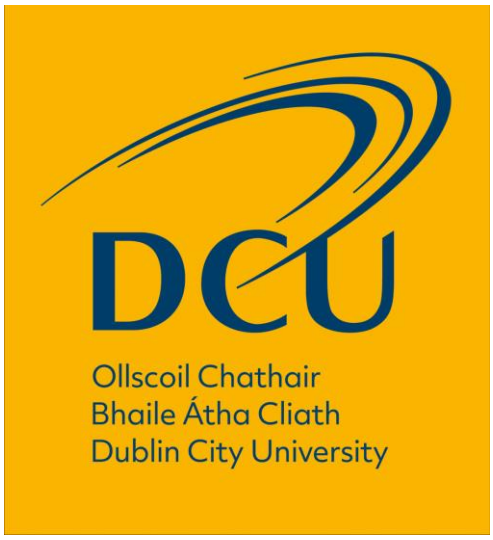
RNA Manipulations to Improve Recombinant Protein Production in Chinese Hamster Ovary Cells

A thesis submitted for the award of Ph.D.



By

Alan Costello, B.Sc. (Hons)

National Institute for Cellular Biotechnology,
School of Biotechnology

The work described herein was conducted under the supervision of
Prof. Martin Clynes & Dr Padraig Doolan
with secondary external supervision from Prof. Niall Barron
National Institute for Bioprocessing Research and Training &
School of Chemical Engineering, University College Dublin.

January 2019

DECLARATION

I hereby certify that this material, which I now submit for assessment on the programme of study leading to the award of PhD is entirely my own work, and that I have exercised reasonable care to ensure that the work is original, and does not to the best of my knowledge breach any law of copyright, and has not been taken from the work of others save and to the extent that such work has been cited and acknowledged within the text of my work.

Signed: _____ ID No.: 10331815 Date: _____

THESIS STRUCTURE

Chapter	Title	Publication status	Authors	Contribution
-	Preface to published chapters	-	-	-
1	Ultra-deep next generation mitochondrial genome sequencing reveals widespread heteroplasmy in Chinese hamster ovary cells	*Metabolic Engineering* Manuscript Accepted, Published.	Kelly, P.S., Clarke, C., **Costello, A.**, Monger, C., Meiller, J., Dhiman, H., Borth, N., Betenbaugh, M.J., Clynes, M. & Barron, N.	Contributions to the initial wet-lab work.
2	Depletion of Endogenous miRNA-378-3p Increases Peak Cell Density of CHO DP12 Cells and is Correlated with Elevated Levels of Ubiquitin Carboxyl-Terminal Hydrolase 14	*Journal of Biotechnology* Manuscript Accepted, Published.	**Costello, A.**, Coleman, O., Lao, T. N., Henry, M., Meleady, P., Barron, N., Clynes, M.	Co-author and primary contributor to all aspects of study excluding LC-MS/ MS sections.
3	Conditional knockdown of endogenous microRNAs in CHO cells using TET-ON-SanDI sponge vectors	*Methods in Molecular Biology* Manuscript Accepted, Published.	**Costello, A.**, Lao, T.N., Clynes, M. & Barron, N.	Manuscript primary author. Primary contributor to experimental design and execution.
4	Leaky Expression of the TET-On System Hinders Control of Endogenous miRNA Abundance	*Biotechnology Journal* Manuscript Accepted, Published.	**Costello, A.**, Lao, N.T., Gallagher, C., Capella Roca, B., Julius, L.A., Suda, S., Ducrée, J., King, D., Wagner, R., Barron, N. & Clynes, M.	Manuscript primary author. Primary contributor to experimental design and execution.
5	Continuous Translation of Circularized mRNA Improves Recombinant Protein Titer	*Metabolic Engineering* Revised manuscript submitted.	**Costello, A.**, Lao, T.N., Barron, N. & Clynes, M.	Manuscript primary author. Primary contributor to experimental design and execution.
6	Sentencing Ribosomes to Tediously Long Messages	*Trends in Biotechnology* Manuscript in draft.	**Costello, A.**, Lao, T.N., Barron, N. & Clynes, M.	Manuscript primary author.
7	Improved Yield of rhEPO in CHO Cells with Synthetic 5' UTR	*Biotechnology Letters* Manuscript Accepted, pending minor revisions.	**Costello, A.**, Lao, T.N., Barron, N. & Clynes, M.	Manuscript primary author. Primary contributor to experimental design and execution.

The role of the candidate as described here is correct.

Signed: _____ Date: _____ (Student)

Signed: _____ Date: _____ (Principal Supervisor)

Signed: _____ Date: _____ (Secondary Supervisor)

Signed: _____ Date: _____ (Secondary External Supervisor)

“Nothing is work unless you'd rather be doing something else.”

- George S. Hallas

ACKNOWLEDGEMENTS

Let's start at the very beginning, after all it is a very good place to start! Thank you to Prof. Martin Clynes for taking a chance on a raw BT-grad from North County Dublin (me) and providing an amazing opportunity to become part of the NICB. Martin has been a great mentor to me over the years in terms of lab work, experimental design and writing. What makes Martin special though is the little nuggets of wisdom he drops to challenge your thinking and rock the boat, always doing so with a touch of mirth. The line "to your objection that I am being inconsistent I can only say what some supposedly wise person said once, consistency is the curse of small minds", from a recent email springs to mind. Lastly, thank you Martin for the confidence you've given me to have in myself and my abilities.

I'd like to thank the molecular biology overload of the CHO group himself, Prof. Niall Barron, for always listening to my unorthodox ideas, encouraging me and embracing the pun culture in the lab. I'm sure I've driven him round the bend on more than one occasion. I feel very fortunate to have had both Niall and Martin to lean on throughout my studies. The meetings and email chains where they had diametrically opposed perspectives on the same piece of work, "is it muck or magic", were the most memorable and enjoyable.

The mentorship I received during my time in NICB was not limited to that of my supervisors. I had the help of some great Postdocs, especially early on (when I was afraid to ask my supervisor). Thanks to Dr. Paul Kelly for showing me the ropes and always lending an ear when I wanted/ needed to bitch about the cells not behaving as I had hoped. Also, a special thanks to Dr. Nga Lao for being an incredible teacher and crutch during my time in NICB. I will miss the random things you brought up in conversation in the molecular lab.

Thanks to all the people who keep NICB running behind the scenes; Mairead, Gillian, Justine and Mick (cough… Dr. Henry) in particular. The building does not function without them.

Thanks to all the lunch crew(s) over the years; Orla (my work sister), Gemma, Shane, Kevin, Laura, Edel, Andrew, Mark and Niamh. Shane's ludicrous anecdotes always kept our minds off the lab for an hour a day. Also, thanks to the CHO and proteomic group members for keeping the office and lab interesting. I can't count the number of times I found myself in a good debate late into the evening with Ric. Thanks to my other science family, the BRS! For all the trips, experiences and friendships.

Finally, I want to thank my parents Sandra and Liam. I would not have been able to get through my studies without your incredible support and patience. I promise I'll try to pay you back when I start my "real" job and hopefully this book looks good on one of the side tables in the sitting room.

RESEARCH OUTPUTS

SCIENTIFIC TALKS

"Rolling Translation of Circularized Open Reading Frames; Fooling the Ribosome", 10th Annual School of Biotechnology Research Day, DCU, Dublin, Ireland, 26th Jan 2018.

"Rolling Translation of Circularized Open Reading Frames; Fooling the Ribosome", 28th Annual ESACT-UK Meeting, Hilton Leeds City, United Kingdom, 10th-11th Jan 2018.

"Rolling Translation of Circularized Open Reading Frames; Fooling the Ribosome", BioProduction Congress, Convention Centre Dublin, Dublin, Ireland, 17th-18th Oct 2017.

"Control of miRNA in Chinese Hamster Ovary Cells", 9th Annual School of Biotechnology Research Day, DCU, Dublin, Ireland, 27th Jan 2017.

POSTER PRESENTATIONS

"Rolling Translation of Circularized Open Reading Frames; Fooling the Ribosome", Cell Culture Engineering XVI, Tampa, Florida, USA, 6-11th May 2018.

"Rolling Translation of Circularized Open Reading Frames; Fooling the Ribosome", Biopharma Ambition, Dublin Castle, Dublin, Ireland, 21st-22nd Feb 2018.

"Rolling Translation of Circularized Open Reading Frames; Fooling the Ribosome", 13th PEACe (Protein Expression in Animal Cells), Valencia, Spain, 24th-28th Sep 2018.

"Process Development by Conditional Manipulation of Endogenous CHO Cell miRNA", 25th ESACT Meeting, EPFL, Lausanne, Switzerland, May 14th-17th 2017.

"CHO Cell Engineering with miRNA Sponges; Endogenous and Otherwise", 8th Annual School of Biotechnology Research Day, DCU, Dublin, Ireland, 29th Jan 2016.

TABLE OF CONTENTS

ABBREVIATIONS

AGO – Argonaut

ATCC – American Tissue Culture Collection

AQ – Absolute quantification

BSJ – Back-splice junction

ceRNA – Competitive endogenous RNA

circRNA – Circular RNA

CHO – Chinese Hamster Ovary

DE – Differential Expression

DNA – Deoxyribonucleic Acid

dsDNA – Double-stranded DNA

ELISA – Enzyme-Linked immunosorbent assay

EPO – Erythropoietin

ER – Endoplasmic reticulum

GOI – Gene of Interest

HCP – Host cell protein

HPLC – High performance liquid chromatography

IRES – Internal Ribosome Entry Site

LC-MS/ MS – Liquid chromatography/Mass spectrometry

LDC – Limited-dilution single cell cloning

lncRNA – Long non-coding RNA

MBS – MicroRNA Binding Site

mRNA – Messenger RNA

miRNA – microRNA

MRE – MicroRNA Response Element

NC – Negative control

ncRNA – Non-coding RNA

nt – Nucleotide

Pre-mRNA – Preliminary mRNA

Pre-miRNA – Preliminary microRNA

Pri-miRNA – Primary microRNA

PVA – Polyvinyl Alcohol

qPCR – Quantitative Polymerase chain reaction

RISC – RNA-Induced Silencing Complex

RNA – Ribonucleic Acid

RNAi – RNA Interference

RNA pol – RNA polymerase

rpm – Revolutions per minute

RT – Reverse transcription

RQ – Relative quantification

SA – Splice acceptor

SD – Splice donor

SEAP – Secreted Alkaline Phosphatase

shRNA – Short Hairpin RNA

spg – Sponge

ssRNA – Single-stranded RNA

TMB – 3,3',5,5' – Tetramethylbenzidine

tTA – Tetracycline responsive trans-activator

UPR – Unfolded protein response

UTR – Untranslated region

Alan Costello

RNA MANIPULATIONS TO IMPROVE RECOMBINANT PROTEIN PRODUCTION IN CHINESE HAMSTER OVARY CELLS

ABSTRACT

To date the major advances in biopharmaceutical production from Chinese Hamster Ovary (CHO) cells has come from culture vessel development and refining growth media formulation. However, it is believed that there is still room for improvement in the advent of media and process optimisation reaching a plateau. Publication of the first CHO-K1 genome in 2011 has resulted in a shift to focus efforts towards engineering of the CHO cell itself. The use of microRNA (miRNA) manipulation is one approach which has been fruitful in boosting cell specific productivity but at a cost of cellular growth rate, ultimately diminishing volumetric titre. Stable transgenic lines were created in a model IgG producing CHO cell line for six prioritized miRNAs robustly associated with growth rate. Depletion of miR-378-3p via miRNA-sponge decoy, was found to significantly enhance cell growth by ~40% but reduced specific productivity. To overcome the inherent antagonistic trade-off of cell energy driven towards growth or recombinant protein production, a tetracycline inducible (TET-ON) system was employed to manipulate the endogenous levels of different miRNAs at defined stages of culture. Absolute quantification of target miRNA and miRNA-sponge mRNA revealed a limitation of current synthetic sponge-decoy designs in the control of endogenous miRNA species. In efforts to re-engineer synthetic sponge-decoys we encountered circular RNAs (circRNA). circRNA is a closed loop structure and naturally occurring circRNAs demonstrate greater stability over their linear mRNA counterparts. The potential of circular transgene open reading frames (ORF) as a means of improving translational output from an RNA molecule was investigated using a model recombinant glycoprotein, Erythropoietin (EPO). Artificial circularization of gene ORF was achieved by splice signals, made proximal via complementary intronic flanking sequence. Rolling circle translation was explored by the removal of the stop codon from the EPO ORF. To ensure homogenous protein product, the addition of a self-cleavage 2A peptide to the infinite ORF was also assessed. These two constructs exhibited a significant increase in cell specific and volumetric productivity. Additional transgene mRNA engineering was investigated by encoding putative methylation sites in the 5’ un-translated region (UTR) of a model protein therapeutic.

