S samples (Figure 5.6).

We further assessed the codon usage among the heavy and light IgG chain in the 3068 and 3080 cell lines and compared this to the tRNA gene copy numbers and codon usage in CHO

cells. The heavy chain sequence showed a statistically significant correlation with tRNA gene copy numbers at r = 0.29, *p*-value = 0.018 in cell line 3080 and r = 0.28, *p*-value = 0.023 in cell line 3068. The correlation with codon usage was much stronger at r = 0.68, *p*-value = 9.5e-10 in cell line 3080 and r = 0.66, *p*-value = 4.6e-09 for cell line 3068. The light chain sequence showed a weaker correlation with the tRNA gene copy numbers with an r = 0.25, p-value = 0.044 in cell line 3080 and r = 0.19, *p*-value = 0.14 in cell line 3068 while the correlation with codon usage was again much stronger at r = 0.66, *p*-value = 4e-09 in cell line 3080 and r = 0.61, *p*-value = 1.1e-07 in cell line 3068. When comparing the IgG codon usage to the sequenced tRNAs at Day 12 of culture, when recombinant protein production is at its peak, we found a better correlation for the heavy chain (r between 0.16 and 0.23, p-value between 0.07 and 0.18) compared to the light chain (r between 0.09 and 0.13, p-value between 0.32 and 0.49) (Summarized in Table 5.5).

In summary, the tRNA analysis in the different cell lines has shown that;

- Symphogen CHO cell line sequenced tRNAs show a positive but not statistically significant correlation with gene copy numbers, while CHO-S cell samples showed a statistically significant correlation.
- CHO-S sequenced tRNAs showed a strong correlation to codon usage while this was non-significant for the Symphogen dataset.
- Correlation between the codon usage among the 3080 and 3068 cell lines for the HC and LC sequences with tRNA gene copy numbers and CHO cell codon usage was statistically significant except for the case of the 3080 cell line LC v tRNA gene copy number.
- The HC codon usage correlated more strongly than LC codon usage with the sequenced tRNA abundancies in the Symphogen dataset.

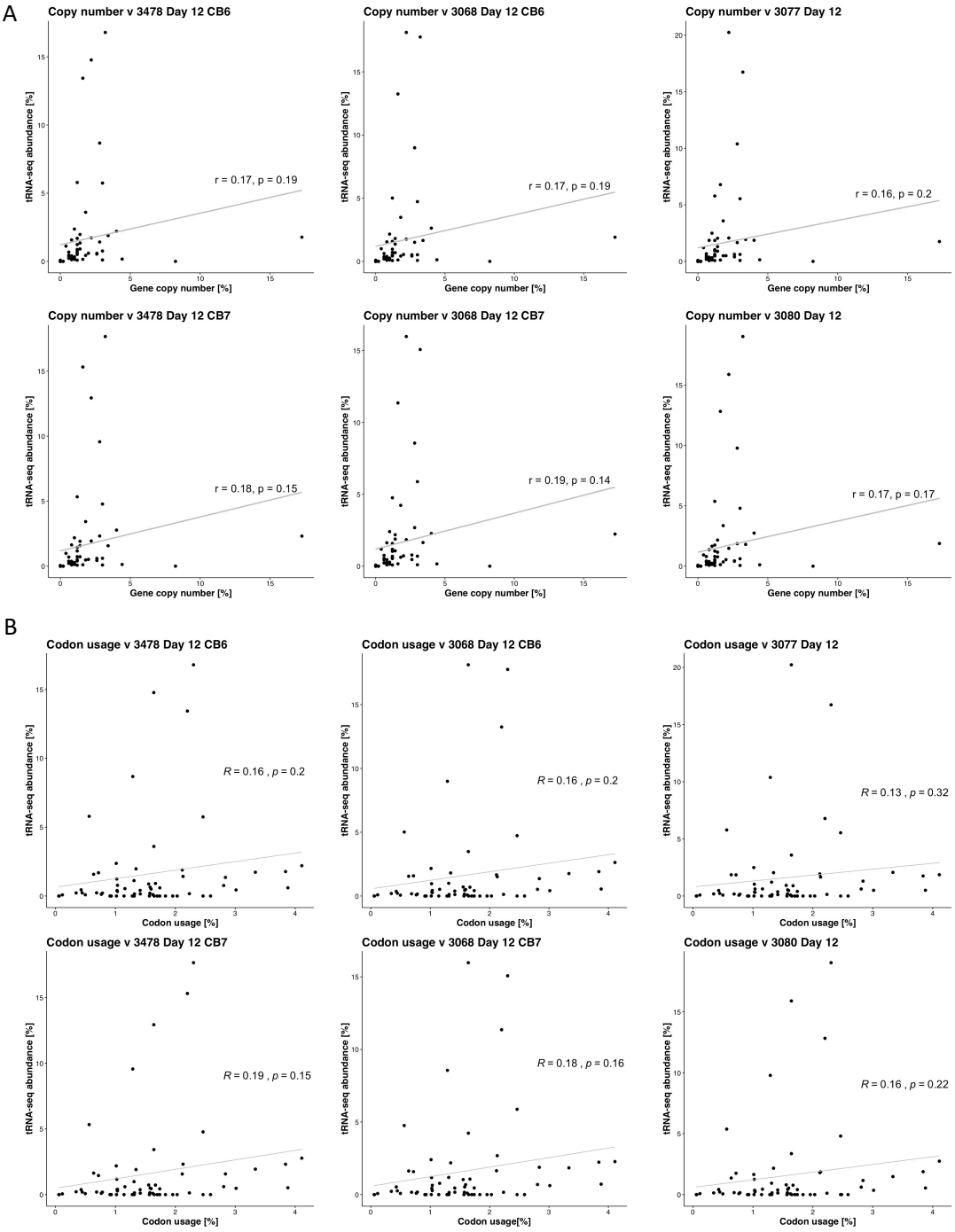


Figure 5.5 - Correlations between sequenced tRNAs, gene copy number and codon usage in the Symphogen dataset

Pearson correlation between the sequenced Symphogen tRNAs sampled at day 12 of culture and the tRNA gene copy numbers (Panel **A**) and the codon usage (Panel **B**) in CHO cells. The coefficients (r = 0.13 - 0.19) coupled with the p-values (p = 0.14 - 0.32) indicate a weak statistically non-significant correlation for all the comparisons.

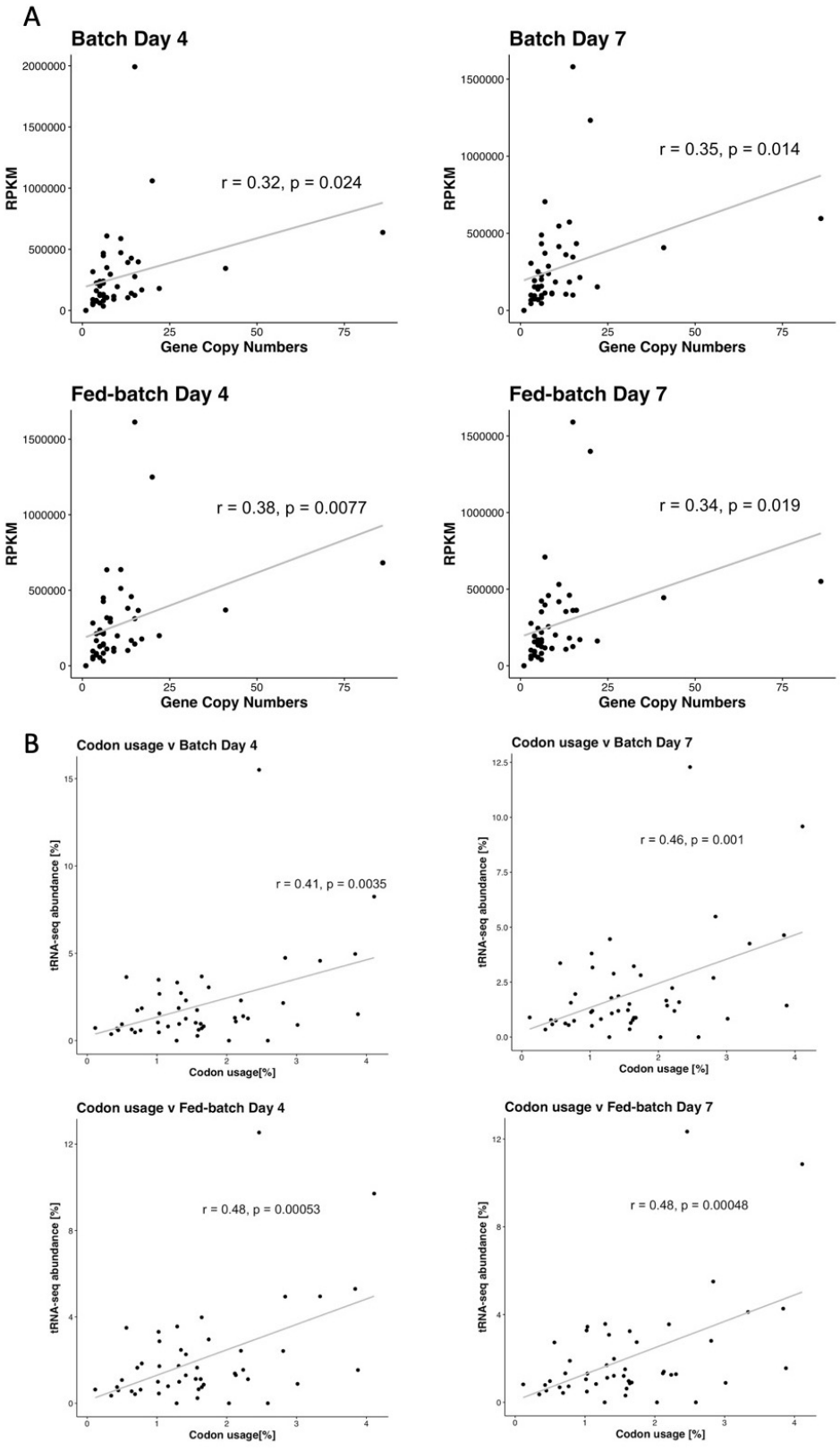


Figure 5.6 - Correlations between sequenced CHO-S tRNAs, gene copy numbers and codon usage

Pearson correlation between the sequenced CHO-S cell tRNAs, sampled at day 4 and day 7 of batch or fed-batch culture, and the tRNA gene copy numbers (Panel **A**) and the codon usage (Panel **B**) in CHO. The coefficients (r = 0.32 - 0.48) indicate a positive correlation for all the comparisons, with a stronger correlation showed between the sequenced tRNAs and the codon usage.

Table 5.5 - Correlation coefficients for the sequenced tRNAs

Sample	Copy Number	Codon Usage	HC 3068	LC 3068	HC 3080	LC 3080
3478 D4 CB6	0.16	0.16	0.19	0.09	0.17	0.07
3478 D12 CB6	0.17	0.16	0.19	0.10	0.17	0.08
3068 D4 CB6	0.18	0.18	0.21	0.11	0.19	0.09
3068 D12 CB6	0.17	0.16	0.20	0.12	0.18	0.10
3478 D4 CB7	0.15	0.15	0.19	0.10	0.17	0.08
3478 D12 CB7	0.18	0.19	0.20	0.11	0.18	0.10
3068 D4 CB7	0.14	0.15	0.19	0.11	0.17	0.08
3068 D12 CB7	0.19	0.18	0.23	0.13	0.20	0.11
3077 D4 CB6	0.20	0.17	0.21	0.10	0.19	0.09
3077 D12 CB6	0.16	0.13	0.20	0.10	0.16	0.09
3080 D4 CB6	0.18	0.18	0.21	0.13	0.19	0.11
3080 D12 CB6	0.17	0.16	0.19	0.10	0.17	0.09
Batch D4	0.33	0.41	0.19	0.01	0.19	0.05
Batch D7	0.35	0.46	0.21	0.04	0.20	0.08
Fed-batch D4	0.38	0.48	0.23	0.07	0.22	0.10
Fed-batch D7	0.32	0.48	0.23	0.06	0.22	0.09

Summary of the correlation coefficients between the sequenced tRNAs and the corresponding gene copy numbers, codon usage in Chinese hamster, codon usage among the 3068 and 3080 IgG1 coding sequences for the heavy and the light chain.

5.2.9 Differential expression of tRNAs among the CHO-S dataset shows variability at Day 7 regardless of the type of culture

Among the CHO-S dataset, comparing batch v fed-batch at Day 4 did not reveal any statistically significant differential expression. On the contrary, the same comparison at Day 7 indicated Phe, Ala, Leu, Gln, Leu, Thr, Arg, Met, His, Arg, Asn, Ile as differentially expressed tRNA isotypes (Table 5.6). When considering the same type of culture at Day 4 v Day 7, batch cultures showed HisGTG, GlyGCC, AsnGTT, LeuTAG, AspGTC, SerGCT, SeC(e)TCA, ProTGG, CysGCA, GluTTC, LeuAAG as being upregulated and TrpCCA, LeuCAA, LeuCAG, GlnCTG, ArgACG, ArgTCG, TyrGTA, ThrTGT, AlaAGC, LysTTT, AlaCGC, AlaTGC, ThrAGT, GlnTTG as being downregulated (Table 5.6). Day 4 v Day 7 among fed-batch cultures presented PheGAA, GlyTCC, SeC(e)TCA, GlyCCC, LeuTAA, SerGCT, LeuTAG, AspGTC as upregulated tRNAs while ThrCGT, ThrAGT, GlnCTG, TyrGTA, ArgTCG, LysCTT, AlaAGC, ArgACG, GlnTTG resulted downregulated (Table 5.6).

Table 5.6 - Differentially expressed tRNAs in CHO-S dataset

Comparison	tRNA	log2FC	adj *p*-value
Batch v Fed-batch D4	-	-	-
Batch v Fed-batch D7	PheGAA	0.65	2.86E-29
	AlaCGC	0.34	7.67E-08
	AlaTGC	0.32	4.52E-08
	LeuTAA	0.15	2.12E-02
	GlnTTG	0.14	4.05E-03
	LeuCAG	0.13	2.08E-05
	ThrAGT	0.10	2.12E-02
	ArgTCT	-0.09	2.01E-02
	ArgTCG	-0.20	4.08E-02
	MetCAT	-0.23	2.45E-03
	HisGTG	-0.25	3.29E-02
	ArgACG	-0.27	1.54E-05
	AsnGTT	-0.30	1.02E-12
	IleTAT	-0.31	1.54E-05
D4 v D7 Batch	HisGTG	0.37	5.01E-09
	GlyGCC	0.34	1.09E-07
	AsnGTT	0.31	8.22E-16
	LeuTAG	0.29	5.77E-05
	AspGTC	0.28	8.65E-07
	SerGCT	0.27	1.43E-06
	SeC(e)TCA	0.26	5.89E-05
	ProTGG	0.22	4.93E-05
	CysGCA	0.19	1.38E-02
	GluTTC	0.17	1.79E-02
	LeuAAG	0.14	6.06E-03
	TrpCCA	-0.09	4.90E-02
	LeuCAA	-0.10	1.79E-02
	LeuCAG	-0.11	4.79E-02
	GlnCTG	-0.13	4.90E-02
	ArgACG	-0.14	1.99E-02
	ArgTCG	-0.17	3.63E-03
	TyrGTA	-0.21	2.07E-04
	ThrTGT	-0.24	5.21E-07
	AlaAGC	-0.26	3.82E-09
	LysTTT	-0.27	4.90E-02
	AlaCGC	-0.32	1.67E-05
	AlaTGC	-0.33	9.52E-07
	ThrAGT	-0.33	1.20E-10
	GlnTTG	-0.41	1.62E-24

Comparison	tRNA	log2FC	adj *p*-value
D4 v D7 Fed-batch	PheGAA	0.47	4.92E-09
	GlyTCC	0.29	1.09E-02
	SeC(e)TCA	0.27	1.61E-02
	GlyCCC	0.25	6.28E-04
	LeuTAA	0.25	8.51E-05
	SerGCT	0.16	4.23E-02
	LeuTAG	0.16	1.25E-02
	AspGTC	0.15	3.77E-02
	ThrCGT	-0.20	2.35E-02
	ThrAGT	-0.22	1.76E-02
	GlnCTG	-0.27	3.46E-02
	TyrGTA	-0.30	1.09E-02
	ArgTCG	-0.31	1.09E-02
	LysCTT	-0.32	6.22E-03
	AlaAGC	-0.33	1.45E-04
	ArgACG	-0.38	1.93E-06
	GlnTTG	-0.38	7.92E-04

Summary of the log2 fold change (log2FC) and adjusted *p*-value calculated with DESeq2 for each tRNA among the indicated comparison for the CHO-S dataset. Only adj *p*-values below 0.05 are reported, with no threshold for fold change.

5.2.10 Differential expression analysis among the Symphogen dataset shows the tRNA pool between cell lines at the same time point is consistent whilst there is a higher variability in the tRNA pool across culture

When comparing the tRNA abundancies from Day 12 of culture against those from Day 4 of culture from the same cell line, I found several differentially expressed tRNAs (Table 5.7). CysGCA was consistently downregulated in all comparisons except in the case of the Symphogen 3080 cell line, the cell line showing the highest Qp. Interestingly, PheGAA was upregulated at Day 12 of culture in the 3068 cell line with the CB7 feeding protocol and in the 3080 cell line, the two cell lines with the highest Qp. The CB6 feeding regime appears to generate wider variations in the tRNAs pool when compared to the CB7 feeding protocol, in particular among the proline codons ProAGG, ProCGG, ProTGG; all of these were upregulated at Day 12 of culture in both the 3478 and 3068 cell lines under CB6 feeding only (not in CB7 and likely reflects the differences in the feeding).

When comparing across the different cell lines at Day 4 of culture there were few changes in tRNA abundance with the exception of a small number of cell line comparisons when fed with the CB6 protocol. For the comparison of the 3478 v 3080 cell line tRNA abundancies under fed-batch CB6 day 4 of culture, the GluCTC, GluTTC, GlyGCC, HisGTG, LeuTAG, LysCTT,

ThrTGT, TyrGTA codons were observed to be differentially expressed. A comparison of the tRNA abundancies for the 3068 v 3080 cell lines on day 4 of culture with CB6 feeding revealed that the GluCTC, HisGTG, PheGAA were up-regulated whilst a comparison of the 3077 v 3080 cell lines with CB6 feeding revealed that the GluCTC and PheGAA codons were upregulated.

An analysis of the differential expression of tRNAs at Day 12 of culture revealed that the variations were even more limited than those observed at day 4 of culture. The PheGAA codon was found to be upregulated in 3478 cell line v 3068 cell line under CB7 feeding conditions, whilst this codon was also upregulated in the 3478 cell line v 3077 cell line on D12 of culture with CB6 feeding, and in the 3077 cell line v 3080 cell line on D12 with CB6 feeding. However, the PheGAA codon was downregulated in the 3068 cell line v 3077 cell line on D12 with CB6 feeding. When comparing the 3068 cell line and 3478 cell line tRNA expression under CB7 v CB6 feeding at both culture time points, only two codons, the GluCTC and GlyGCC, were observed to be differentially regulated. Overall, these results indicate a generalized stability of the available tRNA pool when comparing the same time point among different cell lines, while the 3068 and 3478 cell lines showed a higher variability under the CB6 feeding compared to CB7 feeding protocol.

Table 5.7 - Differentially expressed tRNAs in Symphogen dataset

Comparison	tRNA	log2FC	adj *p*-value
D12 v D4 3478 CB6	ProCGG	0.78	2.45E-22
	ThrCGT	0.76	1.47E-27
	ProTGG	0.73	8.80E-13
	ProAGG	0.69	2.28E-11
	TyrGTA	-0.62	5.84E-04
	GluCTC	-1.06	1.35E-24
	CysGCA	-1.93	1.90E-03
D12 v D4 3478 CB7	ProCGG	0.63	2.11E-10
	CysGCA	-0.7	4.06E-02
	GluCTC	-0.87	1.98E-07
D12 v D4 3068 CB6	ProCGG	0.76	4.43E-50
	ProAGG	0.68	2.24E-17
	ProTGG	0.67	7.57E-18
	ValTAC	0.64	3.75E-46
	ThrTGT	0.62	2.88E-48
	SeC(e)TCA	-0.64	1.58E-27
	eMetCAT	-0.71	2.63E-06
	GlnCTG	-0.71	6.03E-09
	HisGTG	-0.76	7.47E-12
	TyrGTA	-0.88	1.33E-12
	GluCTC	-1.02	6.98E-04
	CysGCA	-1.08	3.34E-85
D12 v D4 3068 CB7	PheGAA	0.85	1.44E-02
	CysGCA	-1.55	3.40E-51
D12 v D4 3080 CB6	PheGAA	0.6	5.37E-09
	AspGTC	-0.72	5.37E-09
D12 v D4 3077 CB6	GlnCTG	-0.65	4.13E-04
	CysGCA	-1.55	8.73E-39
3478 v 3068 D4 CB6	-	-	-
3478 v 3068 D4 CB7	-	-	-
3478 v 3077 D4 CB6	GluCTC	-0.70	1.37E-14
3478 v 3080 D4 CB6	LeuTAG	0.61	9.72E-26
	ThrTGT	0.59	7.33E-16
	GluTTC	-0.69	1.44E-22
	HisGTG	-0.7	4.07E-07
	LysCTT	-0.72	1.56E-09
	TyrGTA	-0.74	3.80E-08
	GlyGCC	-0.92	3.76E-12
	GluCTC	-1.5	2.70E-63
3068 v 3077 D4 CB6	-	-	-
3068 v 3080 D4 CB6	GluCTC	0.97	5.12E-08

Comparison	tRNA	log2FC	adj *p*-value
	HisGTG	0.64	1.29E-05
	PheGAA	0.63	1.12E-03
3077 v 3080 D4 CB6	GluCTC	0.76	1.66E-15
	PheGAA	1.03	3.80E-20
CB7 v CB6 3478 D4	GluCTC	0.70	6.92E-04
CB7 v CB6 3068 D4	-	-	-
CB7 v CB6 3478 D12	GlyGCC	0.87	3.08E-04
CB7 v CB6 3068 D12	-	-	-
3478 v 3068 D12 CB6	-	-	-
3478 v 3068 D12 CB7	PheGAA	0.59	5.00E-03
3478 v 3077 D12 CB6	PheGAA	0.90	3.34E-11
	LeuTAG	0.75	7.06E-20
3478 v 3080 D12 CB6	GlyGCC	-0.59	7.68E-03
3068 v 3077 D12 CB6	PheGAA	-1.06	3.04E-15
3068 v 3080 D12 CB6	-	-	-
3077 v 3080 D12 CB6	PheGAA	1	1.16E-134
	GluCTC	0.64	2.24E-08
	CysGCA	-0.95	1.37E-17

Summary of the log2 fold change (log2FC) and adjusted *p*-value calculated with DESeq2 for each tRNA among the indicated comparison in the Symphogen dataset. Only adj *p*-values below 0.05 and fold changes equal or above an absolute value of 1.5 are reported.

5.2.11 The use of tRNA abundancies to predict elongation speed shows longer decoding times than using tRNA gene copy numbers

In Section 1.5.1, the use of kinetic rate constants coupled with high-throughput techniques to calculate the decoding speed of a given coding sequence through the use of an elongation phase model was described. The sequenced tRNA abundancies were used to calculate the decoding times of each codon under different conditions, then based on either tRNA gene copy numbers or the tRNA abundance, the time required for the elongation of two Symphogen IgG1 molecules (expressed in the clones 3068 and 3080) and Etanercept (sequence reported in Appendix) was predicted using the model for both the heavy and the light chain (Table 5.8). The tRNA abundance model might be considered more accurate as this uses the actual abundancies of the different tRNAs available rather than the gene copy number. The elongation times based on gene copy number resulted in the shortest predicted elongation time for each molecule, predicting an elongation time of 211.3 s for the 3068 heavy chain (HC 3068), 212.8 s for the 3080 heavy chain (HC 3080), 112.5 s for the 3068 light chain (LC 3068) and 114.7 s for the 3080 light chain (LC 3080). When using tRNA abundance as determined in this study and collected for the CHO-S dataset at day 4 and day 7 of either batch or fed-batch culture, the predicted times to decode the full mRNA were between 237.8

and 247.4 s for HC 3068, between 241.0 and 250.1 s for HC 3080, between 127.4 and 132.7 s for LC 3068 and between 129.5 and 135.1 s for LC 3080. Interestingly, Batch D4 consistently yielded the slowest predicted decoding times for each coding sequence.

In addition to calculating the predicted total decoding time of a coding sequence, the model generates the fastest and the slowest possible codon combinations with the same amino acid output for the protein of interest. Once again, the predicted elongation times based on gene copy numbers resulted in the fastest for both the slowest and the fastest possible sequences, with the model calculating the decoding speeds using the gene copy numbers to be respectively 83.4 s for HC 3068, 84.2 s for HC 3080, 46.4 s for LC 3068, 45.5 s for LC 3080 and 374.8 s for HC 3068, 375.8 s for HC 3080, 199.4 s for LC 3068, 198.8 s for LC 3080. Sequenced tRNAs abundance based decoding time predictions when using the CHO-S dataset resulted in predictions of 103.0-109.1 s for HC 3068, 104.1-110.3 s for HC 3080, 54.5-57.2 s for LC 3068 and 54.5-57.5 s for LC 3080 for the fastest possible codon sequence while the times were 419.5-437.2 s for HC 3068, 422.7-440.2 s for HC 3080, 217.1-226.0 s for LC 3068 and 217.9-226.7 s for LC 3080 for the slowest possible sequence.

Collectively these results reveal a slower predicted decoding speed for all the considered coding sequences when using the quantified tRNAs pool as determined by sequencing as opposed to those predicted using gene copy numbers. While Batch culture D4 data consistently generated the slowest overall predicted decoding times, differences between batch and fed-batch culture at day 4 or day 7 were always under 5% of the total time. When focusing on the Etanercept sequence and overall decoding speeds, the model predicted 187.6 s based on tRNA gene copy numbers, 220.6 s based on batch day 7 and 212.6 s based on fed-batch day 7 tRNA abundancies, confirming the trend observed for the IgG1 sequences where the sequenced tRNA pool generated longer predicted decoding times compared to those from gene copy numbers. When investigating each individual tRNA decoding speed as depicted as a single dot in the prediction plot along with a regression trendline makes it possible to follow the decoding speed profile along the original coding sequence, and visualise the fastest and the slowest possible combinations of codons (Figure 5.7, Figure 5.8, Figure 5.9). For every simulation, the presence of short regions with a slower decoding speed followed by faster regions is evident in the original coding sequence, in particular in the first half of the transcript for the heavy chains and Etanercept sequences. This pattern is conserved in the slowest possible sequence, while the use of the fastest codons generates a smooth regression fit line along the entire transcript length.

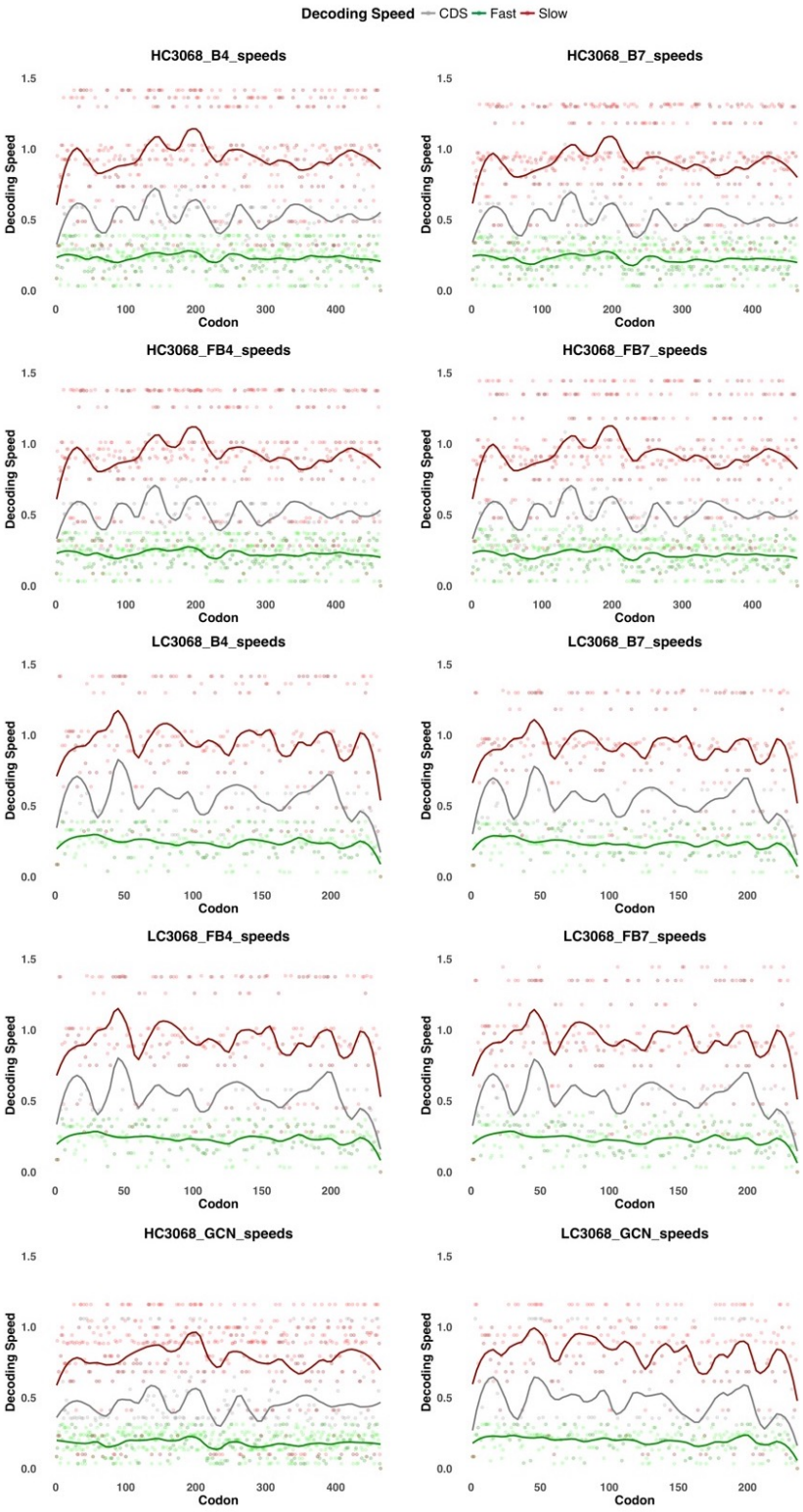


Figure 5.7 - Decoding speed simulations for 3068 IgG1

Decoding speeds in seconds for each codon of the 3068 IgG1 heavy chain (HC3068) and light chain (LC3068) coding sequences, calculated based on tRNA gene copy number (GCN), sequenced tRNA abundance at batch day 4 (B4), batch day 7 (B7), fed-batch day 4 (FB4) and fed-batch day 7 (FB7). For each graph, in green is indicated the fastest and in red the slowest possible combinations of codons generating the same amino acid sequence, while grey represents the original coding sequence. The dots indicate the decoding speed associated with each single codon while the lines are loess fit regression of the individual codons.

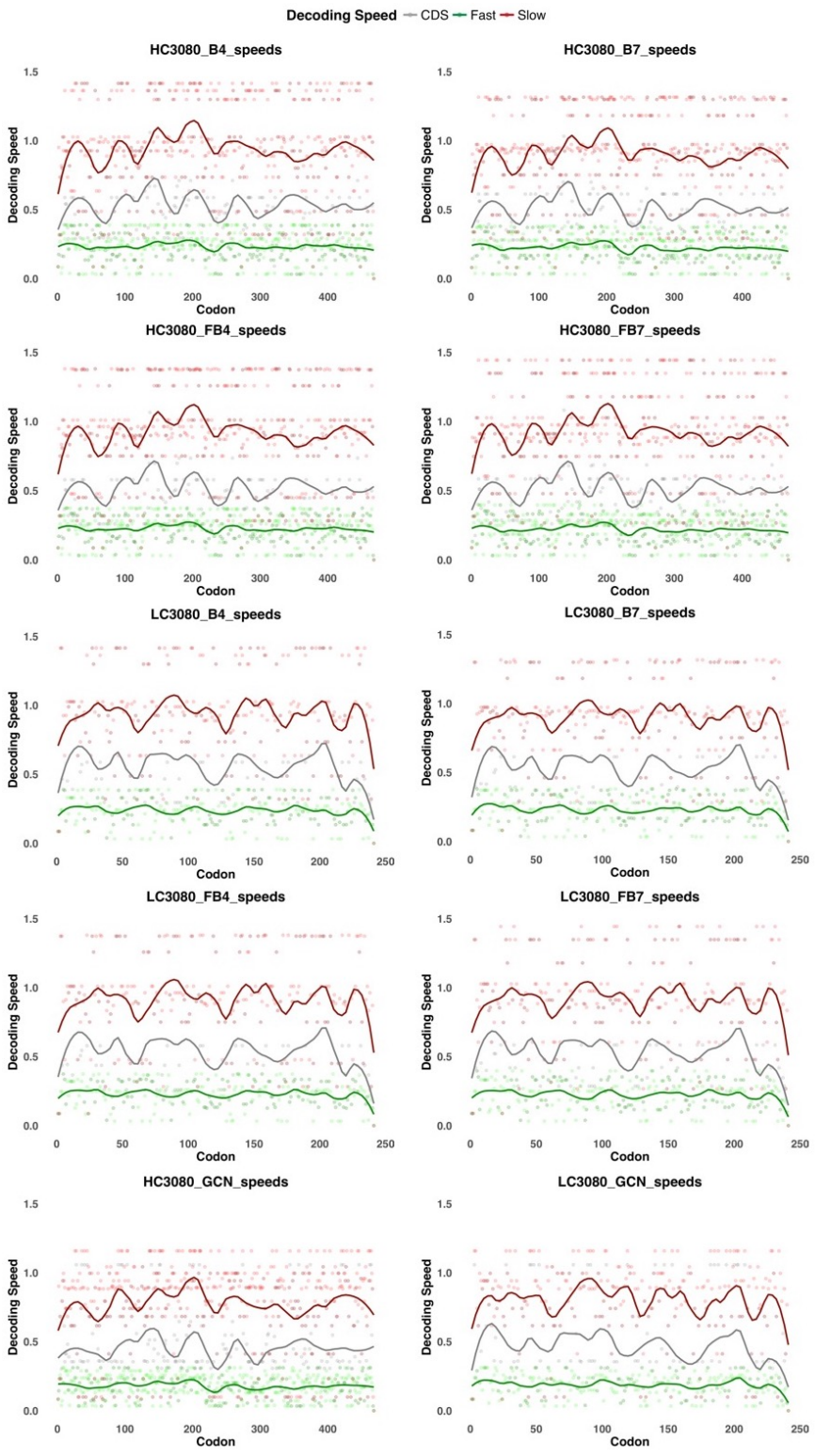


Figure 5.8 - Decoding speed simulations for 3080 IgG1

Decoding speeds in seconds for each codon of the 3080 IgG1 heavy chain (HC3080) and the light chain (LC3080) coding sequences, calculated based on tRNA gene copy number (GCN), sequenced tRNAs at batch day 4 (B4), batch day 7 (B7), fed-batch day 4 (FB4) and fed-batch day 7 (FB7) of culture. For each graph, in green is indicated the fastest and in red the slowest possible combination of codons generating the same amino acid sequence, while in grey is represented the original coding sequence. The dots indicate the decoding speed associated with each single codon while the lines are loess fit regression of the individual codons.

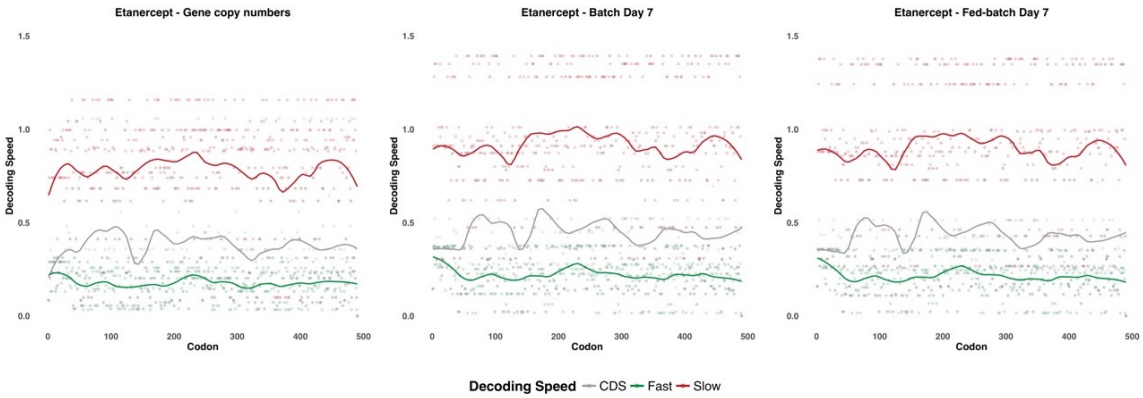


Figure 5.9 - Decoding speed simulations for Etanercept

Decoding speeds in seconds for each codon of the Etanercept coding sequence, calculated based on tRNA gene copy number (GCN), sequenced tRNAs at batch day 7 (B7), and fed-batch day 7 (FB7) of culture. For each graph, in green is indicated the fastest and in red the slowest possible combinations of codons generating the same aminoacidic sequence, while in grey is represented the original coding sequence. The dots indicate the decoding speed associated with each single codon while the lines are loess fit regression of the individual codons.

Table 5.8 - Decoding speeds for Symphogen antibodies and Etanercept sequences

Molecule	Condition	Fast sequence (s)	Original sequence (s)	Slow sequence (s)
HC 3068	GCN	83.4	211.3	374.8
HC 3068	Batch D4	109.1	247.4	437.2
HC 3068	Batch D7	104.1	237.8	419.5
HC 3068	Fed-batch D4	105.5	241.5	427.8
HC 3068	Fed-batch D7	103.0	240.0	428.6
LC 3068	GCN	46.4	112.5	199.4
LC 3068	Batch D4	57.2	132.7	226.0
LC 3068	Batch D7	55.5	127.4	217.1
LC 3068	Fed-batch D4	55.2	128.4	220.1
LC 3068	Fed-batch D7	54.5	127.6	220.6
HC 3080	GCN	84.2	212.8	375.8
HC 3080	Batch D4	110.3	250.1	440.2
HC 3080	Batch D7	105.5	241.0	422.7
HC 3080	Fed-batch D4	106.6	244.4	430.9
HC 3080	Fed-batch D7	104.1	242.8	431.8
LC 3080	GCN	45.5	114.7	198.8
LC 3080	Batch D4	57.5	135.1	226.7
LC 3080	Batch D7	55.4	129.5	217.9
LC 3080	Fed-batch D4	55.4	130.6	220.4
LC 3080	Fed-batch D7	54.5	129.7	221.4
Etanercept	GCN	87.3	187.6	398.3
Etanercept	Batch D7	109.2	220.6	450.2
Etanercept	Fed-batch D7	104.9	212.6	439.2

Summary of the decoding speed times in seconds predicted by the elongation model for the original coding sequence, the fastest and the slowest possible combination of codons with the same amino acid output of the heavy (HC) and light (LC) chains in the 3068 and 3080 IgG1s or Etanercept. Each decoding speed is simulated based on tRNA gene copy numbers, CHO-S cell tRNA abundancies from sequenced data at batch day 4 and day 7 or fed-batch day 4 and day 7.

5.2.12 HEK293 cold shock dataset

In addition to investigating tRNA abundance in CHO cells, during this study I also investigated tRNA abundance using the sequencing method described above in HEK293 cells and whether this changed upon cold-shock in these cells. Indeed, cooling cell cultures for a limited amount of time has been shown to alter translational efficiency, influencing the elongation phase in particular (Bastide et al., 2017). To assess if there were any variations induced on the tRNA pool by a 24 h temperature shift from 37^{o}C to 32^{o}C in Human Embryonic Kidney (HEK293) cells, we therefore applied the same ARM-Seq protocol described for the CHO cell lines. The HEK293 samples showed a correlation between tRNA gene copy numbers and sequenced

tRNAs very similar to the CHO-S samples described in paragraph 1.1.8, with r = 0.35, p-value = 0.0051 at 37^{o}C and r = 0.31, p-value = 0.012 at 32^{o}C (Figure 5.10). On the contrary, the correlation between codon usage in human and sequenced tRNAs was not statistically significant with r = 0.19, p-value = 0.13 at 37^{o}C and r = 0.18, p-value = 0.16 at 32^{o}C (Figure 5.10), mirroring the behaviour of the Symphogen cell lines described in Section 5.2.8. Thus, sequence design and codon usage based upon gene copy number should be improved in terms of elongation speed using tRNA abundancies as opposed to gene copy number based upon these data. Furthermore, both the 37^{o}C and 32^{o}C HEK293 samples showed a strong correlation with the Symphogen dataset (between 0.62 and 0.91) while the correlation with the CHO-S dataset was considerably lower (between 0.24 and 0.27). This difference was mainly due to two single tRNAs, MetCAT, whose quantification can be challenging as a result of the initiator-elongator ratio, and PheGAA. When these two tRNAs were excluded from the analysis, the correlation coefficient with the CHO-S dataset increased to 0.73-0.83, while the correlations measured against codon usage and gene copy numbers did not change significantly (Supplementary Material 5.5).

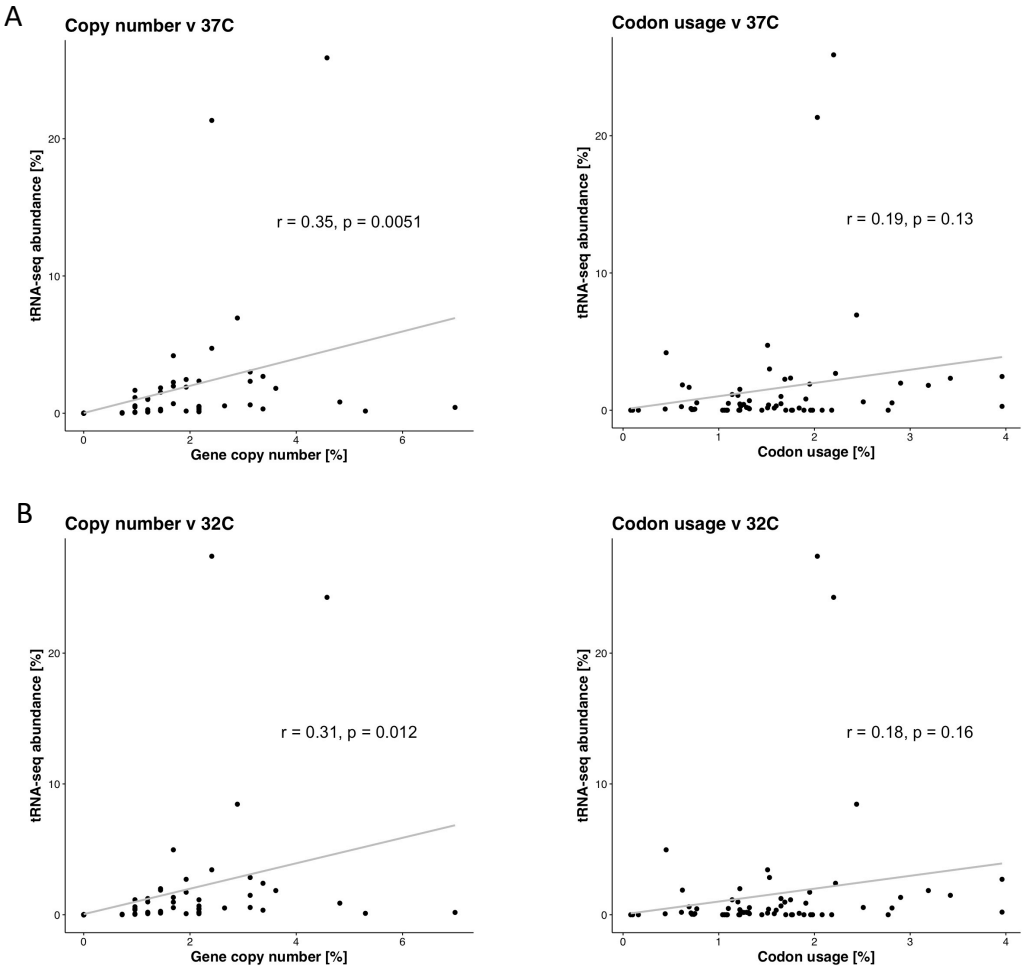


Figure 5.10 - Correlations between sequenced tRNAs, gene copy number and codon usage in HEK293

Pearson correlation between the sequenced HEK293 tRNAs, sampled at 37°C (Panel **A**) or 32°C (Panel **B**), and the tRNA gene copy numbers (left) or codon usage (right) in human. The coefficients for the gene copy number (r = 0.31, 0.35) indicate a positive correlation for the two samples while the correlation with codon usage is not significant (p = 0.13, 0.16).

5.3 Discussion

Translation is a key regulative step composed of three main phases (initiation, elongation, termination) resulting from the interaction of protein factors with RNAs in the translational machinery (Summarized in Figure 1.1). Codon optimization based on tRNA gene copy numbers and codon usage is widely used in recombinant protein production (Mauro, 2018), although the actual abundance of tRNA molecules in the cell is often overlooked due to technical limitations in their detection. The details and application of an adapted ARM-Seq protocol were described here to accurately measure tRNAs in a CHO-S and a variety of industrial CHO cell lines under different culture parameters, in addition to yeast and HEK293 cells. The results showed 4 to 5 fold improved detection of tRNAs compared to standard RNA-Seq (Figure 5.2A) while the detection of the remaining small RNAs was not impacted.

Yeast showed a strong correlation between sequenced tRNAs and gene copy numbers (Figure 5.2B), while only the CHO-S and HEK293 datasets presented statistically significant correlations (Figure 5.6A, Figure 5.10), with the Symphogen dataset suggesting a complete lack of correlation (Figure 5.5A).

When considering the correlation between sequenced tRNAs and codon usage among CHO cells (Figure 5.5A, Figure 5.6A) or human (Figure 5.10), I obtain a comparable outcome, where only CHO-S and HEK293 show statistically significant correlation. These discrepancies between the datasets could be due to the high genetic variability of CHO cells, leading to variable tRNA pools, although using alternative reference genomes to predict tRNA gene copy numbers did not impact the analysis (summarized in Table 5.3). In addition, between 74.3-88.5% of the total reads aligned to tRNAs in the Symphogen datasets compared to 28.9-48.9% in the CHO-S dataset, reflecting an actual biological difference in the tRNA pools or a disparity along the sequencing protocol, at the level of the initial de-methylation or further down during library generation. The expression of an IgG1 in the industrial CHO cell lines is unlikely to be the source of variation in the tRNAs pool as the null-producer 3478 did not deviate from the rest of the Symphogen samples with regard to correlations between sequenced tRNAs with tRNA gene copy numbers or codon usage.

When looking at statistically significant fold-changes, most of the variability arose from comparisons involving later stages in culture, either at Day 7 or Day 12, regardless of the type of feeding strategy. The availability of two CHO datasets allowed us to find common patterns between the heterogenous CHO cell lines considered in this study. We found AspGTC, CysGCA, GlnCTG, GluCTC, HisGTG, PheGAA, ProAGG, ProCGG, ProTGG, SeC(e)TCA, ThrCGT, ThrTGT, TyrGTA as differentially regulated in both datasets at the latest time point. While in the CHO-S dataset all significant log2 fold-changes were in the range between 0 and 0.65 (Table 5.6), in the Symphogen dataset the variability reached -1.93 log2 fold-change for CysGCA in 3478 CB6 Day 12 v Day 4 (Table 5.7) and was in general higher than the aforementioned CHO-S dataset. Although this could be due to the presence of a recombinant protein to produce in the Symphogen cell lines, the absence of a specific deviation of the null-producer 3478 from the rest of the cell lines suggests this does not impact the results. Nonetheless, this higher variability could be due to the specific steps required for industrial cell line selection and the intrinsic genetic instability of CHO cells, suggesting the importance of a tailored approach to codon optimization, particularly in industrial systems.

Indeed, what I investigated next is the impact of variations in the tRNAs pool on the translation of the heavy and light chain of Symphogen IgG1s and Etanercept using a

computational elongation model able to calculate the decoding speed of any coding sequence based on tRNA gene copy numbers or tRNA abundances (introduced in Section 1.5.1). I found a sustained trend towards longer decoding times for every considered coding sequence when sequenced tRNA abundances were used for the calculations (Table 5.8). The main reason behind this trend is the balance among cognate/near-cognate tRNAs, a parameter used in the model to predict the speed of elongation of each codon. Differences of over 35 seconds in the decoding speeds of the heavy chain found for both 3068 and 3080 IgG1s could heavily impact on tRNA availability, ribosome occupancy and recycling, mRNA stability, eventually influencing the overall productivity. The use of tailored approaches for codon optimization based on the tRNA pool of the specific cell line under the particular culture condition of interest could improve quality and quantity of the final product.

Chapter 6 Discussion and Future Work

CHO cell lines are the most widely used mammalian expression system for the production of complex recombinant biopharmaceuticals, due to their ability to grow in suspension and produce correctly folded proteins at a large scale with human compatible post-translational modifications and secrete the final product (Butler & Spearman, 2014; Dumont et al., 2016; Mauro, 2018). Although the CHO cell is generally an effective mammalian cell factory system for the production of secreted biotherapeutic proteins, the final yield of product is influenced by the interplay of environmental factors including the cellular machinery of a given clone and the site of integration of the exogenous recombinant genes, feeding of the cells during culture and culture temperature, the complexity of the recombinant protein and the regulation of biological pathways in the cell; cell growth, effective translation and secretion. mRNA translation in particular is a central regulator of cell proliferation, cellular processes and recombinant protein productivity determined in part by translation factors and ribosomes interacting with the mRNA to translate a given transcript, the availability of transfer RNAs and the regulation of non-coding RNAs, as summarized in Chapter 1. ncRNAs are indeed emerging as key regulators of many biological processes in the cell, where they can act in the nucleus at an epigenetic level to silence or to enhance single genes or even entire genomic regions, and in the cytoplasm, where they compete with miRNAs targets or selectively enhance/repress translation of specific mRNAs (Geisler & Coller, 2013). Hence, the work described in this thesis focused on providing the first landscapes of expression of in particular long non-coding RNAs under different culture conditions in CHO cells and relate the expression of these to functional pathways in the cell and identify new targets for cell manipulation (Chapters 3 and 4). At the same time, I wanted to investigate the tRNA pool response to the environment and how this affects elongation, the central phase of translation and whether a knowledge of tRNA abundance could be used to aid design of specific recombinant mRNA transcripts with decoding speeds (Chapter 5). The findings of each of these previous chapters have been discussed at the end of each chapter, here these are reviewed and collectively considered alongside suggestions for future work and directions.

6.1 Mapping and Investigation of Long non-coding RNA Expression in CHO Cells

6.1.1 CHO-S dataset

In Chapter 3 I presented the first published lncRNA expression profile in CHO cells under batch and fed-batch conditions together with the coding transcriptome using a comprehensive commercially available Mouse microarray, detecting 24603 lncRNAs (68.5% of the total probes) and 19617 mRNAs (78.8% of the total probes). Several hundred lncRNAs showed differential expression on different days of culture and in particular between Batch and Fed-batch culture (Section 3.3.2). While the profile of differential expression was similar between coding and non-coding genes, only 16-28% of the lncRNA probes had a matching transcript in CH, as opposed to 58-80% of the mRNAs. A combined approach was then taken to identify lncRNAs, whereby the Chinese hamster genome, literature search and database mining were jointly considered to find and relate lncRNAs differentially expressed in the CHO cell system with an established biological function. The main findings from this work are discussed in Chapter 3, however it is important to note that the results showed it is possible to successfully use mouse sequences as probes for microarray detection of CHO lncRNAs and, subsequently, identify unannotated non-coding transcripts through secondary structure conservation, homology with model organisms and RT-qPCR experimental validation.

Following the global analysis and RT-qPCR validation of a number of lncRNAs of interest, the study focused on the lncRNAs *MALAT1*, *NEAT1* and *PVT1* among the group of differentially expressed lncRNAs. These three non-coding genes are among the most well characterized in mouse and human, with an association of their expression with increased proliferation and regulation of alternative splicing and DNA repair reported, processes which we found to be strongly regulated in our pathway enrichment analysis. Further follow up work on the single non-coding genes is required to establish any impact the regulation of these has on the CHO cell phenotype, and if so, establish the exact biological mechanism(s) and the potential of these as targets for cell manipulation. Nevertheless, this work provided to the community the first comprehensive coding and non-coding dataset in CHO cells, particularly at the lncRNA level, allowing investigators to further probe the relationships between the expression and regulation of these two classes of RNA.

Additional work would now be to look at over-expression or knockdown of the expression of these lncRNAs and determining if this impacted CHO cell growth or recombinant protein productivity phenotypes. Further, RNA-Seq data can also be analysed for the presence of the mapped lncRNAs we have confirmed using the mouse array to determine if these change in

different CHO host cell lines and recombinant clones and how their expression relates to the observed phenotype. Collectively these studies have the potential to identify new targets for cell engineering and facilitate development of new, engineered CHO host cell systems with enhanced bioprocessing characteristics.

6.1.2 Symphogen datasets

In Chapter 4 I further explored the coding and non-coding expression landscape of 5 IgG1 producing CHO cell lines and 1 null pool (non-IgG expressing control) at different time points of culture cultivated under fed-batch conditions in an automated ambr15™ system. Two main datasets (DAVI and JCE) differentiated by feeding, time point collection and cell lines used were collected for analysis in this project. As opposed to the microarray dataset described in Chapter 3, here I decided to use RNA sequencing and to sample cells at later stages of culture, namely day 4 and 12 for DAVI and day 4 and 9 for JCE. In addition, in the DAVI dataset two different commercially available feeding strategies were compared, Cell Boost 6 and Cell Boost 7, while the JCE dataset included an early time point collected from the seed train flasks used to inoculate the fed-batch cultures.

Among the DAVI dataset, the cultured cell lines showed a range of cell specific productivities (Qp) which was clearly reflected in the PCA clustering of the transcript expression analysis and in the numbers of differentially expressed genes (Figure 4.2,

Figure 4.3). As already noted for the CHO-S cell dataset described in Chapter 3, feeding proved to be a significant source of variability even at an early stage of culture. Indeed, the comparison of the seed train flask data used to inoculate the fed-batch process in JCE shows, for the first time, that there is a significant change in gene expression after only 4 days of culture among the same clone when the cells are in exponential growth phase and before the major phase of production of the protein of interest. Following KEGG functional enrichment analysis of the RNA-Seq datasets, the difference in the Replication and Repair domain pathways confirmed the tendency (already noted for the CHO-S cell line, Chapter 3) to be differentially regulated towards the later stages of culture, with the exception of the 3080 cell line, the clone with the highest Qp.

Based on these observations I suggest a higher productive capacity for those cells that are able to maintain genome integrity for longer periods of culture, considering most of the recombinant protein production happens in the later stages (once the maximum viable cell concentration is achieved) of the fed-batch process. Further comparisons of our dataset with the literature suggested *Cd36*, *Hspa8*, *Serpinh1* and *Vim* as genes negatively correlated with

Qp, with heterogenous mechanisms of function but an overall detrimental effect on Qp. The impact of these could potentially be alleviate using Knock-Out (KO) or Knock-Down (KD) strategies with CRISPR or RNA interference. Along with coding genes, I identified, among many other lncRNAs, *Adapt15*, *GAS5* and *PVT1* as genes of particular interest, the latter being already identified and described in Chapter 3, as associated with higher Qp. Although these lncRNAs are well described in model organisms and human diseases (Colombo et al., 2015b; Hollander et al., 1996; Ma et al., 2016), they lack any form of annotation in CH of CHO cell genomes, or indeed, in any mammalian cell factory related to recombinant protein production. Therefore, the work presented in both Chapter 3 and Chapter 4 provides a base of knowledge for the mammalian recombinant protein production community to evaluate the impact of the long non-coding transcriptome on different systems and start working on single or groups of lncRNAs for cell engineering purposes.

6.1.3 Future work on lncRNAs

The biological significance and mechanisms of each single lncRNA in CHO cells remains to be experimentally investigated as most of our present knowledge comes from studies related to diseases where the setting and the environment are profoundly different. One of the challenges when investigating lncRNAs compared to coding genes is the lack of an open reading frame, which requires dedicated KO and KD approaches where the full-length gene is targeted by RNA interference or CRISPR. The study described here is likely to be of particular interest to the community as it has been made completely publicly available from controlled fed-batch ambr15™ datasets using industrial cell lines and feedings; such reports are rare in the literature, even less so with complete transcriptomics data. The work described here provides a comprehensive detailing of cell culture parameters and transcriptomic analysis and will be fully released and made available to the community for further studies. These studies should involve comparison of the analysis reported here with the expression of lncRNAs in other CHO hosts and recombinant cell lines and then the validation of manipulation of key, differentially regulated lncRNAs on CHO cell phenotypes.

6.2 tRNA abundance and mRNA translation

In Chapter 5 is described the exploration of one of the main parameters influencing translation, specifically elongation of the polypeptide, and the application of an elongation phase model to predict the decoding speed of any coding sequence of choice. Most codon optimization tools commonly use tRNA gene copy numbers and codon usage to select the

'best' sequence for optimal translation elongation and hence the 'best' or highest levels of recombinant protein expression. While this approach can lead to development of a coding sequence with codon usage that results in an effective increase in yield of protein for many protein molecules over that from the 'native' coding sequence, the magnitude and the effects on quantity and product quality are often unpredictable. One of the reasons for this was hypothesised to be the heterogeneity of the tRNA pool among different CHO cell lines used in industry and across culture, a parameter often overlooked due to the lack of information on tRNA abundancies and limitations of standard RNA-Seq protocols to detect and tRNAs.

6.2.1 Quantification of tRNA species through an optimized RNA-Seq protocol

During the studies described here, an adapted ARM-Seq protocol was designed for CHO and HEK293 cell lines to measure tRNA abundance and is described in Section 5.1. These measurements were then related to codon usage tRNA gene copy numbers in CHO cells (Section 5.2.8), showing a statistically significant correlation between tRNA abundance as determined by RNA-Seq and tRNA gene copy number for the CHO-S cell samples but not for the Symphogen dataset. The reasons for the discrepancy between the two datasets could be due to an actual biological difference in the tRNAs pools between the cell lines used in the experiments. Indeed, CHO cells are known to have a high genetic variability, which is likely to reflect at the tRNA level, as well as on the rest of the genome, causing CHO cell host and recombinant cell line specific tRNA expression profiles. A further reason could be the much higher percentage of reads aligned to tRNAs in the Symphogen datasets, ranging between 74.3-88.5% compared to 28.9-48.9% for the data from the CHO-S cells (Table 5.2). While the expression of the IgG1 in the Symphogen dataset could also have been considered a source that impacted on the variation of the tRNAs pool, the null-producer 3478 gave results similar to the rest of the cell clones, suggesting that the presence of large amounts of heavy and light chain transcript, and the translation of these, did not impact upon the tRNA abundancies observed.

The differences in the tRNA abundancies between the CHO-S and Symphogen cell lines with respect to the degree of correlation with the tRNA gene copy number has potential implications of design of sequences and codon selection between different host cells. Simply designing a recombinant sequence based on gene copy number alone for any given CHO cell host is unlikely to deliver the 'optimal' sequence and the data presented here, and discussed in the following section, suggests that the knowledge of an individual host cell lines tRNA

abundancies is likely to facilitate design of recombinant sequences tailored such individual cell lines for 'optimal' elongation and recombinant protein production.

In addition to investigating the CHO cell tRNA abundancies and how this changed throughout culture and with feeding, the HEK293 cell tRNA abundancies were also investigated (Section 5.2.12). The HEK293 cell dataset showed correlations similar to the CHO-S cell line for comparison of RNA-Seq determined tRNA abundancy with tRNA gene copy numbers, while a comparison with codon usage was not significant. Comparing the HEK293, CHO-S and Symphogen datasets showed a solid correlation between sequenced tRNAs, with the exception of the MetCAT and PheGAA tRNAs, which were sensibly higher in the HEK293 samples. While the quantification of MetCAT is likely to be influenced by the alignment of both the initiator and the elongator Met, due to a lack of a clear distinction in the annotation, biases in PheGAA quantification could be due to the presence of specific chemical modifications which impair or particularly enhance the de-methylation reaction efficiency. Further experimental validation on single tRNAs species will be required to elucidate the biological significance of the described results. Nonetheless, this represent the first attempt to quantitatively measure tRNAs in CHO cells in a variety of cell lines and conditions.

6.2.2 mRNA translation decoding speed simulation using an elongation model

Section 1.5.1 introduced the computational elongation model used in Section 5.2.11 to calculate the decoding speed of an IgG1 (both heavy and light chains) and Etanercept based on the CHO-S cell sequenced tRNA pool abundancies. In addition to calculating the predicted elongation decoding speed of a coding sequence for each codon, the model generates a prediction of the fastest and the slowest possible combination of codons that yield the same amino acid sequence. A graphical output of the simulated decoding speeds can be created fitting a Loess regression line along the length of the original sequence and the codon combinations generated by the model (Figure 5.7, Figure 5.8, Figure 5.9). I observed the alternation of regions with a slower decoding speed followed by faster regions in the original and the slowest predicted coding sequences, particularly towards the first half of the transcript for the heavy chains. On-the-other-hand, the predicted slowest possible combination of codons showed a smooth regression fit line along the entire transcript length. While a faster decoding speed should intuitively increase the translation rate of the mRNA and hence the synthesis of the protein of interest, the presence of naturally occurring regions with a slower decoding speed could be important for the correct co-translational folding of the nascent polypeptide. This speculation or hypothesis is strengthened by the prevalent

location of these variable speed regions at the beginning of the heavy chain, where ribosome pausing at particular domains could promote correct folding and avoid ribosome collisions further along the transcript. Experimental validation of the predicted sequences generated by the model will be essential to evaluate what translation rate is optimal for specific IgGs and other recombinant biotherapeutics, fine tuning the presence of regions with different decoding speed rates to increase final yield without impacting quality. This strategy would integrate current codon optimization approaches based on tRNA gene copy numbers and codon usage of the host organism only, allowing the generation of tailored approaches for the cell line specific tRNA pool.

6.2.3 Future work on tRNAs and translation

The work described in this thesis was aimed at integrating current codon optimization approaches with experimental measurement of the tRNA pool under various conditions. While this method allows a tailored optimization strategy for each different cell line, culture condition and molecule of interest, additional work is required to improve our capacity to predict translation elongation speed and the impact on the translation of specific transcripts. Along with tRNA abundance, the charge state of each tRNA is a key parameter to consider. Previous studies have showed most cytosolic tRNAs in HEK293T cells are charged at >80% levels, except for the tRNAs for tRNASer and tRNAThr that are charged at lower levels (Evans et al., 2017). Investigating the charge state of tRNAs in CHO cells would identify which tRNAs show a lower charge level and if this is constant through time and changes with different feed protocols.

Chemical modifications to nucleotides have also been shown to influence tRNA behavior during translation, depending on their position in the tRNA molecule (Section 1.3.5). The use of dedicated mass-spectroscopy techniques, already applied in other organisms to investigate and determine tRNA modifications (Leidel et al., 2009), would improve our understanding of how translation can be modulated without changing the relative abundances of each tRNA in a pool. tRNAs are particularly abundant in the cell, constituting up to 88.5% of the whole small RNA population (Section 5.2.3). Therefore, limited variations in terms of fold-change among a particular tRNA, in the order of FC of around 1.4, can massively alter the balance among the pool and have noticeable effects on cell metabolism (Pavon-Eternod, Gomes, Rosner, & Pan, 2013). This could explain the absence of particularly high FC values in our differential tRNA expression analysis when comparing cell lines with different productivities and feeding strategies (Sections 5.2.9 and 5.2.10). Overexpression of

single tRNAs will be fundamental to understand what level of upregulation is required to have an effect on cell growth or productivity in CHO cells, and if this level is constant for each tRNA species. In addition, the interplay of other factors involved in translation will have to be evaluated. mRNA stability and secondary structures can deeply influence the speed of elongation of particular transcripts (Gorochowski et al., 2015). This parameter becomes particularly important when we modify the codon combination of a given sequence to favour particular codons, potentially creating repetitions and secondary structures which could have a detrimental impact on translation. Changing of a transcript codon sequence, whilst not impacting the amino acid sequence coded for, could also impact on the turnover or half-life of specific transcripts that could subsequently impact on the yield of protein observed.

6.3 Overall Conclusions

Cell engineering in CHO cells, had until recently almost exclusively concentrated on coding genes, with the exception of miRNAs that have been manipulated by a number of groups to show that these can impact the growth and productivity characteristics of CHO cells. Some of the reasons for the lack of studies reporting manipulation of other non-coding RNAs and genes are the lack of sufficient annotation in CH and CHO cell genomes for lncRNAs compared to model organisms and the technical challenges posed by tRNA sequencing. Nonetheless, focusing on ncRNAs for cell engineering in CHO would bring several benefits due to the variety of their mechanisms of action, the absence of a translational burden applied on the cell when manipulating the expression of these molecules as opposed to manipulating coding genes, and the huge number of potential targets unexplored. During this PhD I have generated the first landscape of lncRNA expression in various CHO cell lines under different conditions, identified transcripts with a potential for cell engineering and improved their annotation. In parallel, I have adapted a tRNA sequencing protocol in CHO and HEK293 cells to quantify tRNAs and evaluated their impact on translation using a model of the elongation phase. All of this work has been, or will be, made publicly available to benefit the community and allow/facilitate further studies. I believe this effort will pave the way for both industry and academia to harness ncRNAs for cell engineering and cell line improvement.

Bibliography

Adriaens, C., & Marine, J.-C. (2017). NEAT1-containing paraspeckles: Central hubs in stress response and tumor formation. *Cell Cycle*, *16*(2), 137–138. https://doi.org/10.1080/15384101.2016.1235847

Adriaens, C., Standaert, L., Barra, J., Latil, M., Verfaillie, A., Kalev, P., … Marine, J.-C. (2016). p53 induces formation of NEAT1 lncRNA-containing paraspeckles that modulate replication stress response and chemosensitivity. *Nature Medicine*, *22*(8), 861–868. https://doi.org/10.1038/nm.4135

Alexa, A., & Rahnenfuhrer, J. (2016). topGO: enrichment analysis for gene ontology. R package version 2.28.0.

Ali, M. M., Akhade, V. S., Kosalai, S. T., Subhash, S., Statello, L., Meryet-Figuiere, M., … Kanduri, C. (2018). PAN-cancer analysis of S-phase enriched lncRNAs identifies oncogenic drivers and biomarkers. *Nature Communications*, *9*(1). https://doi.org/10.1038/s41467-018-03265-1

Alsayyari, A. A., Pan, X., Dalm, C., van der Veen, J. W., Vriezen, N., Hageman, J. A., … Martens, D. E. (2018). Transcriptome analysis for the scale-down of a CHO cell fed-batch process. *Journal of Biotechnology*, *279*(May), 61–72. https://doi.org/10.1016/j.jbiotec.2018.05.012

Alves, C. S., & Dobrowsky, T. M. (2017). Strategies and considerations for improving expression of “difficult to express” proteins in CHO cells. In P. Meleady (Ed.), *Heterologous Protein Production in CHO Cells: Methods and Protocols* (pp. 1–23). New York, NY: Springer New York. https://doi.org/10.1007/978-1-4939-6972-2_1

Amann, T., Hansen, A. H., Kol, S., Hansen, H. G., Arnsdorf, J., Nallapareddy, S., … Kildegaard, H. F. (2019). Glyco-engineered CHO cell lines producing alpha-1-antitrypsin and C1 esterase inhibitor with fully humanized N-glycosylation profiles. *Metabolic Engineering*, *52*, 143–152. https://doi.org/10.1016/j.ymben.2018.11.014

Amaral, P. P., Clark, M. B., Gascoigne, D. K., Dinger, M. E., & Mattick, J. S. (2011). lncRNAdb: a reference database for long noncoding RNAs. *Nucleic Acids Research*, *39*(Database issue), D146-51. https://doi.org/10.1093/nar/gkq1138

Ambros, V. (2004). The functions of animal microRNAs. *Nature*, *431*(7006), 350–355. https://doi.org/10.1038/nature02871

Ang, K. S., Kyriakopoulos, S., Li, W., & Lee, D.-Y. (2016). Multi-omics data driven analysis establishes reference codon biases for synthetic gene design in microbial and mammalian cells. *Methods (San Diego, Calif.)*, *102*, 26–35. https://doi.org/10.1016/j.ymeth.2016.01.016

Bachellerie, J. P., Cavaillé, J., & Hüttenhofer, A. (2002). The expanding snoRNA world. *Biochimie*, *84*(8), 775–790.

Baghban, R., Farajnia, S., Rajabibazl, M., Ghasemi, Y., Mafi, A., Hoseinpoor, R., … Aria, M. (2019). Yeast expression systems: Overview and recent advances. *Molecular Biotechnology*, *61*(5), 365–384. https://doi.org/10.1007/s12033-019-00164-8

Bareither, R., & Pollard, D. (2011). A review of advanced small-scale parallel bioreactor technology for accelerated process development: Current state and future need. *Biotechnology Progress*, *27*(1), 2–14. https://doi.org/10.1002/btpr.522

Barnett, J., & Barnett, L. (2011). Yeast research: A historical overview. *American Society of Microbiology*. https://doi.org/10.1128/9781555817152

Barron, N., Kumar, N., Sanchez, N., Doolan, P., Clarke, C., Meleady, P., … Clynes, M. (2011). Engineering CHO cell growth and recombinant protein productivity by overexpression of miR-7. *Journal of Biotechnology*, *151*(2), 204–211. https://doi.org/10.1016/j.jbiotec.2010.12.005

Bastide, A., Peretti, D., Knight, J. R. P., Grosso, S., Spriggs, R. V., Pichon, X., … Willis, A. E. (2017). RTN3 Is a novel cold-induced protein and mediates neuroprotective effects of RBM3. *Current Biology*, *27*(5). https://doi.org/10.1016/j.cub.2017.01.047

Bawa, Z., Bland, C. E., Bonander, N., Bora, N., Cartwright, S. P., Clare, M., … Bill, R. M. (2011). Understanding the yeast host cell response to recombinant membrane protein production. *Biochemical Society Transactions*, *39*(3), 719–723. https://doi.org/10.1042/BST0390719

Becker, J., Hackl, M., Rupp, O., Jakobi, T., Schneider, J., Szczepanowski, R., … Brinkrolf, K. (2011). Unraveling the Chinese hamster ovary cell line transcriptome by next-generation sequencing. *Journal of Biotechnology*, *156*(3), 227–235. https://doi.org/10.1016/j.jbiotec.2011.09.014

Bejerano, G., Pheasant, M., Makunin, I., Stephen, S., Kent, W. J., Mattick, J. S., & Haussler, D. (2004). Ultraconserved elements in the human genome. *Science (New York, N.Y.)*, *304*(5675), 1321–1325. https://doi.org/10.1126/science.1098119

Benson, D. A., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., & Wheeler, D. L. (2004). GenBank: update. *Nucleic Acids Research*, *32*(Database issue), D23-6. https://doi.org/10.1093/nar/gkh045

Betancur, J. G. (2016). Pervasive lncRNA binding by epigenetic modifying complexes — The challenges ahead. *Biochimica et Biophysica Acta*, *1859*(1), 93–101. https://doi.org/10.1016/j.bbagrm.2015.10.009

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*(7), 741–749. https://doi.org/10.1002/bit.20025

Bobrowicz, P., Davidson, R. C., Li, H., Potgieter, T. I., Nett, J. H., Hamilton, S. R., … Wildt, S. (2004). Engineering of an artificial glycosylation pathway blocked in core oligosaccharide assembly in the yeast *Pichia pastoris*: production of complex humanized glycoproteins with terminal galactose. *Glycobiology*, *14*(9), 757–766. https://doi.org/10.1093/glycob/cwh104

Bonander, N., & Bill, R. M. (2012). Optimising yeast as a host for recombinant protein production (Review). In *Methods in molecular biology (Clifton, N.J.)* (Vol. 866, pp. 1–9). https://doi.org/10.1007/978-1-61779-770-5_1

Bort, J. A. H., Hackl, M., Höflmayer, H., Jadhav, V., Harreither, E., Kumar, N., … Borth, N. (2012). Dynamic mRNA and miRNA profiling of CHO-K1 suspension cell cultures. *Biotechnology Journal*, *7*(4), 500–515. https://doi.org/10.1002/biot.201100143

Brookheart, R. T., Michel, C. I., Listenberger, L. L., Ory, D. S., & Schaffer, J. E. (2009). The non-coding RNA gadd7 is a regulator of lipid-induced oxidative and endoplasmic reticulum stress. *The Journal of Biological Chemistry*, *284*(12), 7446–7454. https://doi.org/10.1074/jbc.M806209200

Brule, C. E., & Grayhack, E. J. (2017). Synonymous codons: choose wisely for expression. *Trends in Genetics*, *33*(4), 283–297. https://doi.org/10.1016/j.tig.2017.02.001

Buch, S. C., Diergaarde, B., Nukui, T., Day, R. S., Siegfried, J. M., Romkes, M., & Weissfeld, J. L. (2012). Genetic variability in DNA repair and cell cycle control pathway genes and risk of smoking-related lung cancer. *Molecular Carcinogenesis*, *51*(S1), E11–E20. https://doi.org/10.1002/mc.20858

Burch, P. A., Breen, J. K., Buckner, J. C., Gastineau, D. A., Kaur, J. A., Laus, R. L., … Vuk-Pavlović, S. (2000). Priming tissue-specific cellular immunity in a phase I trial of autologous dendritic cells for prostate cancer. *Clinical Cancer Research*, *6*(6), 2175–2182.

Butler, M., & Spearman, M. (2014). The choice of mammalian cell host and possibilities for glycosylation engineering. *Current Opinion in Biotechnology*, *30*, 107–112. https://doi.org/10.1016/J.COPBIO.2014.06.010

Cabili, M. N., Dunagin, M. C., McClanahan, P. D., Biaesch, A., Padovan-Merhar, O., Regev, A., … Raj, A. (2015). Localization and abundance analysis of human lncRNAs at single-cell and single-molecule resolution. *Genome Biology*, *16*(1), 20. https://doi.org/10.1186/s13059-015-0586-4

Carrieri, C., Cimatti, L., Biagioli, M., Beugnet, A., Zucchelli, S., Fedele, S., … Gustincich, S. (2012). Long non-coding antisense RNA controls Uchl1 translation through an embedded SINEB2 repeat. *Nature*. https://doi.org/10.1038/nature11508

Carrió, M. M., & Villaverde, A. (2002). Construction and deconstruction of bacterial inclusion bodies. *Journal of Biotechnology*, *96*(1), 3–12.

Chan, P. P., & Lowe, T. M. (2016). GtRNAdb 2.0: an expanded database of transfer RNA genes identified in complete and draft genomes. *Nucleic Acids Research*, *44*(D1), D184–D189. https://doi.org/10.1093/nar/gkv1309

Cheever, M. A., & Higano, C. S. (2011). PROVENGE (Sipuleucel-T) in prostate cancer: The first FDA-approved therapeutic cancer vaccine. *Clinical Cancer Research*, *17*(11), 3520–3526. https://doi.org/10.1158/1078-0432.CCR-10-3126

Chen, C., Le, H., & Goudar, C. T. (2017). Evaluation of two public genome references for chinese hamster ovary cells in the context of rna-seq based gene expression analysis. *Biotechnology and Bioengineering*, *114*(7), 1603–1613. https://doi.org/10.1002/bit.26290

Chiu, H.-S., Somvanshi, S., Patel, E., Sood, A. K., Gunaratne, P. H., Correspondence, P. S., … Sumazin, P. (2018). Pan-cancer analysis of lncRNA regulation supports their targeting of cancer genes in each tumor context. *Cell Reports*, *23*, 297–312. https://doi.org/10.1016/j.celrep.2018.03.064

Choi, S. S., Rhee, W. J., Kim, E. J., & Park, T. H. (2006). Enhancement of recombinant protein production in Chinese hamster ovary cells through anti-apoptosis engineering using 30Kc6 gene. *Biotechnology and Bioengineering*, *95*(3), 459–467. https://doi.org/10.1002/bit.21023

Chong, W. P. K., Reddy, S. G., Yusufi, F. N. K., Lee, D.-Y., Wong, N. S. C., Heng, C. K., … Ho, Y. S. (2010). Metabolomics-driven approach for the improvement of Chinese hamster ovary cell growth: Overexpression of malate dehydrogenase II. *Journal of Biotechnology*, *147*(2), 116–121. https://doi.org/10.1016/J.JBIOTEC.2010.03.018

Chong, W. P. K., Sim, L. C., Wong, K. T. K., & Yap, M. G. S. (2009). Enhanced IFNgamma production in adenosine-treated CHO cells: a mechanistic study. *Biotechnology Progress*, *25*(3), 866–873. https://doi.org/10.1002/btpr.118

Chu, D., Zabet, N., & von der Haar, T. (2012). A novel and versatile computational tool to model translation. *Bioinformatics*, *28*(2), 292–293. https://doi.org/10.1093/bioinformatics/btr650

Chu, Dominique, Barnes, D. J., & von der Haar, T. (2011). The role of tRNA and ribosome competition in coupling the expression of different mRNAs in *Saccharomyces cerevisiae*. *Nucleic Acids Research*, *39*(15), 6705–6714. https://doi.org/10.1093/nar/gkr300

Chu, Dominique, Kazana, E., Bellanger, N., Singh, T., Tuite, M. F., & von der Haar, T. (2014). Translation elongation can control translation initiation on eukaryotic mRNAs. *The EMBO Journal*, *33*(1), 21–34. https://doi.org/10.1002/embj.201385651

Chujo, T., Yamazaki, T., & Hirose, T. (2016). Architectural RNAs (arcRNAs): A class of long noncoding RNAs that function as the scaffold of nuclear bodies. *Biochimica et Biophysica Acta*, *1859*(1), 1–8. https://doi.org/10.1016/j.bbagrm.2015.05.007

Chung, B. K.-S., Yusufi, F. N. K., Mariati, Yang, Y., & Lee, D.-Y. (2013). Enhanced expression of codon optimized interferon gamma in CHO cells. *Journal of Biotechnology*, *167*(3), 326–333. https://doi.org/10.1016/j.jbiotec.2013.07.011

Chung, B. K. S., Yusufi, F. N. K., Mariati, Yang, Y., & Lee, D. Y. (2013). Enhanced expression of codon optimized interferon gamma in CHO cells. *Journal of Biotechnology*, *167*(3), 326–333. https://doi.org/10.1016/j.jbiotec.2013.07.011

Clark, M. B., Mercer, T. R., Bussotti, G., Leonardi, T., Haynes, K. R., Crawford, J., … Dinger, M. E. (2015). Quantitative gene profiling of long noncoding RNAs with targeted RNA sequencing. *Nature Methods*, *12*(4), 339–342. https://doi.org/10.1038/nmeth.3321

Clarke, C., Doolan, P., Barron, N., Meleady, P., O'Sullivan, F., Gammell, P., … Clynes, M. (2011). Large scale microarray profiling and coexpression network analysis of CHO cells identifies transcriptional modules associated with growth and productivity. *Journal of Biotechnology*, *155*(3), 350–359. https://doi.org/10.1016/J.JBIOTEC.2011.07.011

Clarke, C., Henry, M., Doolan, P., Kelly, S., Aherne, S., Sanchez, N., … Barron, N. (2012). Integrated miRNA, mRNA and protein expression analysis reveals the role of post-transcriptional regulation in controlling CHO cell growth rate. *BMC Genomics*, *13*(1), 656. https://doi.org/10.1186/1471-2164-13-656

Colombo, T., Farina, L., Macino, G., & Paci, P. (2015a). PVT1: a rising star among oncogenic long noncoding RNAs. *BioMed Research International*, *2015*, 304208. https://doi.org/10.1155/2015/304208

Colombo, T., Farina, L., Macino, G., & Paci, P. (2015b). PVT1: A rising star among oncogenic long noncoding RNAs. *BioMed Research International*, *2015*, 17–21. https://doi.org/10.1155/2015/304208

Contreras-Gómez, A., Sánchez-Mirón, A., García-Camacho, F., Molina-Grima, E., & Chisti, Y. (2014). Protein production using the baculovirus-insect cell expression system. *Biotechnology Progress*, *30*(1), 1–18. https://doi.org/10.1002/btpr.1842

Cost, G. J., Freyvert, Y., Vafiadis, A., Santiago, Y., Miller, J. C., Rebar, E., … Gregory, P. D. (2010). BAK and BAX deletion using zinc-finger nucleases yields apoptosis-resistant CHO cells. *Biotechnology and Bioengineering*, *105*(2), 330–340. https://doi.org/10.1002/bit.22541

Courtes, F. C., Lin, J., Lim, H. L., Ng, S. W., Wong, N. S. C., Koh, G., … Lee, D.-Y. (2013). Translatome analysis of CHO cells to identify key growth genes. *Journal of Biotechnology*, *167*(3), 215–224. https://doi.org/10.1016/j.jbiotec.2013.07.010

Cox, M. M. J., & Hollister, J. R. (2009). FluBlok, a next generation influenza vaccine manufactured in insect cells. *Biologicals*, *37*(3), 182–189. https://doi.org/10.1016/j.biologicals.2009.02.014

Cozen, A. E., Quartley, E., Holmes, A. D., Hrabeta-Robinson, E., Phizicky, E. M., & Lowe, T. M. (2015). ARM-seq: AlkB-facilitated RNA methylation sequencing reveals a complex landscape of modified tRNA fragments. *Nature Methods*, *12*(9), 879–884. https://doi.org/10.1038/nmeth.3508

Crawford, D. R., Schools, G. P., Salmon, S. L., & Davies, K. J. A. (1996). Hydrogen peroxide induces the expression of adapt15,a novel RNA associated with polysomes in hamster HA-1 cells. *Archives of Biochemistry and Biophysics*, *325*(2), 256–264. https://doi.org/10.1006/ABBI.1996.0032

Cunningham, F., Amode, M. R., Barrell, D., Beal, K., Billis, K., Brent, S., … Flicek, P. (2015). Ensembl 2015. *Nucleic Acids Research*, *43*(Database issue), D662-9. https://doi.org/10.1093/nar/gku1010

Daegelen, P., Studier, F. W., Lenski, R. E., Cure, S., & Kim, J. F. (2009). Tracing ancestors and relatives of *Escherichia coli* B, and the derivation of B strains REL606 and BL21(DE3). *Journal of Molecular Biology*, *394*(4), 634–643. https://doi.org/10.1016/j.jmb.2009.09.022

Dall, E., & Brandstetter, H. (2016). Structure and function of legumain in health and disease. *Biochimie*, *122*, 126–150. https://doi.org/10.1016/J.BIOCHI.2015.09.022

Dana, H., Chalbatani, G. M., Mahmoodzadeh, H., Karimloo, R., Rezaiean, O., Moradzadeh, A., … Gharagouzlo, E. (2017). Molecular mechanisms and biological functions of siRNA. *International Journal of Biomedical Science*, *13*(2), 48–57.

Davies, J., Jiang, L., Pan, L.-Z., LaBarre, M. J., Anderson, D., & Reff, M. (2001). Expression of GnTIII in a recombinant anti-CD20 CHO production cell line: Expression of antibodies with altered glycoforms leads to an increase in ADCC through higher affinity for FC?RIII. *Biotechnology and Bioengineering*, *74*(4), 288–294. https://doi.org/10.1002/bit.1119

de Marco, A. (2009). Strategies for successful recombinant expression of disulfide bond-dependent proteins in *Escherichia coli*. *Microbial Cell Factories*, *8*(1), 26. https://doi.org/10.1186/1475-2859-8-26

Derman, A. I., Prinz, W. A., Belin, D., & Beckwith, J. (1993). Mutations that allow disulfide bond formation in the cytoplasm of *Escherichia coli*. *Science*, *262*(5140), 1744–1747.

Deschuyteneer, M., Elouahabi, A., Plainchamp, D., Plisnier, M., Soete, D., Corazza, Y., … Deschamps, M. (2010). Molecular and structural characterization of the L1 virus-like particles that are used as vaccine antigens in *Cervarix* ™, the AS04-adjuvanted HPV-16 and -18 cervical cancer vaccine. *Human Vaccines*, *6*(5), 407–419. https://doi.org/10.4161/hv.6.5.11023

Djebali, S., Davis, C. A., Merkel, A., Dobin, A., Lassmann, T., Mortazavi, A., … Gingeras, T. R. (2012). Landscape of transcription in human cells. *Nature*, *489*(7414), 101–108. https://doi.org/10.1038/nature11233

Doherty, A. J., Connolly, B. A., & Worrall, A. F. (1993). Overproduction of the toxic protein, bovine pancreatic DNaseI, in *Escherichia coli* using a tightly controlled T7-promoter-based vector. *Gene*, *136*(1–2), 337–340.

Dong, H., Nilsson, L., & Kurland, C. G. (1995). Gratuitous overexpression of genes in *Escherichia coli* leads to growth inhibition and ribosome destruction. *Journal of Bacteriology*, *177*(6), 1497–1504. https://doi.org/10.1128/jb.177.6.1497-1504.1995

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*(4), 853–866. https://doi.org/10.1002/bit.22990

Duina, A. A., Miller, M. E., & Keeney, J. B. (2014). Budding yeast for budding geneticists: A primer on the *Saccharomyces cerevisiae* model system. *Genetics*, *197*(1), 33–48. https://doi.org/10.1534/genetics.114.163188

Dumont, J., Euwart, D., Mei, B., Estes, S., & Kshirsagar, R. (2016). Human cell lines for biopharmaceutical manufacturing: history, status, and future perspectives. *Critical Reviews in Biotechnology*, *36*(6), 1110–1122. https://doi.org/10.3109/07388551.2015.1084266

Durocher, Y., & Butler, M. (2009). Expression systems for therapeutic glycoprotein production. *Current Opinion in Biotechnology*, *20*(6), 700–707. https://doi.org/10.1016/j.copbio.2009.10.008

Evans, M. E., Clark, W. C., Zheng, G., & Pan, T. (2017). Determination of tRNA aminoacylation levels by high-throughput sequencing. *Nucleic Acids Research*, *45*(14). https://doi.org/10.1093/nar/gkx514

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*(4), 1007–1015. https://doi.org/10.1002/bit.24365

Faustrup Kildegaard, H., Baycin-Hizal, D., Lewis, N. E., & Betenbaugh, M. J. (2013). The emerging CHO systems biology era: harnessing the 'omics revolution for biotechnology. *Current Opinion in Biotechnology*, *24*, 1102–1107. https://doi.org/10.1016/j.copbio.2013.02.007

Fernandes, F., Teixeira, A. P., Carinhas, N., Carrondo, M. J., & Alves, P. M. (2013). Insect cells as a production platform of complex virus-like particles. *Expert Review of Vaccines*, *12*(2), 225–236. https://doi.org/10.1586/erv.12.153

Fernández, F. J., López-Estepa, M., Querol-García, J., & Vega, M. C. (2016). Production of protein complexes in non-methylotrophic and methylotrophic yeasts. In *Advanced Technologies for Protein Complex Production and Characterization* (pp. 137–153). Springer. https://doi.org/10.1007/978-3-319-27216-0_9

Fernández, F. J., & Vega, M. C. (2016). Choose a Suitable Expression Host: A Survey of Available Protein Production Platforms. In *Advanced Technologies for Protein Complex Production and Characterization* (pp. 15–24). Springer, Cham. https://doi.org/10.1007/978-3-319-27216-0_2

Ferrara, C., Brünker, P., Suter, T., Moser, S., Püntener, U., & Umaña, P. (2006). Modulation of therapeutic antibody effector functions by glycosylation engineering: Influence of Golgi enzyme localization domain and co-expression of heterologous β1, 4-N-acetylglucosaminyltransferase III and Golgi α-mannosidase II. *Biotechnology and Bioengineering*, *93*(5), 851–861. https://doi.org/10.1002/bit.20777

Finer-Moore, J., Czudnochowski, N., O'Connell, J. D., Wang, A. L., & Stroud, R. M. (2015). Crystal structure of the Human tRNA m1A58 methyltransferase–tRNA3Lys complex: refolding of substrate tRNA allows access to the methylation target. *Journal of Molecular Biology*, *427*(24), 3862–3876. https://doi.org/10.1016/j.jmb.2015.10.005

Fischer, S., Handrick, R., & Otte, K. (2015). The art of CHO cell engineering: A comprehensive retrospect and future perspectives. *Biotechnology Advances*, *33*(8), 1878–1896. https://doi.org/10.1016/j.biotechadv.2015.10.015

Fischer, S., Marquart, K. F., Pieper, L. A., Fieder, J., Gamer, M., Gorr, I., … Bradl, H. (2017). miRNA engineering of CHO cells facilitates production of difficult-to-express proteins and increases success in cell line development. *Biotechnology and Bioengineering*, *114*(7), 1495–1510. https://doi.org/10.1002/bit.26280

Fogolín, M. B., Wagner, R., Etcheverrigaray, M., & Kratje, R. (2004). Impact of temperature reduction and expression of yeast pyruvate carboxylase on hGM-CSF-producing CHO cells. *Journal of Biotechnology*, *109*(1–2), 179–191. https://doi.org/10.1016/J.JBIOTEC.2003.10.035

Foster, K. G., & Fingar, D. C. (2010). Mammalian target of rapamycin (mTOR): conducting the cellular signaling symphony. *The Journal of Biological Chemistry*, *285*(19), 14071–14077. https://doi.org/10.1074/jbc.R109.094003

Fussenegger, M., Mazur, X., & Bailey, J. E. (1997). A novel cytostatic process enhances the productivity of Chinese hamster ovary cells. *Biotechnology and Bioengineering*, *55*(6), 927–939. https://doi.org/10.1002/(SICI)1097-0290(19970920)55:6<927::AID-BIT10>3.0.CO;2-4

Gamble, C. E., Brule, C. E., Dean, K. M., Fields, S., Grayhack Correspondence, E. J., & Grayhack, E. J. (2016). Adjacent codons act in concert to modulate translation efficiency in yeast. *Cell*, *166*, 679–690. https://doi.org/10.1016/j.cell.2016.05.070

Geisler, S., & Coller, J. (2013). RNA in unexpected places: long non-coding RNA functions in diverse cellular contexts. *Nature Reviews. Molecular Cell Biology*, *14*(11), 699–712. https://doi.org/10.1038/nrm3679

Gillespie, D. T. (1977). Exact stochastic simulation of coupled chemical reactions. *The Journal of Physical Chemistry*, *81*(25), 2340–2361. https://doi.org/10.1021/j100540a008

Gingold, H., Tehler, D., Christoffersen, N. R., Nielsen, M. M., Asmar, F., Kooistra, S. M., … Pilpel, Y. (2014). A dual program for translation regulation in cellular proliferation and differentiation. *Cell*, *158*(6), 1281–1292. https://doi.org/10.1016/j.cell.2014.08.011

Godfrey, C. L., Mead, E. J., Daramola, O., Dunn, S., Hatton, D., Field, R., … Smales, C. M. (2017). Polysome profiling of mAb producing CHO cell lines links translational control of cell proliferation and recombinant mRNA loading onto ribosomes with global and recombinant protein synthesis. *Biotechnology Journal*, *12*(8), 1700177. https://doi.org/10.1002/biot.201700177

Gogakos, T., Brown, M., Garzia, A., Meyer, C., Hafner, M., & Tuschl, T. (2017). Characterizing expression and processing of precursor and mature human tRNAs by Hydro-tRNAseq and PAR-CLIP. *Cell Reports*, *20*(6), 1463–1475. https://doi.org/10.1016/j.celrep.2017.07.029

Gomez, J. A., Wapinski, O. L., Yang, Y. W., Bureau, J. F., Gopinath, S., Monack, D. M., … Kirkegaard, K. (2013). The NeST long ncRNA controls microbial susceptibility and epigenetic activation of the interferon-γ locus. *Cell*, *152*(4), 743–754. https://doi.org/10.1016/j.cell.2013.01.015

Gonzalez, I., Munita, R., Agirre, E., Dittmer, T. a, Gysling, K., Misteli, T., & Luco, R. F. (2015). A lncRNA regulates alternative splicing via establishment of a splicing-specific chromatin signature. *Nature Structural & Molecular Biology*, *22*(April), 1–10. https://doi.org/10.1038/nsmb.3005

Gorochowski, T. E., Ignatova, Z., Bovenberg, R. A. L., & Roubos, J. A. (2015). Trade-offs between tRNA abundance and mRNA secondary structure support smoothing of translation elongation rate. *Nucleic Acids Research*, *43*(6), 3022–3032. https://doi.org/10.1093/nar/gkv199

Griffith, A., Kelly, P. S., Vencken, S., Lao, N. T., Greene, C. M., Clynes, M., & Barron, N. (2018). miR-CATCH identifies biologically active miRNA regulators of the pro-survival gene XIAP, in Chinese Hamster Ovary cells. *Biotechnology Journal*, *13*(3), 1700299. https://doi.org/10.1002/biot.201700299

Gromadski, K. B., & Rodnina, M. V. (2004). Kinetic determinants of high-fidelity tRNA discrimination on the ribosome. *Molecular Cell*, *13*(2), 191–200.

Gustafsson, C., Minshull, J., Govindarajan, S., Ness, J., Villalobos, A., & Welch, M. (2012). Engineering genes for predictable protein expression. *Protein Expression and Purification*, *83*(1), 37–46. https://doi.org/10.1016/J.PEP.2012.02.013

Guttman, M., Amit, I., Garber, M., French, C., Lin, M. F., Feldser, D., … Lander, E. S. (2009). Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature*, *458*(7235), 223–227.

Hackl, M., Borth, N., & Grillari, J. (2012, August). miRNAs--pathway engineering of CHO cell factories that avoids translational burdening. *Trends in Biotechnology*. England. https://doi.org/10.1016/j.tibtech.2012.05.002

Han, P., & Chang, C.-P. (2015). Long non-coding RNA and chromatin remodeling. *RNA Biology*, *6286*(July 2015), 00–00. https://doi.org/10.1080/15476286.2015.1063770

Hanson, G., & Coller, J. (2017). Codon optimality, bias and usage in translation and mRNA decay. *Nature Reviews Molecular Cell Biology*, *19*(1), 20–30. https://doi.org/10.1038/nrm.2017.91

Hanson, G., & Coller, J. (2018). Translation and Protein Quality Control: Codon optimality, bias and usage in translation and mRNA decay. *Nature Reviews Molecular Cell Biology*, *19*(1), 20–30. https://doi.org/10.1038/nrm.2017.91

Harreither, E., Hackl, M., Pichler, J., Shridhar, S., Auer, N., Łabaj, P. P., … Borth, N. (2015). Microarray profiling of preselected CHO host cell subclones identifies gene expression patterns associated with increased production capacity. *Biotechnology Journal*, n/a-n/a. https://doi.org/10.1002/biot.201400857

Harrison, R. L., & Jarvis, D. L. (2006). Protein N-Glycosylation in the Baculovirus–insect cell expression system and engineering of insect cells to produce “Mammalianized” recombinant glycoproteins. In *Advances in virus research* (Vol. 68, pp. 159–191). https://doi.org/10.1016/S0065-3527(06)68005-6

Hartley, D. L., & Kane, J. F. (1988). Properties of inclusion bodies from recombinant *Escherichia coli*. *Biochemical Society Transactions*, *16*(2), 101–102. https://doi.org/10.1042/BST0160101

Hetz, C., & Papa, F. R. (2018). The unfolded protein response and cell fate control. *Molecular Cell*, *69*(2), 169–181. https://doi.org/10.1016/j.molcel.2017.06.017

Hezroni, H., Koppstein, D., Bartel, D. P., Ulitsky Correspondence, I., Schwartz, M. G., Avrutin, A., & Ulitsky, I. (2015). Principles of Long noncoding RNA evolution derived from direct comparison of transcriptomes in 17 species. *Cell Reports*, *11*, 1110–1122. https://doi.org/10.1016/j.celrep.2015.04.023

Hollander, M. C., Alamo, I., & Fornace, A. J. (1996). A novel DNA damage-inducible transcript, gadd7, inhibits cell growth, but lacks a protein product. *Nucleic Acids Research*, *24*(9), 1589–1593. https://doi.org/10.1093/nar/24.9.1589

Hollister, J., Grabenhorst, E., Nimtz, M., Conradt, H., & Jarvis, D. L. (2002). Engineering the protein N-glycosylation pathway in insect cells for production of biantennary, complex N-glycans. *Biochemistry*, *41*(50), 15093–15104.