Preface to Published Chapters

This introductory section will provide a general outline to biotherapeutic production and serve to put each of the subsequent published chapters in the context of existing literature. Justifications for the studies themselves, the import and novelty of these works with perspective on their immediate and potential future impact on the field will be discussed. The document is structured with each publication appearing as a separate chapter. Personal contributions to each body of work is detailed at the beginning of each chapter.

A BRIEF HISTORY OF RECOMBINANT PROTEIN THERAPEUTICS

In the late 1960s and early 1970s, Stanley Cohen and Herbert Boyer pioneered efforts in cross-species genetics. Conferring antibiotic resistance by introduction of DNA from one strain of *Escherichia coli* (*E. coli*) to another (Cohen et al., 1973) followed by expression of protein from eukaryotic DNA in a bacterial host (Morrow et al., 1974) laying the foundations of recombinant protein production industry. This technology opened the doors for the development and manufacturing of new medicines. Human insulin (Humulin, Genentech) was the first licensed recombinant protein therapeutic released in 1982 (Butler, 2005).

The simplicity of insulin's structure enabled production to be developed in *E. coli*. However, it quickly became apparent that more complex recombinant therapeutic molecules would require eukaryotic post-translational machinery. In 1986 human tissue plasminogen activator (tPA, Activase, Genentech) became the first therapeutic protein produced in mammalian cells to obtain approval. It is estimated that approximately 70% of all recombinant protein therapeutics are now produced in mammalian cells (Wurm, 2004). Today, the recombinant protein therapeutics or "biologics" market encompasses' a variety of products; monoclonal antibodies (mAbs), hormones, growth factors, fusion proteins, cytokines, blood factors, recombinant vaccines and anti-coagulants (Aggarwal, 2012). mAbs dominate the world of biologics, accounting for ~38% of FDA-approved modalities (Aggarwal, 2012). Protein therapeutics' enabling sister discipline, protein engineering, have seen the evolution of natural mAb formats over the years to become more artificial in structure (Carter, 2011). The advent of expanding panels of bi, tri and tetra-specific antibodies (Klein et al., 2016) along with Fc-fusion proteins and non-mAb formats (Vazquez-Lombardi et al., 2015) sees the field moving in a more synthetic approach. With this comes new challenges in manufacturing of such unnatural products.

Analogous to the discovery of new protein-based drugs has been the development and refinement of robust manufacturing platforms. Protein expression systems used today include; plant, insect, bacterial, yeast and mammalian and cell-free hosts. Cultivated mammalian cells remain the preferred system for production of complex human protein therapeutics (Table 1). Chinese hamster ovary (CHO) derived cell lines dominate the domain of therapeutic protein production. Their dominance is largely attributed to the ability of CHO cells to be cultivated in single cell suspension, resistance to viral infection and human-like post-translational modifications (Dumount et al., 2015).

Table 1 Recombinant protein therapeutics produced in mammalian cells.

This table has been adapted from (Dumont et al., 2015) to include indications of the therapeutic proteins produced in mammalian cell systems.

Host	Product	Indication	FDA	EMA
CHO	Adalimumab	Rheumatoid Arthritis, Psoriatic Arthritis, Ankylosing Spondylitis	Approved	Approved
	Alemtuzumab	Chronic Lymphocytic Leukaemia (CLL), Cutaneous T-Cell Lymphoma (CTCL) And T-Cell Lymphoma	Approved	-
	Bevacizumab	Metastatic Cervical Cancer	Approved	Approved
	Brentuximab Vedotin	Stage 3 Or 4 Classical Hodgkin Lymphoma (Chl)	Approved	Approved
	Denosumab	Osteoporosis	Approved	Approved
	Golimumab	Rheumatoid Arthritis, Psoriatic Arthritis, Ulcerative Colitis	Approved	Approved
	Ibritumomab Tiuxetan	Follicular B-Cell NHL	Approved	Approved
	Ipilimumab	Metastatic Melanoma	Approved	Approved
	Obinutuzumab	Hepatitis B Virus, Progressive Multifocal Leukoencephalopathy	Approved	Approved
	Omalizumab	Reduce Sensitivity to Allergens	Approved	Approved
	Panitumumab	Metastatic Colorectal Cancer	Approved	Approved
	Pertuzumab	Used in Combination with Trastuzumab and Docetaxel to treat Metastatic HER2-Positive Breast Cancer	Approved	Approved
	Rituximab	Rheumatoid Arthritis, B-Cell Non-Hodgkin's Lymphoma	Approved	Approved
	Siltuximab	Multicentric Castleman's Disease	Approved	Approved
	Tocilizumab	Rheumatoid Arthritis	Approved	Approved
	Trastuzumab	HER2-Positive Breast Cancer	Approved	Approved
	Vedolizumab	Crohn's Disease, Ulcerative Colitis	Approved	Approved
	Ado-Trastuzumab Emtansine	HER2-Positive Breast Cancer	Approved	Approved
	Ustekinumab	Crohn's Disease	Approved	Approved
	Darbepoetin Alfa	Chronic Kidney Disease	Approved	Approved
	Interferon Beta-1a	Multiple Sclerosis	Approved	Approved
	Epoetin Alfa	Chronic Kidney Disease	Approved	Approved
	Epoetin Beta	Anemia	-	-
	Epoetin Theta	Symptomatic Anaemia Associated with Chronic Renal Failure	-	Approved
	Agalsidase Beta	Fabry Disease	Approved	Approved
	Alglucosidase Alfa	Pompe Disease	Approved	Approved

Host	Product	Indication	FDA	EMA
	Alteplase	Acute Myocardial Infarction	Approved	Approved
	Elosulfase Alfa	Mucopolysaccharidosis Type IVA	Approved	-
	Imiglucerase	Type 1 Gaucher Disease	Approved	-
	Laronidase	Mucopolysaccharidosis I	Approved	-
	Tenecteplase	Acute Myocardial Infarction	Approved	Approved
	Abatacept	Rheumatoid Arthritis	Approved	Approved
	Aflibercept	Neovascular (Wet) Age-Related Macular Degeneration (AMD). Macular Edema Following Retinal Vein Occlusion (RVO). Diabetic Macular Edema (DME). Diabetic Retinopathy (DR) In Patients with DME	Approved	Approved
	Alefacept	Chronic Plaque Psoriasis	Approved	Approved
	Belatacept	Organ Rejection Prophylaxis in Patients Receiving A Kidney Transplant, In Combination with Basiliximab Induction, Mycophenolate Mofetil, and Corticosteroids	Approved	Approved
	Etanercept	Polyarticular Juvenile Idiopathic Arthritis	Approved	-
	Rilonacept	Familial Cold Autoinflammatory Syndrome, Muckle-Wells Syndrome	Approved	Approved
	Ziv-Aflibercept	Metastatic Colorectal Cancer	Approved	Approved
	Choriogonadotropin Alfa	Female Infertility	Approved	-
	Follitropin Alfa	Female Infertility	Approved	Approved
	Follitropin Beta	Infertility	Approved	Approved
	Luteinizing Hormone	Female Infertility	Approved	Approved
	Thyrotropin Alfa	Thyroid Cancer	Approved	Approved
	Factor VIII	Hemophilia A	Approved	Approved
	Factor IX	Hemophilia B	Approved	Approved
NS0	Belimumab	Systemic Lupus Erythematosus	Approved	Approved
	Natalizumab	Multiple Sclerosis	Approved	Approved
	Ofatumumab	Chronic Lymphocytic Leukemia	Approved	Approved
	Palivizumab	Respiratory Syncytial Virus	Approved	Approved
	Ramucirumab	Metastatic Colorectal Cancer	Approved	-
Sp2/0	Abciximab	Percutaneous Coronary Intervention	Approved	-
	Basiliximab	Renal Transplant Failure	Approved	Approved
	Canakinumab	Familial Cold Autoinflammatory Syndrome, Muckle-Wells Syndrome	Approved	Approved
	Cetuximab	Colorectal Cancer	Approved	Approved
	Infliximab	Rheumatoid Arthritis	Approved	Approved
BHK	Factor Viia	Hemohilia A Or B	Approved	Approved
	Factor VIII	Hemophilia A	Approved	Approved
C127	Somatropin	Growth Hormone Deficiency	Approved	Approved

Host	Product	Indication	FDA	EMA
HEK293	Drotrecogin Alfa	Severe Septicemia/Septic Shock	Approved	Approved
	Rfviiifc	Hemophilia A	Approved	Approved
	Rfixfc	Hemophilia B	Approved	-
	Dulaglutide	Type 2 Diabetes	Approved	Approved
	Human-Cl Rhfviii	Hemophilia A	Approved	Approved
HT-1080	Agalsidase Alfa	Fabry Disease	-	Approved
	Epoetin Delta	Anemia Secondary to Chronic Renal Failure	-	Approved
	Idursulfase	Hunter Syndrome	Approved	Approved
	Velaglucerase Alfa	Type 1 Gaucher Disease	Approved	Approved
PER.C6	CL184	Rabies Virus Infection	Approved	-
	MOR103	Rheumatoid Arthritis, Multiple Sclerosis	Approved	Approved

INDUSTRIAL CHO CELL LINE DEVELOPMENT

The fundamental pipeline undertaken by the biopharmaceutical industry to develop producer CHO cell lines has seen little change in the last thirty years. Although the steps remain the same, technological advances at each stage have garnered much research, in the pursuit of a faster, cheaper, and more reliable processes. Two streams feed the generation of new CHO producer cell lines (Figure 1). Drug and host cell line development come together with the transfer of plasmid DNA encoding the product of interest. Transgene delivery can be achieved by viral transduction or non-viral transfection by one of the following; calcium phosphate, electroporation, lipofection or polymer-mediated (Norton & Pachuk, 2003). There may be multiple rounds of gene delivery to ensure high-producing cells.

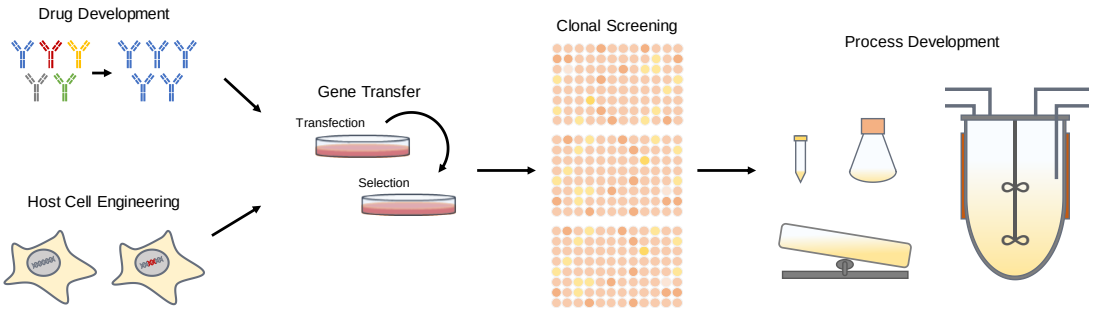


Figure 1. Schematic representation of recombinant CHO cell line development.
Genetically engineered CHO cell lines are used to make stable transgenic mixed pools expressing the drug of choice. Clonal isolation and screening are used to isolate the top candidates. Subsequent process development is required to maximize product output and production efficiency.

Transgene Expression

Expression vectors for recombinant protein production use strong viral or mammalian promoters and enhancers to drive transcription of the product gene (Ringold et al., 1981; Gopalkrishnan et al., 1999). The coding sequence (CDS) of recombinant therapeutic genes are often altered to enhance expression in the choice host, specifically to remove rare codons. The abundance of specific transfer RNAs (tRNA) differs between organisms resulting in codon bias (Bulmer, 1991). Codon adaptation is used to improve the translation elongation efficiency of the transgene in a given host (Makrides, 1999). Shah et al., (2013) illustrated the effects of potential ribosome pausing on transcripts with low (poor) and high (good) codon adaptation in yeast (Figure 2). As recombinant genes do not contain introns, intronic sequence is often incorporated between the promoter and CDS to improve exportation of transgene mRNA from the nucleus to the cytoplasm (Wurm, 2004).

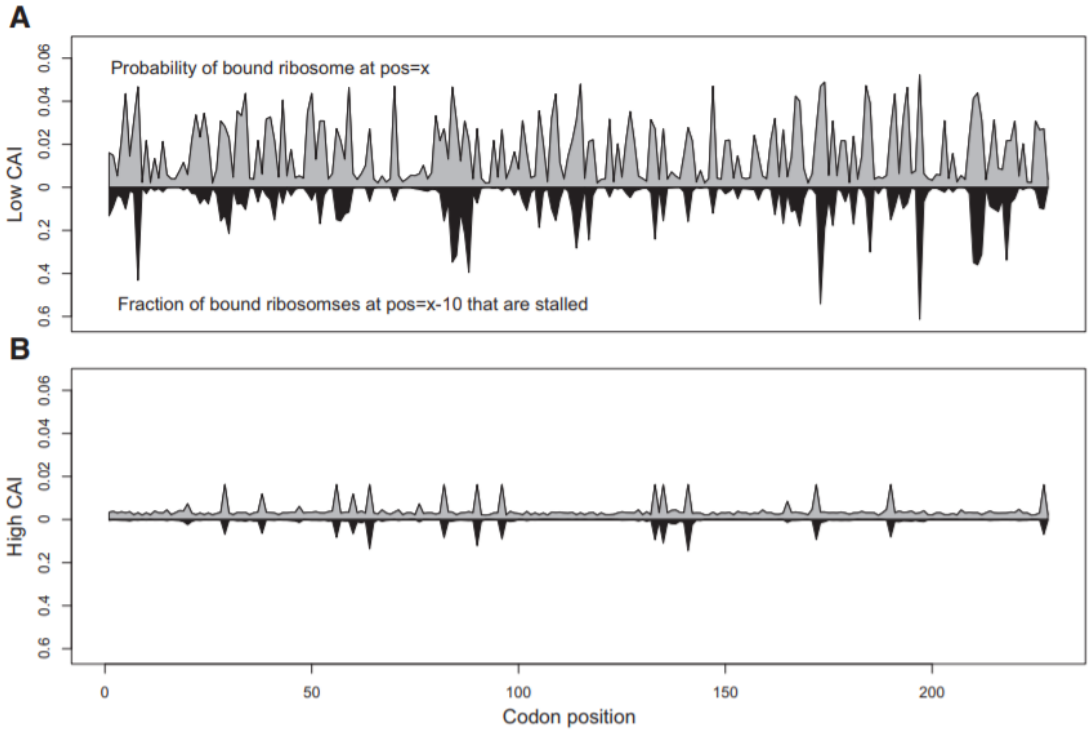


Figure 2. Codon Usage and Ribosomal Pausing and Stalling. (A and B) The influence of codon usage and amino acid sequence on ribosomal pausing and stalling for a simulated transgene with either (A) low or (B) high codon adaptation, expressed at 50% transcript abundance. Gray bars indicate the probability of finding a ribosome bound at a given codon position x, and black bars indicate the probability of finding a ribosome stalled at position x-10 (i.e., a ribosome whose further elongation is obstructed by another ribosome). High codon adaptation reduces both ribosome density and ribosome interference by utilization of the tRNAs in highest abundance. Figure and figure legend are taken from (Shah et al., 2013).

For pilot scale expression of new drugs 10-100 L scale transient expression is routinely used. In the case of drug characterisation, this provides a relatively fast and cost-effective means of generating mg-g quantities of product while negating any need for cell line or process development (Wurm and Bernard, 1999). However, for large scale production of product with a high degree of homology and purity, a stable expression system is still the platform of choice. To ensure stability of transgene expression vectors are equipped with a selectable marker. The selection marker provides transfected cells with the ability to grow in media containing a toxic compound or lacking an essential nutrient. A variety of selection systems have been developed to generate stable producer cell lines including resistance to antibiotics; neomycin, hygromycin and puromycin (Wurm, 2004). Metabolic selection markers are more commonly used for large scale production of biologics. The DHFR system augments expression of the product gene by co-expression of the DHFR gene. DHFR activity is blocked by methotrexate (MTX), cells expressing DHFR to a level exceeding a threshold determined by the MTX concentration used will proliferate. Amplification of transgene expression is commonly done by incrementally increasing the MTX concentration (Noh et al., 2018). The GS system is another metabolic selection system. GS, glutamine synthetase, synthesizes the unstable amino acid, glutamine, from glutamate and ammonia. GS selection is achieved via the inhibition of GS activity by methionine sulfoximine (MSX) (Matasci et al., 2008). The GS system is gaining popularity as it utilizes accumulating, growth inhibitory, ammonia from the culture (Costa et al., 2010). Novel selection methods based on essential vitamins are now being assessed also (Mermod, 2016).

Strategies to confer stable transgene expression with a selection marker can see both the product gene and marker present on a polycistronic mRNA with an internal ribosome entry site (IRES) between the open reading frames. Alternatively, the genes can be driven off separate promoters on the same plasmid. Using separate promoters, selection for the marker gene does not ensure a stable ratio of product gene expression as it does not account for expression vector integrity (Ng et al., 2010). Encoding the selection marker downstream in a polycistronic mRNA, its linked expression is dependent on transcription of upstream product genes (Lai et al., 2013). The site of integration affects the transgene expression as much as, if not more than, the expression cassette itself. Non-homologous recombination or "random integration" is the simplest and most common approach. Site directed homologous re-combination events are achievable though the use of the CRE/Lox P system (Wilson & Kola, 2001). More recently the genome editing phenomenon CRISPR/Cas9 has been utilised in this fashion Lee et al., (2016), outline a method for targeted gene-integration into specific loci in CHO cells, using the homology-directed repair (HDR) pathway (Ran et al., 2013). Identification of transcription hotspots is a prerequisite for the use of HDR targeted integration events. Investigation of active euchromatin with minimal transgene

silencing through histone hypoacetylation and methylation (Schultz et al., 2002; Mutskov et al., 2004; Moniz et al., 2013; Paredes et al., 2013) is an area of ongoing research.

Clonal Isolation and Screening

Clonal isolation and expansion of cells that survive the selection process is still largely done by limited dilution cloning (LDC) (Browne & Al-Rueai, 2007). LDC is a process wherein a cell suspension is diluted to a concentration that will give a probability of <1 cell per well of a micro-titer plate. Nowadays LDC is automated with the use of large liquid handlers, but the principle remains the same. Multiple rounds of LDC is commonly employed and this must be coupled with an image to prove clonality. This process of clonal isolation, propagation and functional vetting is the longest and most labour-intensive part of industrial cell line development. The use of cell sorting to expedite this process has some protential (Kacmar & Srienc, 2005) and novel fully automated micro-fluidic devices are growing in popularity, such as the Berkeley Lights system.

Bioprocess Variables

Top performing clones are scaled from micro-titer plates to 1-10 L for engineering runs. At this stage many process variables for example; stir speed, temperature, DO_2, off-gas analysis, can be manipulated to achieve an optimal bioprocess. The titer and product quality attributes, post-translational modifications (PTM); glycosylation, phosphorylation, lipidation, of each clone is evaluated at this stage. Optimal feeding strategies are achieved by metabolic analysis of each stage of the cell cycle. The carbon source, amino acid and lactate, to name a few, demands of these cells is not the same in exponential growth as in stationary phase (Carrillo-Cocom et al., 2014). Evolution of process analytical technology (PAT), has aided this greatly at both pilot and industrial scale (Teixeira et al., 2009). PAT involves both in-line, in situ, monitoring of culture variables such as; temperature, pH, DO_2, CO_2 and turbidity. Additional off-line measurement of metabolites and cell density are also run at pre-determined regular intervals. Real-time in-line measurement of gas exchange can provide insight to the metabolic switch from growth to production. The oxygen uptake rate (OUR) and carbon dioxide evolution rate (CER) can be used as indicators of cellular metabolism. Real time assessment of the bioprocess enables the manufacturer to identify issues or irregularities and act on them, saving any potential loss in product. At the development stage these technologies are becoming more routinely used. Sensors have been developed for on-line monitoring of cell morphology (Sarró et al., 2015). Along with a number of colorimetric assays for high-throughput of metabolites. The PTM profile of biologics must be within a defined range approved by a regulatory authority. If it is not possible to achieve sufficient titer or maintain pre-determined PTM profile with chemical engineering or media optimization, then more clones must be evaluated. This could revert as far back as generating a new transfected pool, significantly impacting the timeline of production.

ENGINEERING CHO CELL FACTORIES

The first CHO cell line was generated from a spontaneously immortalised population of fibroblast cells from a cultured Chinese Hamster ovarian cells in 1956 (Puck et al., 1958). The ease of cultivating CHO cells in suspension made them popular in the early days of mammalian cell culture and made them the mammalian equivalent to *E. coli*. The low chromosome number (2n=22) of CHO cells also made these cells a good model organism for early cytogenetics studies (Puck et al., 1958). Although CHO cells have very few chromosomes, they have a similar DNA content to human cells. Sequencing the CHO-K1 genome revealed 2.45 Gb encoding ~25k genes over 22 chromosomes, whereas the human genome is 3 Gb encoding ~25k genes over 46 chromosomes (Xu et al., 2011). This large genome and low number of chromosomes suggests CHO cells have a very efficient means of DNA breaking and recombination.

Antagonism between growth and productivity behaviours has been an obstacle for producing large quantities of recombinant protein in living cells (Fussenegger, 2001), manifesting itself as an inherent trade-off in cellular metabolism between growth and recombinant protein production (Hefzi et al., 2016). To minimize the effects of this antagonistic relationship biphasic cultures have been developed. The classic example of biphasic culture is temperature shift, wherein the cultivation temperature is reduced from 37°C to 30-33°C (Kaufmann et al., 1998). Independent studies have shown that cultivating CHO cells at low temperature significantly reduces cell growth rate but increases cell specific productivity and culture longevity (Kaufmann et al., 1998; Trummer et al., 2006). Using a biphasic culture format, the process is split into a biomass accumulation (37°C) phase followed by a cold-induced (30-33°C) low proliferative production phase. This works in principle but constitutive production of recombinant proteins impairing the growth phase remains a limitation.

Single gene functional genomic strategies have been employed for decades now in attempts to engineer CHO cell; growth (Doolan et al., 2010), metabolism (Chong et al., 2010; Fogolin et al., 2004; Kim and Lee, 2007; Le et al., 2013; Park et al., 2000; Tabuchi and Sugiyama, 2013; Tabuchi et al., 2010; Wlaschin and Hu, 2007; Sellick et al., 2015), product quality attributes (Lee et al., 1989; Minch et al., 1995; Weikert et al., 1999; Jeong et al., 2008; Cristea et al., 2013; Yang et al., 2015; Bydlinski et al., 2018), cell cycle (Bi et al., 2004), cell line stability (Betts & Dickson, 2015), the un-folded protein response (Maldonado-Agurto & Dickson, 2018), specific productivity (Urlaub et al., 1983; Kwon et al., 2006; Pybus et al., 2014), difficult to express proteins (Hussain et al., 2017; Hussain et al., 2018) and apoptosis (Chiang and Sisk, 2005; Lee et al., 2013; Tan et al., 2015). A comprehensive retrospective of CHO cell engineering is reviewed in (Fischer et al., 2015). Appreciation for cellular dynamism across bioprocess culture stages has inspired more sophisticated approaches to overcome antagonistic cell growth and protein production behaviours. One such example has seen the use of inducible promoters.