Hon, C. C., Ramilowski, J. A., Harshbarger, J., Bertin, N., Rackham, O. J. L., Gough, J., … Forrest, A. R. R. (2017). An atlas of human long non-coding RNAs with accurate 5' ends. *Nature*, *543*(7644), 199–204. https://doi.org/10.1038/nature21374

Hrabeta-Robinson, E., Marcus, E., Cozen, A. E., Phizicky, E. M., & Lowe, T. M. (2017). High-throughput small RNA sequencing enhanced by AlkB-facilitated RNA de-methylation (ARM-Seq). *Methods in Molecular Biology*, *1562*, 231–243. https://doi.org/10.1007/978-1-4939-6807-7_15

Hwang, S. O., & Lee, G. M. (2009). Effect of Akt overexpression on programmed cell death in antibody-producing Chinese hamster ovary cells. *Journal of Biotechnology*, *139*(1), 89–94. https://doi.org/http://dx.doi.org/10.1016/j.jbiotec.2008.09.008

Iben, J. R., & Maraia, R. J. (2012). tRNAomics: tRNA gene copy number variation and codon use provide bioinformatic evidence of a new anticodon:codon wobble pair in a eukaryote. *RNA*, *18*(7), 1358–1372. https://doi.org/10.1261/rna.032151.111

Itakura, K., Hirose, T., Crea, R., Riggs, A. D., Heyneker, H. L., Bolivar, F., & Boyer, H. W. (1977). Expression in *Escherichia coli* of a chemically synthesized gene for the hormone somatostatin. *Science*, *198*(4321), 1056–1063.

Ito, S., & Nagata, K. (2017). Biology of Hsp47 (Serpin H1), a collagen-specific molecular chaperone. *Seminars in Cell & Developmental Biology*, *62*, 142–151. https://doi.org/10.1016/j.semcdb.2016.11.005

Iwakiri, J., Hamada, M., & Asai, K. (2015). Bioinformatics tools for lncRNA research. *Biochimica et Biophysica Acta*. https://doi.org/10.1016/j.bbagrm.2015.07.014

Janakiraman, H., House, R. P., Gangaraju, V. K., Diehl, J. A., Howe, P. H., & Palanisamy, V. (2018). The long (lncRNA) and short (miRNA) of it: TGFβ-mediated control of RNA-binding proteins and noncoding RNAs. *Molecular Cancer Research*, *16*(4), 567–579. https://doi.org/10.1158/1541-7786.MCR-17-0547

Janakiraman, V., Kwiatkowski, C., Kshirsagar, R., Ryll, T., & Huang, Y.-M. (2015). Application of high-throughput mini-bioreactor system for systematic scale-down modeling, process characterization, and control strategy development. *Biotechnology Progress*, *31*(6), 1623–1632. https://doi.org/10.1002/btpr.2162

Jarvis, D. L. (2009). Baculovirus–insect cell expression systems. In *Methods in enzymology* (Vol. 463, pp. 191–222). https://doi.org/10.1016/S0076-6879(09)63014-7

Jayapal, K. P., Wlaschin, K. F., Hu, W. S., & Yap, M. G. S. (2007). Recombinant Protein Therapeutics from CHO Cells - 20 Years and Counting. *Chemical Engineering Progress*, *103*(10), 40–47.

Jeong, Y. T., Choi, O., Lim, H. R., Son, Y. D., Kim, H. J., & Kim, J. H. (2008). Enhanced sialylation of recombinant erythropoietin in CHO cells by human glycosyltransferase expression. *Journal of Microbiology and Biotechnology*, *18*(12), 1945–1952.

Jiang, L., Shao, C., Wu, Q. J., Chen, G., Zhou, J., Yang, B., … Fu, X. D. (2017). NEAT1 scaffolds RNA-binding proteins and the Microprocessor to globally enhance pri-miRNA processing. *Nature Structural and Molecular Biology*, *24*(10), 816–824. https://doi.org/10.1038/nsmb.3455

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*(12), 2527–2542. https://doi.org/10.1002/bit.25687

Josse, L., Smales, C. M., & Tuite, M. F. (2010). Transient expression of human TorsinA enhances secretion of two functionally distinct proteins in cultured Chinese hamster ovary (CHO) cells. *Biotechnology and Bioengineering*, *105*(3), 556–566. https://doi.org/10.1002/bit.22572

Jossé, L., Xie, J., Proud, C. G., & Smales, C. M. (2016, December). mTORC1 signalling and eIF4E/4E-BP1 translation initiation factor stoichiometry influence recombinant protein productivity from GS-CHOK1 cells. *Biochemical Journal*. https://doi.org/10.1042/BCJ20160845

Jossé, L., Zhang, L., & Smales, C. M. (2018). Application of microRNA targeted 3'UTRs to repress DHFR selection marker expression for development of recombinant antibody expressing CHO cell pools. *Biotechnology Journal*, *13*(10), 1800129. https://doi.org/10.1002/biot.201800129

Kallehauge, T. B., Li, S., Pedersen, L. E., Ha, T. K., Ley, D., Andersen, M. R., … Lewis, N. E. (2017). Ribosome profiling-guided depletion of an mRNA increases cell growth rate and protein secretion. *Scientific Reports*, *7*, 40388. https://doi.org/10.1038/srep40388

Kalvari, I., Argasinska, J., Quinones-Olvera, N., Nawrocki, E. P., Rivas, E., Eddy, S. R., … Petrov, A. I. (2018). Rfam 13.0: shifting to a genome-centric resource for non-coding RNA families. *Nucleic Acids Research*, *46*(D1), D335–D342. https://doi.org/10.1093/nar/gkx1038

Kang, Y.-J., Yang, D.-C., Kong, L., Hou, M., Meng, Y.-Q., Wei, L., & Gao, G. (2017). CPC2: a fast and accurate coding potential calculator based on sequence intrinsic features. *Nucleic Acids Research*, *45*(W1), W12–W16. https://doi.org/10.1093/nar/gkx428

Kapranov, P., Cheng, J., Dike, S., Nix, D. A., Duttagupta, R., Willingham, A. T., … Tammana, H. (2007). RNA maps reveal new RNA classes and a possible function for pervasive transcription. *Science*, *316*(5830), 1484–1488.

Kashi, K., Henderson, L., Bonetti, A., & Carninci, P. (2016). Discovery and functional analysis of lncRNAs: Methodologies to investigate an uncharacterized transcriptome. *Biochimica et Biophysica Acta*, *1859*(1), 3–15. https://doi.org/10.1016/j.bbagrm.2015.10.010

Kawarada, L., Suzuki, T., Ohira, T., Hirata, S., Miyauchi, K., & Suzuki, T. (2017). ALKBH1 is an RNA dioxygenase responsible for cytoplasmic and mitochondrial tRNA modifications. *Nucleic Acids Research*, *45*(12), 7401–7415. https://doi.org/10.1093/nar/gkx354

Kelen, K. Van Der, Beyaert, R., Inzé, D., & Veylder, L. De. (2009). Translational control of eukaryotic gene expression. *Critical Reviews in Biochemistry and Molecular Biology*, *44*(4), 143–168. https://doi.org/10.1080/10409230902882090

Keniry, A., Oxley, D., Monnier, P., Kyba, M., Dandolo, L., Smits, G., & Reik, W. (2012). The H19 lincRNA is a developmental reservoir of miR-675 that suppresses growth and Igf1r. *Nature Cell Biology*, *14*(7), 659–665. https://doi.org/10.1038/ncb2521

Kim, J. Y., Kim, Y. G., & Lee, G. M. (2012). CHO cells in biotechnology for production of recombinant proteins: Current state and further potential. *Applied Microbiology and Biotechnology*. https://doi.org/10.1007/s00253-011-3758-5

Kim, N. S., & Lee, G. M. (2002). Inhibition of sodium butyrate-induced apoptosis in recombinant Chinese hamster ovary cells by constitutively expressing antisense RNA of caspase-3. *Biotechnology and Bioengineering*, *78*(2), 217–228.

Kim, S. H., & Lee, G. M. (2006). Down-regulation of lactate dehydrogenase-A by siRNAs for reduced lactic acid formation of Chinese hamster ovary cells producing thrombopoietin. *Applied Microbiology and Biotechnology*, *74*(1), 152–159. https://doi.org/10.1007/s00253-006-0654-5

Kimchi-Sarfaty, C., Schiller, T., Hamasaki-Katagiri, N., Khan, M. A., Yanover, C., & Sauna, Z. E. (2013). Building better drugs: developing and regulating engineered therapeutic proteins. *Trends in Pharmacological Sciences*, *34*(10), 534–548. https://doi.org/10.1016/j.tips.2013.08.005

Klanert, G., Jadhav, V., Chanoumidou, K., Grillari, J., Borth, N., & Hackl, M. (2014). Endogenous microRNA clusters outperform chimeric sequence clusters in Chinese hamster ovary cells. *Biotechnology Journal*, *9*(4), 538–544. https://doi.org/10.1002/biot.201300216

Klanert, G., Jadhav, V., Shanmukam, V., Diendorfer, A., Karbiener, M., Scheideler, M., … Borth, N. (2016). A signature of 12 microRNAs is robustly associated with growth rate in a variety of CHO cell lines. *Journal of Biotechnology*, *235*, 150–161. https://doi.org/10.1016/j.jbiotec.2016.03.022

Kopp, F., & Mendell, J. T. (2018). Functional classification and experimental dissection of Long noncoding RNAs. *Cell*, *172*(3), 393–407. https://doi.org/10.1016/j.cell.2018.01.011

Kornienko, A. E., Guenzl, P. M., Barlow, D. P., & Pauler, F. M. (2013). Gene regulation by the act of long non-coding RNA transcription. *BMC Biology*, *11*(1), 59. https://doi.org/10.1186/1741-7007-11-59

Koziol, M., & Rinn, J. (2010). RNA traffic control of chromatin complexes. *Current Opinion in Genetics & Development*, *20*(2), 142–148. https://doi.org/10.1016/j.gde.2010.03.003

Kumar, P., Kuscu, C., & Dutta, A. (2016). Biogenesis and function of transfer RNA-related fragments (tRFs). *Trends in Biochemical Sciences*, *41*(8), 679–689. https://doi.org/10.1016/j.tibs.2016.05.004

Kunert, R., & Reinhart, D. (2016). Advances in recombinant antibody manufacturing. *Applied Microbiology and Biotechnology*, *100*(8), 3451–3461. https://doi.org/10.1007/s00253-016-7388-9

Kung, J. T. Y., Colognori, D., & Lee, J. T. (2013, March). Long noncoding RNAs: past, present, and future. *Genetics*. https://doi.org/10.1534/genetics.112.146704

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. *Biotechnology Letters*, *28*(1), 9–15. https://doi.org/10.1007/s10529-005-4680-7

Lafontaine, D. L. J., & Tollervey, D. (2001). The function and synthesis of ribosomes. *Nature Reviews Molecular Cell Biology*, *2*(7), 514–520. https://doi.org/10.1038/35080045

Lai, F., Blumenthal, E., & Shiekhattar, R. (2016). Detection and analysis of Long noncoding RNAs. *Enzymes of Epigenetics* (1st ed., Vol. 573). Elsevier Inc. https://doi.org/10.1016/bs.mie.2016.03.010

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, *9*(4), 357–359. https://doi.org/10.1038/nmeth.1923

Lee, B. D. (2018). Python implementation of Codon Adaptation Index. *Journal of Open Source Software*, *3*(30), 905. https://doi.org/10.21105/joss.00905

Lee, E. U., Roth, J., & Paulson, J. C. (1989). Alteration of terminal glycosylation sequences on N-linked oligosaccharides of Chinese hamster ovary cells by expression of beta-galactoside alpha 2,6-sialyltransferase. *Journal of Biological Chemistry*.

Lee, G. M. (2008). Assessment of cell engineering strategies for improved foreign protein production in cho cells. *Journal of Biotechnology*, *136*, S130–S131. https://doi.org/http://dx.doi.org/10.1016/j.jbiotec.2008.07.275

Lee, J. S., Grav, L. M., Lewis, N. E., & Faustrup Kildegaard, H. (2015). CRISPR/Cas9-mediated genome engineering of CHO cell factories: Application and perspectives. *Biotechnology Journal*, *10*(7), 979–994. https://doi.org/10.1002/biot.201500082

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*(8), 2195–2207. https://doi.org/10.1002/bit.24879

Leidel, S., Pedrioli, P. G. a, Bucher, T., Brost, R., Costanzo, M., Schmidt, A., … Peter, M. (2009). Ubiquitin-related modifier Urm1 acts as a sulphur carrier in thiolation of eukaryotic transfer RNA. *Nature*, *458*(7235), 228–232. https://doi.org/10.1038/nature07643

Leu, J. I.-J., Dumont, P., Hafey, M., Murphy, M. E., & George, D. L. (2004). Mitochondrial p53 activates Bak and causes disruption of a Bak–Mcl1 complex. *Nature Cell Biology*, *6*(5), 443–450. https://doi.org/10.1038/ncb1123

Lewis, N. E., Liu, X., Li, Y., Nagarajan, H., Yerganian, G., O'Brien, E., … Palsson, B. O. (2013). Genomic landscapes of Chinese hamster ovary cell lines as revealed by the Cricetulus griseus draft genome. *Nature Biotechnology*, *31*(8), 759–767. https://doi.org/10.1038/nbt.2624

Li, D., Zhang, J., Wang, M., Li, X., Gong, H., Tang, H., … Liu, Q. (2018). Activity dependent LoNA regulates translation by coordinating rRNA transcription and methylation. *Nature Communications*, *9*(1). https://doi.org/10.1038/s41467-018-04072-4

Li, F., Vijayasankaran, N., Shen, A. (Yijuan), Kiss, R., & Amanullah, A. (2010). Cell culture processes for monoclonal antibody production. *MAbs*, *2*(5), 466–479. https://doi.org/10.4161/mabs.2.5.12720

Liu, Bodu, Sun, L., Liu, Q., Gong, C., Yao, Y., Lv, X., … Song, E. (2015). A cytoplasmic NF-κB interacting Long noncoding RNA blocks IκB phosphorylation and suppresses breast cancer metastasis. *Cancer Cell*, *27*(3), 370–381. https://doi.org/10.1016/j.ccell.2015.02.004

Liu, Botao, & Qian, S.-B. (2016). Characterizing inactive ribosomes in translational profiling. *Translation*, *4*(1), e1138018. https://doi.org/10.1080/21690731.2015.1138018

Liu, F., Clark, W., Luo, G., Klungland, A., Wang, X., Fu, Y., … He, C. (2016). ALKBH1-mediated tRNA demethylation regulates translation. *Cell*, *167*, 816–828. https://doi.org/10.1016/j.cell.2016.09.038

Liu, N., Dai, Q., Zheng, G., He, C., Parisien, M., & Pan, T. (2015). N6-methyladenosine-dependent RNA structural switches regulate RNA–protein interactions. *Nature*, *518*(7540), 560–564. https://doi.org/10.1038/nature14234

Liu, S. J., Horlbeck, M. A., Cho, S. W., Birk, H. S., Malatesta, M., He, D., … Lim, D. A. (2016). CRISPRi-based genome-scale identification of functional long noncoding RNA loci in human cells. *Science*. https://doi.org/10.1126/science.aah7111

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*, 550. https://doi.org/10.1186/s13059-014-0550-8

Lowe, T. M., & Chan, P. P. (2016). tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes. *Nucleic Acids Research*, *44*(W1), W54–W57. https://doi.org/10.1093/nar/gkw413

Lu, Q., Ren, S., Lu, M., Zhang, Y., Zhu, D., Zhang, X., & Li, T. (2013). Computational prediction of associations between long non-coding RNAs and proteins. *BMC Genomics*, *14*(1), 651. https://doi.org/10.1186/1471-2164-14-651

Ma, C., Shi, X., Zhu, Q., Li, Q., Liu, Y., Yao, Y., & Song, Y. (2016). The growth arrest-specific transcript 5 (GAS5): a pivotal tumor suppressor long noncoding RNA in human cancers. *Tumor Biology*, *37*(2), 1437–1444. https://doi.org/10.1007/s13277-015-4521-9

Mabashi-Asazuma, H., Shi, X., Geisler, C., Kuo, C.-W., Khoo, K.-H., & Jarvis, D. L. (2013). Impact of a human CMP-sialic acid transporter on recombinant glycoprotein sialylation in glycoengineered insect cells. *Glycobiology*, *23*(2), 199–210. https://doi.org/10.1093/glycob/cws143

Macauley-Patrick, S., Fazenda, M. L., McNeil, B., & Harvey, L. M. (2005). Heterologous protein production using the *Pichia pastoris* expression system. *Yeast*, *22*(4), 249–270. https://doi.org/10.1002/yea.1208

Magistrelli, G., Poitevin, Y., Schlosser, F., Pontini, G., Malinge, P., Josserand, S., … Fischer, N. (2017). Optimizing assembly and production of native bispecific antibodies by codon de-optimization. *MAbs*, *9*(2), 231–239. https://doi.org/10.1080/19420862.2016.1267088

Mallory, A. C., & Shkumatava, A. (2015). LncRNAs in vertebrates: Advances and challenges. *Biochimie*, *117*, 3–14. https://doi.org/10.1016/j.biochi.2015.03.014

Malphettes, L., Freyvert, Y., Chang, J., Liu, P.-Q., Chan, E., Miller, J. C., … Cost, G. J. (2010). Highly efficient deletion of FUT8 in CHO cell lines using zinc-finger nucleases yields cells that produce completely nonfucosylated antibodies. *Biotechnology and Bioengineering*, *106*(5), 774–783. https://doi.org/10.1002/bit.22751

Marín, M., Fernández-Calero, T., & Ehrlich, R. (2017). Protein folding and tRNA biology. *Biophysical Reviews*. https://doi.org/10.1007/s12551-017-0322-2

Marisch, K., Bayer, K., Scharl, T., Mairhofer, J., Krempl, P. M., Hummel, K., … Striedner, G. (2013). A comparative analysis of industrial *Escherichia coli* K–12 and B strains in high-glucose batch cultivations on process, transcriptome and proteome level. *PLoS ONE*, *8*(8), e70516. https://doi.org/10.1371/journal.pone.0070516

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.Journal*, *17*(1), 10. https://doi.org/10.14806/ej.17.1.200

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*(3), 846–855. https://doi.org/10.1002/btpr.1542

Masterton, R. J., Roobol, A., Al-Fageeh, M. B., Carden, M. J., & Smales, C. M. (2010). Post-translational events of a model reporter protein proceed with higher fidelity and accuracy upon mild hypothermic culturing of Chinese hamster ovary cells. *Biotechnology and Bioengineering*, *105*(1), 215–220. https://doi.org/10.1002/bit.22533

Matera, A. G., Terns, R. M., & Terns, M. P. (2007). Non-coding RNAs: lessons from the small nuclear and small nucleolar RNAs. *Nature Reviews Molecular Cell Biology*, *8*(3), 209–220. https://doi.org/10.1038/nrm2124

Mattick, J. S. (2018). The state of Long non-coding RNA biology. *Non-Coding RNA*, *4*(3), 17. https://doi.org/10.3390/ncrna4030017

Mauro, V. P. (2018). Codon optimization in the production of recombinant biotherapeutics: potential risks and considerations. *BioDrugs*, *32*. https://doi.org/10.1007/s40259-018-0261-x

Mauro, V. P., & Chappell, S. A. (2014). A critical analysis of codon optimization in human therapeutics. *Trends in Molecular Medicine*, *20*(11), 604–613. https://doi.org/10.1016/j.molmed.2014.09.003

McLeod, J., O'Callaghan, P. M., Pybus, L. P., Wilkinson, S. J., Root, T., Racher, A. J., & James, D. C. (2011). An empirical modeling platform to evaluate the relative control discrete CHO cell synthetic processes exert over recombinant monoclonal antibody production process titer. *Biotechnology and Bioengineering*, *108*(9), 2193–2204. https://doi.org/10.1002/bit.23146

Mead, E. J., Chiverton, L. M., Smales, C. M., & Von Haar, T. Der. (2009). Identification of the limitations on recombinant gene expression in CHO cell lines with varying luciferase production rates. *Biotechnology and Bioengineering*. https://doi.org/10.1002/bit.22201

Mead, E. J., Chiverton, L. M., Spurgeon, S. K., Martin, E. B., Montague, G. A., Smales, C. M., & von der Haar, T. (2012). Experimental and in silico modelling analyses of the gene expression pathway for recombinant antibody and by-product production in NS0 cell lines. *PLoS ONE*. https://doi.org/10.1371/journal.pone.0047422

Mead, E. J., Masterton, R. J., Feary, M., Obrezanova, O., Zhang, L., Young, R., & Smales, C. M. (2015). Biological insights into the expression of translation initiation factors from recombinant CHOK1SV cell lines and their relationship to enhanced productivity. *The Biochemical Journal*, *472*(3), 261–273. https://doi.org/10.1042/BJ20150928

Melo, C. A., Drost, J., Wijchers, P. J., van de Werken, H., de Wit, E., Vrielink, J. A. F. O., … Agami, R. (2013). eRNAs are required for p53-dependent enhancer activity and gene transcription. *Molecular Cell*, *49*(3), 524–535. https://doi.org/10.1016/j.molcel.2012.11.021

Mercer, T. R., Clark, M. B., Crawford, J., Brunck, M. E., Gerhardt, D. J., Taft, R. J., … Mattick, J. S. (2014). Targeted sequencing for gene discovery and quantification using RNA CaptureSeq. *Nature Protocols*, *9*(5), 989–1009. https://doi.org/10.1038/nprot.2014.058

Mercer, T. R., Dinger, M. E., Sunkin, S. M., Mehler, M. F., & Mattick, J. S. (2008). Specific expression of long noncoding RNAs in the mouse brain. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(2), 716–721. https://doi.org/10.1073/pnas.0706729105

Mergulhão, F. J. M., Summers, D. K., & Monteiro, G. A. (2005). Recombinant protein secretion in *Escherichia coli*. *Biotechnology Advances*, *23*(3), 177–202. https://doi.org/10.1016/j.biotechadv.2004.11.003

Messens, J., & Collet, J.-F. (2006). Pathways of disulfide bond formation in *Escherichia coli*. *The International Journal of Biochemistry & Cell Biology*, *38*(7), 1050–1062. https://doi.org/10.1016/j.biocel.2005.12.011

Muppirala, U. K., Honavar, V. G., & Dobbs, D. (2011). Predicting RNA-protein interactions using only sequence information. *BMC Bioinformatics*, *12*(1), 489. https://doi.org/10.1186/1471-2105-12-489

Musaelyan, A., Lapin, S., Nazarov, V., Tkachenko, O., Gilburd, B., Mazing, A., … Shoenfeld, Y. (2018). Vimentin as antigenic target in autoimmunity: A comprehensive review. *Autoimmunity Reviews*. https://doi.org/10.1016/J.AUTREV.2018.04.004

Nakagawa, S., & Hirose, T. (2012). Paraspeckle nuclear bodies-useful uselessness? *Cellular and Molecular Life Sciences*, *69*(18), 3027–3036. https://doi.org/10.1007/s00018-012-0973-x

Nett, J. H., Cook, W. J., Chen, M.-T., Davidson, R. C., Bobrowicz, P., Kett, W., … Hamilton, S. R. (2013). Characterization of the *Pichia pastoris* protein-O-mannosyltransferase gene family. *PLoS ONE*, *8*(7), e68325. https://doi.org/10.1371/journal.pone.0068325

Nienow, A. W., Rielly, C. D., Brosnan, K., Bargh, N., Lee, K., Coopman, K., & Hewitt, C. J. (2013). The physical characterisation of a microscale parallel bioreactor platform with an industrial CHO cell line expressing an IgG4. *Biochemical Engineering Journal*, *76*, 25–36. https://doi.org/10.1016/j.bej.2013.04.011

O'Callaghan, P. M., McLeod, J., Pybus, L. P., Lovelady, C. S., Wilkinson, S. J., Racher, A. J., … James, D. C. (2010). Cell line-specific control of recombinant monoclonal antibody production by CHO cells. *Biotechnology and Bioengineering*, *106*(6), 938–951. https://doi.org/10.1002/bit.22769

Ohya, T., Hayashi, T., Kiyama, E., Nishii, H., Miki, H., Kobayashi, K., … Ohtake, H. (2008). Improved production of recombinant human antithrombin III in Chinese hamster ovary cells by ATF4 overexpression. *Biotechnology and Bioengineering*, *100*(2), 317–324. https://doi.org/10.1002/bit.21758

Omasa, T., Takami, T., Ohya, T., Kiyama, E., Hayashi, T., Nishii, H., … Ohtake, H. (2008). Overexpression of GADD34 enhances production of recombinant human antithrombin III in Chinese Hamster Ovary cells. *Journal of Bioscience and Bioengineering*, *106*(6), 568–573. https://doi.org/10.1263/JBB.106.568

Ou, K.-C., Wang, C.-Y., Liu, K.-T., Chen, Y.-L., Chen, Y.-C., Lai, M.-D., & Yen, M.-C. (2014). Optimization protein productivity of human interleukin-2 through codon usage, gene copy number and intracellular tRNA concentration in CHO cells. *Biochemical and Biophysical Research Communications*, *454*, 347–352. https://doi.org/10.1016/j.bbrc.2014.10.097

Pan, T. (2018). Modifications and functional genomics of human transfer RNA. *Cell Research*, *28*(4), 395–404. https://doi.org/10.1038/s41422-018-0013-y

Pan, X., Streefland, M., Dalm, C., Wijffels, R. H., & Martens, D. E. (2017). Selection of chemically defined media for CHO cell fed-batch culture processes. *Cytotechnology*, *69*(1), 39–56. https://doi.org/10.1007/s10616-016-0036-5

Patrucco, L., Chiesa, A., Soluri, M. F., Fasolo, F., Takahashi, H., Carninci, P., … Cotella, D. (2015). Engineering mammalian cell factories with SINEUP noncoding RNAs to improve translation of secreted proteins. *Gene*, *569*(2), 287–293. https://doi.org/10.1016/j.gene.2015.05.070

Pavon-Eternod, M., Gomes, S., Rosner, M. R., & Pan, T. (2013). Overexpression of initiator methionine tRNA leads to global reprogramming of tRNA expression and increased proliferation in human epithelial cells. *RNA*, *19*(4), 461–466. https://doi.org/10.1261/rna.037507.112

Pennock, G. D., Shoemaker, C., & Miller, L. K. (1984). Strong and regulated expression of *Escherichia coli* beta-galactosidase in insect cells with a baculovirus vector. *Molecular and Cellular Biology*, *4*(3), 399–406. https://doi.org/10.1128/mcb.4.3.399

Perry, R. B.-T., & Ulitsky, I. (2016). The functions of long noncoding RNAs in development and stem cells. *Development*, *143*(21), 3882 LP – 3894. https://doi.org/10.1242/dev.140962

Pieper, L. A., Strotbek, M., Wenger, T., Gamer, M., Olayioye, M. A., & Hausser, A. (2017). Secretory pathway optimization of CHO producer cells by co-engineering of the mitosRNA-1978 target genes CerS2 and Tbc1D20. *Metabolic Engineering*, *40*, 69–79. https://doi.org/10.1016/j.ymben.2017.01.003

Plotkin, J. B., Robins, H., & Levine, A. J. (2004). Tissue-specific codon usage and the expression of human genes. *Proceedings of the National Academy of Sciences*, *101*(34), 12588–12591. https://doi.org/10.1073/pnas.0404957101

Podbevšek, P., Fasolo, F., Bon, C., Cimatti, L., Reißer, S., Carninci, P., … Gustincich, S. (2018). Structural determinants of the SINE B2 element embedded in the long non-coding RNA activator of translation AS Uchl1. *Scientific Reports*, *8*(1), 1–13. https://doi.org/10.1038/s41598-017-14908-6

Pope, B., & Kent, H. M. (1996). High efficiency 5 min transformation of *Escherichia coli*. *Nucleic Acids Research*, *24*(3), 536–537. https://doi.org/10.1093/nar/24.3.536

Povey, J. F., O'Malley, C. J., Root, T., Martin, E. B., Montague, G. A., Feary, M., … Smales, C. M. (2014). Rapid high-throughput characterisation, classification and selection of recombinant mammalian cell line phenotypes using intact cell MALDI-ToF mass spectrometry fingerprinting and PLS-DA modelling. *Journal of Biotechnology*, *184*, 84–93. https://doi.org/10.1016/j.jbiotec.2014.04.028

Presnyak, V., Alhusaini, N., Chen, Y.-H., Martin, S., Morris, N., Kline, N., … Coller, J. (2015). Codon optimality is a major determinant of mRNA stability. *Cell*, *160*(6), 1111–1124. https://doi.org/10.1016/j.cell.2015.02.029

Pruitt, K. D., Tatusova, T., & Maglott, D. R. (2007). NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Research*, *35*(Database issue), D61–D65. https://doi.org/10.1093/nar/gkl842

Pybus, L. P., Dean, G., West, N. R., Smith, A., Daramola, O., Field, R., … James, D. C. (2014). Model-directed engineering of "difficult-to-express" monoclonal antibody production by Chinese hamster ovary cells. *Biotechnology and Bioengineering*, *111*(2), 372–385. https://doi.org/10.1002/bit.25116

Raetz, C. R. H., & Whitfield, C. (2002). Lipopolysaccharide endotoxins. *Annual Review of Biochemistry*, *71*(1), 635–700. https://doi.org/10.1146/annurev.biochem.71.110601.135414

Rashid, F., Shah, A., & Shan, G. (2016). Long non-coding RNAs in the cytoplasm. *Genomics, Proteomics & Bioinformatics*, *14*(2), 73–80. https://doi.org/10.1016/j.gpb.2016.03.005

Reid, D. W., & Nicchitta, C. V. (2015). Diversity and selectivity in mRNA translation on the endoplasmic reticulum. *Nature Reviews Molecular Cell Biology*, *16*(4), 221–231. https://doi.org/10.1038/nrm3958

Reinhart, D., Damjanovic, L., Kaisermayer, C., & Kunert, R. (2015). Benchmarking of commercially available CHO cell culture media for antibody production. *Applied Microbiology and Biotechnology*, *99*(11), 4645–4657. https://doi.org/10.1007/s00253-015-6514-4

Resnick, M. A., & Cox, B. S. (2000). Yeast as an honorary mammal. *Mutation Research*, *451*(1–2), 1–11. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10915861

Reuveni, S., Meilijson, I., Kupiec, M., Ruppin, E., & Tuller, T. (2011). Genome-scale analysis of translation elongation with a ribosome flow model. *PLoS Computational Biology*, *7*(9), e1002127. https://doi.org/10.1371/journal.pcbi.1002127

Rinn, J. L., Kertesz, M., Wang, J. K., Squazzo, S. L., Xu, X., Brugmann, S. A., … Chang, H. Y. (2007). Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell*, *129*(7), 1311–1323. https://doi.org/10.1016/j.cell.2007.05.022

Romano, M. C., Thiel, M., Stansfield, I., & Grebogi, C. (2009). Queueing phase transition: theory of translation. *Physical Review Letters*, *102*(19), 198104. https://doi.org/10.1103/PhysRevLett.102.198104

Ronda, C., Pedersen, L. E., Hansen, H. G., Kallehauge, T. B., Betenbaugh, M. J., Nielsen, A. T., & Kildegaard, H. F. (2014). Accelerating genome editing in CHO cells using CRISPR Cas9 and CRISPy, a web-based target finding tool. *Biotechnology and Bioengineering*, *111*(8), 1604–1616. https://doi.org/10.1002/bit.25233

Roobol, A., Roobol, J., Bastide, A., Knight, J. R. P., Willis, A. E., & Smales, C. M. (2015). p58IPK is an inhibitor of the eIF2α kinase GCN2 and its localization and expression underpin protein synthesis and ER processing capacity. *The Biochemical Journal*, *465*(2), 213–225. https://doi.org/10.1042/BJ20140852

Rosano, G. L., & Ceccarelli, E. A. (2014). Recombinant protein expression in *Escherichia coli*: advances and challenges. *Frontiers in Microbiology*, *5*, 172. https://doi.org/10.3389/fmicb.2014.00172

Rouiller, Y., Bielser, J.-M., Brühlmann, D., Jordan, M., Broly, H., & Stettler, M. (2016). Screening and assessment of performance and molecule quality attributes of industrial cell lines across different fed-batch systems. *Biotechnology Progress*, *32*(1), 160–170. https://doi.org/10.1002/btpr.2186

Rudolph, K. L. M., Schmitt, B. M., Villar, D., White, R. J., Marioni, J. C., Kutter, C., & Odom, D. T. (2016). Codon-driven translational efficiency is stable across diverse mammalian cell states. *PLOS Genetics*, *12*(5), e1006024. https://doi.org/10.1371/journal.pgen.1006024

Schmitt, A. M., & Chang, H. Y. (2016). Long noncoding RNAs in cancer pathways. *Cancer Cell*, *29*(4), 452–463. https://doi.org/10.1016/j.ccell.2016.03.010

Schoellhorn, M., Fischer, S., Wagner, A., Handrick, R., & Otte, K. (2017). miR-143 targets MAPK7 in CHO cells and induces a hyperproductive phenotype to enhance production of difficult-to-express proteins. *Biotechnology Progress*, *33*(4), 1046–1058. https://doi.org/10.1002/btpr.2475

Schwanhäusser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., … Selbach, M. (2011). Global quantification of mammalian gene expression control. *Nature*, *473*(7347), 337–342. https://doi.org/10.1038/nature10098

Sezonov, G., Joseleau-Petit, D., & D'Ari, R. (2007). *Escherichia coli* physiology in Luria-Bertani broth. *Journal of Bacteriology*, *189*(23), 8746–8749. https://doi.org/10.1128/JB.01368-07

Sharp, P. M., & Li, W.-H. (1987). The codon adaptation index-a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Research*, *15*(3), 1281–1295. https://doi.org/10.1093/nar/15.3.1281

Shigematsu, M., Honda, S., Loher, P., Telonis, A. G., Rigoutsos, I., & Kirino, Y. (2017). YAMAT-seq: An efficient method for high-throughput sequencing of mature transfer RNAs. *Nucleic Acids Research*, *45*(9), e70. https://doi.org/10.1093/nar/gkx005

Signal, B., Gloss, B. S., & Dinger, M. E. (2016). Computational approaches for functional prediction and characterisation of Long noncoding RNAs. *Trends in Genetics*, *32*(10), 620–637. https://doi.org/10.1016/j.tig.2016.08.004

Singh, A., Kildegaard, H. F., & Andersen, M. R. (2018). An online compendium of CHO RNA-Seq data allows identification of CHO cell line-specific transcriptomic signatures. *Biotechnology Journal*. https://doi.org/10.1002/biot.201800070

Siomi, M. C., Sato, K., Pezic, D., & Aravin, A. A. (2011). PIWI-interacting small RNAs: the vanguard of genome defence. *Nature Reviews Molecular Cell Biology*, *12*(4), 246–258. https://doi.org/10.1038/nrm3089

Smith, C. M., & Steitz, J. A. (1998). Classification of gas5 as a multi-small-nucleolar-RNA (snoRNA) host gene and a member of the 5'-terminal oligopyrimidine gene family reveals common features of snoRNA host genes. *Molecular and Cellular Biology*, *18*(12), 6897–6909. https://doi.org/10.1128/mcb.18.12.6897

Smith, G. E., Summers, M. D., & Fraser, M. J. (1983). Production of human beta interferon in insect cells infected with a baculovirus expression vector. *Molecular and Cellular Biology*, *3*(12), 2156–2165.

Stiefel, F., Fischer, S., Sczyrba, A., Otte, K., & Hesse, F. (2016). miRNA profiling of high, low and non-producing CHO cells during biphasic fed-batch cultivation reveals process relevant targets for host cell engineering. *Journal of Biotechnology*, *225*, 31–43. https://doi.org/10.1016/j.jbiotec.2016.03.028

Stricher, F., Macri, C., Ruff, M., & Muller, S. (2013). HSPA8/HSC70 chaperone protein. *Autophagy*, *9*(12), 1937–1954. https://doi.org/10.4161/auto.26448

Sun, W., Yang, Y., Xu, C., & Guo, J. (2017). Regulatory mechanisms of long noncoding RNAs on gene expression in cancers. *Cancer Genetics*, *216–217*, 105–110. https://doi.org/10.1016/J.CANCERGEN.2017.06.003

Sun, X., Haider Ali, M. S. S., & Moran, M. (2017). The role of interactions of long non-coding RNAs and heterogeneous nuclear ribonucleoproteins in regulating cellular functions. *The Biochemical Journal*, *474*(17), 2925–2935. https://doi.org/10.1042/BCJ20170280

Tabuchi, H. (2013). Novel strategy for a high-yielding mAb-producing CHO strain (overexpression of non-coding RNA enhanced proliferation and improved mAb yield). *BMC Proceedings*, *7*(Suppl 6), O3. https://doi.org/10.1186/1753-6561-7-S6-O3

Tabuchi, H., & Sugiyama, T. (2013). Cooverexpression of alanine aminotransferase 1 in Chinese hamster ovary cells overexpressing taurine transporter further stimulates metabolism and enhances product yield. *Biotechnology and Bioengineering*, *110*(8), 2208–2215. https://doi.org/10.1002/bit.24881

Tamošaitis, L., & Smales, C. M. (2018). Meta-analysis of publicly available Chinese Hamster Ovary (CHO) cell transcriptomic datasets for identifying engineering targets to enhance recombinant protein yields. *Biotechnology Journal*, 1800066. https://doi.org/10.1002/biot.201800066

Tang, H., Wang, S., Wang, J., Song, M., Xu, M., Zhang, M., … Bao, X. (2016). N-hypermannose glycosylation disruption enhances recombinant protein production by regulating secretory pathway and cell wall integrity in *Saccharomyces cerevisiae. Scientific Reports*, *6*(1), 25654. https://doi.org/10.1038/srep25654

Tani, H., Torimura, M., & Akimitsu, N. (2013). The RNA degradation pathway regulates the function of GAS5 a non-coding RNA in mammalian cells. *PloS One*, *8*(1), e55684. https://doi.org/10.1371/journal.pone.0055684

Tarrant, D., & Von Der Haar, T. (2014). Synonymous codons, ribosome speed, and eukaryotic gene expression regulation. *Cellular and Molecular Life Sciences*. https://doi.org/10.1007/s00018-014-1684-2

Tay, Y., Rinn, J., & Pandolfi, P. P. (2014). The multilayered complexity of ceRNA crosstalk and competition. *Nature*, *505*(7483), 344–352. https://doi.org/10.1038/nature12986

Thoring, L., Wustenhagen, D. A., Borowiak, M., Stech, M., Sonnabend, A., & Kubick, S. (2016). Cell-free systems based on CHO cell lysates: Optimization strategies, synthesis of “difficult-to-express” proteins and future perspectives. *PLoS ONE*, *11*(9), 1–21. https://doi.org/10.1371/journal.pone.0163670

Trimarchi, T., Bilal, E., Ntziachristos, P., Fabbri, G., Dalla-Favera, R., Tsirigos, A., & Aifantis, I. (2014). Genome-wide mapping and characterization of notch-regulated long noncoding RNAs in acute leukemia. *Cell*, *158*(3), 593–606. https://doi.org/10.1016/j.cell.2014.05.049

Tripathi, V., Ellis, J. D., Shen, Z., Song, D. Y., Pan, Q., Watt, A. T., … Prasanth, K. V. (2010). The nuclear-retained noncoding RNA MALAT1 regulates aalternative splicing by modulating SR splicing factor phosphorylation. *Molecular Cell*, *39*(6), 925–938. https://doi.org/10.1016/J.MOLCEL.2010.08.011

Tseng, Y.-Y., & Bagchi, A. (2015). The PVT1-MYC duet in cancer. *Molecular & Cellular Oncology*, *2*(2), e974467. https://doi.org/10.4161/23723556.2014.974467

Tzani, I., Monger, C., Kelly, P., Barron, N., Kelly, R. M., & Clarke, C. (2018). Understanding biopharmaceutical production at single nucleotide resolution using ribosome footprint profiling. *Current Opinion in Biotechnology*, *53*, 182–190. https://doi.org/10.1016/j.copbio.2018.01.030

Ulitsky, I. (2016). Evolution to the rescue: using comparative genomics to understand long non-coding RNAs. *Nature Reviews Genetics*, *17*(10), 601–614. https://doi.org/10.1038/nrg.2016.85

Underhill, M. F., Birch, J. R., Smales, C. M., & Naylor, L. H. (2005). eIF2α phosphorylation, stress perception, and the shutdown of global protein synthesis in cultured CHO cells. *Biotechnology and Bioengineering*. https://doi.org/10.1002/bit.20403

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*(7), 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*(2), 405–412. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/6305508

Valdés-Bango Curell, R., & Barron, N. (2018). Exploring the Potential Application of Short Non-Coding RNA-Based Genetic Circuits in Chinese Hamster Ovary Cells. *Biotechnology Journal*, 1700220. https://doi.org/10.1002/biot.201700220

van Oers, M. M., Pijlman, G. P., & Vlak, J. M. (2015). Thirty years of baculovirus-insect cell protein expression: from dark horse to mainstream technology. *Journal of General Virology*, *96*(Pt_1), 6–23. https://doi.org/10.1099/vir.0.067108-0

Vishwanathan, N., Le, H., Le, T., & Hu, W. S. (2014). Advancing biopharmaceutical process science through transcriptome analysis. *Current Opinion in Biotechnology*, *30*, 113–119. https://doi.org/10.1016/j.copbio.2014.06.011

Vito, D., & Smales, C. M. (2018). The Long non-coding RNA transcriptome landscape in CHO cells under batch and fed-batch conditions. *Biotechnology Journal*, *1800122*. https://doi.org/10.1002/biot.201800122

von der Haar, T., & McCarthy, J. E. G. (2002). Intracellular translation initiation factor levels in *Saccharomyces cerevisiae* and their role in cap-complex function. *Molecular Microbiology*, *46*(2), 531–544.

Walsh, G. (2010). Biopharmaceutical benchmarks 2010. *Nature Biotechnology*, *28*(9), 917–924. https://doi.org/10.1038/nbt0910-917

Walsh, G. (2014). Biopharmaceutical benchmarks 2014. *Nature Biotechnology*, *32*(10), 992–1000. https://doi.org/10.1038/nbt.3040

Walsh, G. (2018). Biopharmaceutical benchmarks 2018. *Nature Biotechnology* (Vol. 36). https://doi.org/10.1038/nbt.4305

Wang, F., Yuan, J.-H., Wang, S.-B., Yang, F., Yuan, S.-X., Ye, C., … Sun, S.-H. (2014). Oncofetal long noncoding RNA PVT1 promotes proliferation and stem cell-like property of hepatocellular carcinoma cells by stabilizing NOP2. *Hepatology*, *60*(4), 1278–1290. https://doi.org/10.1002/hep.27239

Wang, K., Zhang, T., Chen, J., Liu, C., Tang, J., & Xie, Q. (2018). The effect of culture temperature on the aggregation of recombinant TNFR-Fc is regulated by the PERK-eIF2a pathway in CHO cells. *Protein & Peptide Letters*, *25*(6), 570–579. https://doi.org/10.2174/0929866525666180530121317

Wang, P., Xue, Y., Han, Y., Lin, L., Wu, C., Xu, S., … Liu, Q. (2014). The STAT3-binding Long noncoding RNA lnc-DC controls uman dendritic cell differentiation. *Science*, *344*(6181), 310–313.