Transgene expression can be modulated with systems controlled by temperature, heavy metal ions, oxygen tension, steroids, aptamer-based, antibiotics and other small molecules (Fusseneger, 2001). In their study, Mazur et al., (1999) designed a tetracycline gene switch, "Tet^{switch}", based on the natural tetracycline (Tc) operon from *E. coli* (Gossen & Bujard, 1992). Presence of the antibiotic Tc in culture medium promoted active repression of transgene expression using a TET-Off promoter. Tc degraded to near undetectable level over the cell growth phase in a dose dependant manner giving rise to the transcription of the product gene encoded in a multicistronic mRNA with a cytostatic cell-cycle arresting gene p27 (Mazur et al., 1999). The use of antibiotics such as Tc in manufacturing biologics is limited nowadays with the perception that unlimited use of antibiotics contributes to multi-drug resistant human-pathogenic prokaryotes (Fusseneger, 2001). Another inducible system for controlled heterologous protein production is now growing in popularity, the cumate inducible system (Mullick et al., 2006) has been used to produce >350 mg/ L of IgG and >900 mg/ L of Fc-fusion proteins in CHO non-optimised fed-batch processes (Poulain et al., 2017).

CH-omics

Publication of the Chinese hamster and CHO-K1 genomes has accelerated the field of CHO cell engineering (Xu et al., 2011; Brinkrolf et al., 2013; Lewis et al., 2014). Additionally, transcriptome, miRome, proteome and epigenomic data has become available (Baycin-Hizal et al., 2012; Becker et al., 2011; Clarke et al., 2012; Courtes et al., 2013; Hackl et al., 2011; Jakobi et al., 2014; Feichtinger et al., 2016). These omics resources have enabled a complete reconstruction of CHO cell metabolism (Hefzi et al., 2016) and illustrated CHO karyotype diversity and instability (Vcelar et al., 2018). Fluctuations of gene, miRNA (Bort et al., 2012) and epigenetic DNA modification (Marx et al., 2018) abundances over bioprocess stages have also indicated new avenues for CHO cell engineering. Profiling global ribosome occupancy of mRNAs in a CHO producer line, Beuchert Kallehauge et al., (2017) found up to 15% of translating mRNA to be recombinant gene (Figure 3). The immediate application of this study was to knock-down the mRNA of the resistance gene NeoR, significantly improving the productivity and growth of the host cells by 35% and 18% respectively. More long-term efforts could be taken to identify and knockout unnecessary host cell proteins freeing valuable energy or engineering transgene mRNA to bias cellular machinery to favor over host cell mRNAs. Large datasets such as these will enable the construction of synthetic gene networks to intelligibly program cellular behavior going forward.

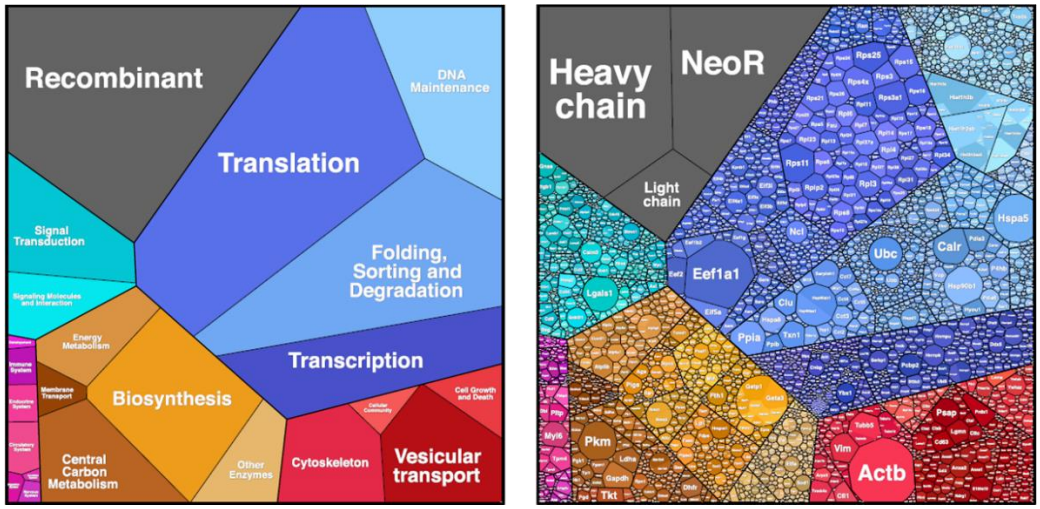


Figure 3. Distribution of translational power across cellular processes for CHO cells in exponential growth. This figure illustrates the relative proportions of cellular ribosome occupancy required for specific cellular functions (left). A more detailed analysis of the genes in each respective polygon is also shown (right). This figure is taken from (Beuchert Kallehauge et al., 2017).

Early in my postgraduate studies I had the opportunity to work on a large collaborative effort to sequence the mitochondrial genome of the Chinese Hamster and twenty-two Chinese Hamster Ovary (CHO) derived cell lines (Chapter 1). This work serves to provide the CHO cell engineering community with a valuable publicly available resource. The study highlighted the innate widespread heteroplasmy in CHO mtDNA offering potential targets for genetic intervention to enhance or refine the powerhouse of industrial cell lines. This study also serves to fuel the growing omics resources which will improve future CHO cell engineering efforts

Engineering CHO cells with miRNAs

The rationale of targeting miRNA as tools in CHO cell engineering is two-fold. Firstly, miRNAs are a class of small (~22nt) non-coding RNAs. Thus, exogenous expression or repression of miRNAs does not directly contribute to host cell translational burden. Second, the small effector or "seed" sequence of each miRNA means a single miRNA can target multiple genes (Bartel, 2009), making manipulation of single miRNAs a more efficient means of effecting global cellular behaviour than single gene strategies (Druz et al., 2013). To that end, numerous miRNAs have been identified as potential "engimiRs" (Hackl et al., 2011). miRNAs were first observed in C. elegans with the discovery of linage-4 (lin-4), a stage-specific gene involved in worm larvae development (Lee et al., 1993). Lee et al., (1993) reported lin-4 to negatively regulate LIN-14

protein expression through a mechanism of complementarity between the LIN-14 mRNA 3'UTR and a 22nt small RNA product of the lin-4 RNA. miRNAs have been identified in almost all metazoan genomes from worms and flies to mammals (He & Hannon, 2004). miRNA genes are transcribed by RNA polymerase II (PolyII) or Pol III, as long primary (pri-miRNA) transcripts (Lee et al., 2004; Steel & Sanghvi, 2012) (Figure 4i). These pri-miRNA transcripts are subject to the microprocessing complex Drosha-DGCR8 to give rise to a (~70nt) precursor (pre-miRNA) (Figure 4ii). Exportin 5 (EXP5) localizes the hairpin to the cytoplasm where it undergoes further processing by DICER (Figure 4iii) prior to mature miRNA association with argonaut (Figure 4iv) and the RNA induced silencing complex (RISC) (Figure 4v).

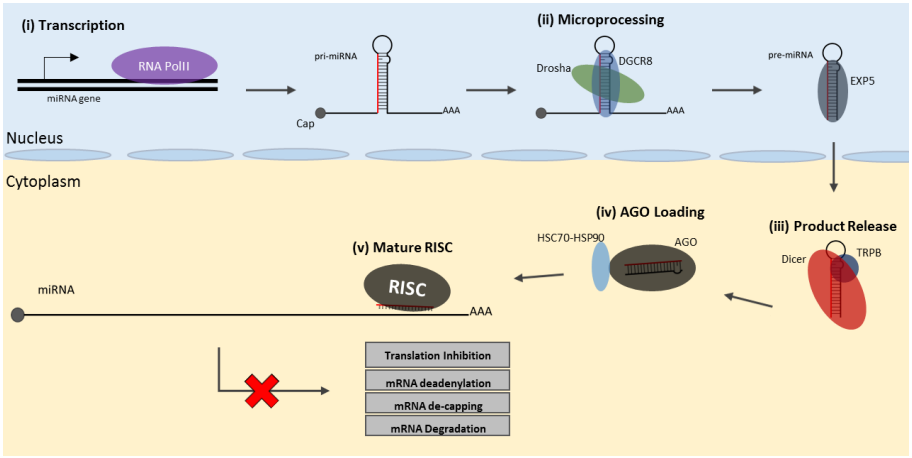


Figure 4. miRNA biogenesis and function. (i) miRNA genes are transcribed by RNA polymerase II (Pol II). (ii) Long primary miRNA transcripts (pri-miRNA) are processed by Drosha and DGCR8 to produce pre-miRNA hairpins. (iii) Localisation to the cytoplasm is done by Exportin (EXP5). (iii) The Dicer-TRPB complex removes the pre-miRNA hairpin loop. (iv) Argonaute (AGO) loading and strand selection. (v) Mature RNA induced silencing complex (RISC) complex performs endogenous functions.

Constitutive exogenous miRNA intervention has proven an effective method for engineering CHO cell behaviours. Finding miRNAs to target for cell engineering can be achieved by profiling differential miRNA abundance and correlating this to desirable cell phenotypes (Clarke et al., 2012) or in a shotgun approach via transient high-content functional screening (Fischer et al., 2014). Constitutive miRNA over-expression or repression has been used to engineer CHO cell specific productivity (Barron et al., 2011; Meleady et al., 2012; Sanchez et al., 2014; Fischer et al., 2015; Kelly et al., 2015; Emmerling et al., 2015; Fischer et al., 2017), inhibit apoptosis (Druz et al., 2011; Kelly et al., 2017) and enhance cell proliferation (Fischer et al., 2014; Strotbek et al., 2013). Several profiling studies correlating miRNA abundance with CHO cell growth have

been conducted without subsequent functional validation (Clarke et al., 2012; Klanert et al., 2016). Improving CHO cell proliferation through miRNA manipulation is an avenue to improving volumetric product yields and remains an neglected area of this niche.

A previous study in our lab found miRNA-candidates with correlation to CHO cell specific growth rate. All prioritized miRNA-candidates being evaluated were found to be low abundant in fast growing clones. This was the starting point for the work described in Chapter 2. Taking six of the top miRNA candidates from the profiling list, miRNA sponges were designed for each. Sponge containing plasmids were then used to generate stable mixed pools. A model IgG producer CHO cell line with a similar growth rate to the low growing clones from the profiling study was used for the functional validation. Enhanced growth was observed in three of the six miRNA candidate depleted pools with respect to negative control sponge and the original cell line. The 50% hit rate could be combinational factors of the study being done in a different cell line to the original profiling, different growth media, also not all differentially expressed candidates from the profiling may be causative but rather a secondary effect of the phenotype. Stable depletion of miR-378-3p was observed to have a biologically significant growth impact. As there are not many miRNAs reported to enhance CHO cell growth in the literature, we sought to further investigate the molecular mechanism by which this miRNA elicits a phenotype.

Label-free LC-MS/ MS was used to find global differential protein abundances between the miR-378 depleted cells and a control. Analysis was performed at two time points during the cell growth phase and sub-cellular protein fractionation used to increase the proteome coverage. miR-378 depletion increased peak cell density of CHO DP12 cultures but reduced cell specific productivity.

Controlling Endogenous miRNA Abundance

The expression of miRNAs in CHO cells is dynamic with fluctuations in specific miRNA abundance between distinct phenotypes and stages of culture. Having seen counter-productive impacts on cellular behaviour with constitutive manipulation of endogenous miRNA levels reported in the literature, we proposed the tuneable regulation of specific miRNAs at defined culture stages. This in theory would outline a strategy for others to mimics with their own cell line specific miRNA signature and culture conditions. Successes in miRNA engineering have not come without the same caveat as the proceeding single gene manipulation strategies. That is, strongly influencing a single desirable cell phenotype comes at the cost of hampering others. To overcome this, we proposed the implementation of a conditional miRNA manipulation strategy. Through reversible control of specific phenotype-driving miRNAs it could be possible to program cellular behaviour as required during a bioprocess. To investigate this, we needed a

means of conditionally regulating specific free miRNA abundance. Thus, we incorporated existing miRNA sponge decoy technology with the most widely used inducible gene expression system, the tetracycline inducible system.

Chapter 3 outlines a method for generating stable mixed pools of CHO cells expressing TET-On inducible miRNA-sponges. The detailed protocols present in this chapter have been implemented in our lab by other postgraduate and postdoctoral researchers. The work outlined in this chapter served as the basis of experimental procedures in Chapter 4 most notably; miRNA-sponge design, miRNA-sponge generation, toxicity of the TET-On induction agent and defining titratable induction of reporter gene expression.

Having generated miRNA-sponge constructs with tuneable expression pertaining to reporter-gene fluorescence, we needed a model miRNA to engineer. In Chapter 2, we observed that stable depletion of miR-378-3p improved CHO DP12 cell growth rate, resulting in a >50% peak cell density but at the cost of cell specific productivity. miR-378 was a prime candidate for conditional engineering as its depletion during exponential growth would improve peak cell density at which point the sponge would be silenced to restore Qp and ultimately improve volumetric titer, simple! Right?

The ambition of the study (Chapter 4) was greater than the results obtained but has hopefully laid foundations for others to build upon. We set out combing two existing molecular tools; the TET-On inducible gene expression system and miRNA-sponge decoy technology. In doing so, numerous unforeseen challenges arose. Firstly, a set of experiments was devised to titrate the TET-On induction agent Doxycycline (Dox) and find a "feeding" strategy for sustaining induction as needed. While this was successful in its outcome, significant deleterious effects on cell growth were observed with the addition of Dox to culture media. This was determined to be a problem worth prioritizing. There was no significant impact on cell growth or viability with the addition of Dox ≤ 10ng/ mL. Unfortunately, 10ng/ mL Dox gave less than desirable induction of just 10-15% of total cells in stable mixed pools. A cell sorting experiment was undertaken to enrich subpopulations that would induce at this concentration. Cells preconditioned with 10ng/ mL Dox for 24hrs were sorted for either medium, mean fluorescence intensity (MFI) ~500, or high MFI >1000. This was successful in defining stable mixed pools with predictable induction characteristics. However, this process did not aid in overcoming the other more pressing challenge in this study.

The most striking finding was the observation of significant target miRNA knockdown in sponge-decoy expressing cells without induction. Absolute quantification of the endogenous levels of target miRNA and the targeting sponge RNA revealed that under basal conditions the copies of sponge present were sufficient to knockdown approximately 50% of endogenous target

miRNA. Two factors were at play here. Firstly, the TET-On inducible gene expression system is just that, a gene expression system. This system is adopted from a naturally occurring gene regulatory sensor. However, from the literature and our own findings in this study the cellular abundance of coding mRNAs is often one or two orders of magnitude higher than that of endogenous miRNAs. While it was achievable to generate stable mixed pools with "negligible" reporter-gene protein expression this was still too high in respect to the target miRNAs. Secondly, the miRNA-sponge design used in this study was derived from previously reported cases of highly effective constitutive sponge-decoy function. We propose in this publication that to gain control of sponge function, a suitable approach could be to sacrifice sponge efficacy or the use of very weak promoters.

Reflecting on the study, evolution of eukaryotic cells has built a system by which low abundant small non-coding RNA molecules regulate the fate of highly expressed protein coding mRNAs. In our study we attempted to regulate cell behaviour by turning this evolutionary conserved mechanism on its head. Controlled synthetic gene circuits relying on endogenous miRNA signatures can work beautifully. Wroblewska et al., (2015) designed simple synthetic gene circuits of two "input sensor molecules" (sponges), each regulating the translation of an output protein. One of these output proteins, L7Ae, is an RNA-binding protein which localizes the 5' UTR of the second output protein product, in their case a GFP reporter. Only under optimal conditions where miR-A regulating L7Ae is high and miR-B controlling GFP is low will the circuit yield translation of the reporter-gene product. This same method could be used in the context of controlling CHO cell protein production. The major limitation of such a strategy is having a robust miRNA signature that is dynamic with culture stages. What must be considered also is that when protein production is not required such as biomass accumulation, the miRNAs being depleted do not have counter-acting phenotypic impact. Tuning the abundance of endogenous miRNAs may be possible with culture media optimization. In this way the reliance on what is a limited number of inducible promoters is negated.

Adaptation of a method developed to reduce exogenous translational burden in a bacterial host, could offer an alternative strategy for optimizing cellular performance. Ceroni et al., (2018) designed a "burden-driven" feedback control system for attenuating the deleterious effects and cellular resource reallocation incurred by transgene expression. RNA-seq was used to identify burden responsive host gene expression and the native promoters driving this. The feedback control circuit contains a product gene under control of a strong constitutive promoter, a non-cutting dCas9 protein constitutively expressed at very low levels and a single guide (sgRNA) targeting the promoter region of the product gene driven by a burden-responsive promoter identified from their RNA-seq study. Should the host become burdened, expression of the sgRNA is triggered and the host automatically adjusts the expression of the product. A miRNA signature

responsive repressor of dCas9 mRNA could be built into such a circuit to eliminate the negative-feedback mechanism following biomass accumulation. In this way, maximum Qp without growth impact is achieved during biomass accumulation followed by a miRNA responsive up-regulation of productivity. As this circuit was designed and used in a bacterial host, mammalian burden-responsive promoters would need to be identified.

The principal benefit for attempting to control cellular behaviour with conditional miRNA manipulation over the approach of tuning transgene expression with endogenous miRNAs is simple. Forced expression or repression of specific miRNAs has demonstrated exaggerated cell behaviours. By simply delaying the translation of transgene mRNA until the cell growth phase is complete cell specific productivity will not be enhanced. Perhaps by incorporating phenotype-driving miRNAs into the design of miRNA-dependant synthetic gene circuits a more favourable outcome would be reached. For this to be achieved, a signature of miRNAs with controllable, predictable endogenous fluctuations would need to be found and the abundance of these miRNAs would need to be causative triggers of cellular behaviour. Alternative miRNA sponge designs driven by weak but tuneable promoters may still be the best strategy and perhaps synthetic sponge designs should better mimic naturally occurring miRNA regulatory molecules.

Natural RNA circles

Trying to find alternative miRNA-sponge designs following the results of the TET-On inducible miRNA work I encountered a study describing naturally-occurring miRNA-sponges in human brain existing as circular RNAs (circRNA) (Hansen et al., 2013). One of these circular sponges happened to contain miRNA responsive elements (MRE) for miR-7, a miRNA which had been used in our lab previously to enhance production of recombinant protein from CHO cells (Sanchez et a., 2013). The MREs of the naturally occurring miR-7 sponge, named “ciRS-7”, resembled those of canonical UTR MREs in eukaryotes with imperfect binding. The expression of synthetic linear lncRNA sponges driven by a U6 poly III promoter have been described previously (Ebert et al., 2007) but require 5’ and 3’ stem loop secondary structures to stabilize the transcripts. While our early results (Figure 6) were promising pertaining to sponge function the main attraction of circular RNA sponges would be RNA stability. Yet, circRNAs are susceptible to the same miRNA-directed turnover as linear sponges (Hansen et al., 2013). During this time several publications (Abe et al., 2015; Yang et al., 2017) were released reporting the translational potential of circular mRNAs. It was thought that this could be a promising avenue to explore in relation to improving the expression of recombinant proteins.

A study by Wesselhoeft et al., (2018) pipped us to the post in becoming the first lab to publish on therapeutic protein production from circular mRNAs but our study goes beyond this work in

two ways. Firstly, the results present in Chapter 5 demonstrate a scalable method for the expression of circular mRNA in mammalian cells. Other works evaluating the translation of circular mRNAs used RNA circularized extracellularly and delivered these circular mRNAs to the cell. Our work used an expression cassette designed to generate a structured intron facilitating circularization of the inserted sequence of interest. This was done with the longer-term vision of potential stable circRNA expressing cell lines. Secondly, ours is the first study to demonstrate the impact of continuous translation coupled with post or co-translational cleavage, permitting the production of secreted, post-translationally modified protein by a rolling circle translation mechanism.

The results in Chapter 5 provide evidence of a novel synthetic mechanism by which recombinant protein yield can be improved in mammalian cells. These results raise some questions that warrant future investigation. Firstly, what mechanism of translation initiation these circular mRNAs are undergoing? Polysome profiling of the linear and circular mRNA variants would be interesting to see. As seen the cells do produce secreted EPO from the IC mRNA and the product is larger in molecular weight, this was an unexpected result. It was anticipated that the cells would produce very high molecular weight poly-EPO proteins. We suspect the secreted EPO is the cleavage of a signal peptide from one EPO product remaining at the C-terminus of another.

The work in Chapter 5 is a novel contribution to the biotechnology community and in particular those focused on recombinant protein production. The work details a mechanism of translation not yet observed in nature and yields significantly improved recombinant protein titer. I suspect there will be much work in the future on the understanding of this mechanism and its potential in the biopharmaceutical industry. In the more immediate future I foresee the employment of circular mRNAs as a tool for delivery of gene therapies. The significant increase in mRNA stability overcomes the main obstacle in the field and could eliminate the use of viral expression vectors. While it may require extensive cell line engineering to efficiently post-translationally cleave nascent polypeptides, I do see potential for the use of continuously translating circular mRNAs in recombinant protein production also.

Chapter 6 is a review of recent findings in relation to circular RNA and circular mRNAs and how these molecules have and can be applied to engineering mammalian cell factories. Chapter 6 will hopefully inspire new research in synthetic circular RNA and circular mRNAs.

Synthetic mRNA Engineering

In generating circular mRNAs to express recombinant EPO for the work in Chapter 5, we created an un-structured intron control vector which had an un-expected improvement on EPO production. This construct contained a splicing intron in the 5'UTR of the EPO mRNA and it has been reported that incorporation of intron sequence between a promoter and the CDS of a transgene improves recombinant protein production. However, the circular mRNAs also harboured putative N^6-methyladenosine (m^6A) motifs. The m^6A motif had been taken from another publication highlighting the abundance of these post-transcriptional RNA epigenetic markers on circular RNAs (Yang et al., 2017). In their study, Yang et al., (2017) constructed three sequence variants; m^6A, and A to T mutant and a no adenosine control, up-stream of GFP open reading frame and observed dramatic effects on circular mRNA translation. Independent studies had reported on variable translation of mRNA with m^6A motifs but none in the context of heterologous recombinant therapeutic protein production.