Welch, M., Villalobos, A., Gustafsson, C., & Minshull, J. (2009, August 6). You're one in a googol: Optimizing genes for protein expression. *Journal of the Royal Society Interface*. https://doi.org/10.1098/rsif.2008.0520.focus

Wildt, S., & Gerngross, T. U. (2005). The humanization of N-glycosylation pathways in yeast. *Nature Reviews Microbiology*, *3*(2), 119–128. https://doi.org/10.1038/nrmicro1087

Wilusz, J. E. (2015). Controlling translation via modulation of tRNA levels. *Wiley Interdisciplinary Reviews: RNA*. https://doi.org/10.1002/wrna.1287

Wilusz, J. E. (2016). Long noncoding RNAs: re-writing dogmas of RNA processing and stability. *Biochimica et Biophysica Acta*, *1859*(1), 128–138. https://doi.org/10.1016/j.bbagrm.2015.06.003

Wilusz, J. E., Freier, S. M., & Spector, D. L. (2008). 3' end processing of a long nuclear-retained noncoding RNA yields a tRNA-like cytoplasmic RNA. *Cell*, *135*(5), 919–932. https://doi.org/10.1016/J.CELL.2008.10.012

Wong, D. C. F., Wong, K. T. K., Lee, Y. Y., Morin, P. N., Heng, C. K., & Yap, M. G. S. (2006). Transcriptional profiling of apoptotic pathways in batch and fed-batch CHO cell cultures. *Biotechnology and Bioengineering*, *94*(2), 373–382. https://doi.org/10.1002/bit.20872

Wu, S.-C. (2009). RNA interference technology to improve recombinant protein production in Chinese hamster ovary cells. *Biotechnology Advances*, *27*(4), 417–422. https://doi.org/10.1016/j.biotechadv.2009.03.002

Wurm, F. M. (2004). Production of recombinant protein therapeutics in cultivated mammalian cells. *Nature Biotechnology*, *22*(11), 1393–1398. https://doi.org/10.1038/nbt1026

Wurm, F. M. (2013). CHO Quasispecies—Implications for Manufacturing Processes. *Processes*, *1*, 296–311. https://doi.org/10.3390/pr1030296

Wurm, F., & Wurm, M. (2017). Cloning of CHO cells, productivity and genetic stability—A discussion. *Processes*, *5*(4), 20. https://doi.org/10.3390/pr5020020

Xu, X, Nagarajan, H., Lewis, N. E., Pan, S., Cai, Z., Liu, X., … Wang, J. (2011). The genomic sequence of the Chinese hamster ovary (CHO)-K1 cell line. *Nat Biotechnol*, *29*(8), 735–741. https://doi.org/10.1038/nbt.1932

Xu, Xun, Nagarajan, H., Lewis, N. E., Pan, S., Cai, Z., Liu, X., … Wang, J. (2011). The genomic sequence of the Chinese hamster ovary (CHO)-K1 cell line. *Nature Biotechnology*. https://doi.org/10.1038/nbt.1932

Yamaji, H. (2014). Suitability and perspectives on using recombinant insect cells for the production of virus-like particles. *Applied Microbiology and Biotechnology*, *98*(5), 1963–1970. https://doi.org/10.1007/s00253-013-5474-9

Yamane-Ohnuki, N., Kinoshita, S., Inoue-Urakubo, M., Kusunoki, M., Iida, S., Nakano, R., … 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*(5), 614–622. https://doi.org/10.1002/bit.20151

Yang, X., Okamura, D. M., Lu, X., Chen, Y., Moorhead, J., Varghese, Z., & Ruan, X. Z. (2017). CD36 in chronic kidney disease: novel insights and therapeutic opportunities. *Nature Reviews Nephrology*, *13*(12), 769–781. https://doi.org/10.1038/nrneph.2017.126

Yao, Y., Jin, S., Long, H., Yu, Y., Zhang, Z., Cheng, G., … Wu, Q. (2015). RNAe: an effective method for targeted protein translation enhancement by artificial non-coding RNA with SINEB2 repeat. *Nucleic Acids Research*, *43*(9), e58–e58. https://doi.org/10.1093/nar/gkv125

Yee, C. M., Zak, A. J., Hill, B. D., & Wen, F. (2018). The coming age of insect cells for manufacturing and development of protein therapeutics. *Industrial & Engineering Chemistry Research*, *57*. https://doi.org/10.1021/acs.iecr.8b00985

Yu, H. (2002). Regulation of APC–Cdc20 by the spindle checkpoint. *Current Opinion in Cell Biology*, *14*(6), 706–714. https://doi.org/10.1016/S0955-0674(02)00382-4

Zerbino, D. R., Achuthan, P., Akanni, W., Amode, M. R., Barrell, D., Bhai, J., … Flicek, P. (2018). Ensembl 2018. *Nucleic Acids Research*, *46*(D1), D754–D761. https://doi.org/10.1093/nar/gkx1098

Zhai, Y., Li, N., Jiang, H., Huang, X., Gao, N., & Tye, B. K. (2017). Unique roles of the non-identical MCM subunits in DNA replication licensing. *Molecular Cell*, *67*(2), 168–179. https://doi.org/10.1016/J.MOLCEL.2017.06.016

Zhao, F., Yu, C., & Liu, Y. (2017). Codon usage regulates protein structure and function by affecting translation elongation speed in Drosophila cells. *Nucleic Acids Research*, *45*(14), 8484–8492. https://doi.org/10.1093/nar/gkx501

Zheng, G., Qin, Y., Clark, W. C., Dai, Q., Yi, C., He, C., … Pan, T. (2015). Efficient and quantitative high-throughput tRNA sequencing. *Nature Methods*, *12*(9), 835–837. https://doi.org/10.1038/nmeth.3478

Zhou, M., Crawford, Y., Ng, D., Tung, J., Pynn, A. F. J., Meier, A., … Shen, A. (2011). Decreasing lactate level and increasing antibody production in Chinese Hamster Ovary cells (CHO) by reducing the expression of lactate dehydrogenase and pyruvate dehydrogenase kinases. *Journal of Biotechnology*, *153*(1–2), 27–34. https://doi.org/10.1016/j.jbiotec.2011.03.003

Zhu, Shikai, Shuai, P., Yang, C., Zhang, Y., Zhong, S., Liu, X., … Zhou, Y. (2017). Prognostic value of long non-coding RNA PVT1 as a novel biomarker in various cancers: a meta-analysis. *Oncotarget*, *8*(68), 113174–113184. https://doi.org/10.18632/oncotarget.22830

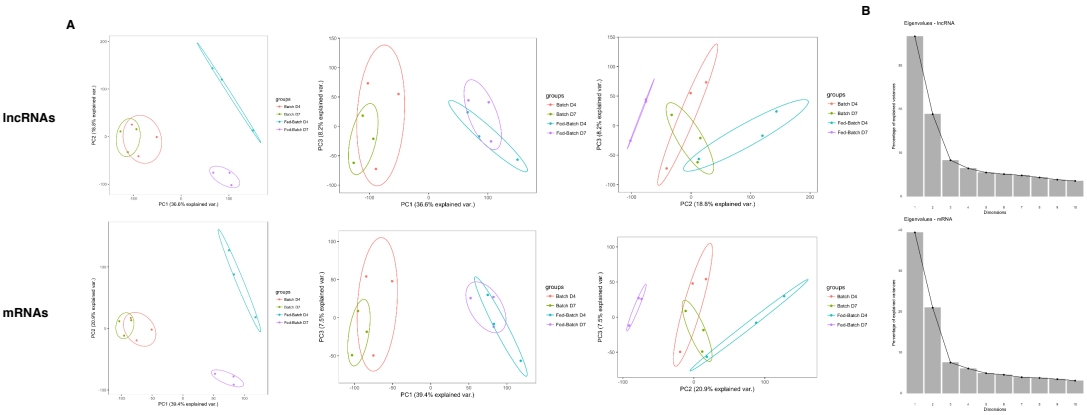
Zhu, Shiyou, Li, W., Liu, J., Chen, C.-H., Liao, Q., Xu, P., … Wei, W. (2016). Genome-scale deletion screening of human long non-coding RNAs using a paired-guide RNA CRISPR–Cas9 library. *Nature Biotechnology*, *34*(12), 1279–1286. https://doi.org/10.1038/nbt.3715

Zouridis, H., & Hatzimanikatis, V. (2007). A model for protein translation: polysome self-organization leads to maximum protein synthesis rates. *Biophysical Journal*, *92*(3), 717–730. https://doi.org/10.1529/biophysj.106.087825

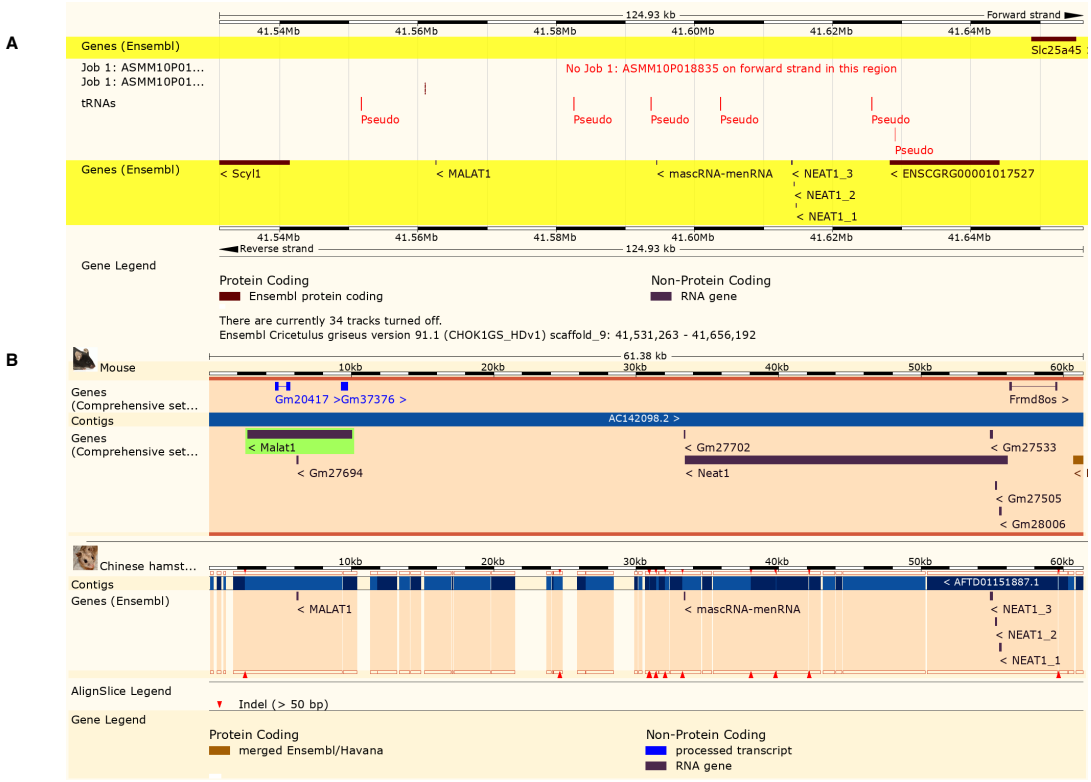
Zucchelli, S., Patrucco, L., Persichetti, F., Gustincich, S., & Cotella, D. (2016). Engineering translation in mammalian cell factories to increase protein yield: The unexpected use of Long non-coding SINEUP RNAs. *Computational and Structural Biotechnology Journal*, *14*, 404–410. https://doi.org/10.1016/j.csbj.2016.10.004

Appendix

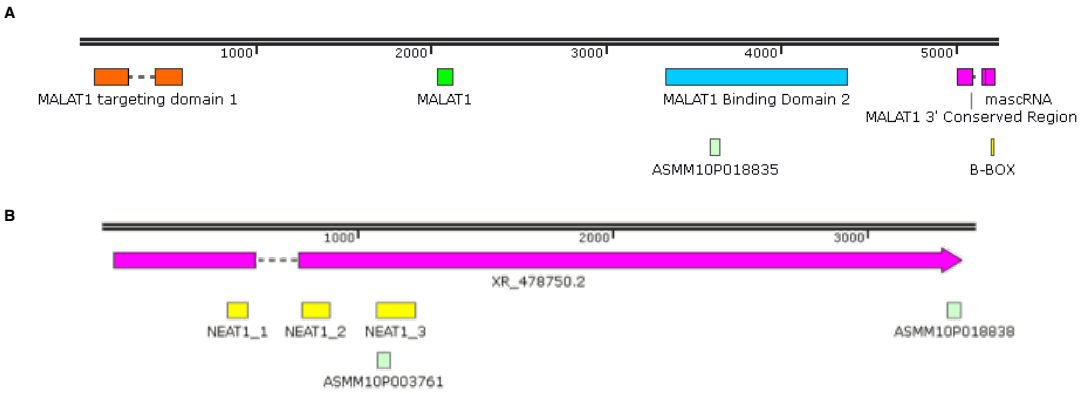
Supplementary figures - Chapter 3



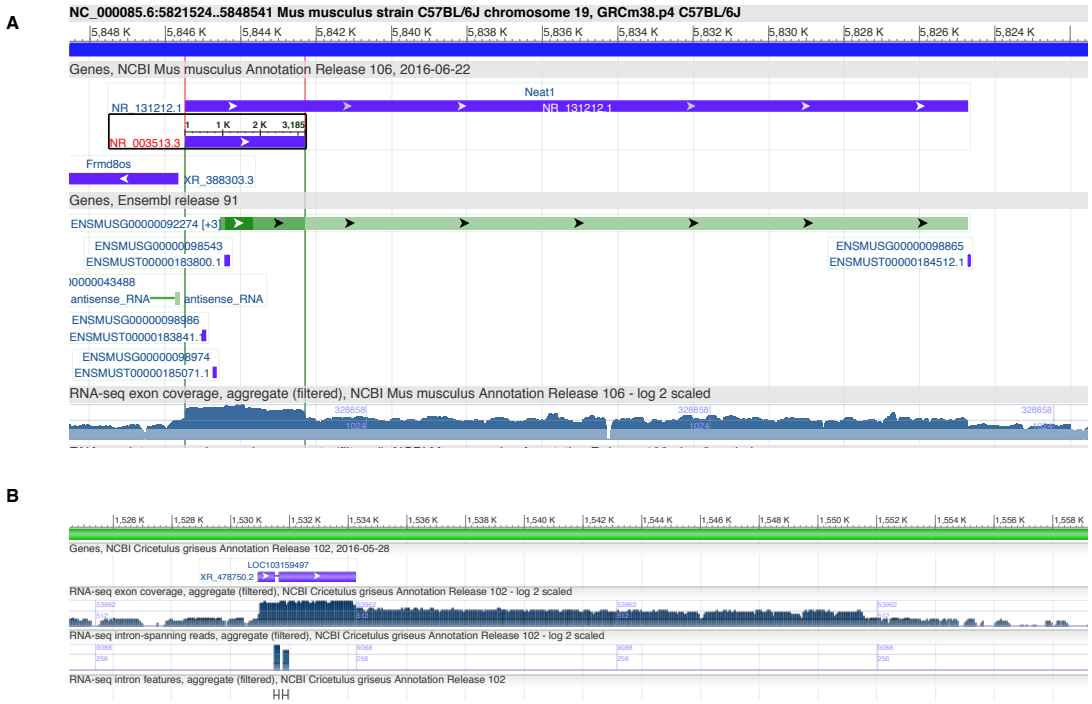
Supplementary Figure 3.1 - (A) The figures show principal component analysis (PCA) for lncRNAs (above) and RNAs (below) based on expression levels of each measured replicate among the four groups, enclosed in corresponding calculated correlation circles. (B) Summary of the first 10 dimensions with corresponding percentage of explained variances for lncRNAs and mRNAs.



Supplementary Figure 3.2 - (A) NEAT1 and MALAT1 loci in Chinese hamster based on Ensembl database, with predicted features and genes. (B) Comparison between the mouse NEAT1 and MALAT1 loci with Chinese hamster identified loci.

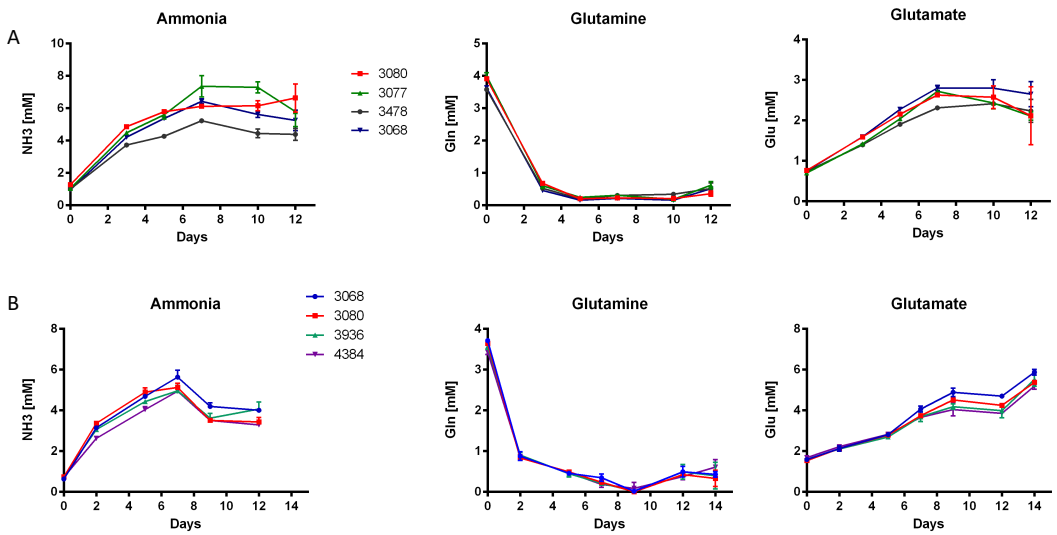


Supplementary Figure 3.3 - A) Schematic representation of the conserved domains based on Rfam and literature found in the putative MALAT1 locus in Chinese hamster. The probe used in the array is indicated as ASMM10P018835 B) Schematic representation of the conserved domains based on Rfam and literature found in the putative NEAT1 transcript in Chinese hamster. The probes used in the array are indicated as ASMM10P003761 and ASMM10P018838.

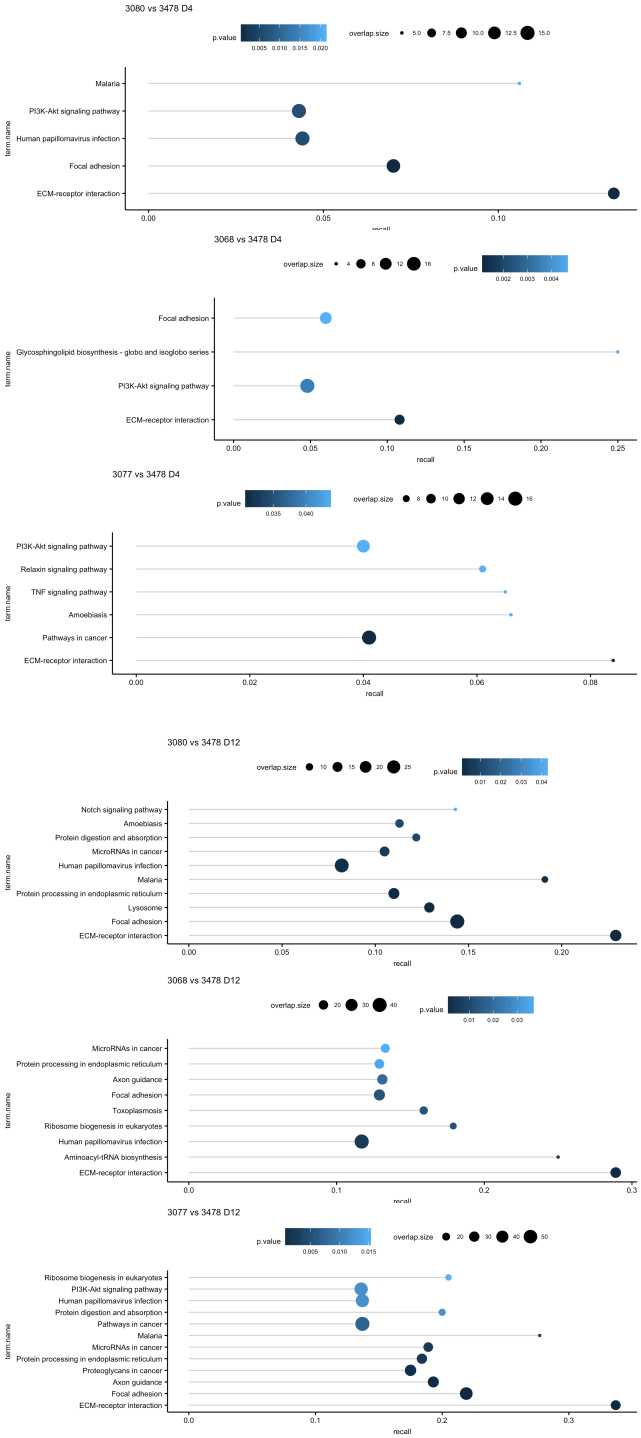


Supplementary Figure 3.4: - (A) Mouse annotation for the NEAT1 locus. (B) Chinese Hamster annotation for the putative NEAT1 locus.

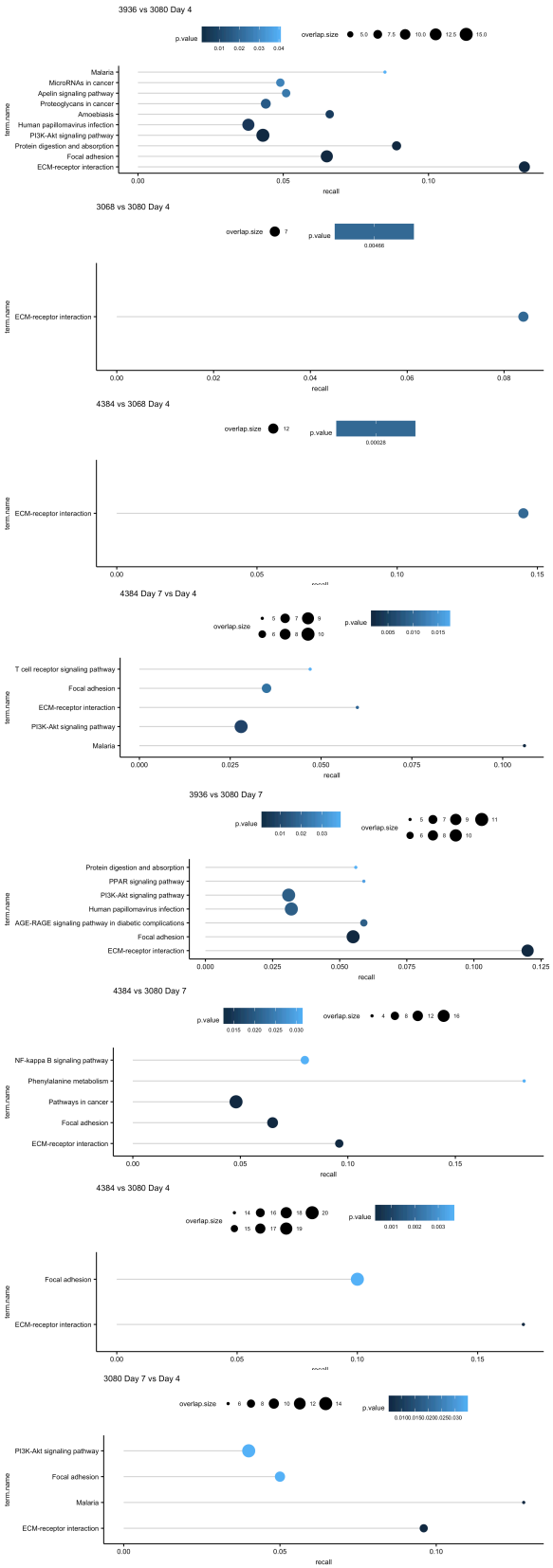
Supplementary figures - Chapter 4



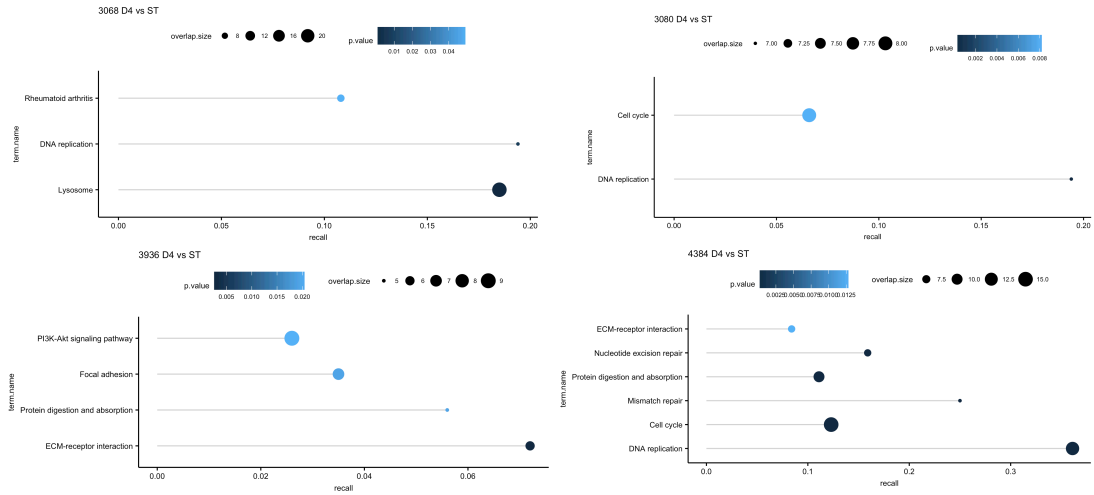
Supplementary Figure 4.1 - Ammonia, glutamine and glutamate concentrations over time for the DAVI dataset (panel A) and the JCE dataset (panel B).



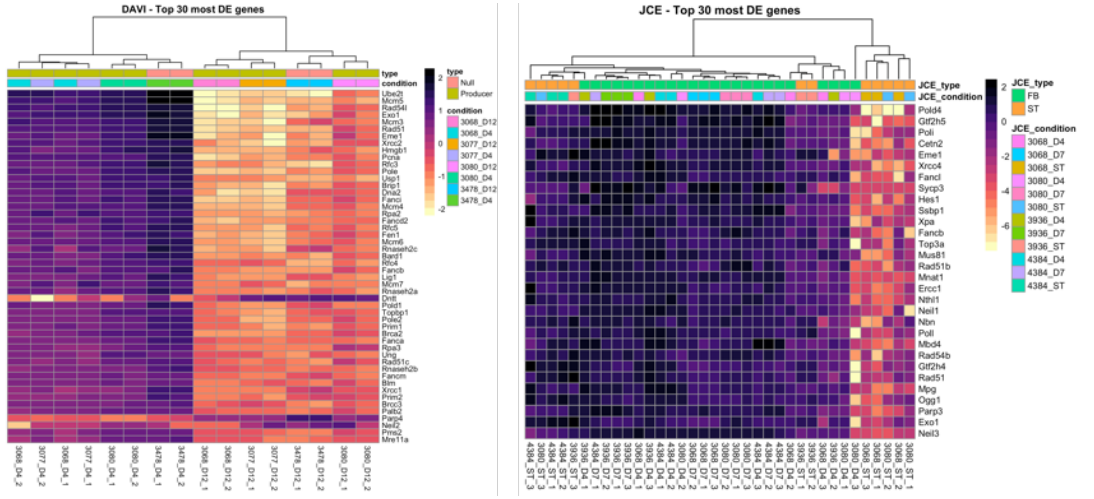
Supplementary Figure 4.2 - Enriched KEGG pathways based on differentially expressed genes for each comparison between different cell lines at the same time point in the DAVI dataset. Each dot represents a pathway, with color shade representing the *p*-value, size proportional to the overlap size (differentially expressed genes in the pathway) and x-coordinate recall (overlap size divided by the total number of genes in the pathway).



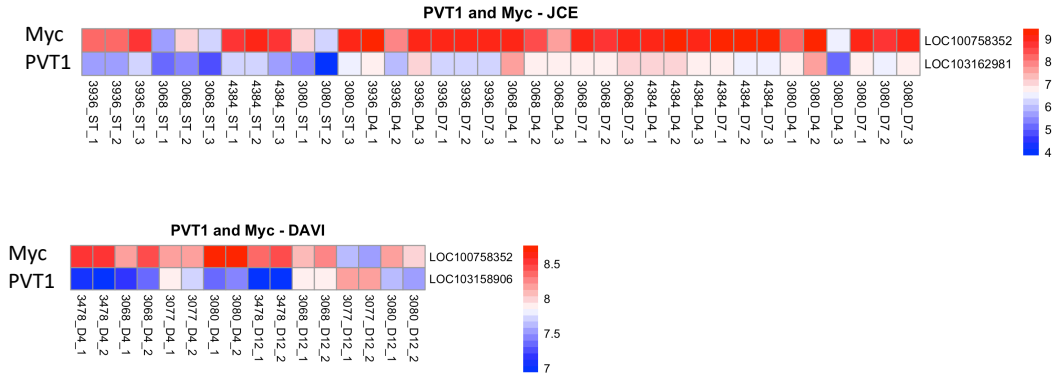
Supplementary Figure 4.3 - Enriched KEGG pathways based on differentially expressed genes for each comparison among different cell lines at day 4 and day 7 in the JCE dataset. Each dot represents a pathway, with color shade representing the *p*-value, size proportional to the overlap size (differentially expressed genes in the pathway) and x-coordinate recall (overlap size divided by the total number of genes in the pathway).



Supplementary Figure 4.4 - Enriched KEGG pathways based on differentially expressed genes for each comparison among the same cell line at day 4 against seed train culture in the JCE dataset. Each dot represents a pathway, with color shade representing the *p*-value, size proportional to the overlap size (differentially expressed genes in the pathway) and x-coordinate recall (overlap size divided by the total number of genes in the pathway).



Supplementary Figure 4.5 - Top 30 differentially expressed genes among the replication and repair domain in the DAVI dataset (A) and JCE dataset (B).



Supplementary Figure 4.6 - PVT1 and Myc expression in the JCE (top) and DAVI (bottom) datasets.

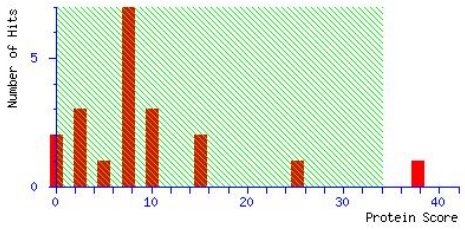
Supplementary figures - Chapter 5

MATRIX SCIENCE **Mascot Search Results**

```
User            : dv78@kent.ac.uk
Email           : dv78@kent.ac.uk
Search title    : AlkB A16
MS data file    : A16 AlkB Mass Spec data.mgf
Database        : SwissProt 2016_06 (551385 sequences; 196948568 residues)
Timestamp       : 4 Jul 2016 at 13:14:41 GMT
Protein hits    : ALKB_ECOLI Alpha-ketoglutarate-dependent dioxygenase AlkB OS=Escherichia coli (strain K12) GN=alkB PE=1 SV=1
```

Mascot Score Histogram

Ions score is -10*Log(P), where P is the probability that the observed match is a random event.
Individual ions scores > 34 indicate identity or extensive homology (p<0.05).
Protein scores are derived from ions scores as a non-probabilistic basis for ranking protein hits.



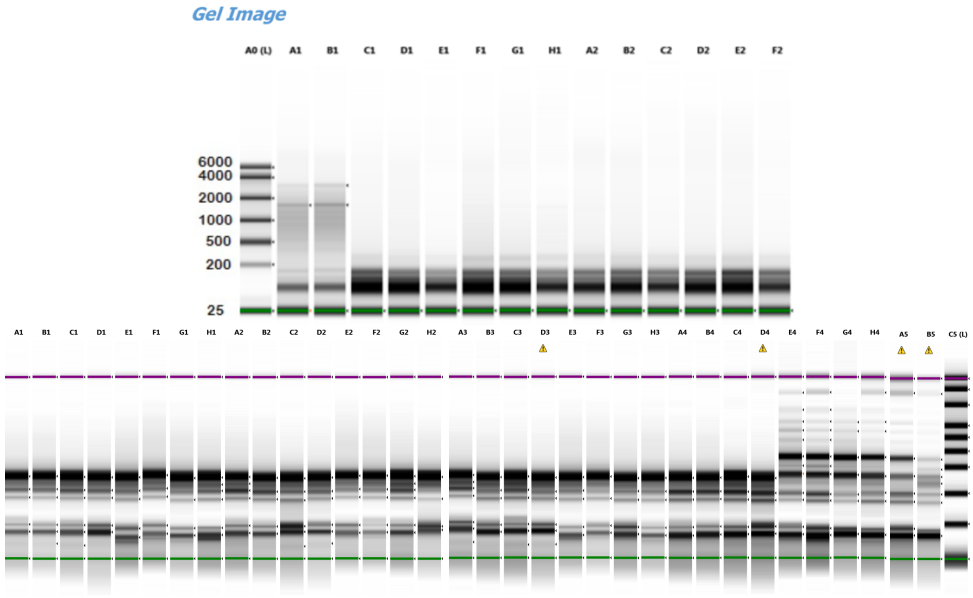
1. ALKB_ECOLI Mass: 24288 Score: 38 Matches: 1(0) Sequences: 1(0)
   Alpha-ketoglutarate-dependent dioxygenase AlkB OS=Escherichia coli (strain K12) GN=alkB PE=1 SV=1
   ☐ Check to include this hit in error tolerant search

Query	Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Rank	Unique	Peptide
☑ 13	2051.0170	2050.0097	2049.9578	25.3	0	38	0.077	1	U	R.AATAAGYPDFQPDACLINR.Y

Peptide matches not assigned to protein hits: (no details means no match)

Query	Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Rank	Unique	Peptide
☑ 10	1766.0911	1765.0838	1765.0025	46.0	0	25	0.19	1		SKPVPPISIPTPEIYK
☑ 8	1348.9235	1347.9162	1347.8966	14.5	1	0	2.9	1		KVILPQVLLGLR
☑ 1	813.4478	812.4405								
☑ 2	848.3050	847.2977								
☑ 3	1215.9546	1214.9473								
☑ 4	1237.9549	1236.9476								
☑ 5	1274.8629	1273.8556								
☑ 6	1279.9610	1278.9537								
☑ 7	1301.9523	1300.9450								
☑ 9	1436.0454	1435.0381								
☑ 11	1788.0771	1787.0698								
☑ 12	2036.2210	2035.2137								
☑ 14	2058.1895	2057.1823								
☑ 15	2065.0241	2064.0168								

Supplementary Figure 5.1 – MS/MS identification of the purified AlkB protein.



Supplementary Figure 5.2 – RNA Screentape performed at the RNA-Seq facilities to check integrity of the de-methylated small RNAs samples before sequencing. The band patterns are clearly visible, with no signs of degradation.

RT-qPCR primers

Transcript\Gene	Forward	Reverse
B2M	ACGGAGTTTACACCCACTGC	CAGACCTCCATGATGCTTGA
B-Actin	AGCTGAGAGGGAAATTGTGCG	GCAACGGAACCGCTCATT
GAPDH	GCCAAGAGGGTCATCATCTC	CCTTCCACAATGCCAAAGTT
XR_001721626.1	CCTCGTCTTGGGTTTCTTGA	GATTGCATTGTGTTGCGTATG
XR_001722284.1	AAGGCACCTTGGGACTTTTT	ACAGCTACATGTTCTTTCTTGACCT
XR_001723852.1	CTGACAGCACCCATTTACTTCA	TTTCTTTGCCCAAATGGCTA
XR_001725088.1	GCACCTTTGTACCCAACCAT	CAGTATCTTCGTGCCAGCAA
XR_001725445.1	AGACTACGGCAGGAGGATCA	AATGGTTTGGGGCTATAGGAA
XR_001725951.1	CCGAATAACTGGGAGCTGAA	AGCAGGTAGAGGTGCTGCTG
XR_001726746.1	GCACTAACTCACACCCCAGAA	TCTTGGCCATTCTCTTGTCC
XR_480869.2	AGCATCTCCCAGTGATGAGC	AGAACCTGATGAGTCTGTTTCTGTT
XR_481007.2	GCTCGCTCACAGAGAACACA	CCCTAGTGGGAGATGCTGAG
XR_482085.2	CAAGAACTCCCTGGAACTGG	GGTGCCTTGGTTGGTCTAAA
XR_482984.2	CTCCAATAGGTTGGACCACAG	GCAGCTTTTCACACATCCAG
XR_483663.2	TCATCTTTATGTCCACCCACAC	AGGGTGACTCTGGCTTCCTT
Gadd45a	GTGACGAACCCACATTCATC	ACTGGAACCCACTGATCCAT
Por	GAGACAGAGAGCACCTGTGGA	AGGTGTTCTGCACGTCTTTGG
Bcl6	GCCGCGGTGTTTGTATTTTGG	TGCACCAAGGGGGAAAAATGA
Ccna2	ACCTACCTCAGAGCAGCACA	GTGTCTCTGGTGGGTTGAGA
Top2a	CTGAGGGAATTGGTACCGGG	GGCAGAGGCTCTTCTCCATC
Aurkb	AGCCACAGTGAGACATACCG	AGCCGATCTGAGGGTTTATG
Rad51	ACACCGAGGGTACGTTTAGG	GTGTTGAACCCTCGAGCATA
Rad51	ACCAACCAGGTAGTAGCCCA	GCCTGGTTGTTGATGCATGG
PVT1	TTGTAGCATCCCAAGGCCAG	CCTCACGTGGGGATTCTCTG
Gas5	TGCTATGTGTTTTGGCCTGC	AAAACCTAACCAAGCCCGGA
Gas5	TGGAATCCAACACCGGCTAT	ACGAGACAACACACGCAGAT
Adapt15	GAATTCAGTGTAGGCCGGGA	TGAACAGAGCTGGCAATGACT
Trmt61a	AGTTCCAGGAGCATCGTGTG	AGAAGACAGCATCAGCCACG

List of primers used in Chapter 3 (upper panel) and Chapter 4 (lower panel) for validation of RNA-Seq results using RT-qPCR.

Etanercept sequence

ATGGCTCCTGTGGCTGTTTGGGCTGCTCTGGCTGTTGGACTGGAACTGTGGGCTGCTGCTCATGCT
CTGCCTGCTCAGGTGGCCTTCACACCTTATGCTCCAGAGCCTGGCTCTACCTGCAGACTGAGAGAGT
ACTACGACCAGACCGCTCAGATGTGCTGCTCCAAGTGTTCTCCTGGCCAGCACGCCAAGGTGTTCT
GCACCAAGACCTCCGATACCGTGTGCGACTCCTGCGAGGACTCCACCTATACTCAGCTGTGGAACT
GGGTGCCCGAGTGCCTGTCTTGTGGCAGCAGATGCTCCTCCGACCAGGTGGAAACCCAGGCCTGTA
CCAGAGAGCAGAACCGGATCTGCACCTGTAGACCCGGCTGGTACTGTGCCCTGTCTAAGCAAGAG
GGCTGTAGACTGTGCGCCCCTCTGAGAAAGTGCAGACCTGGCTTCGGAGTGGCTAGACCTGGCAC
CGAGACATCTGACGTCGTGTGCAAGCCTTGTGCTCCCGGCACCTTCTCCAACACCACCTCCTCTACC
GACATCTGCAGACCCCACCAAATCTGCAACGTGGTGGCTATCCCTGGCAACGCCTCTATGGATGCC
GTGTGCACCTCTACCTCTCCAACTCGGTCTATGGCTCCCGGCGCTGTTCATCTGCCTCAGCCTGTGTC
TACCAGAAGCCAGCACACCCAGCCTACACCTGAGCCTTCTACCGCTCCTTCCACCAGCTTTCTGCTGC
CCATGGGACCATCTCCACCAGCCGAAGGATCTACAGGCGACGAGCCTAAGTCCTGCGACAAGACCC
ATACCTGTCCTCCATGTCCTGCACCTGAGCTGCTCGGAGGCCCTTCCGTGTTTCTGTTCCCTCCAAAG
CCTAAGGACACCCTGATGATCTCTCGGACCCCTGAAGTGACCTGCGTGGTGGTGGATGTGTCTCAC
GAGGACCCAGAAGTGAAGTTCAATTGGTACGTGGACGGCGTGGAAGTGCACAACGCCAAGACCAA
GCCTAGAGAGGAACAGTACAACAGCACCTACAGAGTGGTGTCCGTGCTGACCGTGCTGCACCAGG
ATTGGCTGAACGGCAAAGAGTACAAGTGCAAGGTGTCCAACAAGGCCCTGCCAGCTCCTATCGAA
AAGACCATCAGCAAGGCCAAGGGCCAGCCTAGGGAACCCCAGGTTTACACCTTGCCTCCAAGCCG
GGAAGAGATGACCAAGAACCAGGTGTCCCTGACCTGCCTGGTCAAGGGCTTCTACCCTTCCGACAT
TGCCGTGGAATGGGAGAGCAATGGCCAGCCTGAGAACAACTACAAGACCACACCTCCTGTGCTGG
ACTCCGACGGCTCATTCTTCCTGTACTCCAAGCTGACAGTGGACAAGTCCAGATGGCAGCAGGGCA
ACGTGTTCAGCTGCTCCGTGATGCACGAGGCCCTGCACAATCACTACACACAGAAGTCCCTGTCTCT
GTCCCCTGGCAAGTGATAG

EPO sequence

ATGGGCGTGCACGAGTGTCCTGCTTGGCTGTGGCTGCTGCTGTCCCTGCTGTCTCTGCCTCTGGGAC
TGCCTGTGCTGGGCGCTCCTCCTAGACTGATCTGCGACAGCCGGGTGCTGGAAAGATACCTGCTGG
AAGCCAAAGAGGCCGAGAACATCACCACCGGCTGCGCCGAGCACTGCTCCCTGAACGAGAATATC
ACCGTGCCCGACACCAAAGTGAACTTCTACGCCTGGAAGCGGATGGAAGTGGGCCAGCAGGCTGT
GGAAGTGTGGCAGGGACTGGCTCTGCTGAGCGAGGCTGTGCTGAGAGGACAGGCCCTGCTCGTG
AACTCCTCCCAGCCTTGGGAACCCCTGCAGCTGCACGTGGACAAGGCTGTGTCCGGCCTGAGATCC
CTGACCACCCTGCTGAGAGCACTGGGAGCCCAGAAAGAGGCCATCTCTCCACCTGACGCCGCCTCT
GCTGCTCCTCTGAGAACCATCACCGCCGACACCTTCAGAAAGCTGTTCCGGGTGTACTCCAACTTCC
TGCGGGGCAAGCTGAAGCTGTACACCGGCGAGGCTTGCCGGACAGGCGATAGATGA

Biotechnology Journal
www.biotechnology-journal.com

The Long Non-Coding RNA Transcriptome Landscape in CHO Cells Under Batch and Fed-Batch Conditions

*Davide Vito and Christopher Mark Smales**

The role of non-coding RNAs in determining growth, productivity, and recombinant product quality attributes in Chinese hamster ovary (CHO) cells has received much attention in recent years, exemplified by studies into microRNAs in particular. However, other classes of non-coding RNAs have received less attention. One such class are the non-coding RNAs known collectively as long non-coding RNAs (lncRNAs). The authors have undertaken the first landscape analysis of the lncRNA transcriptome in CHO using a mouse based microarray that also provided for the surveillance of the coding transcriptome. The authors report on those lncRNAs present in a model host CHO cell line under batch and fed-batch conditions on two different days and relate the expression of different lncRNAs to each other. The authors demonstrate that the mouse microarray is suitable for the detection and analysis of thousands of CHO lncRNAs and validated a number of these by qRT-PCR. The authors then further analyzed the data to identify those lncRNAs whose expression changed the most between growth and stationary phases of culture or between batch and fed-batch culture to identify potential lncRNA targets for further functional studies with regard to their role in controlling growth of CHO cells. The authors discuss the implications for the publication of this rich dataset and how this may be used by the community.

1. Introduction

Mammalian expression systems are widely used for the production of recombinant protein biopharmaceuticals, largely due to their ability to correctly fold, assemble, post-translationally modify, and secrete complex human like proteins.[1] Among these, the Chinese hamster ovary (CHO) cell is the most widely utilized expression platform used in industry, especially for the production of monoclonal antibodies (mAbs), and are used for the expression of more than 60% of biotherapeutic proteins made in mammalian cells.[1–5] The most widely used bioprocess for the production of biopharmaceuticals from CHO cells involves fed-batch culture, this offering an advantage over batch culture in terms of cell growth, viability, and product yields due to the supplementation of nutrients resulting in higher biomass accumulation, less build-up of toxic metabolic by-products, and enhanced productivity.[6–8]

The prominence of CHO cell expression systems has driven innovation in the industry such that CHO expression systems and associated bioprocesses have been developed that can deliver yields of mAb in excess of 5–10 g L^{-1} in stably expressing, fed-batch systems.[9,10] Despite this, some biotherapeutic recombinant proteins, and particularly some of the non-mAb novel biotherapeutics in development, remain difficult to express in CHO cell or other mammalian cell expression systems.[11,12] Using a prior knowledge about cellular systems, various approaches have been taken to engineer cells to deliver enhanced product yields and quality including the engineering of chaperones,[13,14] glycosylation machinery,[15–17] and proliferation control strategies, including manipulation of the cell cycle,[18,19] apoptosis,[20,21] and autophagy.[22,23] A long side this approach, there have been studies to further our understanding of the potential cellular constraints on the production of mAbs,[24,25] other recombinant biotherapeutics[26] and difficult to express recombinant proteins[27] in order to identify bottlenecks in the recombinant gene expression pathway and to develop new bioprocesses or adapt/engineer novel hosts for enhanced production and/or quality of such molecules.[28,29] The majority of these studies to date have focussed upon manipulation of coding genes, however with the discovery and improved understanding of non-coding RNA in the control of cellular processes, there has been much interest in the last decade or so in these non-coding RNAs in CHO cells. In particular, the manipulation of microRNAs to enhance the ability of CHO cells to produce biotherapeutic proteins has attracted much attention.[30–33] An advantage of manipulating such non-coding RNAs for modulating CHO cell phenotypes is that such engineering does not place an addition protein synthetic burden upon the host cell,[34] unlike engineering of coding genes, and often such non-coding RNAs can modulate whole pathways rather than individual steps or processes as when manipulating many coding genes.

It has been estimated that at least 75% of transcripts originate from non-coding sequences.[35] Investigations into these transcripts has resulted in the identification of a class of transcript collectively known as Long non-coding RNAs (lncRNAs).[36] LncRNAs are defined as transcripts longer than 200 nucleotides

D. Vito, Prof. C. M. Smales
Industrial Biotechnology Centre and School of Biosciences
University of Kent
Canterbury CT2 7NJ, Kent, UK
E-mail: c.m.smales@kent.ac.uk

The ORCID identification number(s) for the author(s) of this article can be found under https://doi.org/10.1002/biot.201800122.

DOI: 10.1002/biot.201800122

 © 2018 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

that lack a significant open reading frame (ORF), are usually transcribed by RNA polymerase II and spliced, with or without 3′ polyadenylation.[37–39] A number of lncRNA molecules have been shown to play key regulatory roles in various biological processes including epigenetic regulation,[40] transcriptional control,[41] splicing events,[42] and mRNA translation.[43] Indeed, lncRNAs are capable of modulating a wide range of cellular processes and mechanisms both in the nucleus and the cytoplasm.[44] The majority of our understanding into lncRNAs and the mechanism-(s) by which they elicit their responses has come from studies relating to disease[45] and developmental studies.[46] LncRNAs elicit their effects by acting as competing endogenous RNAs (ceRNAs) by binding to and sequestering microRNAs,[47] by acting as architectural RNAs (arcRNAs) whereby they form functional structures,[48] act as *cis* molecules to enhance (eRNAs) coding gene expression[49] or as *trans* protein binding RNA molecules that can recruit chromatin modifying,[50,51] as microRNA precursors,[52] modulators of mRNA stability[53] and to impact upon post-translational modifications.[54,55]

Despite the importance of lncRNAs in controlling cellular processes, and unlike small non-coding RNAs (e.g., siRNAs, microRNAs), the impact(s) of lncRNA expression on CHO cell bioprocessing with regard to growth/proliferation and recombinant protein yields and quality has barely been explored, with only a small number of studies reported.[56,57] However, these two studies show that manipulation of lncRNAs can impact upon recombinant protein production from CHO cells. One of the reasons for the small number of studies on lncRNAs in CHO cells is the lack of a comprehensive annotation of non-coding transcripts in CHO, hampering their identification, genome wide assessment of their expression and modulation during culture, functional studies, and hence identification of target lncRNAs for cell engineering. One way to address this problem is to take advantage of the reported similarities between the Chinese hamster, CHO, and mouse genome,[58] where the number of well annotated non-coding transcripts is much higher. The similarity between CHO cells and mouse has already been utilized for the identification of 416 ncRNAs based on sequenced transcripts from a pooled CHO sample compared to the fRNAdb database of non-coding RNAs using BLAST, with most hits coming from mouse.[59] Here, we report the first analysis in CHO cells of both the coding and the non-coding transcriptome (specifically the lncRNAs) during batch and fed-batch culture at two different time points. We report on the identification of differentially expressed lncRNAs and mRNA, their interconnections and their potential impact on cellular pathways. This has allowed the mapping of the lncRNA landscape in CHO cells and identification of lncRNAs targets in CHO for further manipulation with a view to increase proliferation and to sustain viability throughout batch and fed-batch culture.

2. Experimental Section

2.1. Model Cell Line and Cell Culture Conditions

The CHO—S Freestyle host cell line (ThermoFisher Scientific, MA, USA) was used as a model CHO cell line for this study. Cells were routinely cultured in a Lab-Therm LT-X (Kühner AG, Basel, Switzerland) shaking incubator at 37 °C, 5% CO_2, 125 rpm, and 70% humidity in chemically defined serum-free growth medium (CD CHO, ThermoFisher Scientific). Fed-batch cultures were supplemented with CHO CD Efficient Feed B Liquid Nutrient Supplement (ThermoFisher Scientific). Initial supplementation testing was conducted in duplicate following Conditions 3 and 9 as described in the CHO CD Efficient Feed manual. Cultures with a viability >98% were used to seed initial fed experiments at 3×10^5 viable cells mL^{-1} in a 50 mL working volume in 250 mL polycarbonate Erlenmeyer flasks with vented caps (Corning, Wiesbaden, Germany). From the initial feeding experiments, Condition 3 was used for all future fed-batch cultures and for generating samples analyzed during this study. This adopted feeding strategy consisted of a 15% (v/v) supplementation of CHO CD Efficient Feed B to the CD CHO starting volume immediately on Day 0, followed by 10% (v/v) supplementation on Day 3 and Day 6. Four biological replicates of each culture process (batch and fed-batch) were seeded at 3×10^5 viable cells mL^{-1} from 20 mL cultures with a culture viability $\geq$ 98.5% in 120 mL CD CHO starting working volume in 500 mL polycarbonate Erlenmeyer flasks with vented caps (Corning). Viable cell concentrations and culture viability were determined daily using a Vi-CELL XR Cell Viability Analyzer (Beckman Coulter, Inc., Brea, CA, USA) on 1 mL culture samples.

2.2. Sampling from Cell Cultures and Subsequent RNA Extraction

Samples of 1×10^7 viable cells were taken from each flask after 96 h (Day 4) and 144 h (Day 7) of culture and total RNA immediately extracted using the commercially available mirVana miRNA Isolation Kit (ThermoFisher Scientific) and treated with RapidOut DNA Removal Kit (ThermoFisher Scientific). The RNA quantity and quality were determined using a NanoDrop ND-1000 (ThermoFisher Scientific) instrument and the integrity of RNA assessed by standard denaturing agarose gel electrophoresis.

2.3. lncRNA and Coding RNA Microarray and Data Analysis

2.3.1. Microarray Details

Three of the four cultures were selected for analysis. Analysis of extracted RNA for coding and lncRNAs was undertaken using the commercially available ArrayStar Mouse lncRNA Microarray V3.0 (Rockville, MD, USA). RNA labeling and array hybridization were performed according to the Agilent One-Color Microarray-Based Gene Expression Analysis protocol (Agilent Technologies, Santa Clara, CA, USA) with minor modifications. Ribosomal RNAs were removed from total RNA using the mRNA-ONLY Eukaryotic mRNA Isolation Kit, (Epicentre, Madison, WI, USA). Each sample was then amplified and transcribed into fluorescent cRNA along the entire length of the transcripts without 3′ bias utilizing a mixture of oligo(dT) and random primers using the Arraystar Flash RNA Labeling Kit (Arraystar). The labeled cRNAs were purified by RNeasy Mini Kit (Qiagen, Hilden, Germany). The concentration and specific activity of the labeled cRNAs were determined by NanoDrop ND-1000 (ThermoFisher Scientific). A

© 2018 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

total of 1 μg of each labeled cRNA was then fragmented by adding 5 μL of 10 × Blocking Agent and 1 μL of 25 × Fragmentation Buffer before heating at 60 °C for 30 min. Finally, 25 μL of 2 × GE Hybridization buffer was added to dilute the labeled cRNA. A sample of 50 μL of the hybridization solution was then dispensed into the gasket slide and assembled to the LncRNA expression microarray slide. The slides were incubated for 17 h at 65 °C in a Microarray Hybridization Oven (Agilent Technologies). The hybridized arrays were then washed, fixed and scanned using the G2505C DNA Microarray Scanner (Agilent Technologies).

2.3.2. Microarray Data Analysis

Agilent Feature Extraction software (version 11.0.1.1) was used to analyze acquired array images. Quantile normalization and subsequent data processing were performed using the GeneSpring GX v12.1 software package (Agilent Technologies). After quantile normalization of the raw data, lncRNAs and mRNAs that were present in ≥3 of 12 samples were selected for further data analysis. Raw *p*-values were calculated by unpaired *t*-test, then the differentially expressed lncRNAs and mRNAs with statistical significance between compared groups were filtered for a FC ≥ 2 and a false discovery rate (FDR) ≤0.10 calculated by the Benjamini–Hochberg procedure. GO analysis was performed using the Bioconductor package topGO[60] and with a raw *p*-value cut-off of 0.05 calculated by Fisher's exact test, subsequently filtered for an FDR ≤ 0.10. Pathway analysis for differentially expressed mRNAs was performed based on the KEGG database (http://www.genome.jp/kegg) with a raw *p*-value cut-off of 0.05 calculated by Fisher's exact test, subsequently filtered for an FDR ≤ 0.10.

2.4. RT-qPCR Validation of lncRNAs Identified as Differentially Expressed by Microarray

Primers for RT-qPCR were designed using OligoPerfect Designer (ThermoFisher Scientific) and synthetized by Eurofins Scientific (Luxembourg) (described in Table S1, Supporting Information). RT-qPCR reactions were conducted using a Mastercycler EP Realplex instrument (Eppendorf) following the QuantiFast SYBR Green RT-PCR Kit protocol (Qiagen). Specificity of amplification was checked by the generation of T_m curves and by analysis of the reaction products by 2% agarose gel electrophoresis to confirm the presence of a single amplicon of the expected size. The results were analyzed applying the standard ΔCt method and normalized to GAPDH, β-actin, and B2M housekeeping gene expression.

3. Results

3.1. Growth Characteristics and Sampling of CHO—Cells Throughout Batch and Fed-Batch Culture for (Non)-Coding RNA Analysis

The steps followed to generate the results presented in this work are summarized in **Figure 1**. Initially, growth comparisons and supplementation testing were undertaken using CD CHO and

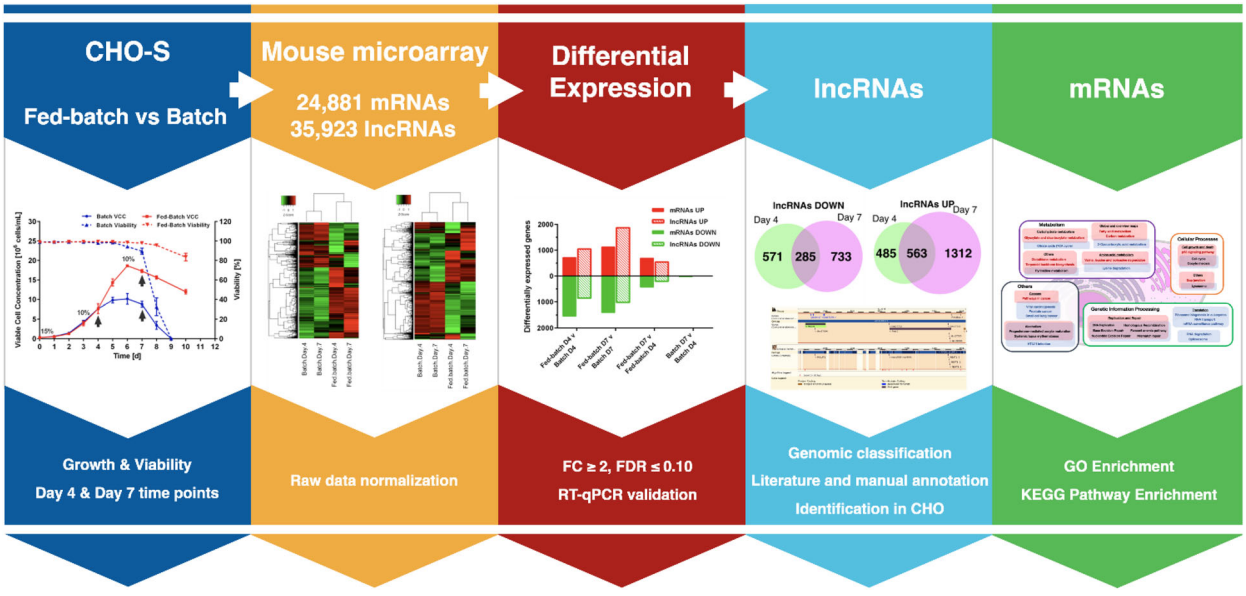


Figure 1. Summary of the experimental workflow. Growth and culture viability of a CHO—S cell line in Batch and Fed-batch cultures were measured for 10 days and samples for RNA extraction taken at Day 4 and at Day 7. The samples were analyzed on a mouse array containing all the coding and non-coding transcripts stored in the main public databases. The measured intensities were log-normalized and differentially expressed transcripts/genes were filtered for a fold change (FC) ≥2 and an FDR ≤ 0.10. A selected group of genes was validated through RT-qPCR (Supplementary Material). Due to the poor annotation of lncRNAs in CHO, the identification of potential targets with a described biological role required the comparison of human and mouse literature and databases, followed by alignment against the Chinese hamster genome, leading to predicted lncRNAs transcripts and previously un-annotated genomic regions (**Table 1**). At the same time, GO and pathway enrichment was implemented on mRNAs.

© 2018 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

the associated commercially available feds (Efficient Feds A and B). From these preliminary experiments, we selected Efficient Feed B Liquid Nutrient Supplement for experimental Fed-batch cultures as this gave the highest viable cell concentrations across a 10-day process from the feed strategies investigated (data not shown). Cells under Fed or Batch culture grew almost identically up until Day 4 of culture whereupon their growth characteristics differed (**Figure 2**A). Whilst growth of the Batch cultures slowed dramatically after Day 4, with a mean peak viable cell concentration on Day 6 of approximately 1×10^7 cells mL^{-1}, in the Fed-batch cultures growth and proliferation continued until Day 6 where a mean peak viable cell concentration of almost 2×10^7 cells mL^{-1} was obtained (Figure 2A). After Day 6 the viable cell number and viability of both cultures decreased with time, however while culture viability rapidly declined in the Batch culture such that by Day 9 culture viability was close to 0%, in the Fed-batch mode cultures had a mean viability of 83.6% at Day 10 (Figure 2A). The first-time point at which samples were harvested was Day 4, when both types of culture had grown in a similar way and were still in the exponential growth phase (late

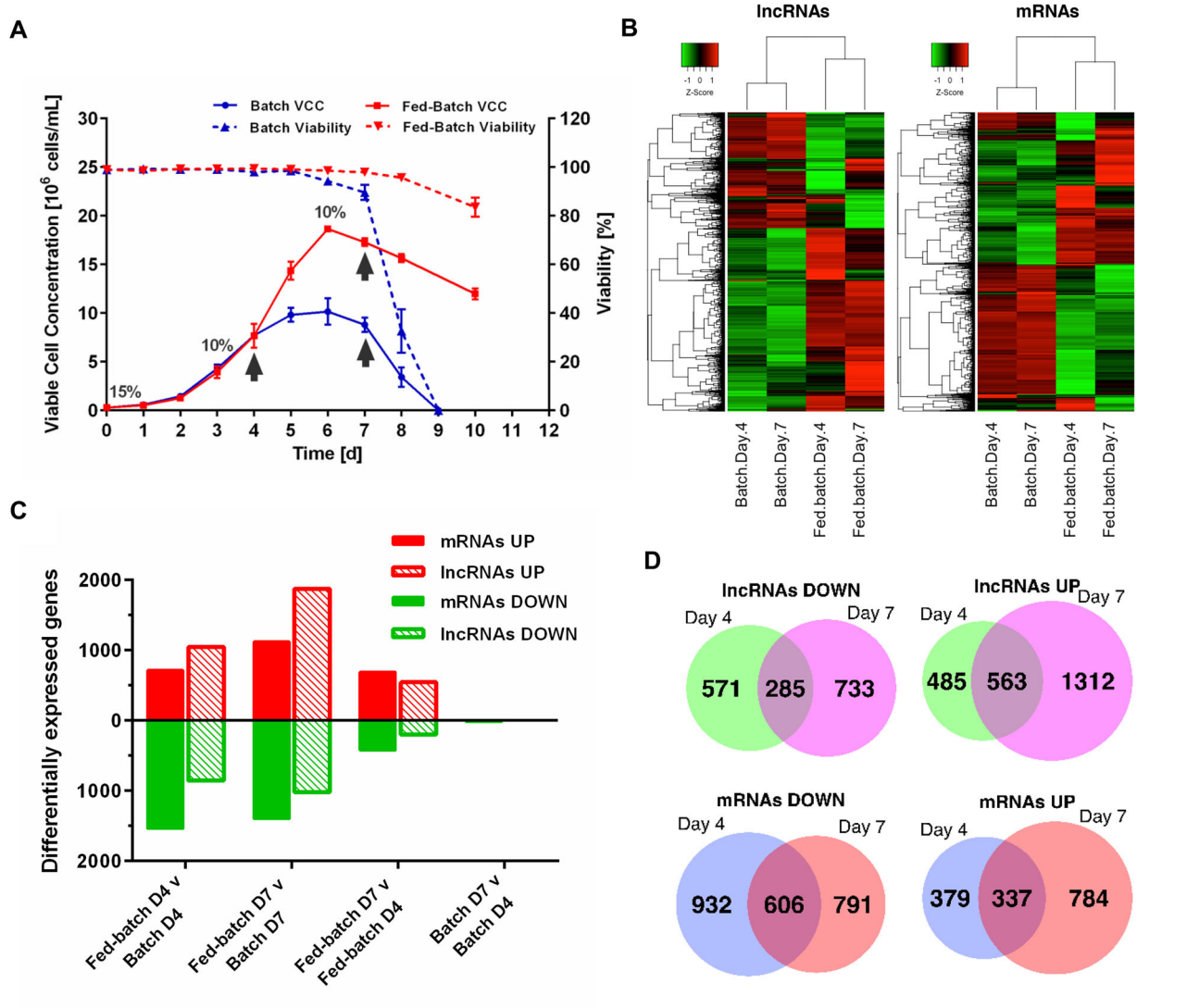


Figure 2. A) Growth profiles for the model CHO—S host cell line throughout Batch and Fed-batch culture. Viable cell concentration (VCC) and culture viability are shown (red for Batch and blue for Fed-batch). The percentage of CHO CD Efficient Feed B Liquid Nutrient Supplement added to the existing working volume are shown for each feeding Day (Days 0, 3, and 6). The arrows indicate when samples for microarray analysis were harvested (Day 4 and Day 7). B) Hierarchical clustering heatmaps arranging samples into groups based on their averaged log-normalized expression levels. Only transcripts/genes with an expression variance between each group above the 80th percentile are shown. The dendrogram shows the relationships for lncRNAs (left panel) and mRNAs (right panel). C) Down-regulated (green bars) and up-regulated (red bars) mRNAs (full bars) and lncRNAs (textured bars) in CHO—S cells during batch and fed-batch culture for each of the compared pairs showing a fold change ≥ 2 and an FDR ≤ 0.10. D) Venn diagrams showing the number of lncRNAs and mRNAs differentially expressed (DE) in Fed-batch vs Batch (FC ≥ 2, FDR ≤ 0.10). Genes DE at both Day 4 and Day 7 are represented in the overlaps while genes DE only at one time point are represented inside their corresponding circle.

© 2018 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

exponential for batch), whilst the second time point at which cells were harvested for RNA analysis was on Day 7, corresponding to the end of the stationary phase/first day of decline for both cultures. In the case of the Fed-batch cultures, these two sampling points were also 24 h after addition of the Efficient Feed B.

3.2. Microarray Analysis of mRNA and lncRNA Transcripts in Batch and Fed-Batch Culture

3.2.1. Differential Expression of lncRNAs and mRNAs in Fed-Batch v Batch Comparisons

The Arraystar Mouse LncRNA Microarray V3.0 used in this study is based on publicly available databases and publications, allowing for the potential simultaneous surveillance of 35923 lncRNAs and 24881 coding transcripts. The lncRNAs collection is based on the NCBI Refseq, UCSC Known Gene 6.0, Ensembl 38.71, Fantom3, RNAdb 2.0, NRED databases, a number of literature publications, T-UCRs, and evolutionary constrained lncRNAs.[61–68] Positive probes for housekeeping genes and negative probes are printed onto the array for hybridization quality control. After quantile normalization of the raw data, 24603 unique lncRNAs and 19617 mRNAs were selected for analysis. Firstly, we implemented Principal Component Analysis (PCA) on the log2 transformed intensities and plotted the first three components against each other, where lncRNAs and mRNAs showed very comparable results, especially where for the first two components samples from Batch cultures at both days grouped closely compared to Fed-batch samples, which showed a wider separation even between the same condition (Figure S1, Supporting Information). Then we applied hierarchical clustering, arranging samples into groups based on their averaged log-normalized expression levels to show the relationships among gene expression patterns of samples for both lncRNAs and mRNAs (Figure 2B). We continued performing differential expression analysis (DE) where we compared Fed-batch against Batch at Day 4, Fed-batch against Batch at Day 7, Fed-batch at Day 7 against Fed-batch at Day 4, and Batch at Day 7 against Batch at Day 4. At this point, the expression of a selected number of the identified DE transcripts was confirmed by RT-qPCR, that showed these transcripts were present but that lower fold-changes were observed between samples than from the array analysis (Table S1, Supporting Information). The genes identified as DE were then filtered based on a threshold fold change (FC) $\geq$2 and a false discovery rate (FDR) $\leq$0.10 (Table S2 and S3, Supporting Information). When we compared Batch Day 7 to Batch Day 4 samples, we found surprisingly low numbers of mRNAs or lncRNAs where the expression changed beyond the set thresholds, with 0 mRNAs being up and 19 down regulated between the 2 days whilst for the lncRNAs there were no transcripts that changed above the thresholds set (Figure 2C). To check whether this lack of identified transcripts changing was due to the FC $\geq$ 2 threshold, we repeated the analysis for FC $\geq$ 1.5 and FC $\geq$ 1.25, FDR $\leq$0.10 (Table S4–S11, Supporting Information). For a FC $\geq$ 1.25, 99 mRNAs were down and 30 were up-regulated while none met the threshold for the lncRNAs. We then carried on with the FC $\geq$ 2, FDR $\leq$ 0.10 thresholds for the remaining comparisons. Fed-batch Day 7 vs. Day 4 revealed 693 mRNAs up-regulated and 421 down-regulated whilst for the lncRNAs we found 545 up-regulated and 200 down-regulated genes (Figure 2C). When we compared Fed-batch against Batch at Day 4 and Day 7, 1048 and 1875 lncRNAs were up-regulated respectively in addition to 856 down-regulated at Day 4 and 1018 down-regulated at Day 7. For the same comparisons we saw more down-regulated mRNAs at Day 4 (1538) and Day 7 (1397) while there were 716 up-regulated mRNAs at Day 4 and 1121 at Day 7 (Figure 2C).

3.3. Identification of lncRNAs Differentially Expressed as Potential Engineering Targets for Modulation of Cell Growth

As one of the aims of this work was to identify new transcripts for manipulation in CHO, we reduced the list of potential targets to a manageable group for further experimental validation. Firstly, we aligned all the 60 nucleotide mouse probes corresponding to the differentially expressed genes identified for each comparison against the Chinese hamster (CH) genome using the discontiguous megablast algorithm to check how many transcripts had an annotation in CHO. Only 16–28% of the probes corresponding to differentially expressed lncRNAs had a matching transcript in CH, as opposed to 58–80% of the mRNAs (Table S12, Supporting Information). This is likely due to the poor annotation of the CH transcriptome compared to mouse, where the number of described lncRNAs is significantly higher. If we consider the ENSEMBL 91 database release alone, 2563 lncRNAs and 446 pseudogenes are listed for Chinese hamster as opposed to 9308 lncRNAs and 12363 pseudogenes in mouse. To address this issue we complemented this approach with literature and databases mining (NCBI, ENSEMBL, lncRNAdb, LNCipedia) to identify differentially expressed lncRNAs already described in mouse but not in Chinese hamster, revealing genomic positions that align with potential lncRNAs (Figure S2–S4, Supporting Information). The identified sequences were then examined using the Rfam database[69] to assess their resemblance to existing non-coding RNA families. A list of potential lncRNAs found using the described approaches is reported in Table 1, along with literature references describing their biological function.

The role of lncRNAs in diseases is well established, especially linked to cell proliferation in cancer[70] and several of those identified here have been investigated in such systems. NEAT1 (nuclear-enriched abundant transcript 1) and MALAT1 (also known as NEAT2), found to be up-regulated at both days in our Fed-batch cultures (Table 1), are among the most well characterized lncRNAs and have been reported to promote cell proliferation through regulation of gene expression at the nuclear level. MALAT1 is a mostly un-spliced transcript around 6.7 kb in mouse, with a long half-life due to the tRNA-like structure adopted at the 3′ end.[37] This structure is cleaved to generate a 61 nt mascRNA (MALAT1-associated small cytoplasmic RNA) exported to the cytoplasm where it is subjected to canonical CCA nucleotides addition and accumulates in the cytoplasm.[71,72] We were able to identify the JH002628.1 genomic region in Chinese hamster using the above mentioned

© 2018 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

Table 1. Summary of the lncRNAs identified using the described approaches, consisting in a direct alignment of the differentially expressed gene probe against the Chinese hamster genome to literature search and RNA families conservation in RFAM.

Gene symbol	NCBI ref	FC Day 4	FC Day 7	FDR Day 4	FDR Day 7	Rfam family	Rfam ID	*E*-value	References
MALAT1	JH002628.1	3.5	4.5	2.4E-02	1.0E-02	MALAT1	RF01871	7.5E-13	Sun et al.[70]; Wilusz[37]
						mascRNA-menRNA	RF01684	7.0E-16	
MEG3	JH001208.1	2.9	2.8	5.9E-02	4.7E-02	MEG3_2	RF01872	3.5E-14	He et al.[106]
MIAT	XM_007625231.2	2.6	0.0	2.6E+00	1.0E-02	MIAT_exon1	RF01874	7.0E-12	Liao et al.[107]
NEAT1	XR_478750.2	6.1	3.3	3.9E-02	6.3E-02	NEAT1_1	RF01955	5.4E-18	Adriaens et al.[75]; Hirose et al.[108]
						NEAT1_2	RF01956	1.2E-17	
						NEAT1_3	RF01957	2.2E−19	
PVT1	XR_478426.2	2.5	3.5	1.2E-01	2.7E-02	PVT1_3	RF02166	1.7E-16	Colombo et al.[79]; Zhu et al.[76]
TERC	AF221928.1	–	6.3	–	4.0E-02	Telomerase-vert	RF00024	5.7E-58	Engreitz et al.[109]
TUG1	XR_483407.1	–	2.1	–	8.0E-02	TUG1_3	RF01891	1.6E-46	Li et al.[110]
						TUG1_1	RF01882	1.3E-21	
						TUG1_4	RF01892	2.1E-20	
						TUG1_2	RF01883	3.2E-17	

From left to right, the columns show the gene symbol, the NCBI accession reference, the fold-change measured at Day 4 and Day 7 with the corresponding FDRs, the ID of the non-coding RNA family found for the sequence in Rfam database with the corresponding ID and *E*-value and the references describing the biological function of the gene.

complementary approaches and we could recognize all of its main conserved domains (Figure S2, Supporting Information). The structure of this genomic region is conserved in the most recent CHOK1GS_HDv1 genome and resembles the pattern found in human and mouse, where close to MALAT1 is the NEAT1 locus. The NEAT1 locus is regulated by alternative 3′ end processing, where the primary transcript can be cleaved and polyadenylated to generate a 3.2 kb long MENε isoform or cleaved and non-polyadenylated to generate the 20.8 kb long MENβ isoform.[37] We were able to identify a predicted 3.2 kb long non-coding transcript (NCBI Ref XR_478750.2) arising from a locus (LOC103159497, C_griseus_v1.0) with similar primary structure to both the 3.2 kb and the 20.8 kb isoforms (*E*-value 0.0, Identity 80%, algorithm megablast). MALAT1 and NEAT1 both localize to the paraspeckles stress-responsive nuclear bodies in the cell,[73] where they are reported to influence the splicing machinery[70,74] and the DNA repair machinery.[75] The Plasmacytoma Variant Translocation 1 (PVT1) lncRNA is considered a biomarker for various cancers due to its ability to promote cell proliferation with a range of proposed mechanisms.[76–79] We identified a predicted non-coding RNA in Chinese hamster (NCBI Ref XR_478426.2) which contained the PVT1_3 RFAM domain and was up-regulated at both Day 4 (FDR = 1.2E-01) and Day 7 in the Fed-batch compared to Batch (Table 1).

3.4. Comparison of mRNAs with Existing Datasets

We compared our Fed-batch against Batch dataset with previous work to identify coding genes related to growth in CHO. Although all these works used different approaches and cell lines, we saw the opportunity to find common patterns of expression in CHO across various conditions. The first dataset compared was Clarke 2011,[80] where we observed 10 of their reported down-regulated genes at Day 4 and 7 at Day 7 of our dataset (Table S13, Supporting Information). On comparing our data to Clarke 2012,[81] only the kinesin family member C1 (KIFC1) gene, up-regulated at both Day 4 and Day 7, was observed to behave in the same way. Lastly, when we considered the translatome analysis in Courtes 2013 we identified 3 genes at Day 4 and 8 genes at Day 7 among our down-regulated transcripts, together with 1 gene at Day 4 and 2 genes at Day 7 among the up-regulated transcripts.[82] Interestingly, a group of the common genes between the selected works and our dataset (CDC20, MAD2L1, MCM7, MCM4, GTF2H4) are involved in cell cycle and DNA replication, supporting the findings in the pathway enrichment analysis (described in section 3.5).[83–85] In addition, we found single genes in consistently enriched pathways in our pathway analysis such as HNRNPC, involved in RNA molecule binding or LGMN, participating in protein degradation in the lysosome.[86] Interestingly, HNRNPs proteins are well known to bind and mediate the functions of lncRNAs, and HNRNPC in particular has been reported to interact with MALAT1 in a tightly regulated N^6-methyladenosine-dependent manner.[87,88]

3.5. GO Analysis and Pathway Enrichment

We performed GO term analysis on the differentially expressed mRNAs with an FDR cut-off of 0.10 calculated by the BH method, followed by pathway enrichment analysis based on the KEGG database (Table S14–S29, Supporting Information), allowing the determination of the significantly enriched biological pathways filtered for an FDR $\leq$ 0.10 and grouped based on KEGG class annotation (**Table 2**). The pathway enrichment identified 22 pathways containing down-regulated

© 2018 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

Table 2. Summary of the enriched pathways based on the list of up-regulated (lysosome) and down-regulated (lower tables) transcripts in Fed-batch vs Batch comparison at Day 4 (left panel) and Day 7 (right panel).

KEGG ID	Pathway	FDR	Diff. expressed (DE) genes	Total genes (TG)	DE:TG ratio
		UP fed D4 v batch D4			
04142	Lysosome	6.3E-02	14	124	0.11
		UP fed D7 v batch D7			
04142	Lysosome	7.1E-04	21	124	0.17
		DOWN fed D4 v batch D4			
05034	Alcoholism	8.5E-11	48	203	0.24
04110	Cell cycle	2.3E-09	34	125	0.27
05322	Systemic lupus erythematosus	6.8E-07	33	147	0.22
03030	DNA replication	3.5E-04	12	35	0.34
03460	Fanconi anemia pathway	8.5E-04	14	51	0.27
05203	Viral carcinogenesis	1.1E-03	37	243	0.15
03410	Base excision repair	1.3E-03	11	35	0.31
03430	Mismatch repair	4.4E-03	8	22	0.36
01200	Carbon metabolism	4.5E-03	21	117	0.18
00480	Glutathione metabolism	5.7E-03	13	56	0.23
00900	Terpenoid backbone biosynthesis	6.4E-03	8	24	0.33
04114	Oocyte meiosis	8.5E-03	20	116	0.17
05200	Pathways in cancer	1.3E-02	48	397	0.12
03440	Homologous recombination	1.6E-02	8	28	0.29
04914	Progesterone-mediated oocyte maturation	1.8E-02	16	90	0.18
01212	Fatty acid metabolism	2.1E-02	11	51	0.22
04540	Gap junction	2.8E-02	15	86	0.17
00280	Valine, leucine and isoleucine degradation	4.2E-02	11	56	0.20
04115	p53 signaling pathway	6.4E-02	12	68	0.18
00240	Pyrimidine metabolism	6.4E-02	16	104	0.15
03420	Nucleotide excision repair	6.5E-02	9	44	0.20
00630	Glyoxylate and dicarboxylate metabolism	7.6E-02	7	30	0.23
		DOWN fed D7 v batch D7			
04110	Cell cycle	7.9E-14	38	125	0.30
03030	DNA replication	1.2E-09	17	35	0.49
03013	RNA transport	1.2E-09	38	170	0.22
03040	Spliceosome	1.3E-07	30	134	0.22
05203	Viral carcinogenesis	2.8E-06	40	243	0.16
03430	Mismatch repair	2.2E-05	10	22	0.45
03460	Fanconi anemia pathway	2.3E-05	15	51	0.29
03410	Base excision repair	3.3E-04	11	35	0.31
05322	Systemic lupus erythematosus	9.3E-04	24	147	0.16
03440	Homologous recombination	1.4E-03	9	28	0.32
03420	Nucleotide excision repair	2.5E-03	11	44	0.25
05034	Alcoholism	3.5E-03	28	203	0.14
04114	Oocyte meiosis	4.2E-03	19	116	0.16
03008	Ribosome biogenesis in eukaryotes	2.0E-02	14	83	0.17
00240	Pyrimidine metabolism	2.3E-02	16	104	0.15
03015	mRNA surveillance pathway	2.6E-02	15	96	0.16

(*Continued*)

© 2018 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

Table 2. (*Continued*)

KEGG ID	Pathway	FDR	Diff. expressed (DE) genes	Total genes (TG)	DE:TG ratio
00310	Lysine degradation	3.1E-02	10	52	0.19
04914	Progesterone-mediated oocyte maturation	3.4E-02	14	90	0.16
03018	RNA degradation	4.4E-02	13	83	0.16
05166	HTLV-I infection	4.8E-02	32	294	0.11
05222	Small cell lung cancer	4.9E-02	13	85	0.15
00020	Citrate cycle (TCA cycle)	5.5E-02	7	32	0.22
05215	Prostate cancer	6.7E-02	13	89	0.15
01210	2-Oxocarboxylic acid metabolism	1.0E-01	5	20	0.25

The columns show from left to right the ID from KEGG, the pathway name, the FDR associated with the enrichment (FDR $\leq$ 0.10 threshold), the number of differentially expressed genes in the pathway, the number of total genes listed in the pathway, the differentially expressed genes over total genes ratio.

genes at Fed-batch vs. Batch Day 4 and 24 at Fed-batch vs. Batch Day 7 while only one contained up-regulated genes (Lysosome, ID: 04142) at either day. This is most likely a reflection of the lower number of up-regulated genes compared to down-regulated (**Figure 3**). The Metabolism domain includes seven enriched pathways at Day 4 while only one of these was still enriched at Day 7, suggesting a central role of metabolism together with p53 signalling predominantly during the exponential growth phase. On the contrary, toward Day 7 we see the prevalent enrichment of pathways related to translation regulation and RNA interaction at different levels, from transport to splicing. The most evident pattern of enrichment between Day 4 and Day 7 is represented by the Replication and Repair class, where the majority of the pathways involved in genome maintenance and diverse repair mechanisms are consistently enriched, indicating an early and sustained regulation of these genes throughout culture. DNA damage is reported to stimulate the expression of NEAT1 and, together with MALAT1, to promote the formation of paraspeckles, which regulate alternative splicing and promote proliferation.[73,75,89] We then compared our results with relevant KEGG pathway enrichment datasets available for CHO[90] and found four common pathways, which were enriched exclusively at Day 7 in our dataset: RNA Transport (ID: 03013), mRNA surveillance (ID: 03015), RNA degradation (ID: 03018), Spliceosome (ID: 03040). Taken together, these results suggest the importance of cell cycle

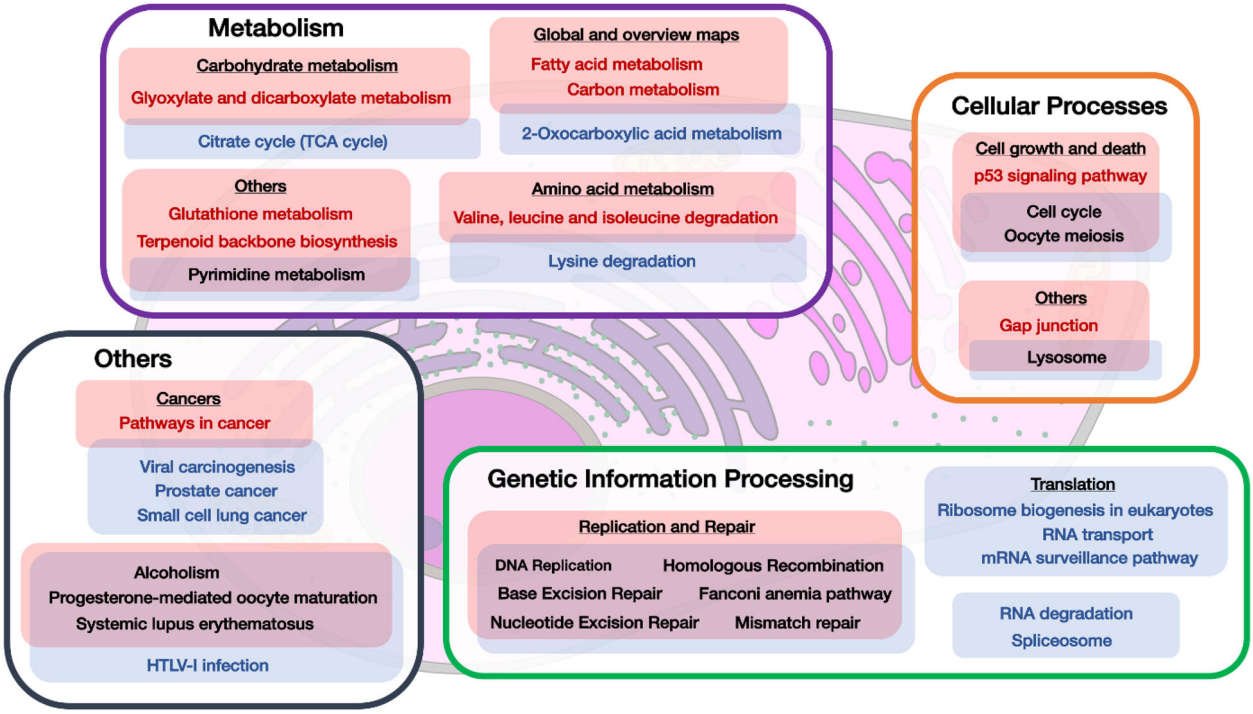


Figure 3. Graphical representation of the enriched KEGG pathways listed in **Table 2**, hierarchically grouped based on KEGG Pathway Maps. Each filled rectangle (in red for Day 4, in blue for Day 7) contains the corresponding enriched pathways, with pathways enriched at both Day 4 and Day 7 enclosed in the overlaps between the filled rectangles. The enrichment is based on genes differentially expressed in Fed-batch vs. Batch at Day 4 and Day 7 with an FDR $\leq$ 0.10.

© 2018 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

and genome repair mechanism control likely due to the high proliferation of the fed-batch system. In addition, the active regulation of RNA transport, RNA maintenance and splicing seems to be particularly important toward the later stages of our Fed-batch cultures.

4. Discussion

By taking advantage of a comprehensive commercially available mouse microarray containing 35923 lncRNAs and 24881 mRNAs, we have provided the first mapping of the CHO lncRNA landscape, together with the coding transcriptome. Previous reports have shown that more than 70% of the assembled CHO transcriptome is similar to mouse (*Mus musculus*) and closely related to rat (*Rattus norvegicus*) transcriptomes[58] suggesting that this approach was likely to be valid for lncRNAs as well. Due to the species and tissue-specificity of lncRNAs compared to mRNAs, the number of detectable lncRNAs in CHO is likely to be lower than the 35923 probes included in the array. Nevertheless, using this approach we were able to detect 24603 lncRNAs (68.5% of the total probes) and 19617 mRNAs (78.8% of the total probes), and found that the several hundreds of lncRNAs exhibit changing expression profiles on different days of culture and between Batch and Fed-batch culture in a model CHO—S system. This was especially true for the Fed system, where comparing Day 4 and Day 7 we observed 1114 differentially expressed mRNAs and 745 lncRNAs, as opposed to the Batch, were we saw only 19 differentially expressed mRNAs and no lncRNA for the same comparison, suggesting a prevalent variability induced by the Fed supplement as compared to time only.

Among the differentially expressed genes, between 16–28% of the lncRNAs probes had a matching transcript in CH, as opposed to 58–80% of the mRNAs. This required a specific approach where the Chinese hamster genome, literature search, and databases mining were combined to detect lncRNAs differentially expressed in our system with an established biological function. Within these lncRNAs, we focused on MALAT1, NEAT1, and PVT1 to provide a comparison between the mouse gene and the Chinese hamster putative homologues. The number of lncRNAs with a fully understood role in the cell remains small, however these three non-coding genes are among the most well characterized.[91] MALAT1 and NEAT1 in particular are associated with increased proliferation and participate in the regulation of alternative splicing and DNA repair, which we found to be strongly regulated in our pathway enrichment analysis. This suggests a potential role for these lncRNAs in CHO, although further experimental studies on the single genes are now required to assess the actual effects on the cell under different conditions.

Several approaches to investigate and confirm the functional annotation of lncRNAs in other organisms have been described, including the perturbation of lncRNA expression by over-expression, knockout or knockdown,[92,93] in addition to complementary strategies.[39] Future developments in CHO will have to proceed with a mix of functional prediction tools to assess the properties of the transcriptome and evaluate the degree of conservation with other species[94–96] and of targeted approaches to ameliorate the annotation and propose mechanisms of action for the single transcripts.[97–100] Further, the data reported here is for both lncRNAs and mRNAs and hence will allow investigators to further probe the relationships between the expression and regulation of these two classes of RNA. As the majority of the lncRNAs reported in the literature are discussed and related to human or model organism systems, our work aimed at unveiling the role of lncRNAs in CHO under industrially relevant conditions to identify new targets for manipulation to sustain proliferation. Examples of successful cell engineering of lncRNAs to selectively enhance translation[101] and product yield[57] have already been reported in CHO, demonstrating the potential of manipulation of lncRNAs for enhancing industrial processes. Moreover, since it was reported up to 15% of the total ribosome occupancy can be occupied by a single recombinant mRNA, the intrinsic characteristics of lncRNAs place them as ideal candidates for cell line engineering of protein production cell factories, as they do not add any translational burden on top of the coding gene of interest.[102]

Our work has identified potential lncRNA targets differentially expressed in Fed-batch compared to Batch culture from which we selected a group of molecules to be experimentally studied (Table 1). In addition to the expression of lncRNAs we also looked at expression of mRNAs (coding transcripts) and found a consistent change in differentially expressed mRNAs when comparing batch and fed-batch cultures. Pathway enrichment analysis (Figure 3) underlined the importance of genes involved in cell cycle and genome maintenance pathways along with the regulation of lysosome formation as potential targets for cell engineering to enhance proliferation. Our approach allowed the identification of previously undescribed lncRNAs in CHO along with mRNAs to identify the connections between them and compared these with existent literature. This network of reciprocal interactions is beginning to be unveiled in other organisms[103–105] and our work will help pave the way for the definition of new layers of regulation involving single transcripts or even entire pathways in CHO.

5. Conclusions

He we report on the lncRNA landscape and how this changes in CHO cells, presenting a full dataset of those lncRNAs present as determined from an array study and how these change through a Batch and Fed-batch culture and between the two culture systems. From analysis of the data, we have determined those lncRNAs whose expression changes the most between 2 days in culture and between fed and batch culture that are attractive targets for cell engineering. This resource will now provide the community with the opportunity to undertake functional validation studies by undertaking single or multiple knock downs/outs, or by the up-regulation of target lncRNAs, and determine the impact on growth, and productivity, characteristics of CHO cells. Ultimately, we anticipate such a resource will be incorporated into wider genome analysis datasets including coding mRNAs and other non-coding RNAs to develop a wider appreciation of the role of RNAs in controlling recombinant CHO cell line growth and productivity characteristics.

© 2018 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

Abbreviations

arcRNAs, architectural RNAs; ceRNAs, competing endogenous RNAs; CH, Chinese hamster; CHO, Chinese hamster ovary cells; DE, differentially expressed; eRNAs, enhance; FC, fold change; FDR, false discovery rate; HNRNPs, Heterogeneous nuclear ribonucleoproteins; KIFC1, kinesin family member C1; lncRNAs, Long non-coding RNAs; mAbs, monoclonal antibodies; ORF, open reading frame; PCA, principal component analysis; PVT1, Plasmacytoma Variant Translocation 1.

Supporting Information

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

Acknowledgements

The authors acknowledge the European Commission for funding for DV and supporting this work (EC – Horizon 2020 MSCA ITN 2014 – 642663). DV undertook experiments, analyzed the data and co-wrote the paper; CMS devised the project and with DV devised experiments, aided in experimental design, analyzed data, and co-wrote the paper. The authors acknowledge the Noun Project for sharing the icons used in the graphical abstract.

Conflict of Interest

The authors declare no commercial or financial conflict of interest.

Keywords

Chinese hamster ovary (CHO) cells, long non-coding RNAs (lncRNAs), microarray, recombinant protein production, transcriptome

Received: February 26, 2018
Revised: April 16, 2018
Published online:

[1] G. Walsh, *Nat. Biotechnol*. **2014**, *32*, 992.
[2] J. Y. Kim, Y. G. Kim, G. M. Lee, *Appl. Microbiol. Biot*. **2012**, *93*, 917.
[3] K. P. Jayapal, K. F. Wlaschin, W. S. Hu, M. G. S. Yap, *Chem. Eng. Prog*. **2007**, *103*, 40.
[4] J. Dumont, D. Euwart, B. Mei, S. Estes, R. Kshirsagar, *Crit. Rev. Biotechnol*. **2015**, *1*, 1.
[5] R. Kunert, D. Reinhart, *Appl. Microbiol. Biotechnol*. **2016**, *100*, 3451.
[6] X. Pan, M. Streefland, C. Dalm, R. H. Wijffels, D. E. Martens, *Cytotechnology* **2017**, *69*, 39.
[7] Y. Durocher, M. Butler, *Curr. Opin. Biotechnol*. **2009**, *20*, 700.
[8] D. C. F. Wong, K. T. K. Wong, Y. Y. Lee, P. N. Morin, C. K. Heng, M. G. S. Yap, *Biotechnol. Bioeng*. **2006**, *94*, 373.
[9] J. F. Povey, C. J. O'Malley, T. Root, E. B. Martin, G. A. Montague, M. Feary, C. Trim, D. A. Lang, R. Alldread, A. J. Racher, C. M. Smales, *J. Biotechnol*. **2014**, *184*, 84.
[10] D. Reinhart, L. Damjanovic, C. Kaisermayer, R. Kunert, *Appl. Microbiol. Biotechnol*. **2015**, *99*, 4645.
[11] C. S. Alves, T. M. Dobrowsky, in *Heterologous Protein Production in CHO Cells: Methods and Protocols* (Ed: P. Meleady), Springer, New York **2017**, Ch.1.
[12] L. Jossé, J. Xie, C. G. Proud, C. M. Smales, *Biochem. J*. **2016**, *473*, 4651.
[13] L. Josse, C. M. Smales, M. F. Tuite, *Biotechnol. Bioeng*. **2010**, *105*, 556.
[14] G. M. Lee, *J. Biotechnol*. **2008**, *136*, S130.
[15] C. Ferrara, P. Brünker, T. Suter, S. Moser, U. Püntener, P. Umaña, *Biotechnol. Bioeng*. **2006**, *93*, 851.
[16] N. Yamane-Ohnuki, S. Kinoshita, M. Inoue-Urakubo, M. Kusunoki, S. Iida, R. Nakano, M. Wakitani, R. Niwa, M. Sakurada, K. Uchida, K. Shitara, M. Satoh, *Biotechnol. Bioeng*. **2004**, *87*, 614.
[17] L. Malphettes, Y. Freyvert, J. Chang, P.-Q. Liu, E. Chan, J. C. Miller, Z. Zhou, T. Nguyen, C. Tsai, A. W. Snowden, T. N. Collingwood, P. D. Gregory, G. J. Cost, *Biotechnol. Bioeng*. **2010**, *106*, 774.
[18] J.-X. Bi, J. Shuttleworth, M. Al-Rubeai, *Biotechnol. Bioeng* **2004**, *85*, 741.
[19] M. Fussenegger, X. Mazur, J. E. Bailey, *Biotechnol. Bioeng*. **1997**, *55*, 927.
[20] N. S. Kim, G. M. Lee, *Biotechnol. Bioeng*. **2002**, *78*, 217.
[21] S. S. Choi, W. J. Rhee, E. J. Kim, T. H. Park, *Biotechnol. Bioeng*. **2006**, *95*, 459.
[22] S. O. Hwang, G. M. Lee, *J. Biotechnol*. **2009**, *139*, 89.
[23] J. S. Lee, T. K. Ha, J. H. Park, G. M. Lee, *Biotechnol. Bioeng*. **2013**, *110*, 2195.
[24] M. Mason, B. Sweeney, K. Cain, P. Stephens, S. T. Sharfstein, *Biotechnol. Prog*. **2012**, *28*, 846.
[25] L. P. Pybus, G. Dean, N. R. West, A. Smith, O. Daramola, R. Field, S. J. Wilkinson, D. C. James, *Biotechnol. Bioeng*. **2014**, *111*, 372.
[26] Y. B. Johari, S. D. Estes, C. S. Alves, M. S. Sinacore, D. C. James, *Biotechnol. Bioeng*. **2015**, *112*, 2527.
[27] L. Thoring, D. A. Wustenhagen, M. Borowiak, M. Stech, A. Sonnabend, S. Kubick, *PLoS ONE* **2016**, *11*, e0163670.
[28] E. J. Mead, L. M. Chiverton, S. K. Spurgeon, E. B. Martin, G. A. Montague, C. M. Smales, T. von der Haar, *PLoS ONE* **2012**, *7*, e47422.
[29] E. J. Mead, R. J. Masterton, M. Feary, O. Obrezanova, L. Zhang, R. Young, C. M. Smales, *Biochem. J*. **2015**, *472*, 261.
[30] F. Stiefel, S. Fischer, A. Sczyrba, K. Otte, F. Hesse, *J. Biotechnol*. **2016**, *225*, 31.
[31] J. A. H. Bort, M. Hackl, H. Höflmayer, V. Jadhav, E. Harreither, N. Kumar, W. Ernst, J. Grillari, N. Borth, *Biotechnol. J*. **2012**, *7*, 500.
[32] S. Fischer, K. F. Marquart, L. A. Pieper, J. Fieder, M. Gamer, I. Gorr, P. Schulz, H. Bradl, *Biotechnol. Bioeng*. **2017**, *114*, 1495.
[33] N. Barron, N. Kumar, N. Sanchez, P. Doolan, C. Clarke, P. Meleady, F. O'Sullivan, M. Clynes, *J. Biotechnol*. **2011**, *151*, 204.
[34] M. Hackl, N. Borth, J. Grillari, *Trends Biotechnol*. **2012**, *30*, 405.
[35] S. Djebali, C. A. Davis, A. Merkel, A. Dobin, T. Lassmann, A. Mortazavi, A. Tanzer, J. Lagarde, W. Lin, F. Schlesinger, C. Xue, G. K. Marinov, J. Khatun, B. A. Williams, C. Zaleski, J. Rozowsky, M. Röder, F. Kokocinski, R. F. Abdelhamid, T. Alioto, I. Antoshechkin, M. T. Baer, N. S. Bar, P. Batut, K. Bell, I. Bell, S. Chakrabortty, X. Chen, J. Chrast, J. Curado, T. Derrien, J. Drenkow, E. Dumais, J. Dumais, R. Duttagupta, E. Falconnet, M. Fastuca, K. Fejes-Toth, P. Ferreira, S. Foissac, M. J. Fullwood, H. Gao, D. Gonzalez, A. Gordon, H. Gunawardena, C. Howald, S. Jha, R. Johnson, P. Kapranov, B. King, C. Kingswood, O. J. Luo, E. Park, K. Persaud, J. B. Preal, P. Ribeca, B. Risk, D. Robyr, M. Sammeth, L. Schaffer, L. H. See, A. Shahab, J. Skancke, A. M. Suzuki, H. Takahashi, H. Tilgner, D. Trout, N. Walters, H. Wang, J. Wrobel, Y. Yu, X. Ruan, Y. Hayashizaki, J. Harrow, M. Gerstein, T. Hubbard, A. Reymond, S. E. Antonarakis, G. Hannon, M. C. Giddings, Y. Ruan, B. Wold, P. Carninci, R. Guigó, T. R. Gingeras, *Nature* **2012**, *489*, 101.
[36] P. Kapranov, J. Cheng, S. Dike, D. A. Nix, R. Duttagupta, A. T. Willingham, P. F. Stadler, J. Hertel, J. Hackermuller,