The presences of methylation sites in mRNA has been reported to influence translation. We decided to evaluate the effects on translation of the three sequence variants from (Yang et al., 2017) on the expression of linear mRNA (Chapter 7). We had anticipated that there may be an increase in protein yield as there was in the circular mRNAs but the mechanism by which this occurred was very surprising. We observed of a 4-5-fold increase in EPO mRNA abundance with the putative m^6A and an A to T mutant over the no adenosine control. By analysing the secondary and tertiary structures of the 5'UTR for each variant we found the m^6A and A to T mutant to have a similar strong hairpin structure, whereas the no adenosine control was an open confirmation. Digging deep into the literature of 5'UTR engineering, similar studies had been investigated in the mid-1980s (Kozak, 1986). Kozak reported that highly stable RNA duplexes in the 5'UTR of a gene open reading frame inhibited protein synthesis as the structure was too strong for a scanning ribosome 40S subunit to melt. What this study neglected was the possibility of cap-independent translation initiation, as it was not widely agreed this form of translation existed in eukaryotes at the time. Also, the inability to quantify relative mRNA abundance was a limitation of Kozak's study. m^6A-mediated translation initiation has been uncovered to result from direct interaction of m^6A motifs and the cap-independent initiation factor eIF3 (Meyer et al., 2015). We believe the stable hairpin to be the cause of the increased mRNA abundance and the methylation of adenosine in the 5'UTR to be the cause of increased translation efficiency over the A to T mutant. While 5'UTR engineering has not been a hot research area in the last two decades it is seeing a resurgence. A pioneering study in this regard was published this year. Cambray and colleagues generated a synthetic library of 244,000 sequence variants to investigate principles of optimal translation in *E. coli* (Cambray et al., 2018). Interestingly, their study concluded that RNA secondary structure near the start codon had the greatest effect on translation

efficiency (Gorochowshi & Ellis, 2018). While the study in Chapter 7 is not comparable to the metadata produced by (Cambray et al., 2018) it does highlight the dramatic effects 5'UTR structure has on mammalian translation. Also demonstrating the novel combination of mRNA structure with RNA modifications to further improve translation. Another study has just been published reporting predictable translational output through rational 5'UTR hairpin design in yeast (Weenink et al., 2018). I foresee a large synthetic 5'UTR library screen being a route to enhance recombinant protein production from mammalian hosts in the not too distant future.

SUMMARY AND CONCLUSIONS

This section is a succinct summation of key findings from the subsequent published chapters. Many of the points made in this section are expounded upon in the discussion sections of the respective chapter.

Chapter 1: Ultra-deep next generation mitochondrial genome sequencing reveals widespread heteroplasmy in Chinese hamster ovary cells

- The results of this study illustrate the heterogeneity of CHO cell mitochondrial DNA (mtDNA).
- Without single-cell RNA-seq analysis it is difficult to draw conclusions about cellular mtDNA heterogeneity or if these mutations are confined to sub-populations of the analysed pool.
- Enriching for or engineering specific mutations in mtDNA could improve CHO cell metabolism.

Chapter 2: Stable Depletion of miR-378-3p Enhances CHO DP12 cell Growth via Up-regulation of the Ubiquitin-specific protease 14 (USP14)

- Inhibition of endogenous miR-378-3p function by miRNA-sponge-decoy improved peak cell density by 59%.
- Quantitative label free LC-MS/MS proteomic analysis of the fractionated cell cultures at day 4 and 8 of batch culture found 216 cytosolic and 114 membrane-associated proteins differentially expressed with stable miR-378-3p depletion.
- The Ubiquitin carboxyl-terminal hydrolase 14 (Usp14) protein was identified in the cytosolic fractions at both timepoints as differentially expressed with an increased abundance of 1.58-fold in the miR-378-3p depleted cells on day 8.
- Overexpression of Usp14 in CHO cells had significant effects on cell growth supporting a role of Usp14 in the increased peak cell density seen with miR-378-3p depletion.
- This study highlights miR-378-3p as a novel engineering candidate for improving CHO cell growth.

Chapter 3: Conditional Knockdown of Endogenous MicroRNAs in CHO Cells Using TET-ON-SanDI Sponge Vectors

- Concatemerization of DNA oligo duplexes is a fast and reliable method for generating specific miRNA sponges of varying sizes.
- Titration of the TET-On induction agent Doxycycline corresponded to the % of cell population induction and the intensity of protein expression.
- The addition of >100 ng/ mL Doxycycline to cell culture media had significant deleterious effects on CHO cell growth and viability.

Chapter 4: Leaky Expression of the TET-On System Hinders Control of Endogenous miRNA Abundance

- The deleterious effects of the TET-On induction agent can be augmented by enriching sub-populations of cells with a greater capacity for induction.
- This enrichment may serve as a method for creating highly inducible cells pertaining to product gene expression. However, this only exacerbated the effects of leaky miRNA-sponge expression of non-conditional target miRNA depletion.
- High levels of reporter-gene expression also had significant detrimental growth effects, with induction of these cells leading to significantly reduced growth and compromised viability.
- We correlated the absolute levels of a specific miRNA, and MREs with mere leaky expression. Showing significant stable knockdown of target miRNAs without need for induction, we recognize that the transcript levels from basal leaky expression are negligible pertaining to controlled gene expression. However, we suggest that with the potency of their function, and volatile expression patterns, miRNAs are more sensitive than protein coding genes and controlled expression requires a more refined approach.
- The findings of this study will hopefully serve to inspire new efforts to engineer CHO cells in dynamic and controllable/ programable ways.

Chapter 5: Continuous Translation of Circularized mRNA Improves Recombinant Protein Titer

- It is possible to robustly produce circular mRNA in live mammalian cells from pDNA using a structured intron expression vector.
- Alterations to the coding sequence do not influence RNA circularization.
- Rolling circle translation of a continuous circular mRNA produced multimeric EPO variants.

- High molecular weight EPO variants were abolished with the inclusion of a viral 2A self-cleavage peptide.
- Continuous translation of circular mRNAs yielded increased secreted EPO titer by co and post-translational cleavage mechanisms.
- This novel mechanism of continuous translation of a circular mRNA yielding monomeric product could be improved further by understanding of how the cells produce post-translationally cleaved products. Or by introduction of known post-translational cleavage sites into the script augmented by co-expression of targeting protein.

Chapter 6: Sentencing Ribosomes to Tediously Long Messages

- CircRNA and circular mRNAs pose novel means of enhancing the biotherapeutic productive potential of mammalian cell and cell-free expression platforms.
- The abundance of natural RNA circles has only come to light of late.
- The complexity and diversity of endogenous circRNA expression and function could perhaps make them favourable targets for host cell engineering.
- Application of the desirable attributes of these molecules; stability, translational potential and form synthetic circRNA or circular mRNAs has shown promise.
- Process intensification through platform or chemical engineering has done wonders for improving protein titres. Synthetic circRNA and circular mRNA could see process intensification on the molecular level by rounding out the cell engineering toolbox.

Chapter 7: Improved Yield of rhEPO in CHO Cells with Synthetic 5' UTR

- The impact of local structure on mRNA translation is not well-defined pertaining the 5' UTR. Reports suggest structural remodelling of the 5' UTR can significantly influence mRNA translation both in cis and trans however a new layer of complexity has been applied to this model with the now known reversible post-transcriptional chemical modification of RNA.
- The addition of putative m6A motifs to the 5' UTR of a model recombinant human therapeutic glycoprotein, Erythropoietin (EPO), yielded significantly improved EPO titer in transient batch culture over no adenosine and A to T mutant controls by 2.84 and 2.61-fold respectively.
- Strong local secondary structure between the 5'UTR and start of the gene coding sequence reduced the rate of transgene mRNA loss in transient batch culture.
- This study highlights that greater nuanced appreciation for eukaryotic mRNA translation can yield significant improvement in transgene mRNA metabolism.

FUTURE WORK AND PERSPECTIVE

This section will give indications of immediate studies that could support the work of respective chapters and outline possible future directions to follow these works.

Chapter 2: Engineering CHO Cells with miR-378-3p Depletion

- Depletion of endogenous miR-378-3p significantly improved peak cell density of CHO DP12 cells. It would be interesting to investigate if miR-378-3p depletion in other CHO producer cell lines has the same impact.
- It would have been interesting to do a comparative coverage study with LC-MS/ MS of whole cell lysates versus the cytosolic and membrane enriched fractions.
- The miRNA – sponge construct used in this case expressed a GFP-reporter. The translational burden associated would need to be negated in an industrial context. This could be done using a truncated GFP sequence or a strictly non-coding sponge cassette under control of an RNA Pol III promoter.
- Some un-published data from our lab suggests that CRISPR mediated miRNA knockout can have s significantly different impact than sponge mediated knockdown as both arms of the miRNA hairpin are affected by the knockout. A homology driven repair strategy could be devised to try and knockout one arm without disrupting the biogenesis of the second arm.

Chapter 4: Leaky Expression of the TET-On System Hinders Control of Endogenous miRNA Abundance

- The deleterious effects of the TET-On induction agent can be augmented by enriching sub-populations of cells with a greater capacity for induction. Other inducible expression systems such as the Cumate system are being evaluated for industrial contexts. This FACS enrichment strategy could be used with the Cumate system to significantly reduce costs of induction agent.
- To minimize the effect of leaky miRNA-sponge function, several strategies are outlined in the discussion of the manuscript. These include the use of active repression of miRNA-sponge-report RNA through RNAi, alterations to miRNA sponge design to reduce efficacy, and the use of very weak promoters. Additional control could be gained by the use of multiple miRNAs to evoke a desired phenotype.
- Relying on the dynamic expression of endogenous miRNA may be a risky strategy and this makes reproducibility across cell lines difficult. A combinatorial approach using endogenous and exogenous miRNAs could be a viable strategy. In this case miRNAs would be conditionally regulated by the user and its impact could regulate the expression

of the product gene of interest. For example, a miRNA sponge sequence targeting a growth increasing miRNA (such as miR-378-3p) could be cloned in the 3' UTR of the product gene of interest, this is construct one and is regulated by an endogenous miRNA. Construct two would co-express a Qp impacting miRNA and a sponge targeting the first endogenous miRNA component. Conditional expression of this construct would first divert the endogenous miRNA from targeting the product gene of interest and secondly impact the cell's translational machinery by forced expression of a phenotype driving miRNA.

Chapter 5: Continuous Translation of Circularized mRNA Improves Recombinant Protein Titer

- Evaluate the viability of producing other therapeutic proteins from circular mRNAs. The single gene ORF of EPO lends itself as a perfect model candidate. However, the therapeutic pipeline is flooded with a variety of antibody formats. Expression of IgG type molecules could be explored by either incorporation of both heavy chain (HC) and light chains (LC) encoded on a single circular mRNA separated by a 2A motif or a short IRES-like site. Alternatively, HC and LC ORFs could be expressed as separate circular mRNAs.
- Further validation of rolling circle translation could be conducted with a targeted LC-MS/ MS study. Continuous translation of a circling ribosome complex will produce a unique peptide through translation of the BSJ. It could be possible to use a synthetic peptide of the predicted BSJ product to validate this mechanism of translation.
- A knockdown study targeting signal peptidases or lysosomal machinery could be used to evaluate the ability of mammalian cells to produce long multimeric polyproteins.
- Rolling circle translation of continuous circular mRNA yielded ~100-fold increase of heterogenous product when evaluated in a bacterial cell-free system. Using synthesized mRNA molecules circularized with T4RNA ligase, the potential of co-translationally cleaved continuous translation of circular mRNA could be evaluated in a mammalian cell-free system. The limitation of increase ribosomal occupancy would be minimized in a cell-free system with the increased stability of circular RNAs overcoming the current bottleneck of this platform.
- Inclusion of an imperfect miRNA responsive element in the UTR of a circular mRNA. Imperfect miRNA – sponge pairing can result in a accumulation of miRNA abundance as seen by the phenomenon of TMMP. In this strategy endogenous miRNA could be used to protect circular mRNA and allow transcript accumulation but not translation. At a defined point of culture a conditional competing miRNA sponge could be expressed to

divert the endogenous miRNAs from the circular mRNAs and up-regulate translation of the gene of interest.

Chapter 7: Improved Yield of rhEPO in CHO Cells with Synthetic 5' UTR

- Validation of m6A methylation by pulldown enrichment with anti-m6A antibody.
- Inclusion of consensus Kozak motif in the 5'UTR design and compare mRNA translation efficiency.
- A meta-analysis of 5' UTR elements could be conducted to find relationships responsible for high or low translational efficiency in host cell transcripts. These rules could then be used to inform design of transgene mRNA UTR elements.