 © 2018 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

I. L. Hofacker, I. Bell, E. Cheung, J. Drenkow, E. Dumais, S. Patel, G. Helt, M. Ganesh, S. Ghosh, A. Piccolboni, V. Sementchenko, H. Tammana, T. R. Gingeras, *Science* **2007**, *316*, 1484.
[37] J. E. Wilusz, *Biochim. Biophys. Acta* **2016**, *1859*, 128.
[38] J. T. Y. Kung, D. Colognori, J. T. Lee, *Genetics* **2013**, *193*, 651.
[39] K. Kashi, L. Henderson, A. Bonetti, P. Carninci, *Biochim. Biophys. Acta* **2016**, *1859*, 3.
[40] J. G. Betancur, *Biochim. Biophys. Acta* **2016**, *1859*, 93.
[41] T. Trimarchi, E. Bilal, P. Ntziachristos, G. Fabbri, R. Dalla-Favera, A. Tsirigos, I. Aifantis, *Cell* **2014**, *158*, 593.
[42] I. Gonzalez, R. Munita, E. Agirre, T. A. Dittmer, K. Gysling, T. Misteli, R. F. Luco, *Nat. Struct. Mol. Biol*. **2015**, *22*, 1.
[43] C. Carrieri, L. Cimatti, M. Biagioli, A. Beugnet, S. Zucchelli, S. Fedele, E. Pesce, I. Ferrer, L. Collavin, C. Santoro, A. R. R. Forrest, P. Carninci, S. Biffo, E. Stupka, S. Gustincich, *Nature* **2012**, *491*, 454.
[44] S. Geisler, J. Coller, *Nat. Rev. Mol. Cell Biol*. **2013**, *14*, 699.
[45] A. M. Schmitt, H. Y. Chang, *Cancer Cell* **2016**, *29*, 452.
[46] R. B. T. Perry, I. Ulitsky, *Development* **2016**, *143*, 3882.
[47] Y. Tay, J. Rinn, P. P. Pandolfi, *Nature* **2014**, *505*, 344.
[48] T. Chujo, T. Yamazaki, T. Hirose, *Biochim. Biophys. Acta* **2016**, *1859*, 1.
[49] C. A. Melo, J. Drost, P. J. Wijchers, H. van de Werken, E. de Wit, J. A. F. O. Vrielink, R. Elkon, S. A. Melo, N. Léveillé, R. Kalluri, W. de Laat, R. Agami, *Mol. Cell* **2013**, *49*, 524.
[50] M. Koziol, J. Rinn, *Curr. Opin. Genet. Dev*. **2010**, *20*, 142.
[51] J. A. Gomez, O. L. Wapinski, Y. W. Yang, J. F. Bureau, S. Gopinath, D. M. Monack, H. Y. Chang, M. Brahic, K. Kirkegaard, *Cell* **2013**, *152*, 743.
[52] A. Keniry, D. Oxley, P. Monnier, M. Kyba, L. Dandolo, G. Smits, W. Reik, *Nat. Cell Biol*. **2012**, *14*, 659.
[53] F. Rashid, A. Shah, G. Shan, *Genomics Proteomics Bioinformatics* **2016**, *14*, 73.
[54] B. Liu, L. Sun, Q. Liu, C. Gong, Y. Yao, X. Lv, L. Lin, H. Yao, F. Su, D. Li, M. Zeng, E. Song, *Cancer Cell* **2015**, *27*, 370.
[55] P. Wang, Y. Xue, Y. Han, L. Lin, C. Wu, S. Xu, Z. Jiang, J. Xu, Q. Liu, *Science* **2014**, *344*, 310.
[56] L. Patrucco, A. Chiesa, M. F. Soluri, F. Fasolo, H. Takahashi, P. Carninci, S. Zucchelli, C. Santoro, S. Gustincich, D. Sblattero, D. Cotella, *Gene* **2015**, *569*, 287.
[57] S. Zucchelli, L. Patrucco, F. Persichetti, S. Gustincich, D. Cotella, *Comput. Struct. Biotechnol. J*. **2016**, *14*, 404.
[58] J. Becker, M. Hackl, O. Rupp, T. Jakobi, J. Schneider, R. Szczepanowski, T. Bekel, N. Borth, A. Goesmann, J. Grillari, C. Kaltschmidt, T. Noll, A. Pühler, A. Tauch, K. Brinkrolf, *J. Biotechnol*. **2011**, *156*, 227.
[59] X. Xu, H. Nagarajan, N. E. Lewis, S. Pan, Z. Cai, X. Liu, W. Chen, M. Xie, W. Wang, S. Hammond, M. R. Andersen, N. Neff, B. Passarelli, W. Koh, H. C. Fan, J. Wang, Y. Gui, K. H. Lee, M. J. Betenbaugh, S. R. Quake, I. Famili, B. O. Palsson, J. Wang, *Nat. Biotechnol*. **2011**, *29*, 735.
[60] A. Alexa, J. Rahnenfuhrer, topGO: Enrichment Analysis for Gene Ontology. R package version 2.28.0. **2016**.
[61] F. Cunningham, M. R. Amode, D. Barrell, K. Beal, K. Billis, S. Brent, D. Carvalho-Silva, P. Clapham, G. Coates, S. Fitzgerald, L. Gil, C. G. Giron, L. Gordon, T. Hourlier, S. E. Hunt, S. H. Janacek, N. Johnson, T. Juettemann, A. K. Kahari, S. Keenan, F. J. Martin, T. Maurel, W. McLaren, D. N. Murphy, R. Nag, B. Overduin, A. Parker, M. Patricio, E. Perry, M. Pignatelli, H. S. Riat, D. Sheppard, K. Taylor, A. Thormann, A. Vullo, S. P. Wilder, A. Zadissa, B. L. Aken, E. Birney, J. Harrow, R. Kinsella, M. Muffato, M. Ruffier, S. M. Searle, G. Spudich, S. J. Trevanion, A. Yates, D. R. Zerbino, P. Flicek, *Nucleic Acids Res*. **2015**, *43*, D662.
[62] K. D. Pruitt, T. Tatusova, D. R. Maglott, *Nucleic Acids Res*. **2007**, *35*, D61.
[63] P. P. Amaral, M. B. Clark, D. K. Gascoigne, M. E. Dinger, J. S. Mattick, *Nucleic Acids Res*. **2011**, *39*, D146.
[64] G. Bejerano, M. Pheasant, I. Makunin, S. Stephen, W. J. Kent, J. S. Mattick, D. Haussler, *Science* **2004**, *304*, 1321.
[65] J. L. Rinn, M. Kertesz, J. K. Wang, S. L. Squazzo, X. Xu, S. A. Brugmann, L. H. Goodnough, J. A. Helms, P. J. Farnham, E. Segal, H. Y. Chang, *Cell* **2007**, *129*, 1311.
[66] T. R. Mercer, M. E. Dinger, S. M. Sunkin, M. F. Mehler, J. S. Mattick, *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 716.
[67] M. Guttman, I. Amit, M. Garber, C. French, M. F. Lin, D. Feldser, M. Huarte, O. Zuk, B. W. Carey, J. P. Cassady, M. N. Cabili, R. Jaenisch, T. S. Mikkelsen, T. Jacks, N. Hacohen, B. E. Bernstein, M. Kellis, A. Regev, J. L. Rinn, E. S. Lander, *Nature* **2009**, *458*, 223.
[68] D. A. Benson, I. Karsch-Mizrachi, D. J. Lipman, J. Ostell, D. L. Wheeler, *Nucleic Acids Res*. **2004**, *32*, D23.
[69] I. Kalvari, J. Argasinska, N. Quinones-Olvera, E. P. Nawrocki, E. Rivas, S. R. Eddy, A. Bateman, R. D. Finn, A. I. Petrov, *Nucleic Acids Res*. **2018**, *46*, D335.
[70] W. Sun, Y. Yang, C. Xu, J. Guo, *Cancer Genet*. **2017**, *216-217*, 105.
[71] J. E. Wilusz, *Wiley Interdiscip. Rev. RNA* **2015**, *6*, 453.
[72] J. E. Wilusz, S. M. Freier, D. L. Spector, *Cell* **2008**, *135*, 919.
[73] S. Nakagawa, T. Hirose, *Cell. Mol. Life Sci*. **2012**, *69*, 3027.
[74] V. Tripathi, J. D. Ellis, Z. Shen, D. Y. Song, Q. Pan, A. T. Watt, S. M. Freier, C. F. Bennett, A. Sharma, P. A. Bubulya, B. J. Blencowe, S. G. Prasanth, K. V. Prasanth, *Mol. Cell* **2010**, *39*, 925.
[75] C. Adriaens, L. Standaert, J. Barra, M. Latil, A. Verfaillie, P. Kalev, B. Boeckx, P. W. G. Wijnhoven, E. Radaelli, W. Vermi, E. Leucci, G. Lapouge, B. Beck, J. van den Oord, S. Nakagawa, T. Hirose, A. A. Sablina, D. Lambrechts, S. Aerts, C. Blanpain, J-C. Marine, *Nat. Med*. **2016**, *22*, 861.
[76] S. Zhu, P. Shuai, C. Yang, Y. Zhang, S. Zhong, X. Liu, K. Chen, Q. Ran, H. Yang, Y. Zhou, S. Zhu, P. Shuai, C. Yang, Y. Zhang, S. Zhong, X. Liu, K. Chen, Q. Ran, H. Yang, Y. Zhou, *Oncotarget* **2017**, *8*, 113174.
[77] F. Wang, J.-H. Yuan, S.-B. Wang, F. Yang, S.-X. Yuan, C. Ye, N. Yang, W.-P. Zhou, W.-L. Li, W. Li, S.-H. Sun, *Hepatology* **2014**, *60*, 1278.
[78] Y.-Y. Tseng, A. Bagchi, *Mol. Cell. Oncol*. **2015**, *2*, e974467.
[79] T. Colombo, L. Farina, G. Macino, P. Paci, *Biomed Res. Int*. **2015**, *2015*, 304208.
[80] C. Clarke, P. Doolan, N. Barron, P. Meleady, F. O'Sullivan, P. Gammell, M. Melville, M. Leonard, M. Clynes, *J. Biotechnol*. **2011**, *155*, 350.
[81] C. Clarke, M. Henry, P. Doolan, S. Kelly, S. Aherne, N. Sanchez, P. Kelly, P. Kinsella, L. Breen, S. F. Madden, L. Zhang, M. Leonard, M. Clynes, P. Meleady, N. Barron, *BMC Genomics* **2012**, *13*, 656.
[82] F. C. Courtes, J. Lin, H. L. Lim, S. W. Ng, N. S. C. Wong, G. Koh, L. Vardy, M. G. S. Yap, B. Loo, D. Y. Lee, *J. Biotechnol*. **2013**, *167*, 215.
[83] Y. Zhai, N. Li, H. Jiang, X. Huang, N. Gao, B. K. Tye, *Mol. Cell* **2017**, *67*, 168.
[84] H. Yu, *Curr. Opin. Cell Biol*. **2002**, *14*, 706.
[85] S. C. Buch, B. Diergaarde, T. Nukui, R. S. Day, J. M. Siegfried, M. Romkes, J. L. Weissfeld, *Mol. Carcinog*. **2012**, *51*, E11.
[86] E. Dall, H. Brandstetter, *Biochimie* **2016**, *122*, 126.
[87] X. Sun, M. S. S. Haider Ali, M. Moran, *Biochem. J*. **2017**, *474*, 2925.
[88] N. Liu, Q. Dai, G. Zheng, C. He, M. Parisien, T. Pan, *Nature* **2015**, *518*, 560.
[89] C. Adriaens, J.-C. Marine, *Cell Cycle* **2017**, *16*, 137.
[90] E. Harreither, M. Hackl, J. Pichler, S. Shridhar, N. Auer, P. P. Łabaj, M. Scheideler, M. Karbiener, J. Grillari, D. P. Kreil, N. Borth, *Biotechnol. J*. **2015**, *10*, 1625.
[91] F. Kopp, J. T. Mendell, *Cell* **2018**, *172*, 393.
[92] S. Zhu, W. Li, J. Liu, C.-H. Chen, Q. Liao, P. Xu, H. Xu, T. Xiao, Z. Cao, J. Peng, P. Yuan, M. Brown, X. S. Liu, W. Wei, *Nat. Biotechnol*. **2016**, *34*, 1279.

© 2018 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

[93] S. J. Liu, M. A. Horlbeck, S. W. Cho, H. S. Birk, M. Malatesta, D. He, F. J. Attenello, J. E. Villalta, M. Y. Cho, Y. Chen, M. A. Mandegar, M. P. Olvera, L. A. Gilbert, B. R. Conklin, H. Y. Chang, J. S. Weissman, D. A. Lim, *Science* **2017**, *355*, 6320.
[94] B. Signal, B. S. Gloss, M. E. Dinger, *Trends Genet*. **2016**, *32*, 620.
[95] J. Iwakiri, M. Hamada, K. Asai, *Biochim. Biophys. Acta* **2016**, *1859*, 23.
[96] I. Ulitsky, *Nat. Rev. Genet*. **2016**, *17*, 601.
[97] M. B. Clark, T. R. Mercer, G. Bussotti, T. Leonardi, K. R. Haynes, J. Crawford, M. E. Brunck, K.-A. L. Cao, G. P. Thomas, W. Y. Chen, R. J. Taft, L. K. Nielsen, A. J. Enright, J. S. Mattick, M. E. Dinger, *Nat. Methods* **2015**, *12*, 339.
[98] T. R. Mercer, M. B. Clark, J. Crawford, M. E. Brunck, D. J. Gerhardt, R. J. Taft, L. K. Nielsen, M. E. Dinger, J. S. Mattick, *Nat. Protoc*. **2014**, *9*, 89.
[99] F. Lai, E. Blumenthal, R. Shiekhattar, *Methods Enzymol*. **2016**, *573*, 421.
[100] M. N. Cabili, M. C. Dunagin, P. D. McClanahan, A. Biaesch, O. Padovan-Merhar, A. Regev, J. L. Rinn, A. Raj, *Genome Biol*. **2015**, *16*, 20.
[101] S. Zucchelli, L. Patrucco, F. Persichetti, S. Gustincich, D. Cotella, *Comput. Struct. Biotechnol. J*. **2016**, *14*, 404.
[102] T. B. Kallehauge, S. Li, L. E. Pedersen, T. K. Ha, D. Ley, M. R. Andersen, H. F. Kildegaard, G. M. Lee, N. E. Lewis, *Sci. Rep*. **2017**, *7*, 40388.
[103] P. Han, C.-P. Chang, *RNA Biol*. **2015**, *12*, 1094.
[104] A. E. Kornienko, P. M. Guenzl, D. P. Barlow, F. M. Pauler, *BMC Biol*. **2013**, *11*, 59.
[105] A. C. Mallory, A. Shkumatava, *Biochimie* **2015**, *117*, 3.
[106] Y. He, Y. Luo, B. Liang, L. Ye, G. Lu, W. He, Y. He, Y. Luo, B. Liang, L. Ye, G. Lu, W. He, Y. He, Y. Luo, B. Liang, L. Ye, G. L. W. He, *Oncotarget* **2017**, *8*, 73282.
[107] J. Liao, Q. He, M. Li, Y. Chen, Y. Liu, J. Wang, *Gene* **2016**, *578*, 158.
[108] T. Hirose, G. Virnicchi, A. Tanigawa, T. Naganuma, R. Li, H. Kimura, T. Yokoi, S. Nakagawa, M. Bénard, A. H. Fox, G. Pierron, *Mol. Biol. Cell* **2014**, *25*, 169.
[109] J. M. Engreitz, N. Ollikainen, M. Guttman, *Nat. Rev. Mol. Cell Biol*. **2016**, *17*, 756.
[110] Z. Li, J. Shen, M. T. V. Chan, W. K. K. Wu, *Cell Prolif*. **2016**, *49*, 471.

© 2018 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Chemical Engineering

Engineering of the cellular translational machinery and non-coding RNAs to enhance CHO cell growth, recombinant product yields and quality

Davide Vito and C Mark Smales

Chinese hamster ovary cells are the main mammalian cell expression system currently used for the production of recombinant protein biopharmaceuticals. One of the key processes determining the achievable biomass of cells in the bioreactor and the yield and quality of recombinant protein from such systems is mRNA translation. Translation is the process by which ribosomes and associated cellular machinery decode an mRNA to produce a polypeptide. In recent years the roles of different classes of non-coding RNAs in controlling global and transcript specific mRNA translation has also come to light. Here we review approaches to engineer the translational machinery and non-coding RNAs, particularly long non-coding RNAs and tRNAs in CHO cells and then outline the challenges and potential of such approaches to revolutionize the yields and quality of recombinant protein from CHO and other mammalian cell expression systems.

Address
Industrial Biotechnology Centre and School of Biosciences, University of Kent, Canterbury, Kent, CT2 7NJ, UK

Corresponding author: Smales, C Mark (c.m.smales@kent.ac.uk)

Current Opinion in Chemical Engineering 2018, **22**:199–208

This review comes from a themed issue on **Biotechnology and bioprocess engineering: biomanufacturing**

Edited by **Michael Betenbaugh** and **Nigel Titchener-Hooker**

https://doi.org/10.1016/j.coche.2018.11.002

2211-3398/© 2018 Elsevier Ltd. All rights reserved.

Introduction

For the production of biopharmaceuticals, Chinese hamster ovary (CHO) cells are the most widely used mammalian cell expression system, able to produce secretory yields of monoclonal antibody in fed-batch culture in excess of 5 g/L [1]. The secretory yield from such an expression system is governed by the number of cells in the bioreactor across the culture (the integral of viable cell concentration or IVC) and the average amount of material expressed by each cell, usually referred to as the cell specific productivity (qP) and expressed as pg of protein/per cell/per day [2]. mRNA translation is a key cellular process that is involve in determining global and protein specific synthesis, and hence control of the abundance of proteins that constitute the cellular machinery, cell growth, division and the IVC of culture. Likewise, mRNA translation plays a key role in determining the qP of a given cell line and hence is a key regulatory process impacting on the yields and quality of recombinant protein from CHO cells [3].

mRNA translation is the process by which the ribosome and associated cellular machinery decodes a target mRNA to yield a polypeptide. Translation is a key step in the gene expression pathway and is the predominant process by which protein cellular abundance is controlled [4]. Over the last few decades it has been established that the control in mammalian cells of mRNA translation, and hence protein synthesis, is not only determined by the translational machinery, modulation of the activity of various translation factors by phosphorylation, and the abundance, availability and makeup of a given mRNA, but also by availability, abundance and activity of non-coding RNAs [5]. Non-coding RNAs are generally described as either long non-coding RNAs of >200 nucleotides in length or small non-coding RNAs <200 nucleotides and includes microRNAs (also referred to as miRs) and tRNAs. The discovery of the mechanism(s) by which non-coding RNAs exert an influence on gene expression has opened up new opportunities for the engineering of cells to manipulate cell processes that underpin cell growth and recombinant protein production and quality. Further, manipulation of such non-coding RNAs offers the advantage of not placing an additional translational burden on the cell that over expression of coding mRNAs does. Here we briefly review our understanding of the control of mRNA translation in CHO cells, describe approaches and outcomes to engineer the translational machinery and non-coding RNAs in CHO cells, and discuss current and future cell engineering opportunities and challenges such approaches present (Summarized in Table 1).

The translational machinery, mRNA analysis and manipulation

As mRNA translation is a key process in defining cell growth, biomass accumulation and recombinant protein yields and quality from cultured CHO cells [3,6], the translational machinery and the abundance and availability of global and recombinant mRNAs between cell lines

Table 1

Summary of non-coding RNA cell engineering studies discussed in this review grouped by type of approach, with the reported experimental effect and the appropriate reference

Method	Effect	Reference
siRNA	LDH-A activities were decreased by 75–89%, while the specific glucose consumption rates reduced by 54–87% and the specific lactate production rates reduced to 45–79% of the control cell line level.	[32]
	siRNA mediated inhibition of PDHKs and LDH-A in CHO cells expressing a therapeutic monoclonal antibody reduced lactate production, increased specific productivity and volumetric antibody production by 90%, 75% and 68%, respectively.	[33]
	Ribosome profiling identified NeoR as a highly transcribed and translated gene in an IgG-producing CHO cell line. Viable cell concnetration was increased by 35% upon siRNA knock-down of NeoR, which was accompanied by an 18% increase in product titer.	[34]
	Combined transient siRNA-mediated knockdown of the expression of the endoplasmic reticulum localized proteins CerS2 and Tbc1D20 resulted in a 50–66% increase in specific productivity of CHO-IgG cells.	[35]
miRNA	Co-expression of miR-557 and a difficult-to-express antibody resulted in a two-fold increase in product titer.	[27]
	miR-143 overexpression resulted in a 20% final increase in mAb productivity.	[28]
	Addition of a synthetic 3'UTR to destabilize DHFR expression allowed the generation of stable DG44-derived cell pools expressing a model monoclonal antibody (mAb) with low MTX concentrations.	[29]
	Inhibition of miR-124-3p and miR-19b-3p in CHO increased X-linked inhibitor of apoptosis protein levels, enhancing CHO cell growth and prolonged culture longevity while additionally boosting productivity.	[31•]
lncRNA	Engineering of CHO cells with SINEUP long non-coding RNAs resulted in a 150% increase in periostin levels in cell supernatant at 72 h post-transfection.	[50••]
	The development of a 'universal' protein expression enhancer tool based upon long non-coding RNAs gave expression enhancement in various mammalian cells of recombinant proteinsin the order of 50-1000%, with more than 200% enhancement in most cases.	[51•]
Codon optimization	Modification of human interleukin-2 (IL-2) through codons with high gene copy number and high codon usage bias significantly increased protein productivity in CHO-K1 cells.	[52]
	Codon de-optimization of a bispecific antibody sequence through the introduction of less frequently occurring codons in CHO gave a 2-fold final yield increase.	[55]
	In vivo expression of various codon context (CC) optimized IFN-γ in CHO cells exhibited at least 13-fold increase in expression compared to the wild-type IFN-γ while a maximum of 10-fold increase was observed for the individual codon usage (ICU) optimized genes.	[60]

and process conditions has been investigated using a variety of approaches. For example, the phosphorylation of the translation initiation factor eIF2α and attenuation of global protein synthesis during recombinant protein production in CHO cells is known to occur [7]. Culture temperature has been shown to impact mRNA translation and the quality of recombinant product produced [8] and the PERK-eIF2α pathway was reported to impact upon the aggregation of a recombinant TNFR-Fc fusion protein [9]. Indeed the activity and availability of translation factors has been shown to change during culture and under different culture conditions, where for example under reduced temperature translation elongation factor 2 (eEF2) becomes phosphorylated and a reprogramming of translation occurs that means transcripts with particular codon usage can escape the general global attenuation of translation under such conditions and the translation of these transcripts is actually enhanced [10].

High producing antibody cell lines have been shown to maintain translation initiation factors at levels that allow such cells to maintain enhanced recombinant protein synthesis above that of lower producing cells [3]. With regard to monoclonal antibody synthesis in CHO and other cells, investigations have shown that recombinant antibody production is limited by translational efficiency [3,11–13]. Manipulation of the cellular translational machinery is however, not straightforward. One global regulator of ribosome biogenesis and translation is mTORC1, which coordinates cellular responses to signaling pathways involved in sensing growth factors, nutrient availability, intracellular energy status and other perceived cell stresses and modulates translation and ribosome biogenesis in response [14]. In particular, mTORC1 can influence translation initiation via phosphorylation of eukaryotic initiation factor 4E binding protein (4E-BP1), which when phosphorylated at multiple sites promotes dissociation of 4E-BP1 from the initiation factor eIF4E. Increased phosphorylation of 4E-BP1 has been correlated with increased interferon-γ production [15] whilst the stoichiometry of 4E-BP1 to eIF4E is reported to relate to recombinant antibody productivity [16]. Exogenous mTOR expression has also been shown to enhance recombinant protein expression in CHO cells by improving cell viability, growth, proliferation and cell specific productivity [17].

One approach applied to investigate mRNA amounts, and hence determine gene expression profiles of high producing or fast growing recombinant cell lines is

transcriptomics. A general assumption of most such studies is that the amount of mRNA present at a given timepoint reflects the 'state' or 'need' of a cell with regard to the proteins that these mRNAs encode for. As such, transcriptomic profiling has been applied to identify mRNAs whose abundance correlates with cell growth and recombinant protein productivity and quality with a view to using the identification of such targets to engineer the cell for improved performance. Many of these initial studies were hampered by the lack of the CHO genome and appropriate arrays, however the elucidation of the genome [18,19] and advent of RNA-Seq has made it possible to undertake such studies with a higher degree of precision. Despite this, there has been little consensus across transcriptomic studies to date with regard to those mRNAs that correlate with cell growth and recombinant protein productivity [20]. Further, previous correlation analysis has shown that transcript amounts and translation efficiency are uncoupled for around 95% of investigated genes [21], providing strong evidence that global and mRNA specific translational control needs to be understood and determined to evaluate the impact of mRNAs on phenotype rather than simple mRNA abundance itself.

In order to address this issue, investigators have begun to apply ribosome footprint profiling or RiboSeq analysis to unravel the fine detail of translational control in CHO cells [22]. This powerful approach allows genome wide, but also transcript specific, detail on initiation and elongation stages of mRNA translation to be studied and identification of those mRNAs that are being translated at any given time (as opposed to just their abundance), the efficiency of mRNA translation and how this changes during a process or between cell lines to identify targets for cell line engineering [22]. Indeed, any given mRNA in the cell may be translated by one or multiple ribosomes (so call polysomes) at any one time [23]. In some cases the number of ribosomes per transcript has been used to estimate translational efficiency of a transcript assuming that more ribosomes on a transcript indicates greater translational efficiency [24], but this does not account for elongation speed that RiboSeq analysis can. RNA-Seq approaches can also potentially be used to investigate translational activity at the single cell level. The application of such approaches is certain to provide a more detailed understanding of mRNA translation and its control in recombinant protein producing CHO cells, at the population and single cell level and at a global and transcript specific level, revealing new engineering approaches by which translation can be modulated to enhance protein production.

MicroRNAs and siRNAs

mRNA translation can also be tuned by non-coding RNAs. One such class of non-coding RNA that has been applied to reprogramming translation in CHO cells is that of microRNAs (also known as miRs or miRNAs). The potential application of microRNAs to CHO cell engineering has recently been reviewed elsewhere [25]. These RNAs are transcribed as long primary transcripts but then processed to yield small (20–23 nucleotide) non-coding RNAs and were first described in *Caenorhabditis elegans*. MicroRNAs tend to act as repressors of translation of target mRNAs by interacting with the 3′untranslated regions (3′UTRs) of such mRNAs. A given microRNA can in theory target multiple mRNAs via base pairing and hence modulate multiple mRNAs and pathways without placing an additional translational burden on the cell [25].

Early microRNA studies were limited by the lack of available Chinese hamster sequence annotation of microRNA primary transcripts, and hence chimeric microRNAs that contained the mature miR sequence but flanking sequences in the primary transcript from other species were used. Subsequent studies showed that endogenous CHO microRNA flanking sequences gave rise to higher expression when over expressing microRNAs [26]. Recent engineering approaches harnessing microRNAs include studies that look to enhance the ability of CHO cells to produce so called 'difficult to express proteins'. For example, one such study showed that a CHO cell line constitutively over expressing miR-557 and a difficult to express antibody produced twice the antibody yield of cells engineered to express a negative control microRNA [27]. A further study reported that both transient and stable miR-143 over expression resulted in enhanced difficult to express protein production and targeted MAPK7 in CHO [28]. The natural repertoire of microRNAs has also been harnessed to repress expression of the DHFR selection marker during cell line construction and allow the generation of higher producing cell pools [29]. Others have shown that microRNA fingerprints or signatures can be correlated with growth rate across a number of different CHO cell lines [30].

However, although microRNA engineering appears an attractive approach by which to tune translation of multiple mRNAs and translation of specific targets, the potential large number of predicted targets of any given microRNA means that the outcome of such engineering approaches can be difficult to predict as is identifying which targets a given microRNA interacts with. Barron and colleagues have described a system termed 'miR-CATCH' that allows the investigator to identify those microRNAs that interact with a given target and thus validate these for potential cell engineering approaches [31•]. The authors had identified that the overexpression of the X-linked inhibitor of apoptosis (XIAP) enhanced CHO cell productivity, growth and culture longevity. To avoid overexpressing this gene and placing an additional translational burden on the cell, microRNA regulators of XIAP were identified using a biotin-labelled antisense DNA for XIAP resulting in the capture of interacting

microRNAs. Inhibition of two of these microRNAs resulted in increased XIAP protein expression, validating the microRNA catch approach and the utility of this for identifying cell engineering targets.

The use of small interfering RNAs (siRNAs) has also proved to be an effective strategy for CHO cell line engineering to selectively knockdown expression of target genes detrimental to cell growth or productivity. The advantage of knockdown, as opposed to knockout strategies, is that essential genes can be reduced in their expression and the impact on cell phenotype assessed when knockout proves fatal. One successful application of siRNA engineering has been the inhibition of lactate dehydrogenase-A (LDH-A) on its own [32], or in combination with pyruvate dehydrogenase kinases (PDHKs) [33] to significantly reduce lactate accumulation in cultured CHO cells without negatively impacting cell growth and enhancing cell specific productivity. A further group undertook ribosomal profiling in CHO cells and identified the resistance marker NeoR as being highly transcribed and translated, and as expression of this exogenous gene in CHO cells is not required, used siRNA knockdown to reduce its expression with a resultant improvement in production and growth of the host observed [34]. Finally, an siRNA approach was used to knockdown the expression of the endoplasmic reticulum localized proteins ceramide synthase 2 (CerS2) and Rab1 GAP Tbc domain family member 20 (Tbc1D20) in CHO IgG producing cells with a subsequent observed increase in recombinant protein specific productivity and enhanced cell growth [35].

Long non−coding RNAs (lncRNAs) and their manipulation

Recent genome wide analysis in mammalian cells estimates that 75% of the transcriptome is composed of non-coding sequences [36] and led to the identification of a heterogeneous class of transcripts known as Long non-coding RNAs (lncRNAs) [37]. LncRNAs are defined as transcripts longer than 200 nucleotides that lack a significant open reading frame (ORF) and are usually transcribed by RNA polymerase II and spliced, with or without, 3′ polyadenylation [38]. These molecules are emerging as key regulators in various biological processes both in the nucleus and the cytoplasm [39], including epigenetic regulation, transcriptional control, splicing events, and mRNA translation. While most of our current understanding into lncRNAs and the underlying mechanism(s) by which they elicit their responses has come from studies relating to disease and developmental studies, their potential as targets for cell engineering in mammalian cell factories remains largely unexplored.

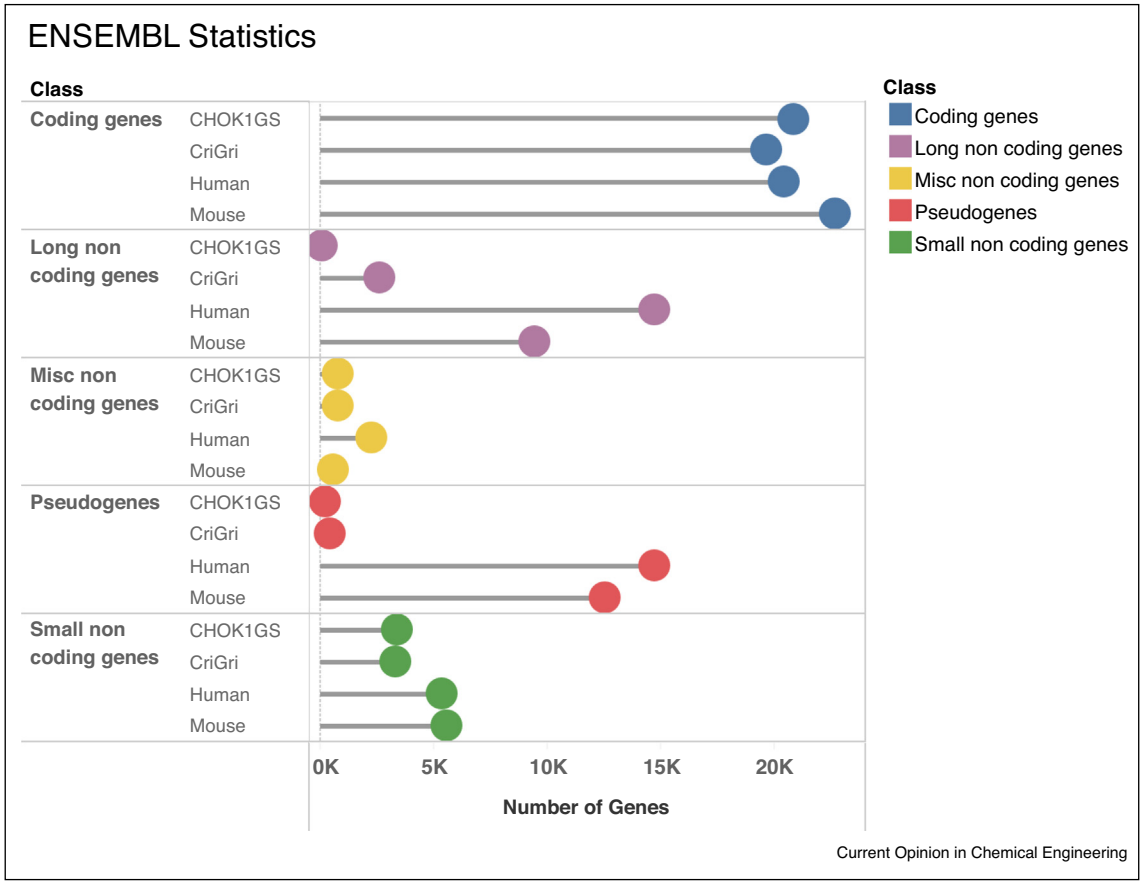
The first analysis of the non-coding transcriptome in CHO cells under batch and fed-batch conditions has recently been published, unveiling a number of differentially regulated lncRNAs depending on feed and culture time which could be targets for cell engineering [40•]. One of the main challenges in identifying lncRNAs is the low sequence conservation between species. This, coupled with incomplete genome sequences and partial annotations of coding and non-coding genes of most vertebrates including Chinese hamster, have impaired an effective lncRNAs annotation outside from model organisms (Figure 1).

A recent study compared lncRNAs among 16 vertebrates and the echinoid sea urchin finding thousands of human lincRNAs homologs with conserved genomic position sharing 5′-biased patches of sequence nested in rewired exonic architectures [41]. The FANTOM consortium applied a cap analysis of gene expression (CAGE) to data obtaining more than 27,919 human lncRNA genes with high-confidence 5′ ends and expression profiles across 1829 samples from the major human primary cell types and tissues [42]. Through the incorporation of conservation and expression data, the consortium was able to identify 19,175 potentially functional lncRNAs in the human genome. Because of the tissue-specificity of lncRNAs, comparing the expression among several cell types has led to a more robust identification of functional targets. By modelling their effects on the activity of transcription factors, RNA-binding proteins, and microRNAs in 5185 TCGA tumors and 1019 ENCODE assays, it was possible to identify potential lncRNAs involved in dysregulated cancer pathways. This approach indicated OIP5-AS1, TUG1, NEAT1, MEG3, and TSIX, as synergic lncRNAs leading to dysregulated cancer pathways in multiple tumor contexts [43]. A similar effort using nascent RNA capture sequencing identified 1145 temporally expressed S-phase-enriched lncRNAs across TCGA data sets in several cancer models showing effects on pathways including FGF/FGFR and its downstream PI3K/AKT and MAPK pathways [44].

The NEAT1 lncRNA is a central component of paraspeckles, nuclear bodies that regulate multiple aspects of gene expression, promoting their formation through ATR signaling in response to replication stress and p53 activation [45]. The RNA-binding NONO–PSF heterodimer binds a large number of expressed pri-miRNAs in the paraspeckles to promote processing by the Drosha–DGCR8 Microprocessor. NEAT1 thus regulates efficient processing of potentially an entire class of small non-coding RNAs in the nucleus by interaction with the NONO–PSF heterodimer as well as other ribosome binding proteins (RBPs) [46].

The relationship between lncRNAs and the translational machinery was further elucidated with the discovery of a long nucleolus-specific lncRNA (LoNA) [47•]. LoNA is expressed at high levels at resting state suppressing rRNAs transcription in the nucleoli through the

Figure 1



The figure shows the number of genes for each annotated class in the ENSEMBL database for Chinese hamster (CriGri_1.0, GCA_000223135.1 and CHOK1GS_HDv1, GCA_900186095.1), mouse (GRCm38.p6, GCA_000001635.8) and human (GRCh38.p12, GCA_000001405.27). While the number of coding genes is comparable between the four organisms, human contains 14,720 and mouse 9443 annotated long non-coding genes as compared to 2563 for CriGri.

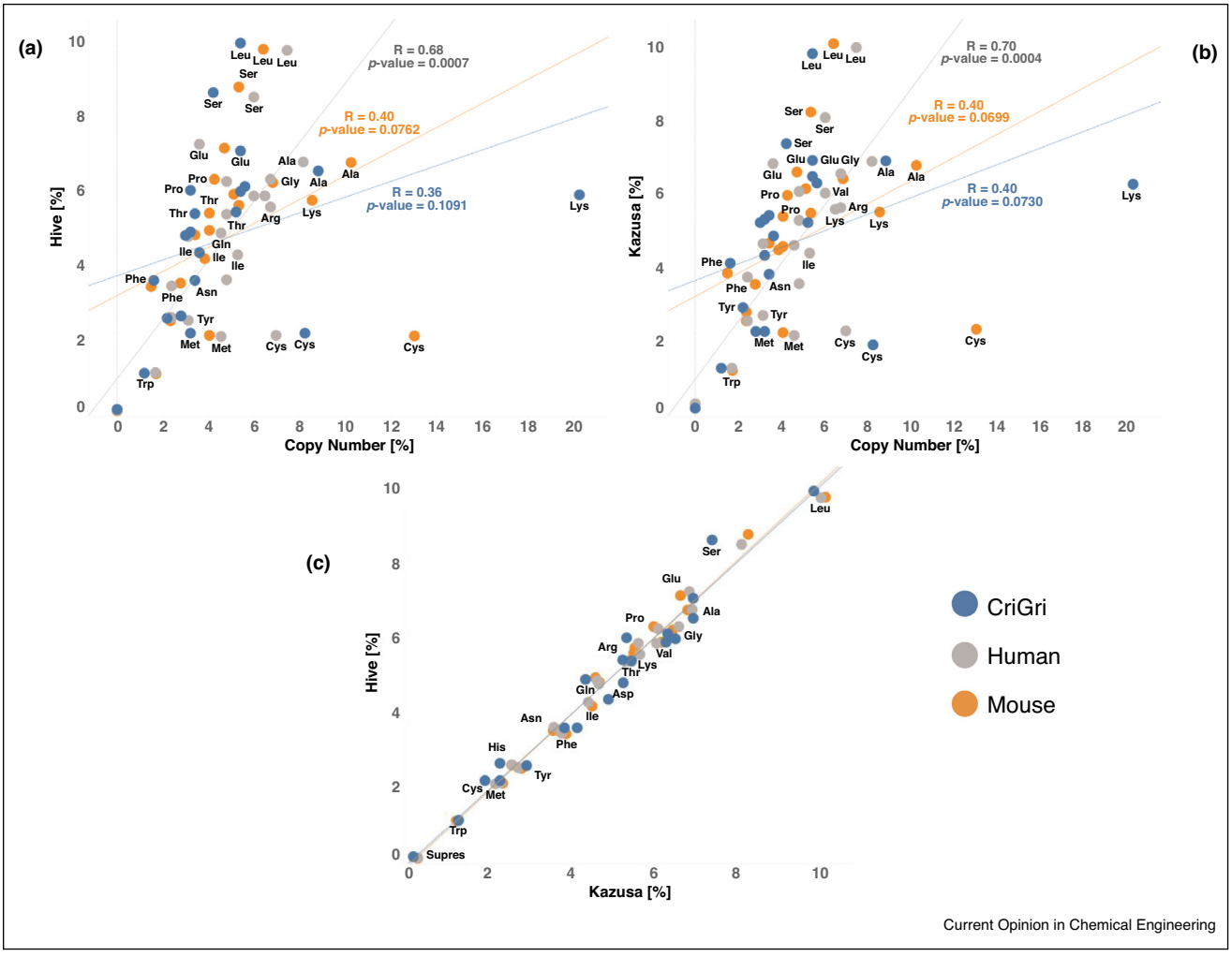
combined effects of its 5′ portion, which binds and sequesters nucleolin, and its snoRNA like 3′ end, which recruits and diminishes fibrillarin activity to reduce rRNA methylation. When the cell needs to sustain an elevated translational load, LoNA expression decreases leading to elevated rRNA and ribosome levels, an increased proportion of polysomes, mRNA polysome loading, and eventually protein synthesis.

The first successful engineering of lncRNAs for enhanced recombinant protein production involved manipulation of SINEUPs, natural and synthetic antisense lncRNAs that can activate translation in a gene-specific manner using an inverted SINEB2 sequence [48]. A Binding Domain (BD) located towards the 5′ region of the SINEUP overlaps a target mRNA of choice conferring specificity, while an inverted SINEB2 element defined as the Effector Domain (ED) provides the translation activation function [49]. Synthetic SINEUPs have been used to increase translation and secretion of recombinant proteins in a range of mammalian cell lines, including CHO [50••] and HEK293 [51•]. As further studies define those lncRNAs present in CHO cells and how these influence cell growth, fate and recombinant protein production, engineering of these non-coding RNAs is sure to offer potential to further tune and enhance mRNA specific and global mRNA translational efficiency.

tRNAs and translation

The use of specific codons with high gene copy number and high codon bias coupled with the modulation of intracellular tRNA concentration has been shown to improve protein production in CHO cells [52]. However, despite translational efficiency often being considered the mere result of codon optimization based on the correlation between codon bias and tRNA gene copy numbers (Figure 2), recent evidence suggests a considerably more intricate picture where ribosome collisions, co-translational folding, mRNA stability, composition, charge status and post-transcriptional modifications of

Figure 2



The gene copy numbers for each codon were retrieved from the genomic tRNA database (GtRNAdb) while the codon usage was sourced from either the Codon Usage Database (kazusa.or.jp/codon/), indicated as 'Kazusa', or the Codon Usage Table Database (hive.biochemistry.gwu.edu/review/codon), indicated as 'Hive'. For each aminoacid, the summary of every single codon coding for the same aminoacid is shown as a % of the total. Cricetulus griseus (CriGri) is shown in blue, Human in grey, and Mouse in orange. The figure shows in panel **(a)** the correlation between the Hive and gene copy number, panel **(b)** shows the correlation between the Kazusa and gene copy number and panel **(c)** shows the correlation between Kazusa and Hive. The two databases perform similarly, giving a significantly higher correlation (R = 0.68–0.70) with a lower p-value (0.0007 and 0.0004) for Human, while Mouse and CriGri have lower correlations and higher p-values.

the tRNA pool all contribute to finely tune protein production in response to the environment [53]. Controlling the translational capacity of an expression system through the use of alternative codon combinations modulates ribosome decoding speed, impacting protein quality as well as final yield [54]. The use of suboptimal codons has been reported to slow translation at key structural motifs in order to facilitate correct co-translational polypeptide folding and signal recognition particle (SRP) recognition, which assists in protein translocation across membranes [55]. Thus, although recombinant genes are often 'codon optimized', we do not currently have all the information required around codon usage, context, tRNA abundance, modifications and charging to fully harness codon usage in recombinant sequences or to engineer tRNA abundance.

Codon bias has been referred to as a secondary genetic code that impacts on the fidelity of translation, efficiency of translation, polypeptide/protein folding and mRNA stability/half-life [56]. The cell utilizes such codon effects to tailor the proteome and allow reprogramming, such as under cold stress whereby reprogramming and synthesis of specific proteins is enhanced through codon bias [10]. Codon bias or optimization is also linked to tuning mRNA stability and stable mRNAs are found to be enriched in

codons that are considered optimal whilst also impacting on ribosome translocation [57]. Specific combinations of adjacent codons in yeast and mammalian cells can have an effect on translation efficiency resulting in reduced expression, proving how the focus must be on global translation efficiency and codon context as opposed to single codons optimality [58•,59,60]. On top of this, mRNA secondary structure combined with tRNA abundance modulate translational elongation speed among different regions of the same transcript to avoid excessively slow or fast ribosome movement [61]. As such, there remains enormous potential to enhance recombinant protein yields from further manipulation of codon usage.

In order to further enhance recombinant protein yields by manipulation of codon usage it is necessary to further understand the abundance and modifications of tRNAs and the role these play in their activity. Determination of tRNA copy numbers can now be undertaken using RNA-Seq approaches. tRNA secondary structure and nucleotide modifications, mainly methylations, impair the efficiency of standard sequencing. Dedicated protocols based on an initial de-methylation step were recently developed to overcome this limitation, allowing for direct measurement of each tRNA abundance and detailed mapping of modifications [62,63]. While some methods focus exclusively on mature tRNAs [64], partial alkaline RNA hydrolysis complemented with tRNA precursors enrichment identified tRNA leaders, trailers, and introns and showed that around half of all predicted tRNA genes are transcribed in human cells [65]. While tRNA abundance is a major modulator of translational elongation, the aminoacylation state has to also be considered. The addition of chemical steps that specifically remove the 3′A residue in uncharged tRNA coupled with the aforementioned demethylation RNA-Seq protocols showed most cytosolic tRNAs in HEK293T cells are charged at >80% levels, whereas tRNASer and tRNAThr are charged at lower levels [66].

An additional layer of regulation during elongation is chemical modification of nucleotides among tRNAs [67]. One of the key enzymes to regulate the methylation state of tRNAs is the demethylase ALKBH1, which acts dynamically in response to specific conditions such as variations in glucose availability to impact translation at both the initiation and the elongation phases [68]. These modifications can have different effects depending on the target tRNA and the position in the transcript, as it was shown ALKBH1 is required for the formation of essential methylations at position 34 of anticodon in cytoplasmic tRNALeu and mitochondrial tRNAMet [69]. Advances in high-throughput sequencing and data analysis have also allowed the identification of new classes of small non-coding RNAs derived from tRNAs: stress-induced tRNA halves (tiRs) and tRNA-related fragments (tRFs). These RNAs act on cell proliferation, priming of viral reverse transcriptase, regulation of gene expression, RNA processing, modulation of the DNA damage response, tumor suppression, and stress response [70]. The application of such approaches to study tRNAs in CHO cells will further elucidate the mechanism(s) by which tRNAs and their modifications modulate translation.

Future directions and challenges

Our understanding and ability to manipulate the translational machinery and harness non-coding RNAs to enhance global and recombinant protein synthesis in CHO cells has advanced rapidly in the last decade. Further, the advent of the Chinese hamster and CHO cell line genomes has helped in the identification of non-coding RNAs such that these can be studied and manipulated. The ability of non-coding RNAs, in particular microRNAs, siRNAs, lncRNAs and tRNAs to tune both global and transcript specific translation, and hence protein synthesis, offers enormous opportunities to use these to enhance cell growth and proliferation, extend culture lifetimes, and increase recombinant protein yields and quality. However, our ability to harness these non-coding RNAs by engineering of CHO cells is currently limited by our knowledge of the mechanisms and targets by which many of these non-coding RNAs elicit their responses. The manipulation of microRNAs that can, in theory, tune multiple target transcripts appears an appealing approach, however in our view this approach alone is unlikely to deliver new commercially viable host cells with dramatically enhanced phenotypes due to the fact these are 'tuning' molecules and tend to be negative regulators and off target approaches can be difficult to control. Where these might be more applicable is for the tuning of transcript targets with a specific role, such as enzymes involved in glycosylation or to harness modulation of the cells own endogenous microRNA pool as inducible controllers of exogenous gene circuits. The potential of lncRNA engineering is very much in its infancy and would appear to offer the potential to act as negative and positive regulators of gene expression. The limitation here is that many of these are, as the name suggests, long RNAs and thus the manipulation is more challenging and we do not yet understand what, if any, role many of these play in the cell. The control of gene expression via the elongation step of mRNA translation and tRNA availability, charging and modification, linked with improved predictive models for how such changes in abundance or modification change elongation rates of target mRNAs is likely to offer advances that can be directly applied industrially to engineering of the target recombinant gene (s) and of pathways in the cell to deliver new engineered host cell lines with improved growth, productivity and post-translational modification abilities. However, the major challenge will be to unravel the mechanisms by which the control on gene expression that these different non-coding RNAs provide are coordinated together, in

order to reprogram the translational efficiency of current CHO cell chassis, under appropriate bioprocessing conditions (including continuous processes) to generate new chassis with enhance bioprocessing properties.

Conflict of interest statement

Nothing declared.

Acknowledgement

The authors acknowledge the European Commission for funding this work (EC - Horizon 2020 MSCA ITN 2014 - 642663).

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Povey JF, O'Malley CJ, Root T, Martin EB, Montague GA, Feary M, Trim C, Lang DA, Alldread R, Racher AJ, Smales CM: **Rapid high-throughput characterisation, classification and selection of recombinant mammalian cell line phenotypes using intact cell MALDI-ToF mass spectrometry fingerprinting and PLS-DA modelling**. *J Biotechnol* 2014, **184**:84-93 http://dx.doi.org/10.1016/j.jbiotec.2014.04.028.

2. Kunert R, Reinhart D: **Advances in recombinant antibody manufacturing**. *Appl Microbiol Biotechnol* 2016, **100**:3451-3461 http://dx.doi.org/10.1007/s00253-016-7388-9.

3. Mead EJ, Masterton RJ, Feary M, Obrezanova O, Zhang L, Young R, Smales CM: **Biological insights into the expression of translation initiation factors from recombinant CHOK1SV cell lines and their relationship to enhanced productivity**. *Biochem J* 2015, **472**:261-273 http://dx.doi.org/10.1042/BJ20150928.

4. Schwanhäusser B, Busse D, Li N, Dittmar G, Schuchhardt J, Wolf J, Chen W, Selbach M: **Global quantification of mammalian gene expression control**. *Nature* 2011, **473**:337-342 http://dx.doi.org/10.1038/nature10098.

5. Janakiraman H, House RP, Gangaraju VK, Diehl JA, Howe PH, Palanisamy V: **The long (lncRNA) and short (miRNA) of it: TGFβ-mediated control of RNA-binding proteins and noncoding RNAs**. *Mol Cancer Res* 2018, **16**:567-579 http://dx.doi.org/10.1158/1541-7786.MCR-17-0547.

6. McLeod J, O'Callaghan PM, Pybus LP, Wilkinson SJ, Root T, Racher AJ, James DC: **An empirical modeling platform to evaluate the relative control discrete CHO cell synthetic processes exert over recombinant monoclonal antibody production process titer**. *Biotechnol Bioeng* 2011, **108**:2193-2204 http://dx.doi.org/10.1002/bit.23146.

7. Underhill MF, Birch JR, Smales CM, Naylor LH: **eIF2α phosphorylation, stress perception, and the shutdown of global protein synthesis in cultured CHO cells**. *Biotechnol Bioeng* 2005, **89**:805-814 http://dx.doi.org/10.1002/bit.20403.

8. Masterton RJ, Roobol A, Al-Fageeh MB, Carden MJ, Smales CM: **Post-translational events of a model reporter protein proceed with higher fidelity and accuracy upon mild hypothermic culturing of Chinese hamster ovary cells**. *Biotechnol Bioeng* 2010, **105**:215-220 http://dx.doi.org/10.1002/bit.22533.

9. Wang K, Zhang T, Chen J, Liu C, Tang J, Xie Q: **The effect of culture temperature on the aggregation of recombinant TNFR-Fc is regulated by the PERK-eIF2a pathway in CHO cells**. *Protein Pept Lett* 2018, **25**:570-579 http://dx.doi.org/10.2174/0929866525666180530121317.

10. Bastide A, Peretti D, Knight JRP, Grosso S, Spriggs RV, Pichon X, Sbarrato T, Roobol A, Roobol J, Vito D *et al.*: **RTN3 is a novel cold-induced protein and mediates neuroprotective effects of RBM3**. *Curr Biol* 2017, **27**:638-650 http://dx.doi.org/10.1016/j.cub.2017.01.047.

11. Mead EJ, Chiverton LM, Spurgeon SK, Martin EB, Montague GA, Smales CM, von der Haar T: **Experimental and in silico modelling analyses of the gene expression pathway for recombinant antibody and by-product production in NS0 cell lines**. *PLoS One* 2012, **7** http://dx.doi.org/10.1371/journal.pone.0047422 e47422.

12. Roobol A, Roobol J, Bastide A, Knight JRP, Willis AE, Smales CM: **p58IPK is an inhibitor of the eIF2α kinase GCN2 and its localization and expression underpin protein synthesis and ER processing capacity**. *Biochem J* 2015, **465**:213-225 http://dx.doi.org/10.1042/BJ20140852.

13. O'Callaghan PM, McLeod J, Pybus LP, Lovelady CS, Wilkinson SJ, Racher AJ, Porter A, James DC: **Cell line-specific control of recombinant monoclonal antibody production by CHO cells**. *Biotechnol Bioeng* 2010, **106**:938-951 http://dx.doi.org/10.1002/bit.22769.

14. Foster KG, Fingar DC: **Mammalian target of rapamycin (mTOR): conducting the cellular signaling symphony**. *J Biol Chem* 2010, **285**:14071-14077 http://dx.doi.org/10.1074/jbc.R109.094003.

15. Chong WPK, Sim LC, Wong KTK, Yap MGS: **Enhanced IFNgamma production in adenosine-treated CHO cells: a mechanistic study**. *Biotechnol Prog* 2009, **25**:866-873 http://dx.doi.org/10.1002/btpr.118.

16. Jossé L, Xie J, Proud CG, Smales CM: **mTORC1 signalling and eIF4E/4E-BP1 translation initiation factor stoichiometry influence recombinant protein productivity from GS-CHOK1 cells**. *Biochem J* 2016, **473**:4651-4664 http://dx.doi.org/10.1042/BCJ20160845.

17. Dreesen IAJ, Fussenegger M: **Ectopic expression of human mTOR increases viability, robustness, cell size, proliferation, and antibody production of chinese hamster ovary cells**. *Biotechnol Bioeng* 2011, **108**:853-866 http://dx.doi.org/10.1002/bit.22990.

18. Xu X, Nagarajan H, Lewis NE, Pan S, Cai Z, Liu X, Chen W, Xie M, Wang W, Hammond S *et al.*: **The genomic sequence of the Chinese hamster ovary (CHO)-K1 cell line**. *Nat Biotechnol* 2011 http://dx.doi.org/10.1038/nbt.1932.

19. Lewis NE, Liu X, Li Y, Nagarajan H, Yerganian G, O'Brien E, Bordbar A, Roth AM, Rosenbloom J, Bian C *et al.*: **Genomic landscapes of Chinese hamster ovary cell lines as revealed by the Cricetulus griseus draft genome**. *Nat Biotechnol* 2013 http://dx.doi.org/10.1038/nbt.2624.

20. Tamoaitis L, Smales CM: **Meta-analysis of publicly available Chinese Hamster Ovary (CHO) cell transcriptomic datasets for identifying engineering targets to enhance recombinant protein yields**. *Biotechnol J* 2018 http://dx.doi.org/10.1002/biot.201800066. e1800066.

21. Courtes FC, Lin J, Lim HL, Ng SW, Wong NSC, Koh G, Vardy L, Yap MGS, Loo B, Lee D-Y: **Translatome analysis of CHO cells to identify key growth genes**. *J Biotechnol* 2013, **167**:215-224 http://dx.doi.org/10.1016/j.jbiotec.2013.07.010.

22. Tzani I, Monger C, Kelly P, Barron N, Kelly RM, Clarke C: **Understanding biopharmaceutical production at single nucleotide resolution using ribosome footprint profiling**. *Curr Opin Biotechnol* 2018, **53**:182-190 http://dx.doi.org/10.1016/j.copbio.2018.01.030.

23. Godfrey CL, Mead EJ, Daramola O, Dunn S, Hatton D, Field R, Pettman G, Smales CM: **Polysome profiling of mAb producing CHO cell lines links translational control of cell proliferation and recombinant mRNA loading onto ribosomes with global and recombinant protein synthesis**. *Biotechnol J* 2017, **12** http://dx.doi.org/10.1002/biot.201700177 1700177.

24. Liu B, Qian S-B: **Characterizing inactive ribosomes in translational profiling**. *Transl (Austin, Tex)* 2016, **4** http://dx.doi.org/10.1080/21690731.2015.1138018 e1138018.

25. Valdés-Bango Curell R, Barron N: **Exploring the potential application of short non-coding RNA-based genetic circuits in Chinese hamster ovary cells**. *Biotechnol J* 2018 http://dx.doi.org/10.1002/biot.201700220. 1700220.

26. Klanert G, Jadhav V, Chanoumidou K, Grillari J, Borth N, Hackl M: **Endogenous microRNA clusters outperform chimeric**

sequence clusters in Chinese hamster ovary cells. *Biotechnol J* 2014, **9**:538-544 http://dx.doi.org/10.1002/biot.201300216.

27. Fischer S, Marquart KF, Pieper LA, Fieder J, Gamer M, Gorr I, Schulz P, Bradl H: **miRNA engineering of CHO cells facilitates production of difficult-to-express proteins and increases success in cell line development**. *Biotechnol Bioeng* 2017, **114**:1495-1510 http://dx.doi.org/10.1002/bit.26280.

28. Schoellhorn M, Fischer S, Wagner A, Handrick R, Otte K: **miR-143 targets MAPK7 in CHO cells and induces a hyperproductive phenotype to enhance production of difficult-to-express proteins**. *Biotechnol Prog* 2017, **33**:1046-1058 http://dx.doi.org/10.1002/btpr.2475.

29. Jossé L, Zhang L, Smales CM: **Application of microRNA targeted 3'UTRs to repress DHFR selection marker expression for development of recombinant antibody expressing CHO cell pools**. *Biotechnol J* 2018 http://dx.doi.org/10.1002/biot.201800129.

30. Klanert G, Jadhav V, Shanmukam V, Diendorfer A, Karbiener M, Scheideler M, Bort JH, Grillari J, Hackl M, Borth N: **A signature of 12 microRNAs is robustly associated with growth rate in a variety of CHO cell lines**. *J Biotechnol* 2016, **235**:150-161 http://dx.doi.org/10.1016/j.jbiotec.2016.03.022.

31. • Griffith A, Kelly PS, Vencken S, Lao NT, Greene CM, Clynes M, Barron N: **miR-CATCH identifies biologically active miRNA regulators of the pro-survival gene XIAP, in Chinese hamster ovary cells**. *Biotechnol J* 2018, **13** http://dx.doi.org/10.1002/biot.201700299 1700299.

The authors developed a system to identify and validate microRNAs that target a specific mRNA and used this to identify microRNAs that negatively regulate the pro-survival gene XIAP. They subseqeuntly show that reducing or inhibiting two of the microRNAs that target this transcript resulted in enhanced XIAP expression and prolonged culture duration.

32. Kim SH, Lee GM: **Down-regulation of lactate dehydrogenase-A by siRNAs for reduced lactic acid formation of Chinese hamster ovary cells producing thrombopoietin**. *Appl Microbiol Biotechnol* 2007, **74**:152-159 http://dx.doi.org/10.1007/s00253-006-0654-5.

33. Zhou M, Crawford Y, Ng D, Tung J, Pynn AFJ, Meier A, Yuk IH, Vijayasankaran N, Leach K, Joly J, Snedecor B, Shen A: **Decreasing lactate level and increasing antibody production in Chinese Hamster ovary cells (CHO) by reducing the expression of lactate dehydrogenase and pyruvate dehydrogenase kinases**. *J Biotechnol* 2011, **153**:27-34 http://dx.doi.org/10.1016/j.jbiotec.2011.03.003.

34. Kallehauge TB, Li S, Pedersen LE, Ha TK, Ley D, Andersen MR, Kildegaard HF, Lee GM, Lewis NE: **Ribosome profiling-guided depletion of an mRNA increases cell growth rate and protein secretion**. *Sci Rep* 2017, **7**:40388 http://dx.doi.org/10.1038/srep40388.

35. Pieper LA, Strotbek M, Wenger T, Gamer M, Olayioye MA, Hausser A: **Secretory pathway optimization of CHO producer cells by co-engineering of the mitosRNA-1978 target genes CerS2 and Tbc1D20**. *Metab Eng* 2017, **40**:69-79 http://dx.doi.org/10.1016/j.ymben.2017.01.003.

36. Djebali S, Davis CA, Merkel A, Dobin A, Lassmann T, Mortazavi A, Tanzer A, Lagarde J, Lin W, Schlesinger F *et al.*: **Landscape of transcription in human cells**. *Nature* 2012, **489**:101-108 http://dx.doi.org/10.1038/nature11233.

37. Kapranov P, Cheng J, Dike S, Nix DA, Duttagupta R, Willingham AT, Stadler PF, Hertel J, Hackermuller J, Hofacker IL *et al.*: **RNA maps reveal new RNA classes and a possible function for pervasive transcription**. *Science* 2007, **316**:1484-1488 http://dx.doi.org/10.1126/science.1138341.

38. Wilusz JE: **Long noncoding RNAs: re-writing dogmas of RNA processing and stability**. *Biochim Biophys Acta* 2016, **1859**:128-138 http://dx.doi.org/10.1016/j.bbagrm.2015.06.003.

39. Geisler S, Coller J: **RNA in unexpected places: long non-coding RNA functions in diverse cellular contexts**. *Nat Rev Mol Cell Biol* 2013, **14**:699-712 http://dx.doi.org/10.1038/nrm3679.

40. • Vito D, Smales CM: **The long non-coding RNA transcriptome landscape in CHO cells under batch and fed-batch conditions**. *Biotechnol J* 2018 http://dx.doi.org/10.1002/biot.201800122. e1800122.

This is the first report of the global lncRNA transcritpome landscape in CHO cells, using an array approach to identify the presence of lncRNAs, qPCR to confirm the array results on selected transcripts and demonstrates how these change during batch and fed-batch culture.

41. Hezroni H, Koppstein D, Bartel DP, Ulitsky I: **Principles of long noncoding RNA evolution derived from direct comparison of transcriptomes in 17 species**. *Cell Rep* 2015, **11**:1110-1122 http://dx.doi.org/10.1016/j.celrep.2015.04.023 Correspondence, M.G. Schwartz, A. Avrutin, I. Ulitsky.

42. Hon CC, Ramilowski JA, Harshbarger J, Bertin N, Rackham OJL, Gough J, Denisenko E, Schmeier S, Poulsen TM, Severin J *et al.*: **An atlas of human long non-coding RNAs with accurate 5′ ends**. *Nature* 2017, **543**:199-204 http://dx.doi.org/10.1038/nature21374.

43. Chiu H-S, Somvanshi S, Patel E, Sood AK, Gunaratne PH, Correspondence PS, Chen T-W, Singh VP, Zorman B, Patil SL *et al.*: **Pan-Cancer analysis of lncRNA regulation supports their targeting of cancer genes in each tumor context**. *Cell Rep* 2018, **23**:297-312 http://dx.doi.org/10.1016/j.celrep.2018.03.064.

44. Ali MM, Akhade VS, Kosalai ST, Subhash S, Statello L, Meryet-Figuiere M, Abrahamsson J, Mondal T, Kanduri C: **PAN-cancer analysis of S-phase enriched lncRNAs identifies oncogenic drivers and biomarkers**. *Nat Commun* 2018, **9** http://dx.doi.org/10.1038/s41467-018-03265-1.

45. Adriaens C, Standaert L, Barra J, Latil M, Verfaillie A, Kalev P, Boeckx B, Wijnhoven PWG, Radaelli E, Vermi W *et al.*: **p53 induces formation of NEAT1 lncRNA-containing paraspeckles that modulate replication stress response and chemosensitivity**. *Nat Med* 2016, **22**:861-868 http://dx.doi.org/10.1038/nm.4135.

46. Jiang L, Shao C, Wu QJ, Chen G, Zhou J, Yang B, Li H, Gou LT, Zhang Y, Wang Y *et al.*: **NEAT1 scaffolds RNA-binding proteins and the microprocessor to globally enhance pri-miRNA processing**. *Nat Struct Mol Biol* 2017, **24**:816-824 http://dx.doi.org/10.1038/nsmb.3455.

47. • Li D, Zhang J, Wang M, Li X, Gong H, Tang H, Chen L, Wan L, Liu Q: **Activity dependent LoNA regulates translation by coordinating rRNA transcription and methylation**. *Nat Commun* 2018, **9** http://dx.doi.org/10.1038/s41467-018-04072-4.

The authors demonstrate how a specific long non-coding RNA plays a pivitol role in regulating translation by coordinating rRNA transcription and methylation making this a target for cell engineering.

48. Carrieri C, Cimatti L, Biagioli M, Beugnet A, Zucchelli S, Fedele S, Pesce E, Ferrer I, Collavin L, Santoro C: **Long non-coding antisense RNA controls Uchl1 translation through an embedded SINEB2 repeat**. *Nature* 2012 http://dx.doi.org/10.1038/nature11508.

49. Podbevek P, Fasolo F, Bon C, Cimatti L, Reißer S, Carninci P, Bussi G, Zucchelli S, Plavec J, Gustincich S: **Structural determinants of the SINE B2 element embedded in the long non-coding RNA activator of translation AS Uchl1**. *Sci Rep* 2018, **8**:1-13 http://dx.doi.org/10.1038/s41598-017-14908-6.

50. •• Patrucco L, Chiesa A, Soluri MF, Fasolo F, Takahashi H, Carninci P, Zucchelli S, Santoro C, Gustincich S, Sblattero D, Cotella D: **Engineering mammalian cell factories with SINEUP noncoding RNAs to improve translation of secreted proteins**. *Gene* 2015, **569**:287-293 http://dx.doi.org/10.1016/j.gene.2015.05.070.

The authors report the first engineering of CHO cells with SINEUP long non-coding RNAs. They subsequently show that this resulted in enhanced translation of recombinant proteins and increased secretory yields, demonstrating the potential for cell engineering using this class of non-coding RNA.

51. • Yao Y, Jin S, Long H, Yu Y, Zhang Z, Cheng G, Xu C, Ding Y, Guan Q, Li N *et al.*: **RNAe: An effective method for targeted protein translation enhancement by artificial non-coding RNA with SINEB2 repeat**. *Nucleic Acids Res* 2015, **43** http://dx.doi.org/10.1093/nar/gkv125.

The authors develop a 'universal' protein expression enhancer tool based upon long non-coding RNA elements. The authors identifed a minimal RNAe element which consists of a pairing sequence for specificity, a short non-pairing interspersed nuclear element that enhnaces ribosome

recruitment and a polyA tail. The authors claim expression enhancement in various mammalian cells of recombinant proteins in the order of 50–1000%.

52. Ou K-C, Wang C-Y, Liu K-T, Chen Y-L, Chen Y-C, Lai M-D, Yen M-C: **Optimization protein productivity of human interleukin-2 through codon usage, gene copy number and intracellular tRNA concentration in CHO cells**. *Biochem Biophys Res Commun* 2014, **454**:347-352 http://dx.doi.org/10.1016/j.bbrc.2014.10.097.

53. Hanson G, Coller J: **Translation and protein quality control: codon optimality, bias and usage in translation and mRNA decay**. *Nat Rev Mol Cell Biol* 2018, **19**:20-30 http://dx.doi.org/10.1038/nrm.2017.91.

54. Zhao F, Yu C-H, Liu Y: **Codon usage regulates protein structure and function by affecting translation elongation speed in Drosophila cells**. *Nucleic Acids Res* 2017, **45**:8484-8492 http://dx.doi.org/10.1093/nar/gkx501.

55. Magistrelli G, Poitevin Y, Schlosser F, Pontini G, Malinge P, Josserand S, Corbier M, Fischer N: **Optimizing assembly and production of native bispecific antibodies by codon de-optimization**. *MAbs* 2017, **9**:231-239 http://dx.doi.org/10.1080/19420862.2016.1267088.

56. Hanson G, Coller J: **Codon optimality, bias and usage in translation and mRNA decay**. *Nat Rev Mol Cell Biol* 2017, **19**:20-30 http://dx.doi.org/10.1038/nrm.2017.91.

57. Presnyak V, Alhusaini N, Chen Y-H, Martin S, Morris N, Kline N, Olson S, Weinberg D, Baker KE, Graveley BR, Coller J: **Codon optimality is a major determinant of mRNA stability**. *Cell* 2015, **160**:1111-1124 http://dx.doi.org/10.1016/j.cell.2015.02.029.

58. • Gamble CE, Brule CE, Dean KM, Fields S, Grayhack Correspondence EJ, Grayhack EJ: **Adjacent codons act in concert to modulate translation efficiency in yeast**. *Cell* 2016, **166**:679-690 http://dx.doi.org/10.1016/j.cell.2016.05.070.
The authors demonstrate in yeast how codon context is important in modulating translational efficiency and the imporatnce of considering codon usage in the context of the local environment and not as single codons alone.

59. Ang KS, Kyriakopoulos S, Li W, Lee D-Y: **Multi-omics data driven analysis establishes reference codon biases for synthetic gene design in microbial and mammalian cells**. *Methods* 2016, **102**:26-35 http://dx.doi.org/10.1016/j.ymeth.2016.01.016.

60. Chung BKS, Yusufi FNK, Mariati YYang, Lee DY: **Enhanced expression of codon optimized interferon gamma in CHO cells**. *J Biotechnol* 2013, **167**:326-333 http://dx.doi.org/10.1016/j.jbiotec.2013.07.011.

61. Gorochowski TE, Ignatova Z, Bovenberg RAL, Roubos JA: **Trade-offs between tRNA abundance and mRNA secondary structure support smoothing of translation elongation rate**. *Nucleic Acids Res* 2015, **43**:3022-3032 http://dx.doi.org/10.1093/nar/gkv199.

62. Cozen AE, Quartley E, Holmes AD, Hrabeta-Robinson E, Phizicky EM, Lowe TM: **ARM-seq: AlkB-facilitated RNA methylation sequencing reveals a complex landscape of modified tRNA fragments**. *Nat Methods* 2015, **12**:879-884 http://dx.doi.org/10.1038/nmeth.3508.

63. Zheng G, Qin Y, Clark WC, Dai Q, Yi C, He C, Lambowitz AM, Pan T: **Efficient and quantitative high-throughput tRNA sequencing**. *Nat Methods* 2015, **12**:835-837 http://dx.doi.org/10.1038/nmeth.3478.

64. Shigematsu M, Honda S, Loher P, Telonis AG, Rigoutsos I, Kirino Y: **YAMAT-seq: an efficient method for high-throughput sequencing of mature transfer RNAs**. *Nucleic Acids Res* 2017, **45**:e70 http://dx.doi.org/10.1093/nar/gkx005.

65. Gogakos T, Brown M, Garzia A, Meyer C, Hafner M, Tuschl T: **Characterizing expression and processing of precursor and mature human tRNAs by hydro-tRNAseq and PAR-CLIP**. *Cell Rep* 2017, **20**:1463-1475 http://dx.doi.org/10.1016/j.celrep.2017.07.029.

66. Evans ME, Clark WC, Zheng G, Pan T: **Determination of tRNA aminoacylation levels by high-throughput sequencing**. *Nucleic Acids Res* 2017, **45** http://dx.doi.org/10.1093/nar/gkx514.

67. Pan T: **Modifications and functional genomics of human transfer RNA**. *Cell Res* 2018, **28**:395-404 http://dx.doi.org/10.1038/s41422-018-0013-y.

68. Liu F, Clark W, Luo G, Klungland A, Wang X, Fu Y, Wei J, Wang X, Hao Z, Dai Q *et al.*: **ALKBH1-mediated tRNA demethylation regulates translation**. *Cell* 2016, **167**:816-828 http://dx.doi.org/10.1016/j.cell.2016.09.038.

69. Kawarada L, Suzuki T, Ohira T, Hirata S, Miyauchi K, Suzuki T: **ALKBH1 is an RNA dioxygenase responsible for cytoplasmic and mitochondrial tRNA modifications**. *Nucleic Acids Res* 2017, **45**:7401-7415 http://dx.doi.org/10.1093/nar/gkx354.

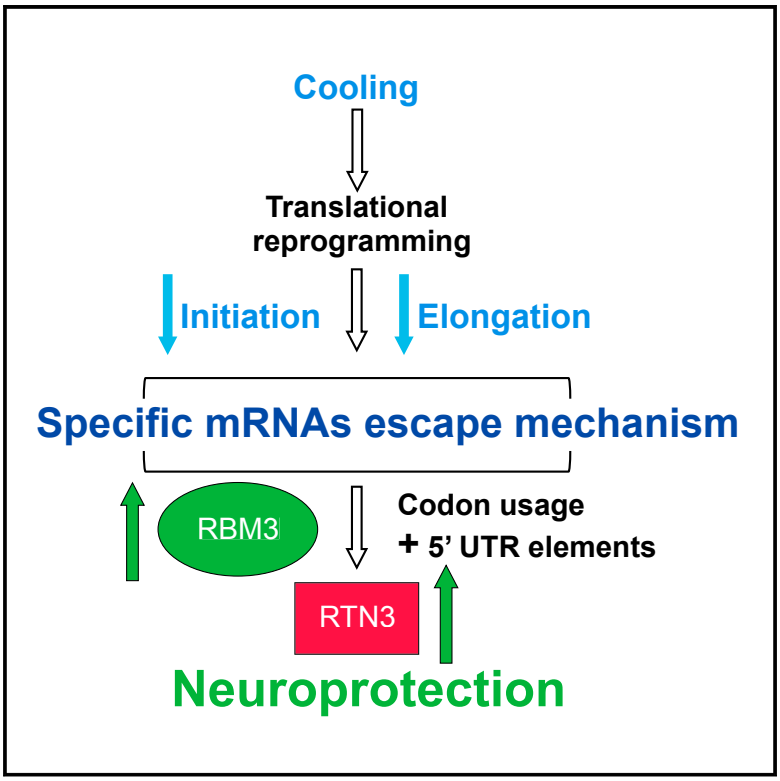
70. Kumar P, Kuscu C, Dutta A: **Biogenesis and function of transfer RNA-related fragments (tRFs)**. *Trends Biochem Sci* 2016, **41**:679-689 http://dx.doi.org/10.1016/j.tibs.2016.05.004.

Article

Current Biology

RTN3 Is a Novel Cold-Induced Protein and Mediates Neuroprotective Effects of RBM3

Graphical Abstract



Authors

Amandine Bastide, Diego Peretti, John R.P. Knight, ..., C. Mark Smales, Giovanna R. Mallucci, Anne E. Willis

Correspondence

c.m.smales@kent.ac.uk (C.M.S.), gm522@cam.ac.uk (G.R.M.), aew5@le.ac.uk (A.E.W.)

In Brief

Therapeutic hypothermia is neuroprotective, and the cold shock protein RMB3 plays a critical role in mediating synaptic repair processes that accompany cooling. Bastide and Peretti et al. show that cooling selectively reprograms the translatome and identify RTN3 as a cold-induced protein that acts downstream of RBM3 in the neuroprotection pathway.

Highlights

- Cooling-induced reprogramming of the translatome increases synthesis of *RTN3*
- The neuroprotective protein RBM3 binds *RTN3* mRNA and drives its expression
- RTN3 overexpression prevents synaptic loss in mice with prion disease
- RTN3 expression is a mediator of RBM3-induced neuroprotection

Accession Numbers

E-MTAB-5437

Bastide et al., 2017, Current Biology *27*, 638–650
March 6, 2017 © 2017 The Author(s). Published by Elsevier Ltd.
http://dx.doi.org/10.1016/j.cub.2017.01.047

CellPress

Current Biology
Article

RTN3 Is a Novel Cold-Induced Protein and Mediates Neuroprotective Effects of RBM3

Amandine Bastide,[1,4] Diego Peretti,[1,2,4] John R.P. Knight,[1] Stefano Grosso,[1] Ruth V. Spriggs,[1] Xavier Pichon,[1] Thomas Sbarrato,[1] Anne Roobol,[3] Jo Roobol,[3] Davide Vito,[3] Martin Bushell,[1] Tobias von der Haar,[3] C. Mark Smales,[3,*] Giovanna R. Mallucci,[1,2,*] and Anne E. Willis[1,5,*]

[1]Medical Research Council Toxicology Unit, Lancaster Road, Leicester LE1 9HN, UK
[2]Department of Clinical Neurosciences, University of Cambridge, Cambridge Biomedical Campus, Cambridge CB2 0AH, UK
[3]Centre for Molecular Processing and School of Biosciences, University of Kent, Canterbury, Kent CT2 7NJ, UK
[4]Co-first author
[5]Lead Contact
*Correspondence: c.m.smales@kent.ac.uk (C.M.S.), gm522@cam.ac.uk (G.R.M.), aew5@le.ac.uk (A.E.W.)
http://dx.doi.org/10.1016/j.cub.2017.01.047

SUMMARY

Cooling and hypothermia are profoundly neuroprotective, mediated, at least in part, by the cold shock protein, RBM3. However, the neuroprotective effector proteins induced by RBM3 and the mechanisms by which mRNAs encoding cold shock proteins escape cooling-induced translational repression are unknown. Here, we show that cooling induces reprogramming of the translatome, including the upregulation of a new cold shock protein, RTN3, a reticulon protein implicated in synapse formation. We report that this has two mechanistic components. Thus, *RTN3* both evades cooling-induced translational elongation repression and is also bound by RBM3, which drives the increased expression of RTN3. In mice, knockdown of RTN3 expression eliminated cooling-induced neuroprotection. However, lentivirally mediated RTN3 overexpression prevented synaptic loss and cognitive deficits in a mouse model of neurodegeneration, downstream and independently of RBM3. We conclude that RTN3 expression is a mediator of RBM3-induced neuroprotection, controlled by novel mechanisms of escape from translational inhibition on cooling.

INTRODUCTION

Therapeutic hypothermia is a powerful neuroprotectant, acting through multiple mechanisms, although the underlying pathways are not fully understood [1, 2]. We recently showed that the cold shock RNA-binding protein, RBM3, plays a critical role in mediating synaptic repair processes essential for neuroprotection in mouse models of neurodegenerative disease [3]. Thus, the inability to induce RBM3 expression in early disease results in defective structural synaptic plasticity and hence reduced regenerative capacity, leading to synapse loss and eventually neuronal loss. Inducing endogenous RBM3 expression in vivo through cooling, or by lentiviral-mediated overexpression, prevented synapse loss in prion and Alzheimer-type mice, rescued memory deficits, protected against neurodegeneration, and significantly prolonged survival [3]. How RBM3 mediates these effects is unknown, although it is likely to be related to its RNA chaperone function, as it facilitates selective mRNA translation following a number of cellular stresses, including cooling [4–6]. RBM3 is also implicated in protection against cell death [7, 8] and increases local protein synthesis at dendrites [9].

Upon cooling, the changes to the protein synthesis machinery are similar to those observed with other stress inducers (e.g., UVB exposure) [10, 11] and include phosphorylation of eukaryotic initiation factor 2 (eIF2) on the alpha subunit [12, 13]. However, cold shock differs because relieving eIF2α-mediated inhibition of translation is insufficient to restore protein synthesis rates [12]. Instead, the rapid reduction in protein synthesis that accompanies cooling [12–15] is a result of a decrease in translational elongation mediated by the phosphorylation of elongation factor 2 (eEF2) by elongation factor 2 kinase (eEF2K; a negative regulator of eEF2) [12]. Importantly, suppression of eEF2K and subsequent reactivation of eEF2 significantly increases the rate of protein synthesis rates in cooled cells [12], consistent with the concept that elongation is a major regulatory node under specific pathophysiological conditions [16, 17].

To examine the relationship between cold stress, RBM3 induction, and the modulation of mRNA translation for the synthesis of putative neuroprotective proteins, we investigated the post-transcriptional response to hypothermia in vitro and validated the data in vivo in a mouse model of neurodegeneration, in which we know cooling is protective, mediated by RBM3. We show that, following cold shock/cooling, the global decrease in protein synthesis rates is associated with selective reprogramming of the translatome. We find enhanced synthesis of specific proteins: not only the cold shock protein RBM3 (as predicted) but also of a number of proteins with a role in development and function of the nervous system, including reticulon 3 (RTN3). RTN3 has a role in synaptic plasticity and synapse formation [18, 19] and is thus a compelling candidate for the neuroprotective effects mediated by increased RBM3 expression. We find that both *RBM3* and *RTN3* evade translational repression and that

 © 2017 The Author(s). Published by Elsevier Ltd.
This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

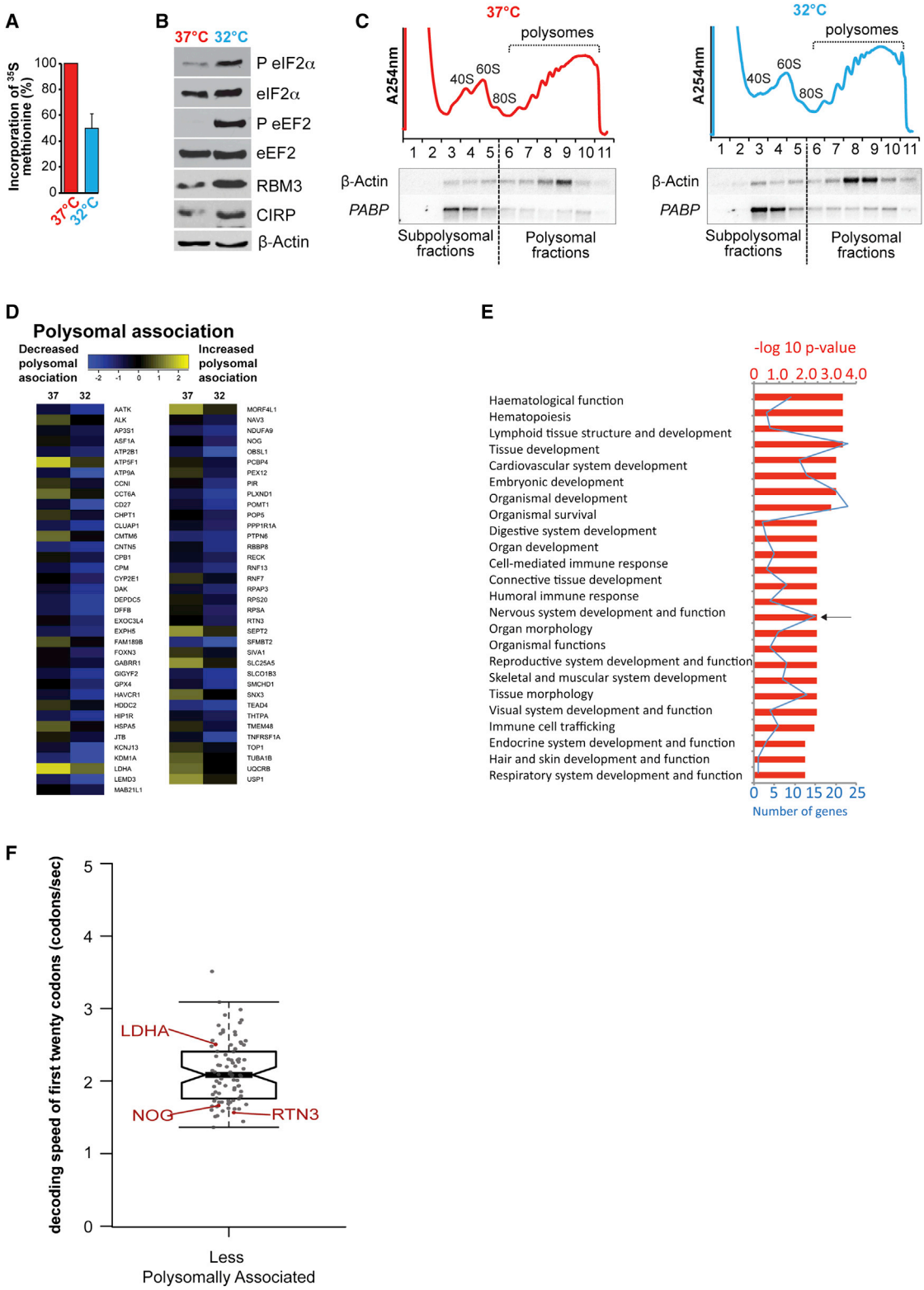


Figure 1. Mild Cooling Results in Translation Regulation of a Defined Set of Transcripts

(A) Protein synthesis rates determined by [^{35}S]-methionine label incorporation after 24 hr incubation of HEK293 cells at 32°C. Values were normalized to cells incubated at 37°C. Error bars represent SE within three independent experiments.

(B) HEK293 cells were incubated at 37°C or at 32°C for 24 hr and immunoblotted for RBM3 and CIRP, eIF2 alpha eEF2, and β-actin.

(legend continued on next page)

RBM3 binds *RTN3* mRNA and plays a major role in driving cooling-induced upregulation of RTN3 expression. Finally, we show that RTN3 expression, downstream of RBM3 induction, mediates cooling-induced neuroprotection in mice with neurodegenerative disease and importantly is neuroprotective even in the absence of cooling.

RESULTS

Cooling Induces Reprogramming of the Translatome

Both transcriptional and post-transcriptional control mechanisms are required for the overall response to cell stress [20]. In order to examine the genome-wide changes accompanying cooling, we incubated HEK293 cells at 32°C for 24 hr. This resulted in reduction in protein synthesis (Figure 1A) and phosphorylation of the translation initiation and elongation factors eIF2α and eEF2 (Figure 1B; in agreement with previous studies [12]). We chose HEK293 cells because the response to cooling is well documented [12, 15, 21] and, in addition, they express many markers associated with neuronal lineage [22]; thus, using this cell line increases the potential for the identification of cold-induced putative neuroprotective proteins. Transcriptional analyses of the cooled HEK293 cells showed that 119 genes were downregulated at the transcriptional level, with no significant increases in transcription of any mRNAs (Figure S1; Table S1). In addition, no differences greater than 2-fold were identified in the expression of microRNAs (miRNAs) (Table S2). These data support regulation of protein synthesis as an important mechanism for control of gene expression following cooling. We have shown previously that cooling reduces global rates of protein synthesis and importantly that elongation repression is the driver of this process (Figures 1A and 1B [12]). We hypothesize that, during cooling, specific mRNAs are able to evade a global reduction in translation elongation so that the expression of the corresponding proteins is maintained or even increased. However, identifying such mRNAs represents a technical challenge. Under conditions in which the initiation of translation is inhibited, the number of actively translating ribosomes decreases [10, 23] and polysome profiling can be used to identify those mRNAs that remain polysomally associated; this generally correlates with increased synthesis of the corresponding proteins [20]. However, under conditions in which elongation is inhibited, the number of polysomally associated ribosomes will stay the same or increase, and therefore it is difficult to identify mRNAs that either escape or are relatively insensitive to elongation slow down. Therefore, to identify proteins whose synthesis was increased during cooling, computational modeling was used in conjunction with polysome profiling.

Sucrose density gradient analysis was carried out on cooled samples to separate polysomes and the associated transcripts (Figure 1C), which were subsequently purified and analyzed by cDNA microarray. By microarray 71, mRNAs were identified that were associated with a decreased number of polysomes at 32°C (Figures 1D and S2; Tables S3 and S4). Importantly, ingenuity pathway analysis showed that there was significant enrichment for mRNAs that encode proteins that function in the nervous system and its development (15/71; marked by an arrow, Figure 1E). To predict which of these neuronal-related mRNAs were translated in cooled cells, we used a computational model of elongation control [24] generated by defining the intracellular concentration of ribosomes, translation factors, and tRNAs (Table S5). The model allows the speed of decoding to be estimated for any given open reading frame [25], assuming that decoding speed is not significantly limited by tRNA-independent parameters, such as mRNA secondary structure or modifications. We have previously shown that, despite this assumption, the model can be used to rank expression levels from multiple elongation-controlled mRNAs reliably [25]. The model predicts that, under eEF2 ablation, fast codons (which are decoded by abundant tRNAs) change their elongation rate by an order of magnitude, whereas slow codons are relatively unaffected (see Figures S3A–S3C).

We applied the model to the transcripts that encoded mRNAs with roles in neuronal processes identified by polysome profiling (Figure 1D) to analyze elongation over the initial 20 codons (Figure 1F; Table S6). Our analysis showed that a subset of mRNAs, including *RTN3* and *Noggin*, contained codons that require less abundant tRNAs in the 5′ end of the transcripts (Figure 1F), and our model predicted that these could escape the repression of elongation.

mRNAs Encoding a Subset of Neuronal Proteins Overcome Elongation Inhibition on Cooling

Western analysis showed that expression of Noggin and RTN3, and as expected RBM3, increased at 32°C, whereas GBBR1 and LDHA levels, which are encoded by mRNAs that contain codons requiring abundant tRNAs, were unchanged in both cell lines (Figure 2A). To confirm a post-transcriptional response, we examined mRNA expression changes of RTN3 and Noggin using qPCR (Figure 2B); there was a small reduction in the levels of Noggin, consistent with the transcriptional profiling data (Table S1; Figure S1) and no change for RTN3. There was an increase in the expression of RBM3 mRNA, in agreement with previous studies [4].

To confirm escape of translation elongation repression and identify a strong candidate protein with a neuroprotective function,

(C) Sucrose density gradient ultracentrifugations were performed from HEK293 cells incubated at 37°C or 32°C for 24 hr. Plots show the distribution of RNA within subpolysomes (40S, 60S, and 80S) and polysomes. Northern analysis was carried out on individual fractions, which were probed for β-actin or PABP.
(D) mRNAs from gradient fractions were pooled and subjected to cDNA microarray. The color scale represents the ratio of mRNA in subpolysome and polysome fractions, normalized $\log_2$ (polysome/subpolysome) value; yellow is polysome- and blue subpolysome-associated mRNAs.
(E) mRNAs that showed significant change in polysome/subpolysome (P/S) ratio on cooling were clustered into functional groups. Biological functions associated with decreased polysomal-associated transcripts, obtained from the ingenuity pathways analysis. Fisher's exact test was used to calculate a p value (threshold $p < 0.05$) for each biological function represented in the red bar chart. The blue line represents number of proteins per category.
(F) Predictive modeling of transcript-decoding speed was performed on the initial 20 codons of human transcript sequences from those that showed decreased polysomal association. The boxplot shows mRNAs that have a decrease in polysomal association and contain an initial 20 "slow" codons (e.g., RTN3 and Noggin [NOG]) compared to those that contain "fast" codons, such as LDHA.
See also Figures S1–S3 and Tables S1–S6.

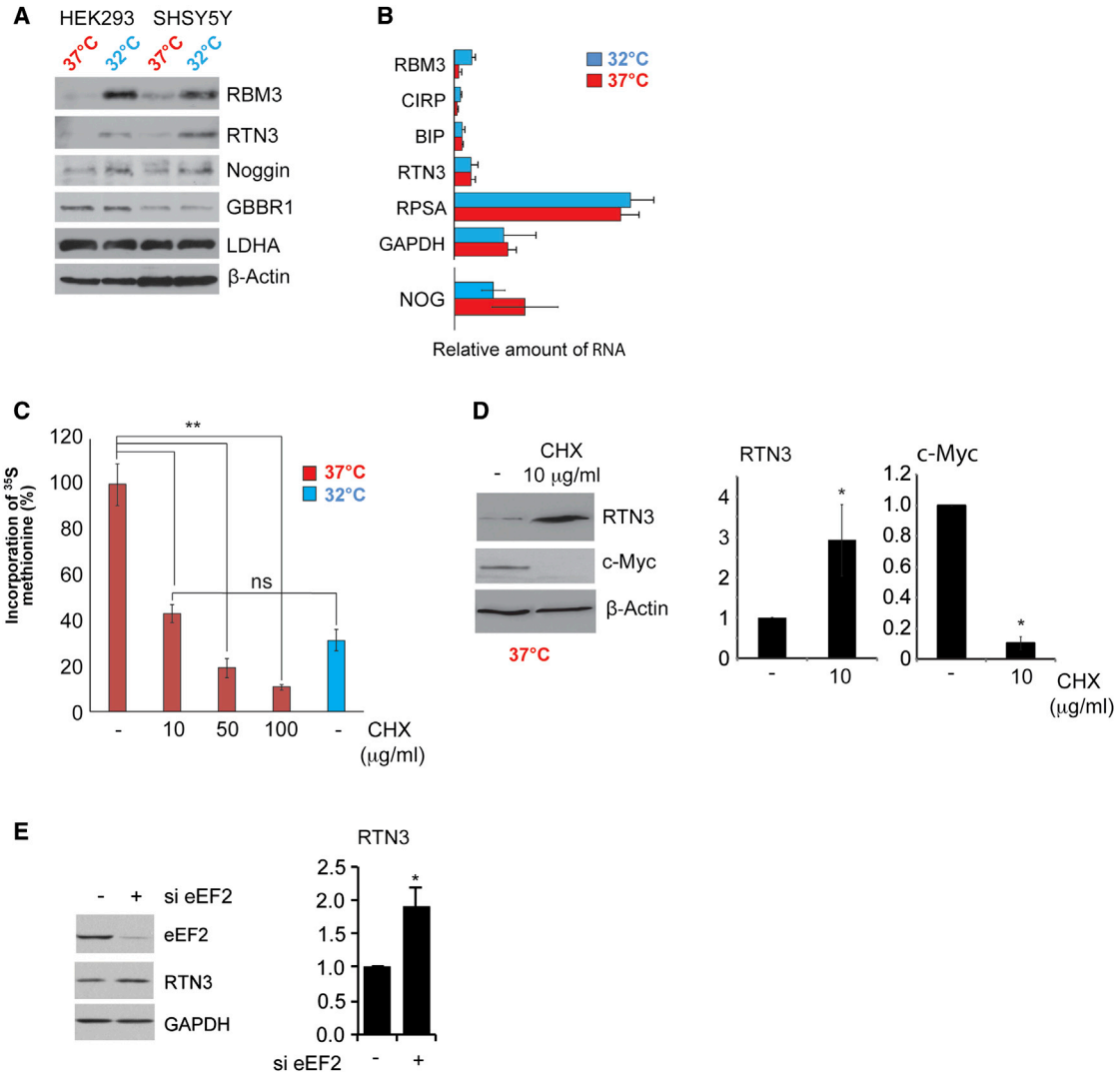


Figure 2. RTN3 Is Subject to Elongation Control

(A) Extracts from control or cooled HEK293 or SHSY5Y cells were immunoblotted for RTN3 and Noggin, GBBR1, and LDHA. β-actin is used as a loading control.

(B) qRT-PCR was performed on corresponding transcripts. Error bars represent 1 SD from the mean within three independent experiments. *GAPDH* was used as a control.

(C) Protein synthesis rates determined by [^{35}S]-methionine label incorporation after 24 hr incubation of HEK293 cells at 37°C with cycloheximide. Values were normalized to untreated cells. A two-tailed paired Student's t test was used to calculate statistical significance. Error bars represent SE within three independent experiments. **$p < 0.001$, all three conditions.

(D) Extracts from cells exposed to 10 μg/mL cycloheximide at 37°C were immunoblotted for RTN3, c-Myc, and β-actin. Error bars represent 1 SD from the mean within three independent experiments. *$p < 0.01$.

(E) eEF2 expression was reduced by siRNA, proteins extracted and immunoblotted with the antibodies shown. *GAPDH* was used as a loading control. Error bars represent 1 SD from the mean within three independent experiments. *$p < 0.01$.

See also Figures S4 and S5A.

we focused on RTN3, a protein that has known function in synaptogenesis in the adult nervous system [26–28] and a role in neuroprotection. To reduce elongation by an alternative method, we treated cells grown at 37°C with cycloheximide, which stalls the translocation step in the elongation cycle [29]. As expected, incubation with cycloheximide decreased global protein synthesis rates (Figure 2C) and reduced expression of c-Myc, which is known to have a short half-life of 20 min. However, there was increased RTN3 expression, consistent with our model's prediction that slowed elongation enhances synthesis of this protein (Figures 2D and S4).

To mimic the cooling-induced elongation block, we reduced eEF2 expression by RNAi (Figure 2E). This resulted in an increase in RTN3 levels, suggesting that the rate of elongation along *RTN3* mRNA is relatively unaffected by reduced availability of eEF2, in agreement with our model (Figure S3).

Whereas we have shown previously there is a small effect of cooling on the stability of specific proteins [21], RTN3 is

a relatively stable protein with a half-life of at least 24 hr, and therefore, any effects of turnover in the time frame of the experiment will be minimal (Figures S5A and S5B).

RBM3 Binds to RTN3 mRNA and Increases Its Translation through *trans*-Acting Effects on Initiation

Cooling also reduces the rate of translation initiation, via inhibitory phosphorylation of eIF2α (Figure 1A) [12], which compensates for the reduction in translation elongation, as fewer ribosomes will be available for initiation while they are limited by globally reduced elongation speeds. A similar phenomenon has been suggested to occur previously in yeast [25, 30]. Thus, we predict that, in order to increase their translation, transcripts also overcome the cooling-induced initiation block. This is likely to be driven by *trans*-acting factors acting upon *cis* elements within the transcripts. Because RBM3 is an RNA chaperone whose expression is increased in cooled cells [31, 32] and is known to mediate the neuroprotective effects of cooling [3], we hypothesized that cooling-induced RBM3 may act as a *trans*-acting factor promoting *RTN3* translation and that some of the neuroprotective effects of RBM3 may be mediated through RTN3.

To address whether RBM3 interacted with RTN3 mRNA, we carried out immunoprecipitation reactions, and data showed that RBM3 binds to *RTN3* mRNA in both HEK293 cells and in hippocampus of wild-type mice (Figures 3A and 3B). We then asked whether RBM3 expression affects RTN3 levels in HEK293 cells and mouse brain by transfecting with RBM3-expressing constructs or lentiviruses, respectively. In each case (Figures 3C and 3D), the data show that overexpression of RBM3 resulted in a dramatic increase in RTN3 protein expression, with no corresponding increase in *RTN3* transcript levels in vitro (Figure 3C) or in vivo (Figure 3D).

cis-Acting Elements in RTN3 Contribute to Evasion of the Initiation Block

To examine the role of *cis*-acting elements in 5′ UTR of *RTN3* in its post-transcriptional regulation by RBM3, we used a luciferase reporter construct containing the RTN3 5′ UTR (Figure 3E). At 25°C, there was a 5-fold increase in translation of messages containing the 5′ UTR of *RTN3*, compared to the control (Figure 3F). Further, overexpression of RBM3 resulted in a 5-fold induction of luciferase expression at 37°C (Figure 3Gi). In contrast, RNAi of RBM3 produced a small but significant reduction in luciferase activity (Figure 3Gii). Taken together, the data support a role for RBM3 in controlling RTN3 expression through the *RTN3* 5′ UTR cis-regulatory sequence.

Cooling Induces RTN3 Expression In Vivo through Post-transcriptional Mechanisms

Given the role of RBM3 in regulating RTN3 (Figures 2 and 3), we focused on RTN3 as a potential mediator of the neuroprotective effects of RBM3 induction. RTN3 is a strong candidate for this role. It is a member of the reticulon family of proteins, with multiple functions in the nervous system, including axon and neurite outgrowth [18, 19] and synapse formation [26–28]. It also has an indirect role in synaptic plasticity through its inhibition of BACE1, a secretase involved in cleavage of APP and a negative modulator of pCREB levels [28, 33, 34].

We tested whether RTN3 expression was increased in brain on cooling in vivo by inducing hypothermia in wild-type FVB mice using 5′ AMP, as described [3, 35]. We found the levels of both RBM3 and RTN3 increased on cooling (Figure 4Ai) without corresponding changes in respective mRNA levels (Figure 4Aii), consistent with post-transcriptional upregulation. Further, cooling induced a reduction in global protein synthesis rates to ~40% of control levels (Figure 4B), as observed in vitro (Figure 1A) (although the abundance of polysomes was not reduced to an equivalent extent [Figure 4C], again consistent with in vitro findings [Figure 1C] [12]). Knockdown of RBM3 in mice via lentivirally mediated RNAi significantly reduced the RTN3 increase on cooling (Figure 4D), confirming physiological relevance of this functional interaction and suggesting that RTN3 expression is downstream of RBM3 in cooling in vivo.

RTN3 Mediates Cooling-Induced Neuroprotective Effects of RBM3

We next asked to what extent RTN3 is neuroprotective in mice with neurodegenerative disease, using mice with prion disease, specifically tg37 mice [36] inoculated with Rocky Mountain Laboratory (RML) strain as in our previous studies [3, 37–41]. These mice overexpress prion protein (PrP) and have a rapid disease course, succumbing to disease in 12 weeks [36]. In these mice, synaptic loss is associated with the failure to induce RBM3 on cooling early in the disease course at 6 weeks post-inoculation (w.p.i.) [3], developing behavioral deficits at 9 w.p.i., with neuronal loss from 10 w.p.i. In general, terminal clinical signs appear at around 12 w.p.i.