REFERENCES

Abe, N., Matsumoto, K., Nishihara, M., Nakano, Y., Shibata, A., Maruyama, H., Shuto, S., Matsuda, A., Yoshida, M. & Ito, Y. 2015, "Rolling circle translation of circular RNA in living human cells", *Scientific Reports,* vol. 5, pp. 16435.

Aggarwal, S. 2007, "What's fueling the biotech engine?", *Nature Biotechnology,* vol. 25, no. 10, pp. 1097-1104.

Barron, N., Kumar, N., Sanchez, N., Doolan, P., Clarke, C., Meleady, P., O'Sullivan, F. & Clynes, M. 2011, "Engineering CHO cell growth and recombinant protein productivity by overexpression of miR-7", *Journal of Biotechnology,* vol. 151, no. 2, pp. 204-211.

Bartel, D.P. 2009, "MicroRNAs: target recognition and regulatory functions", *Cell,* vol. 136, no. 2, pp. 215-233.

Baycin-Hizal, D., Tabb, D.L., Chaerkady, R., Chen, L., Lewis, N.E., Nagarajan, H., Sarkaria, V., Kumar, A., Wolozny, D. & Colao, J. 2012, "Proteomic analysis of Chinese hamster ovary cells", *Journal of Proteome Research,* vol. 11, no. 11, pp. 5265-5276.

Becker, J., Hackl, M., Rupp, O., Jakobi, T., Schneider, J., Szczepanowski, R., Bekel, T., Borth, N., Goesmann, A. & Grillari, J. 2011, "Unraveling the Chinese hamster ovary cell line transcriptome by next-generation sequencing", *Journal of Biotechnology,* vol. 156, no. 3, pp. 227-235.

Betts, Z. & Dickson, A.J. 2016, "Ubiquitous chromatin opening elements (UCOEs) effect on transgene position and expression stability in CHO cells following methotrexate (MTX) amplification", *Biotechnology Journal,* vol. 11, no. 4, pp. 554-564.

Betts, Z. & Dickson, A.J. 2015, "Assessment of UCOE on recombinant EPO production and expression stability in amplified Chinese hamster ovary cells", *Molecular Biotechnology,* vol. 57, no. 9, pp. 846-858.

Bi, J., 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,* vol. 85, no. 7, pp. 741-749.

Bort, J.A.H., Hackl, M., Höflmayer, H., Jadhav, V., Harreither, E., Kumar, N., Ernst, W., Grillari, J. & Borth, N. 2012, "Dynamic mRNA and miRNA profiling of CHO-K1 suspension cell cultures", *Biotechnology Journal,* vol. 7, no. 4, pp. 500-515.

Brinkrolf, K., Rupp, O., Laux, H., Kollin, F., Ernst, W., Linke, B., Kofler, R., Romand, S., Hesse, F. & Budach, W.E. 2013, "Chinese hamster genome sequenced from sorted chromosomes", *Nature Biotechnology,* vol. 31, no. 8, pp. 694-695.

Browne, S.M. & Al-Rubeai, M. 2007, "Selection methods for high-producing mammalian cell lines", *Trends in Biotechnology,* vol. 25, no. 9, pp. 425-432.

Bulmer, M. 1991, "The selection-mutation-drift theory of synonymous codon usage", *Genetics,* vol. 129, no. 3, pp. 897-907.

Butler, M. 2005, "Animal cell cultures: recent achievements and perspectives in the production of biopharmaceuticals", *Applied Microbiology and Biotechnology,* vol. 68, no. 3, pp. 283-291.

Bydlinski, N., Maresch, D., Schmieder, V., Klanert, G., Strasser, R. & Borth, N. 2018, "The contributions of individual galactosyltransferases to protein specific N-glycan processing in Chinese Hamster Ovary cells", *Journal of Biotechnology,* vol. 282, pp. 101-110.

Cambray, G., Guimaraes, J.C. & Arkin, A.P. 2018, "Evaluation of 244,000 synthetic sequences reveals design principles to optimize translation in *Escherichia coli*", *Nature Biotechnology,* 36(10), p.1005.

Carrillo-Cocom, L., Genel-Rey, T., Araiz-Hernandez, D., Lopez-Pacheco, F., Lopez-Meza, J., Rocha-Pizana, M., Ramirez-Medrano, A. & Alvarez, M. 2015, "Amino acid consumption in naive and recombinant CHO cell cultures: producers of a monoclonal antibody", *Cytotechnology,* vol. 67, no. 5, pp. 809-820.

Carter, P.J. 2011, "Introduction to current and future protein therapeutics: a protein engineering perspective", *Experimental Cell Research,* vol. 317, no. 9, pp. 1261-1269.

Ceroni, F., Boo, A., Furini, S., Gorochowski, T.E., Borkowski, O., Ladak, Y.N., Awan, A.R., Gilbert, C., Stan, G. & Ellis, T. 2018, "Burden-driven feedback control of gene expression", *Nature Methods,* vol. 15, no. 5, pp. 387.

Chiang, G.G. & Sisk, W.P. 2005, "Bcl-xL mediates increased production of humanized monoclonal antibodies in Chinese hamster ovary cells", *Biotechnology and Bioengineering,* vol. 91, no. 7, pp. 779-792.

Chong, W.P., Reddy, S.G., Yusufi, F.N., Lee, D., Wong, N.S., Heng, C.K., Yap, M.G. & Ho, Y.S. 2010, "Metabolomics-driven approach for the improvement of Chinese hamster ovary cell growth: overexpression of malate dehydrogenase II", *Journal of Biotechnology,* vol. 147, no. 2, pp. 116-121.

Clarke, C., Henry, M., Doolan, P., Kelly, S., Aherne, S., Sanchez, N., Kelly, P., Kinsella, P., Breen, L. & Madden, S.F. 2012, "Integrated miRNA, mRNA and protein expression analysis reveals the role of post-transcriptional regulation in controlling CHO cell growth rate", *BMC genomics,* vol. 13, no. 1, pp. 656.

Cohen, S.N., Chang, A.C., Boyer, H.W. & Helling, R.B. 1973, "Construction of biologically functional bacterial plasmids in vitro", *Proceedings of the National Academy of Sciences of the United States of America,* vol. 70, no. 11, pp. 3240-3244.

Costa, A.R., Rodrigues, M.E., Henriques, M., Azeredo, J. & Oliveira, R. 2010, "Guidelines to cell engineering for monoclonal antibody production", *European Journal of Pharmaceutics and Biopharmaceutics,* vol. 74, no. 2, pp. 127-138.

Courtes, F.C., Lin, J., Lim, H.L., Ng, S.W., Wong, N.S., Koh, G., Vardy, L., Yap, M.G., Loo, B. & Lee, D. 2013, "Translatome analysis of CHO cells to identify key growth genes", *Journal of Biotechnology*, vol. 167, no. 3, pp. 215-224.

Cristea, S., Freyvert, Y., Santiago, Y., Holmes, M.C., Urnov, F.D., Gregory, P.D. & Cost, G.J. 2013, "In vivo cleavage of transgene donors promotes nuclease-mediated targeted integration", *Biotechnology and Bioengineering*, vol. 110, no. 3, pp. 871-880.

Dadehbeigi, N. & Dickson, A.J. 2015, "Chemical manipulation of the mTORC1 pathway in industrially relevant CHOK1 cells enhances production of therapeutic proteins", *Biotechnology Journal*, vol. 10, no. 7, pp. 1041-1050.

Doolan, P., Meleady, P., Barron, N., Henry, M., Gallagher, R., Gammell, P., Melville, M., Sinacore, M., McCarthy, K. & Leonard, M. 2010, "Microarray and proteomics expression profiling identifies several candidates, including the valosin-containing protein (VCP), involved in regulating high cellular growth rate in production CHO cell lines", *Biotechnology and Bioengineering*, vol. 106, no. 1, pp. 42-56.

Druz, A., Son, Y., Betenbaugh, M. & Shiloach, J. 2013, "Stable inhibition of mmu-miR-466h-5p improves apoptosis resistance and protein production in CHO cells", *Metabolic Engineering*, vol. 16, pp. 87-94.

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*, vol. 36, no. 6, pp. 1110-1122.

Ebert, M.S., Neilson, J.R. & Sharp, P.A. 2007, "MicroRNA sponges: competitive inhibitors of small RNAs in mammalian cells", *Nature Methods*, vol. 4, no. 9, pp. 721-726.

Emmerling, V.V., Fischer, S., Stiefel, F., Holzmann, K., Handrick, R., Hesse, F., Hörer, M., Kochanek, S. & Otte, K. 2016, "Temperature-sensitive miR-483 is a conserved regulator of recombinant protein and viral vector production in mammalian cells", *Biotechnology and Bioengineering*, vol. 113, no. 4, pp. 830-841.

Feichtinger, J., Hernández, I., Fischer, C., Hanscho, M., Auer, N., Hackl, M., Jadhav, V., Baumann, M., Krempl, P.M. & Schmidl, C. 2016, "Comprehensive genome and epigenome characterization of CHO cells in response to evolutionary pressures and over time", *Biotechnology and Bioengineering*, vol. 113, no. 10, pp. 2241-2253.

Fischer, S., Marquart, K.F., Pieper, L.A., Fieder, J., Gamer, M., Gorr, I., Schulz, P. & 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*, vol. 114, no. 7, pp. 1495-1510.

Fischer, S., Paul, A.J., Wagner, A., Mathias, S., Geiss, M., Schandock, F., Domnowski, M., Zimmermann, J., Handrick, R. & Hesse, F. 2015, "miR-2861 as novel HDAC5 inhibitor in CHO cells enhances productivity while maintaining product quality", *Biotechnology and Bioengineering*, vol. 112, no. 10, pp. 2142-2153.

Fischer, S., Handrick, R. & Otte, K. 2015, "The art of CHO cell engineering: A comprehensive retrospect and future perspectives", *Biotechnology Advances*, vol. 33, no. 8, pp. 1878-1896.

Fogolın, 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*, vol. 109, no. 1-2, pp. 179-191.

Fussenegger, M. 2001, "The impact of mammalian gene regulation concepts on functional genomic research, metabolic engineering, and advanced gene therapies", *Biotechnology Progress*, vol. 17, no. 1, pp. 1-51.

Gopalkrishnan, R., Christiansen, K., Goldstein, N., DePinho, R. & Fisher, P. 1999, "Use of the human EF-1α promoter for expression can significantly increase success in establishing stable cell lines with consistent expression: a study using the tetracycline-inducible system in human cancer cells", *Nucleic Acids Research*, vol. 27, no. 24, pp. 4775-4782.

Gorochowski, T.E. & Ellis, T. 2018, "Designing efficient translation", *Nature Biotechnology*, vol. 36, no. 10, pp. 934.

Gossen, M. & Bujard, H. 1992, "Tight control of gene expression in mammalian cells by tetracycline-responsive promoters", *Proceedings of the National Academy of Sciences of the United States of America*, vol. 89, no. 12, pp. 5547-5551.

Griffith, A., Kelly, P.S., Vencken, S., Lao, N.T., Greene, C.M., Clynes, M. & Barron, N. 2017, "miR-CATCH Identifies Biologically Active miRNA Regulators of the Pro-Survival Gene XIAP, in Chinese Hamster Ovary Cells", *Biotechnology Journal*, .

Hackl, M., Jakobi, T., Blom, J., Doppmeier, D., Brinkrolf, K., Szczepanowski, R., Bernhart, S.H., zu Siederdissen, C.H., Bort, J.A.H. & Wieser, M. 2011, "Next-generation sequencing of the Chinese hamster ovary microRNA transcriptome: Identification, annotation and profiling of microRNAs as targets for cellular engineering", *Journal of Biotechnology*, vol. 153, no. 1-2, pp. 62-75.

Hansen, T.B., Jensen, T.I., Clausen, B.H., Bramsen, J.B., Finsen, B., Damgaard, C.K. & Kjems, J. 2013, "Natural RNA circles function as efficient microRNA sponges", *Nature*, vol. 495, no. 7441, pp. 384-388.