We first confirmed that increased RTN3 expression is downstream of RBM3 in prion-diseased mice. Lentivirally mediated RNAi of *RTN3* reduced RTN3 levels in hippocampi (Figure 5A), but not RBM3 levels (Figure 5A). Further, RNAi of RTN3 did not affect high RBM3 levels induced by cooling while preventing, as predicted, the cold-induced rise in RTN3 (Figure 5B). This allowed us to address whether and to what extent RTN3 mediates the neuroprotective effects of RBM3. We found that lentivirally mediated RNAi of RTN3 in prion-infected mice abolished the protective effects of cooling on behavioral impairments (Figure 5C), accelerated neuronal loss (Figure 5D), and abolished the cooling-associated increase in survival in prion-infected mice (Figure 5E). Thus, reducing levels of RTN3 largely abolished the protective effects of high levels of cold-induced RBM3, supporting the conclusion that RTN3 is a major mediator of the effects of RBM3. Even in the absence of cooling, RNAi of RTN3 accelerated synapse loss in disease (Figures S5C and S5D), supporting a role for RTN3 in synapse maintenance and formation/plasticity. However, the exact mechanism remains unknown.

RTN3 Is Neuroprotective in Neurodegenerative Disease

RBM3 induction is profoundly neuroprotective in both prion and 5XFAD mice, effects that are abrogated if animals undergo knockdown of RBM3 [3] or of RTN3 (Figures 5C–5E). To address whether RTN3 is neuroprotective in the absence of cooling, we injected prion-infected tg37 mice with lentiviruses overexpressing RTN3 (LV-RTN3). This increased expression of RTN3 as expected (Figure 6A), importantly without increasing RBM3 expression, consistent with RTN3 being downstream of RBM3

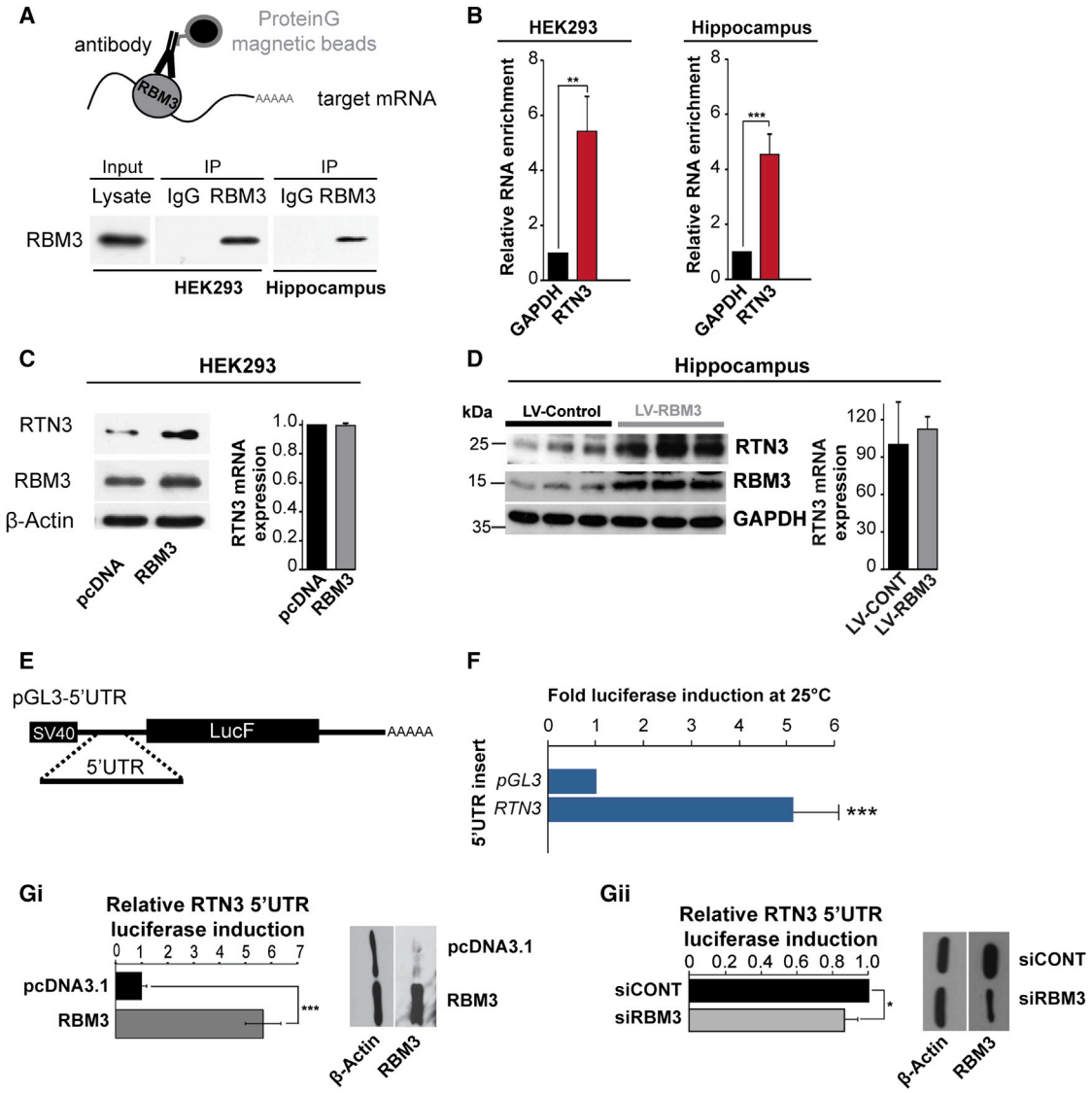


Figure 3. RTN3 Expression Is Downstream of RBM3

(A) Schematic representation of an RNA immunoprecipitation (RNA-IP) assay. Immunoblots of input lysate from HEK293 cells or hippocampus, immunoprecipitated with either rabbit IgG or RBM3 antibody, are shown.

(B) qRT-PCR was performed on RNA-IP samples using primers specific for human (HEK293) or mouse (hippocampus) samples. All values are normalized with respect to the initial RNA input material, and the enrichment is plotted relative to *GAPDH*. A two-tailed paired Student's t test was used to calculate statistical significance. Error bars represent 1 SD from the mean within three independent experiments. $^{**}p < 0.01$; $^{***}p < 0.001$.

(C) HEK293 cells were transfected with an expression plasmid construct encoding RBM3 or a control plasmid, and extracts were immunoblotted for RTN3. β-actin was used as a loading control. RNA expression of RTN3 was assessed by qRT-PCR.

(D) Mouse hippocampi stereotaxically injected with lentivirus containing a construct to overexpress RBM3 and extracts were immunoblotted for RTN3 and GAPDH. qRT-PCR was used to assess the expression of RTN3.

(E) Schematic representation of the RTN3 containing plasmid constructs encoding firefly luciferase.

(F) HEK293 cells were transfected with construct containing the 5′ UTR of RTN3 and a Renilla luciferase control and incubated at either 37°C or 25°C for 24 hr. Firefly luciferase activity was calculated relative to Renilla luciferase for each condition and expressed as the fold induction from 37°C to 25°C.

(Gi) HEK293 cells were transfected with either control (pcDNA3.1) or RBM3 expression plasmid (pcDNA-RBM3) and then transfected with either RTN3 5′ UTR pGL3 or pGL3 and Renilla luciferase constructs and luciferase activity determined. A two-tailed paired Student's t test was used to calculate error. Error bars represent 1 SD from the mean within three independent experiments. $^{***}p < 0.001$.

(Gii) HEK293 cells were transfected with either control siRNA (siCONT) or *RBM3* siRNA (siRBM3) and then transfected with pGL3 and Renilla luciferase constructs. The fold repression from *RTN3* 5′ UTR pGL3 compared to the control pGL3 was calculated and normalized to siCONT transfection. A two-tailed paired Student's t test was used to calculate statistical significance. Error bars represent 1 SD from the mean within three independent experiments. $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

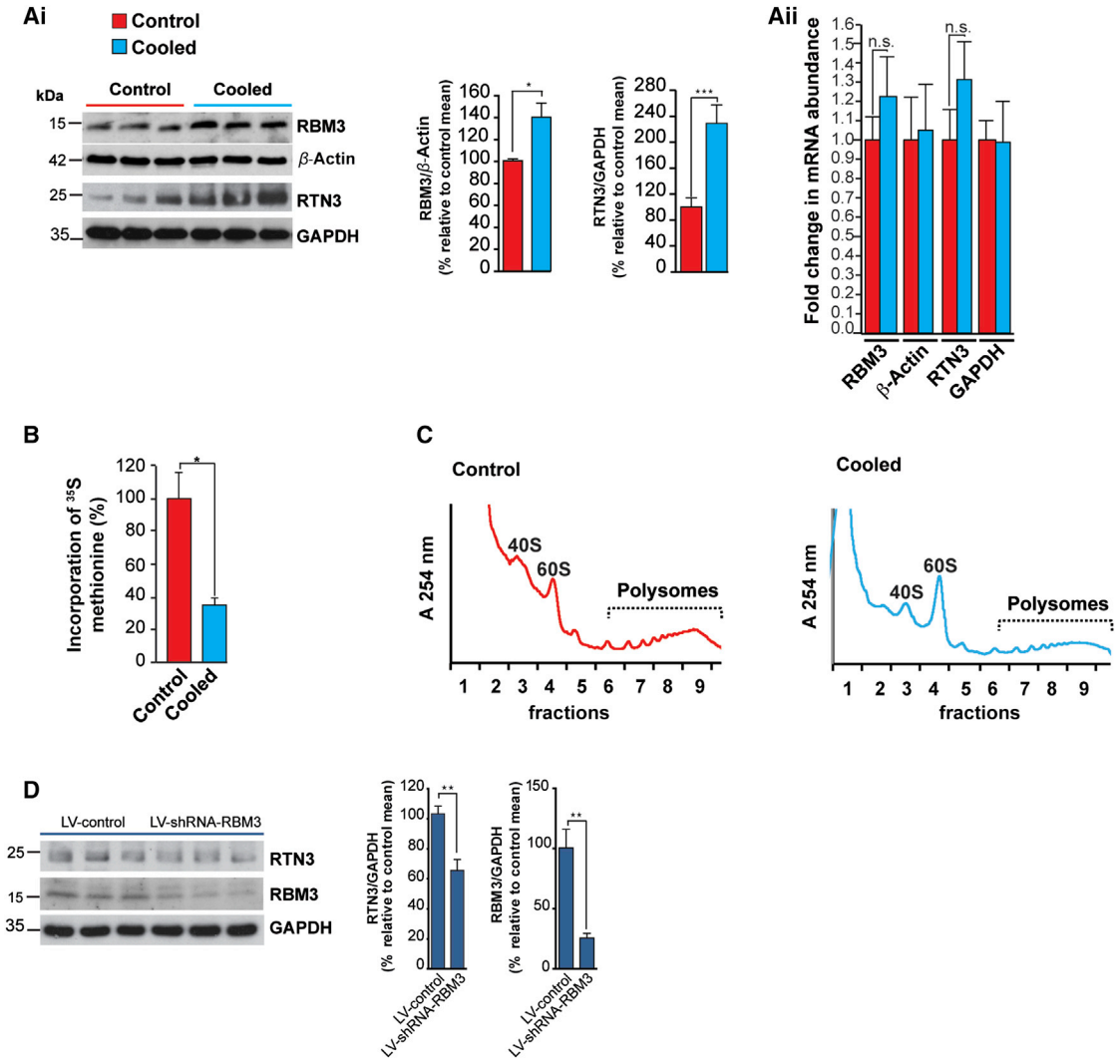


Figure 4. RBM3 Regulates RTN3 In Vivo

(Ai) Cooling increases RBM3 and RTN3 levels in hippocampi of wild-type mice. Representative western blots are shown. Bar graphs show quantification of RBM3 and RTN3 levels relative to β-actin and GAPDH, respectively. n = 5 control and 5 cooled mice for RBM3. n = 9 control mice and 9 cooled mice for RTN3. *$p < 0.05$; ***$p < 0.001$

(Aii) qRT-PCR of RNA isolated from hippocampi of cooled mice showed no significant change in the abundance of RBM3 or RTN3 mRNAs following cooling.

(B) Protein synthesis rates were determined by ^{35}S methionine incorporation into nascent protein using ex vivo hippocampus slices from cooled mice and control mice. A two-tailed paired Student's t test was used to calculate statistical significance. Error bars represent 1 SD from the mean within three independent experiments. *$p < 0.05$.

(C) Sucrose density gradient ultracentrifugation performed of cytoplasmic extracts from hippocampi from control and cooled mice. Absorbance plots show the distribution of RNA within subpolysomes (40S, 60S, and 80S) and polysomes.

(D) RTN3 induction on in vivo cooling is dependent on RBM3 protein expression. Knockdown of RBM3 resulted in a 38% decrease in RTN3 induction. n = 6 LV-control and 6 LV-shRNA-RBM3 mice. **$p < 0.01$.

(Figure 6A). To assess neuroprotective effects of RTN3 overexpression, we measured synapse number (Figure 6B), burrowing activity (Figure 6C), and neuron number in CA1 of hippocampus from diseased animals treated with LV-RTN3 compared to controls (Figure 6D) over the time course of disease. All of these parameters decline in the course of prion disease in the absence of intervention. Importantly, RTN3 overexpression restored synapse number to wild-type levels, markedly above levels seen in untreated mice at the same and later time points (Figure 6B), and prevented the decline in burrowing behavior (Figure 6C). RTN3 overexpression also conferred marked neuroprotection at the histological level, with very notable conservation of the CA1 pyramidal neuron layer (Figure 6D). Importantly, LV-RTN3 significantly increased survival of prion-infected mice (Figure 6E), recapitulating the effects of LV-RBM3 we previously described [3]. In our study of RBM3, we showed that onset of synaptic failure correlates with failure of induction of RBM3 at 6 w.p.i. [3]. And interestingly, RTN3 induction at this time point

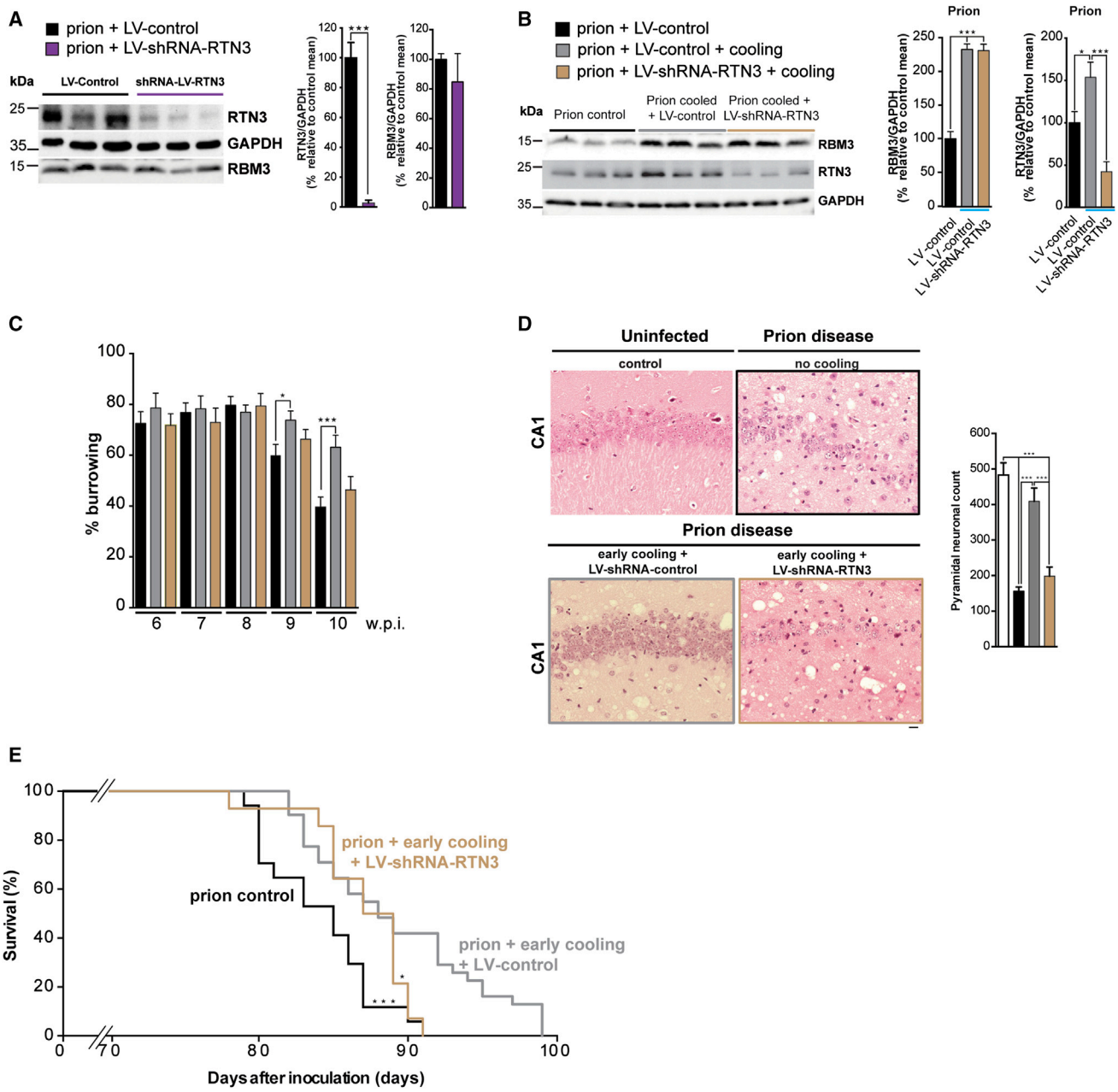


Figure 5. Lentiviral-Mediated Downregulation of Endogenous RTN3 Prevents Cooling-Induced RBM3-Mediated Neuroprotection

(A) LV-shRNA-RTN3 injected into hippocampi of prion-infected mice significantly reduces RTN3 protein levels compared to control shRNA (LV-shCONT). n = 6 prion+LV-control mice and 6 prion+LV-shRNA-RTN3 mice. Representative western blots and bar graphs quantification are shown.

(B) Western blot of RBM3 in LV-shRNA-RTN3-treated early-cooled prion mice shows no change in expression. n = 6 mice per experimental condition.

(C) The early-cooling-induced protection in burrowing behavior declines in LV-shRNA-RTN3 mice. Food pellet remaining in the tube measured after 2 hr is expressed in percentage burrowed. Graph bar with prion (black bars; n = 12 mice), prion + early cooling (gray bars; n = 20 mice), and prion + early cooling + LV-shRNA-RTN3 (light brown bars; n = 12 mice) is shown. One-way ANOVA with Tukey's post-test was used for multiple comparisons. *p < 0.05; ***p < 0005.

(D) Representative images of H&E-stained hippocampal sections from uninfected control (top left-hand panel), prion-infected mice (top right-hand panel), prion-infected mice treated with early cooling and LV-control (bottom left-hand panel), and prion-infected mice treated with early cooling and LV-shRNA-RTN3 (bottom right-hand panel). Prion-infected mice show extensive neuronal loss, with associated spongiosis, whereas early cooling treatment prevents neurodegeneration. This protection is abrogated with LV-shRNA-RTN3. The graph bar shows quantification of the average number of neurons for each condition in the CA1 area of hippocampus. n = 3 mice (white bar), 7 mice (black bar), 7 prion mice (bar), and 9 mice (light brown bar). One-way ANOVA and Brown-Forsythe test with Tukey's post hoc analysis for multiple comparisons were used. ***p < 0.001. The scale bar represents 50 μm.

(legend continued on next page)

is also lost (Figure 6F), supporting a functional relationship between these proteins in the context of disease and loss of synaptic structural plasticity.

Misfolded PrP levels were not affected by cooling or RTN3 expression, and levels were equivalent in all mice, precluding a mechanism of action via prion protein aggregation, consistent with our previous findings with RBM3 [3] (Figure S5D).

Thus, RTN3 overexpression results in neuroprotection at the level of synapse number, behavior, neuronal numbers, and increased survival, downstream and independently of cooling-mediated RBM3 induction.

DISCUSSION

The metabolic response to cooling is highly conserved [3, 35, 42]. The neuroprotective effects of hypothermia are essential for healthy brain function after hibernation and are widely exploited medically [43, 44]. However, relatively little is known about how global gene expression changes bring about these protective effects.

We have examined the genome-wide changes induced by cold stress by carrying out transcriptional, miRNA, and translational profiling on cells that were subjected to cooling. Our data show that specific induction of gene expression during cold stress is regulated at the level of translation with no significant transcriptional upregulation or changes in miRNA expression (Figures 1, S1, and S2; Tables S1 and S2).

Elongation rate control is the major determinant of global protein synthesis suppression upon cooling [12], but which transcripts are controlled in this way was unknown. We therefore generated a computational model to predict those messages that were particularly sensitive to regulation at this stage (Figure S3). According to our model, expression from transcripts that require abundant tRNAs would be dependent on eEF2 to maintain efficient elongation and protein expression. In contrast, mRNAs requiring rare tRNAs would be proportionally less affected by reductions in eEF2 availability and would be predicted to display either a small decrease or exhibit no net change in polysomal association upon cooling (Figure S3). In support of this hypothesis, the *RTN3* mRNA is decoded by rare tRNAs and, following cooling, exhibited reduced polysomal-associated yet increased protein expression (Figure 2). We showed that *cis*- and *trans*-acting factors were required for *RTN3* to overcome cooling-induced translation inhibition (Figure 3) and that the RNA chaperone RBM3 [6, 31, 32] was required (Figures 3 and 4).

We were interested in the functional consequence of RTN3 upregulation in response to cold shock. We have previously reported that RBM3 mediates the neuroprotective effects of cooling in mouse models of neurodegeneration and is necessary for maintenance of synaptic structural plasticity [3]. How it does this was not understood. We considered RTN3 to be a candidate neuroprotective protein specifically upregulated by RBM3 induction on cooling. In support of this, knockdown of RTN3 reduced synapse number at an earlier stage and abolished the neuroprotective effects of cooling in prion-diseased mice, despite cooling-induced increase in RBM3 levels (Figures 5 and S5C). Conversely, lentivirally mediated overexpression of RTN3 prevented synapse loss and neurodegeneration in prion-diseased mice (Figure 6), recapitulating the neuroprotective effects observed during RBM3 expression [3].

In conclusion, we propose that, following cooling, there is translational reprogramming, leading to the overexpression of specific cold-inducible proteins, including the known cold shock protein, RBM3, but also of RTN3. Critically, we show that induction of RTN3 is downstream of RBM3 expression, and our data suggest that RTN3 is a mediator of the RBM3-driven neuroprotective effects of cooling in prion-diseased mice, most likely through its multiple roles in the regulation of neurite outgrowth and regulation of synaptic plasticity. It is likely that RTN3 induction would mediate a similar neuroprotective role in other neurodegenerative conditions. Further, its inhibition of BACE1-mediated cleavage of APP could also contribute to neuroprotection in Alzheimer's pathology. We propose that the control of RTN3 expression through escape from inhibition of translation on cooling at the levels of initiation and elongation provides new targets for neuroprotective therapies in neurodegenerative disease.

EXPERIMENTAL PROCEDURES

All animal work was conducted according to UK Home Office Regulations. For details of antibodies, plasmids, and oligonucleotides, please see the Supplemental Experimental Procedures.

Cell Culture and Transfections

HEK293 cells were cultured under standard conditions in DMEM supplemented with 10% fetal bovine serum (FBS). For induction of cold stress, cells were incubated at 32°C or at 25°C for 24 hr before harvesting. Small interfering RNAs (siRNAs) were transfected using DharmaFECT 1 and plasmids with Lipofectamine 2000 (Invitrogen).

SDS-PAGE and Western Blotting

Western blots were performed as described previously [45] with modifications described in the Supplemental Experimental Procedures.

Determination of Protein Synthesis Rates

The rate of protein synthesis was measured by the incorporation of ^{35}S-methionine into trichloroacetic acid (TCA)-insoluble material as described previously [10]. Further details are provided in the Supplemental Experimental Procedures.

Sucrose Gradient Density Centrifugation and RNA Detection

Sucrose density gradient analysis was carried out as described [10]. Full details are provided in the Supplemental Experimental Procedures.

RNA Extraction

Total RNA extraction was performed using Trizol reagent (Invitrogen) according to the manufacturer's instructions.

Microarray Hybridization

The human cDNA microarrays were manufactured in Nottingham University. This custom cDNA microarray consists of 29,593 (32,448 total, including 2,855 control probes) oligonucleotide probes derived from MWG Human 30K slides A, B, and C. RNA from sucrose density gradient fractionation was

(E) Early cooling prolongs survival in prion-infected mice but is abolished by knockdown of RTN3. Kaplan-Meier plot; n = 25 cooled mice (gray line); n = 17 not cooled (black line); n = 14 cooled + shRNA of RTN3 (light brown line). One-way ANOVA with Tukey's post-test was used for multiple comparisons; not cooled versus cooled mice ***$p < 0.001$; not cooled versus cooled + shRNA of RTN3 n.s.; cooled mice versus cooled + shRNA of RTN3 *$p < 0.05$.
See also Figures 5C and 5D.

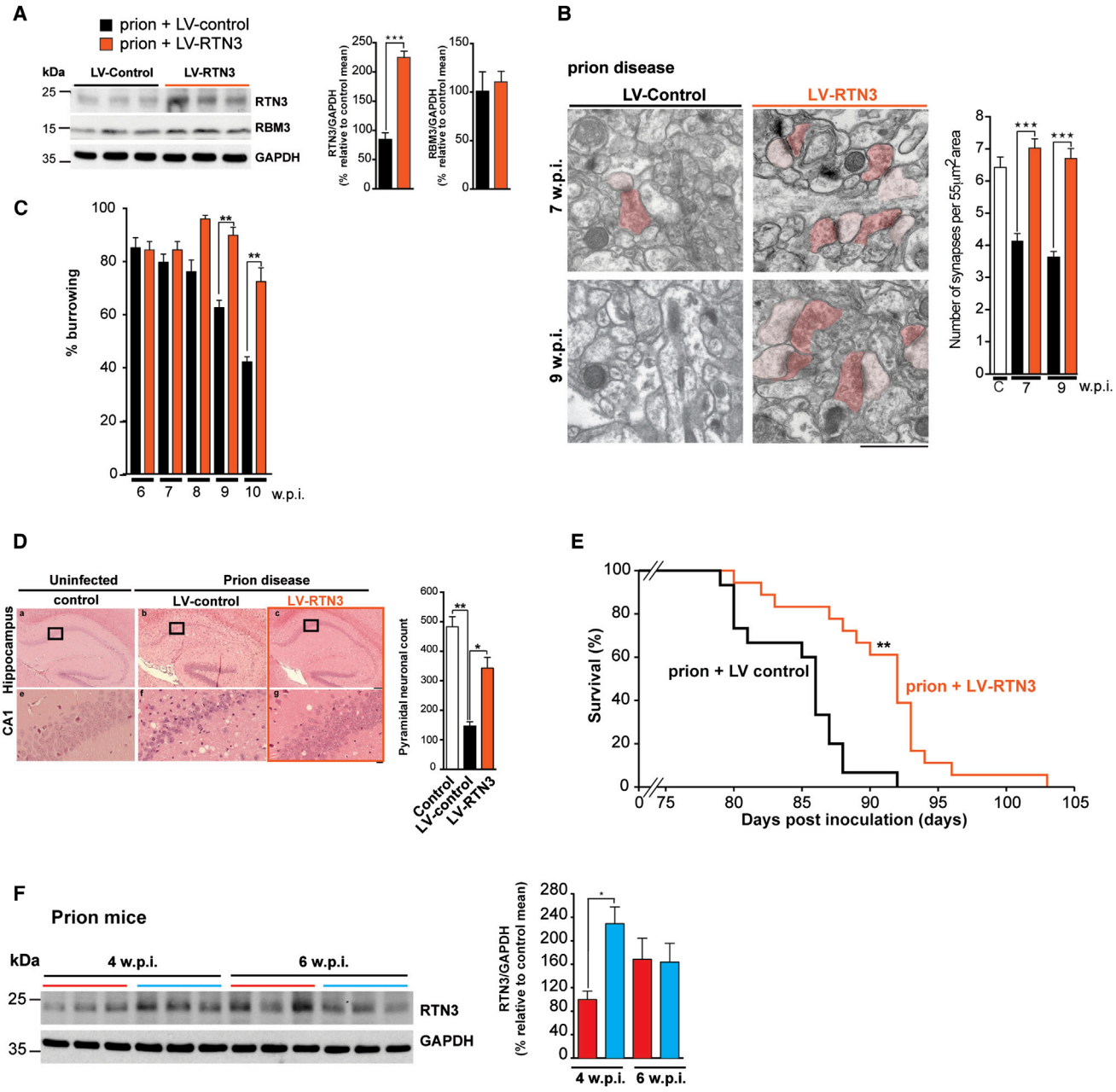


Figure 6. Enhanced RTN3 Expression Is Sufficient to Protect against Prion Disease in the Absence of Cooling

(A) LV-RTN3 delivery to hippocampi of prion-infected mice increases RTN3 in the absence of cooling compared to control lentiviral treatment (LV-control) and endogenous RBM3 (remain constant). n = 6 mice LV-control and 6 mice LV-shRNA-RTN3. Representative western blots and bar graphs quantification are shown.

(B) LV-RTN3 protected the deficit in synapse loss in prion-infected mice at 7 and 9 w.p.i. Representative electron micrographs are shown, pseudo-colored for ease of synapse identification: presynaptic, dark pink; postsynaptic, light pink. Bar chart quantification is shown. n = 93 images from three animals per condition. Data represent mean ± SEM; t test; ***$p < 0.0001$. The scale bar represents 1 μm.

(C) RTN3 overexpression prevented the decline in burrowing behavior of prion–infected mice. Food pellets remaining in the tube were measured after 2 hr and are expressed as percentage burrowed. Graph bar with prion (black bars; n = 14 mice) and prion + LV-RTN3 (orange bars; n = 20 mice) is shown. Kruskal-Wallis test with Dunn's multiple comparisons test; **$p < 0.01$.

(D) Representative images of H&E-stained hippocampal sections from uninfected control, prion-infected mice with LV-control, and prion-infected mice with LV-RTN3. The graph bar shows quantification of the average number of neurons for each condition in the CA1 area of hippocampus. n = 3 mice (white bar), 5 mice (black bar), and 15 mice (orange bar). One-way ANOVA and Brown-Forsythe test with Tukey's post hoc analysis for multiple comparisons were used. ***$p < 0.001$. The scale bars represent 400 μm (top row) and 50 μm (bottom row).

(legend continued on next page)

pooled into subpolysomal or polysomal fractions, labeled, and hybridized to the arrays as described previously [10]. Microarray slides were scanned using a GenePix 4200B microarray scanner and GenePix Pro 6.0 software (Axon Instruments).

Analysis of Microarray Data

GenePix Pro 6.0 was used to quantify fluorescence intensities for individual spots on the microarray. All statistical analysis was performed in the statistical environment R, version 2.6.1, and the Limma package [46].

Northern Blot

Northern analysis was performed as described previously [10]. Visualization and quantification of northern blot analysis was performed using a Molecular Imager FX phosphoimager and ImageJ software.

RNA-Protein Complex Immunoprecipitation

Post-nuclear extracts were incubated with either anti-RBM3 antibody or immunoglobulin G (IgG)-coated protein G magnetic beads and processed as described in the Supplemental Information.

Reverse Transcription and qPCR

Reverse transcription was carried out using random primers and Superscript III Reverse Transcriptase (Invitrogen) according to manufacturer instructions. qPCR was carried out using SensiFAST SYBR Lo-ROX Kit (Bioline) according to manufacturer instructions. Primers used are in the Supplemental Information.

Prion Infection of Mice

As described previously [37], hemizygous tg37 mice of both sexes were inoculated with 1% brain homogenate of Chandler/RML prions at 3–6 weeks of age. Control mice received 1% normal brain homogenate.

Induction of Hypothermia in Mice

FVB wild-type (WT) or hemizygous Tg37 mice weighing more than 20 g were cooled using 5′ AMP as described [3]. Prion-infected mice were injected with lentiviruses at 2 w.p.i. and subsequently cooled at 3 and 4 w.p.i.

Lentiviruses and Mice Stereotaxic Surgery

GenTarget generated lentiviral plasmids. Viruses were injected stereotaxically into the CA1 region of the hippocampus as described [3]; additional information is provided in the Supplemental Information.

Burrowing

Burrowing was performed as described [3]. Briefly, mice were placed in a cage with a tube full of pellets, which they "burrowed." The extent of burrowing was assessed by the weight of pellets displaced in 2 hr.

Histology

Paraffin-embedded brains and pancreases were sectioned at 5 μm and stained with H&E as described [38, 39]. Neuronal counts were determined by quantifying NeuN-positive pyramidal CA1 neurons as described [39]. Synapses were counted in electron microscopy (EM) images of the stratum radiatum of the hippocampal CA1 region, blind. A synapse was defined as a structure with synaptic vesicles, synaptic cleft, and post-synaptic density, as described [3].

Computational Modeling

Translation elongation rates on human mRNAs were estimated using a published computational model [24]. The model was re-parameterized for the human decoding system using relative total tRNA abundances from [47] and a total tRNA concentration in HEK293 cells of 5.6 pg per cell, which was determined by comparing the staining intensity of the tRNA band in total RNA preps from HEK293 cells to the intensity of bands generated with known amounts of commercial yeast tRNA. Individual tRNA selection and translocation reactions were modeled in PRISM [48] using rate constants [49] and tRNA ratios for individual codons [24] as published.

ACCESSION NUMBERS

The accession number for the microarray data reported in this paper is ArrayExpress: E-MTAB-5437.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, five figures, and six tables and can be found with this article online at http://dx.doi.org/10.1016/j.cub.2017.01.047.

AUTHOR CONTRIBUTIONS

A.B., D.P., J.R.P.K., S.G., X.P., J.R., A.R., and D.V. conducted the experiments. R.V.S. and T.S. analyzed the data. T.v.d.H. generated the computational model and analyzed decoding speeds. A.E.W., M.B., G.R.M., and C.M.S. designed the experiments and wrote the paper.

ACKNOWLEDGMENTS

The work was funded by the BBSRC (BB/I019790/1; A.B., A.R., J.R., J.R.P.K., and R.V.S.), MRC (D.P., programme grant, 5TR50 to G.R.M.; S.G., programme funding, 5TR00 to A.E.W.), and Wellcome Trust (2010847/Z/16/Z; S.G., collaborative award to A.E.W., G.R.M., C.M.S., and T.v.d.H.).

Received: October 17, 2016
Revised: December 21, 2016
Accepted: January 23, 2017
Published: February 23, 2017

REFERENCES

1. Delhaye, C., Mahmoudi, M., and Waksman, R. (2012). Hypothermia therapy: neurological and cardiac benefits. J. Am. Coll. Cardiol. *59*, 197–210.
2. Yenari, M.A., and Han, H.S. (2012). Neuroprotective mechanisms of hypothermia in brain ischaemia. Nat. Rev. Neurosci. *13*, 267–278.
3. Peretti, D., Bastide, A., Radford, H., Verity, N., Molloy, C., Martin, M.G., Moreno, J.A., Steinert, J.R., Smith, T., Dinsdale, D., et al. (2015). RBM3 mediates structural plasticity and protective effects of cooling in neurodegeneration. Nature *518*, 236–239.
4. Wellmann, S., Bührer, C., Moderegger, E., Zelmer, A., Kirschner, R., Koehne, P., Fujita, J., and Seeger, K. (2004). Oxygen-regulated expression of the RNA-binding proteins RBM3 and CIRP by a HIF-1-independent mechanism. J. Cell Sci. *117*, 1785–1794.
5. Fedorov, V.B., Goropashnaya, A.V., Tøien, O., Stewart, N.C., Chang, C., Wang, H., Yan, J., Showe, L.C., Showe, M.K., and Barnes, B.M. (2011). Modulation of gene expression in heart and liver of hibernating black bears (Ursus americanus). BMC Genomics *12*, 171.
6. Liu, X., Wang, M., Chen, H., Guo, Y., Ma, F., Shi, F., Bi, Y., and Li, Y. (2013). Hypothermia protects the brain from transient global ischemia/reperfusion by attenuating endoplasmic reticulum response-induced apoptosis through CHOP. PLoS ONE *8*, e53431.

(E) LV-RTN3 significantly lengthened survival of prion-infected mice. Kaplan-Meier plot (orange; n = 18) compared to LV-control prion-infected mice (black; n = 15); t test; **p < 0.01.
(F) Induction of RTN3 fails at 6 w.p.i. in prion-diseased mice. Blue lines above the western blots are samples from cooled mice, whereas red lines denote control mice. Bar graphs show quantification of RTN3 levels relative to GAPDH at 4 and 6 w.p.i. prion disease, blue bars represent quantification from cooled mice, and red bars from control mice. n = 9 mice per condition. *p < 0.05. Data are mean ± SEM.

7. Chip, S., Zelmer, A., Ogunshola, O.O., Felderhoff-Mueser, U., Nitsch, C., Bührer, C., and Wellmann, S. (2011). The RNA-binding protein RBM3 is involved in hypothermia induced neuroprotection. Neurobiol. Dis. *43*, 388–396.
8. Tong, G., Endersfelder, S., Rosenthal, L.M., Wollersheim, S., Sauer, I.M., Bührer, C., Berger, F., and Schmitt, K.R. (2013). Effects of moderate and deep hypothermia on RNA-binding proteins RBM3 and CIRP expressions in murine hippocampal brain slices. Brain Res. *1504*, 74–84.
9. Smart, F., Aschrafi, A., Atkins, A., Owens, G.C., Pilotte, J., Cunningham, B.A., and Vanderklish, P.W. (2007). Two isoforms of the cold-inducible mRNA-binding protein RBM3 localize to dendrites and promote translation. J. Neurochem. *101*, 1367–1379.
10. Powley, I.R., Kondrashov, A., Young, L.A., Dobbyn, H.C., Hill, K., Cannell, I.G., Stoneley, M., Kong, Y.W., Cotes, J.A., Smith, G.C., et al. (2009). Translational reprogramming following UVB irradiation is mediated by DNA-PKcs and allows selective recruitment to the polysomes of mRNAs encoding DNA repair enzymes. Genes Dev. *23*, 1207–1220.
11. Somers, J., Wilson, L.A., Kilday, J.P., Horvilleur, E., Cannell, I.G., Pöyry, T.A., Cobbold, L.C., Kondrashov, A., Knight, J.R., Puget, S., et al. (2015). A common polymorphism in the 5′ UTR of ERCC5 creates an upstream ORF that confers resistance to platinum-based chemotherapy. Genes Dev. *29*, 1891–1896.
12. Knight, J.R., Bastide, A., Roobol, A., Roobol, J., Jackson, T.J., Utami, W., Barrett, D.A., Smales, C.M., and Willis, A.E. (2015). Eukaryotic elongation factor 2 kinase regulates the cold stress response by slowing translation elongation. Biochem. J. *465*, 227–238.
13. Roobol, A., Roobol, J., Bastide, A., Knight, J.R., Willis, A.E., and Smales, C.M. (2015). p58IPK is an inhibitor of the eIF2alpha kinase GCN2 and its localization and expression underpin protein synthesis and ER processing capacity. Biochem. J. *465*, 213–225.
14. Roobol, A., Carden, M.J., Newsam, R.J., and Smales, C.M. (2009). Biochemical insights into the mechanisms central to the response of mammalian cells to cold stress and subsequent rewarming. FEBS J. *276*, 286–302.
15. Roobol, A., Roobol, J., Carden, M.J., Bastide, A., Willis, A.E., Dunn, W.B., Goodacre, R., and Smales, C.M. (2011). ATR (ataxia telangiectasia mutated- and Rad3-related kinase) is activated by mild hypothermia in mammalian cells and subsequently activates p53. Biochem. J. *435*, 499–508.
16. Richter, J.D., and Coller, J. (2015). Pausing on polyribosomes: make way for elongation in translational control. Cell *163*, 292–300.
17. Faller, W.J., Jackson, T.J., Knight, J.R., Ridgway, R.A., Jamieson, T., Karim, S.A., Jones, C., Radulescu, S., Huels, D.J., Myant, K.B., et al. (2015). mTORC1-mediated translational elongation limits intestinal tumour initiation and growth. Nature *517*, 497–500.
18. Kumamaru, E., Kuo, C.H., Fujimoto, T., Kohama, K., Zeng, L.H., Taira, E., Tanaka, H., Toyoda, T., and Miki, N. (2004). Reticulon3 expression in rat optic and olfactory systems. Neurosci. Lett. *356*, 17–20.
19. Matsuzaki, F., Shirane, M., Matsumoto, M., and Nakayama, K.I. (2011). Protrudin serves as an adaptor molecule that connects KIF5 and its cargoes in vesicular transport during process formation. Mol. Biol. Cell *22*, 4602–4620.
20. Spriggs, K.A., Bushell, M., and Willis, A.E. (2010). Translational regulation of gene expression during conditions of cell stress. Mol. Cell *40*, 228–237.
21. Knight, J.R., Bastide, A., Peretti, D., Roobol, A., Roobol, J., Mallucci, G.R., Smales, C.M., and Willis, A.E. (2016). Cooling-induced SUMOylation of EXOSC10 down-regulates ribosome biogenesis. RNA *22*, 623–635.
22. Stepanenko, A.A., and Dmitrenko, V.V. (2015). HEK293 in cell biology and cancer research: phenotype, karyotype, tumorigenicity, and stress-induced genome-phenotype evolution. Gene *569*, 182–190.
23. Bushell, M., Stoneley, M., Kong, Y.W., Hamilton, T.L., Spriggs, K.A., Dobbyn, H.C., Qin, X., Sarnow, P., and Willis, A.E. (2006). Polypyrimidine tract binding protein regulates IRES-mediated gene expression during apoptosis. Mol. Cell *23*, 401–412.
24. Chu, D., and von der Haar, T. (2012). The architecture of eukaryotic translation. Nucleic Acids Res. *40*, 10098–10106.
25. Chu, D., Kazana, E., Bellanger, N., Singh, T., Tuite, M.F., and von der Haar, T. (2014). Translation elongation can control translation initiation on eukaryotic mRNAs. EMBO J. *33*, 21–34.
26. Chang, K., Seabold, G.K., Wang, C.Y., and Wenthold, R.J. (2010). Reticulon 3 is an interacting partner of the SALM family of adhesion molecules. J. Neurosci. Res. *88*, 266–274.
27. Laurén, J., Hu, F., Chin, J., Liao, J., Airaksinen, M.S., and Strittmatter, S.M. (2007). Characterization of myelin ligand complexes with neuronal Nogo-66 receptor family members. J. Biol. Chem. *282*, 5715–5725.
28. Deng, M., He, W., Tan, Y., Han, H., Hu, X., Xia, K., Zhang, Z., and Yan, R. (2013). Increased expression of reticulon 3 in neurons leads to reduced axonal transport of β site amyloid precursor protein-cleaving enzyme 1. J. Biol. Chem. *288*, 30236–30245.
29. Schneider-Poetsch, T., Ju, J., Eyler, D.E., Dang, Y., Bhat, S., Merrick, W.C., Green, R., Shen, B., and Liu, J.O. (2010). Inhibition of eukaryotic translation elongation by cycloheximide and lactimidomycin. Nat. Chem. Biol. *6*, 209–217.
30. Shah, P., Ding, Y., Niemczyk, M., Kudla, G., and Plotkin, J.B. (2013). Rate-limiting steps in yeast protein translation. Cell *153*, 1589–1601.
31. Sureban, S.M., Ramalingam, S., Natarajan, G., May, R., Subramaniam, D., Bishnupuri, K.S., Morrison, A.R., Dieckgraefe, B.K., Brackett, D.J., Postier, R.G., et al. (2008). Translation regulatory factor RBM3 is a proto-oncogene that prevents mitotic catastrophe. Oncogene *27*, 4544–4556.
32. Liu, Y., Hu, W., Murakawa, Y., Yin, J., Wang, G., Landthaler, M., and Yan, J. (2013). Cold-induced RNA-binding proteins regulate circadian gene expression by controlling alternative polyadenylation. Sci. Rep. *3*, 2054.
33. He, W., Lu, Y., Qahwash, I., Hu, X.Y., Chang, A., and Yan, R. (2004). Reticulon family members modulate BACE1 activity and amyloid-beta peptide generation. Nat. Med. *10*, 959–965.
34. Chen, Y., Huang, X., Zhang, Y.W., Rockenstein, E., Bu, G., Golde, T.E., Masliah, E., and Xu, H. (2012). Alzheimer's β-secretase (BACE1) regulates the cAMP/PKA/CREB pathway independently of β-amyloid. J. Neurosci. *32*, 11390–11395.
35. Zhang, J., Kaasik, K., Blackburn, M.R., and Lee, C.C. (2006). Constant darkness is a circadian metabolic signal in mammals. Nature *439*, 340–343.
36. Mallucci, G.R., Ratté, S., Asante, E.A., Linehan, J., Gowland, I., Jefferys, J.G.R., and Collinge, J. (2002). Post-natal knockout of prion protein alters hippocampal CA1 properties, but does not result in neurodegeneration. EMBO J. *21*, 202–210.
37. Mallucci, G., Dickinson, A., Linehan, J., Klöhn, P.C., Brandner, S., and Collinge, J. (2003). Depleting neuronal PrP in prion infection prevents disease and reverses spongiosis. Science *302*, 871–874.
38. Moreno, J.A., Halliday, M., Molloy, C., Radford, H., Verity, N., Axten, J.M., Ortori, C.A., Willis, A.E., Fischer, P.M., Barrett, D.A., and Mallucci, G.R. (2013). Oral treatment targeting the unfolded protein response prevents neurodegeneration and clinical disease in prion-infected mice. Sci. Transl. Med. *5*, 206ra138.
39. Moreno, J.A., Radford, H., Peretti, D., Steinert, J.R., Verity, N., Martin, M.G., Halliday, M., Morgan, J., Dinsdale, D., Ortori, C.A., et al. (2012). Sustained translational repression by eIF2α-P mediates prion neurodegeneration. Nature *485*, 507–511.
40. Mallucci, G.R., White, M.D., Farmer, M., Dickinson, A., Khatun, H., Powell, A.D., Brandner, S., Jefferys, J.G., and Collinge, J. (2007). Targeting cellular prion protein reverses early cognitive deficits and neurophysiological dysfunction in prion-infected mice. Neuron *53*, 325–335.
41. White, M.D., Farmer, M., Mirabile, I., Brandner, S., Collinge, J., and Mallucci, G.R. (2008). Single treatment with RNAi against prion protein rescues early neuronal dysfunction and prolongs survival in mice with prion disease. Proc. Natl. Acad. Sci. USA *105*, 10238–10243.

42. Daniels, I.S., Zhang, J., O'Brien, W.G., 3rd, Tao, Z., Miki, T., Zhao, Z., Blackburn, M.R., and Lee, C.C. (2010). A role of erythrocytes in adenosine monophosphate initiation of hypometabolism in mammals. J. Biol. Chem. *285*, 20716–20723.

43. van der Worp, H.B., Sena, E.S., Donnan, G.A., Howells, D.W., and Macleod, M.R. (2007). Hypothermia in animal models of acute ischaemic stroke: a systematic review and meta-analysis. Brain *130*, 3063–3074.

44. Jena, A.B., Romley, J.A., Newton-Cheh, C., and Noseworthy, P. (2012). Therapeutic hypothermia for cardiac arrest: real-world utilization trends and hospital mortality. J. Hosp. Med. *7*, 684–689.

45. West, M.J., Stoneley, M., and Willis, A.E. (1998). Translational induction of the c-myc oncogene via activation of the FRAP/TOR signalling pathway. Oncogene *17*, 769–780.

46. Smyth, G.K. (2004). Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Stat. Appl. Genet. Mol. Biol. *3*, Article3.

47. Dittmar, K.A., Goodenbour, J.M., and Pan, T. (2006). Tissue-specific differences in human transfer RNA expression. PLoS Genet. *2*, e221.

48. Kwiatkowska, M., Norman, G., and Parker, D. (2011). PRISM 4.0: verification of probabilistic real-time systems. In Computer Aided Verification, *Volume 6806*, G. Gopalakrishnan, and S. Qadeer, eds. (Springer), pp. 585–591.

49. Fluitt, A., Pienaar, E., and Viljoen, H. (2007). Ribosome kinetics and aa-tRNA competition determine rate and fidelity of peptide synthesis. Comput. Biol. Chem. *31*, 335–346.