He, L. & Hannon, G.J. 2004, "MicroRNAs: small RNAs with a big role in gene regulation", *Nature Reviews Genetics*, vol. 5, no. 7, pp. 522.

Hefzi, H., Ang, K.S., Hanscho, M., Bordbar, A., Ruckerbauer, D., Lakshmanan, M., Orellana, C.A., Baycin-Hizal, D., Huang, Y. & Ley, D. 2016, "A consensus genome-scale reconstruction of Chinese hamster ovary cell metabolism", *Cell Systems*, vol. 3, no. 5, pp. 434-443. e8.

Hussain, H., Fisher, D.I., Abbott, W.M., Roth, R.G. & Dickson, A.J. 2017, "Use of a protein engineering strategy to overcome limitations in the production of “Difficult to Express” recombinant proteins", *Biotechnology and Bioengineering*, vol. 114, no. 10, pp. 2348-2359.

Hussain, H., Fisher, D.I., Roth, R.G., Mark Abbott, W., Carballo-Amador, M.A., Warwicker, J. & Dickson, A.J. 2018, "A protein chimera strategy supports production of a model “difficult-to-express” recombinant target", *FEBS Letters*, vol. 592, no. 14, pp. 2499-2511.

Jakobi, T., Brinkrolf, K., Tauch, A., Noll, T., Stoye, J., Pühler, A. & Goesmann, A. 2014, "Discovery of transcription start sites in the Chinese hamster genome by next-generation RNA sequencing", *Journal of Biotechnology*, vol. 190, pp. 64-75.

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*, vol. 18, no. 12, pp. 1945-1952.

Kacmar, J. & Srienc, F. 2005, "Dynamics of single cell property distributions in Chinese hamster ovary cell cultures monitored and controlled with automated flow cytometry", *Journal of Biotechnology*, vol. 120, no. 4, pp. 410-420.

Kallehauge, T.B., Li, S., Pedersen, L.E., Ha, T.K., Ley, D., Andersen, M.R., Kildegaard, H.F., Lee, G.M. & Lewis, N.E. 2017, "Ribosome profiling-guided depletion of an mRNA increases cell growth rate and protein secretion", *Scientific Reports*, vol. 7, pp. 40388.

Kaufmann, H., Mazur, X., Fussenegger, M. & Bailey, J.E. 1999, "Influence of low temperature on productivity, proteome and protein phosphorylation of CHO cells", *Biotechnology and Bioengineering*, vol. 63, no. 5, pp. 573-582.

Kelly, P.S., Breen, L., Gallagher, C., Kelly, S., Henry, M., Lao, N.T., Meleady, P., O'Gorman, D., Clynes, M. & Barron, N. 2015, "Re-programming CHO cell metabolism using miR-23 tips the balance towards a highly productive phenotype", *Biotechnology Journal*, vol. 10, no. 7, pp. 1029-1040.

Kim, S.H. & Lee, G.M. 2007, "Functional expression of human pyruvate carboxylase for reduced lactic acid formation of Chinese hamster ovary cells (DG44)", *Applied Microbiology and Biotechnology*, vol. 76, no. 3, pp. 659-665.

Klanert, G., Jadhav, V., Shanmukam, V., Diendorfer, A., Karbiener, M., Scheideler, M., Bort, J.H., Grillari, J., Hackl, 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*, .

Klein, C., Schaefer, W. & Regula, J.T. 2016, "The use of CrossMAb technology for the generation of bi-and multispecific antibodies", *MAbs*Taylor & Francis, , pp. 1010.

Kozak, M. 1986, "Influences of mRNA secondary structure on initiation by eukaryotic ribosomes", *Proceedings of the National Academy of Sciences of the United States of America*, vol. 83, no. 9, pp. 2850-2854.

Kwon, R., Kim, S.K., Lee, S., Hwang, S., Lee, G.M., Kim, J. & Seol, W. 2006, "Artificial transcription factors increase production of recombinant antibodies in Chinese hamster ovary cells", *Biotechnology Letters*, vol. 28, no. 1, pp. 9-15.

Lai, T., Yang, Y. & Ng, S.K. 2013, "Advances in mammalian cell line development technologies for recombinant protein production", *Pharmaceuticals*, vol. 6, no. 5, pp. 579-603.

Lee, J.S., Grav, L.M., Pedersen, L.E., Lee, G.M. & Kildegaard, H.F. 2016, "Accelerated homology-directed targeted integration of transgenes in Chinese hamster ovary cells via CRISPR/Cas9 and fluorescent enrichment", *Biotechnology and Bioengineering*, .

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*, vol. 110, no. 8, pp. 2195-2207.

Lee, R.C., Feinbaum, R.L. & Ambros, V. 1993, "The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14", *Cell*, vol. 75, no. 5, pp. 843-854.

Lee, Y.S., Nakahara, K., Pham, J.W., Kim, K., He, Z., Sontheimer, E.J. & Carthew, R.W. 2004, "Distinct roles for Drosophila Dicer-1 and Dicer-2 in the siRNA/miRNA silencing pathways", *Cell*, vol. 117, no. 1, pp. 69-81.

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", *The Journal of Biological Chemistry*, vol. 264, no. 23, pp. 13848-13855.

Lewis, N.E., Liu, X., Li, Y., Nagarajan, H., Yerganian, G., O'Brien, E., Bordbar, A., Roth, A.M., Rosenbloom, J. & Bian, C. 2013, "Genomic landscapes of Chinese hamster ovary cell lines as revealed by the Cricetulus griseus draft genome", *Nature Biotechnology*, vol. 31, no. 8, pp. 759-765.

Makrides, S.C. 1999, "Components of vectors for gene transfer and expression in mammalian cells", *Protein Expression and Purification*, vol. 17, no. 2, pp. 183-202.

Maldonado-Agurto, R. & Dickson, A.J. 2018, "Multiplexed Digital mRNA Expression Analysis Profiles System-Wide Changes in mRNA Abundance and Responsiveness of UPR-Specific Gene Expression Changes During Batch Culture of Recombinant Chinese Hamster Ovary Cells", *Biotechnology Journal*, vol. 13, no. 3, pp. 1700429.

Marx, N., Grünwald-Gruber, C., Bydlinski, N., Dhiman, H., Ngoc, L.N., Klanert, G. & Borth, N. 2018, "CRISPR-based targeted epigenetic editing enables gene expression modulation of the silenced beta-galactoside alpha-2, 6-sialyltransferase 1 in CHO cells", *Biotechnology Journal*, , pp. 1700217.

Matasci, M., Hacker, D.L., Baldi, L. & Wurm, F.M. 2008, "Recombinant therapeutic protein production in cultivated mammalian cells: current status and future prospects", *Drug Discovery Today: Technologies*, vol. 5, no. 2-3, pp. e37-e42.

Mazur, X., Eppenberger, H.M., Bailey, J.E. & Fussenegger, M. 1999, "A novel autoregulated proliferation-controlled production process using recombinant CHO cells", *Biotechnology and Bioengineering*, vol. 65, no. 2, pp. 144-150.

Meleady, P., Gallagher, M., Clarke, C., Henry, M., Sanchez, N., Barron, N. & Clynes, M. 2012, "Impact of miR-7 over-expression on the proteome of Chinese hamster ovary cells", *Journal of Biotechnology*, vol. 160, no. 3-4, pp. 251-262.

Mermod, N., Pourcel, L., Girod, P. & LE, F.V. 2016, *Use of vitamins and vitamin metabolic genes and proteins for recombinant protein production in mammalian cells, Patent*.

Meyer, K.D., Patil, D.P., Zhou, J., Zinoviev, A., Skabkin, M.A., Elemento, O., Pestova, T.V., Qian, S. & Jaffrey, S.R. 2015, "5′ UTR m 6 A promotes cap-independent translation", *Cell*, vol. 163, no. 4, pp. 999-1010.

Minch, S.L., Kallio, P.T. & Bailey, J.E. 1995, "Tissue plasminogen activator coexpressed in Chinese Hamster Ovary cells with α (2, 6)-sialyltransferase contains NeuAcα (2, 6) Galβ (1, 4) Gl-N-AcR Linkages", *Biotechnology Progress*, vol. 11, no. 3, pp. 348-351.

Moniz, S., Martinho, O., Pinto, F., Sousa, B., Loureiro, C., Oliveira, M.J., Moita, L.F., Honavar, M., Pinheiro, C., Pires, M., Lopes, J.M., Jones, C., Costello, J.F., Paredes, J., Reis, R.M. & Jordan, P. 2013, "Loss of WNK2 expression by promoter gene methylation occurs in adult gliomas and triggers Rac1-mediated tumour cell invasiveness", *Human Molecular Genetics*, vol. 22, no. 1, pp. 84-95.

Morrow, J.F., Cohen, S.N., Chang, A.C., Boyer, H.W., Goodman, H.M. & Helling, R.B. 1974, "Replication and transcription of eukaryotic DNA in Escherichia coli", *Proceedings of the National Academy of Sciences of the United States of America*, vol. 71, no. 5, pp. 1743-1747.

Mullick, A., Xu, Y., Warren, R., Koutroumanis, M., Guilbault, C., Broussau, S., Malenfant, F., Bourget, L., Lamoureux, L. & Lo, R. 2006, "The cumate gene-switch: a system for regulated expression in mammalian cells", *BMC Biotechnology*, vol. 6, no. 1, pp. 43.

Mutskov, V. & Felsenfeld, G. 2004, "Silencing of transgene transcription precedes methylation of promoter DNA and histone H3 lysine 9", *The EMBO Journal*, vol. 23, no. 1, pp. 138-149.

Ng, S.K., Lin, W., Sachdeva, R., Wang, D.I. & Yap, M.G. 2010, "Vector fragmentation: characterizing vector integrity in transfected clones by Southern blotting", *Biotechnology Progress*, vol. 26, no. 1, pp. 11-20.

Noh, S.M., Shin, S. & Lee, G.M. 2018, "Comprehensive characterization of glutamine synthetase-mediated selection for the establishment of recombinant CHO cells producing monoclonal antibodies", *Scientific Reports*, vol. 8, no. 1, pp. 5361.

Norton, P.A. & Pachuk, C.J. 2003, "Methods for DNA introduction into mammalian cells", *New Comprehensive Biochemistry*, vol. 38, pp. 265-277.

Paredes, V., Park, J.S., Jeong, Y., Yoon, J. & Baek, K. 2013, "Unstable expression of recombinant antibody during long-term culture of CHO cells is accompanied by histone H3 hypoacetylation", *Biotechnology Letters*, vol. 35, no. 7, pp. 987-993.

Park, H., Kim, I., Kim, I., Kim, K. & Kim, H. 2000, "Expression of carbamoyl phosphate synthetase I and ornithine transcarbamoylase genes in Chinese hamster ovary dhfr-cells decreases accumulation of ammonium ion in culture media", *Journal of Biotechnology,* vol. 81, no. 2-3, pp. 129-140.

Poulain, A., Perret, S., Malenfant, F., Mullick, A., Massie, B. & Durocher, Y. 2017, "Rapid protein production from stable CHO cell pools using plasmid vector and the cumate gene-switch", *Journal of Biotechnology,* vol. 255, pp. 16-27.

Puck, T.T., Cieciura, S.J. & Robinson, A. 1958, "Genetics of somatic mammalian cells. III. Long-term cultivation of euploid cells from human and animal subjects", *The Journal of Experimental Medicine,* vol. 108, no. 6, pp. 945-956.

Pybus, L.P., Dean, G., West, N.R., Smith, A., Daramola, O., Field, R., Wilkinson, S.J. & James, D.C. 2014, "Model-d irected engineering of “difficult-to-e xpress” monoclonal antibody production by Chinese hamster ovary cells", *Biotechnology and Bioengineering,* vol. 111, no. 2, pp. 372-385.

Ran, F.A., Hsu, P.D., Wright, J., Agarwala, V., Scott, D.A. & Zhang, F. 2013, "Genome engineering using the CRISPR-Cas9 system", *Nature Protocols,* vol. 8, no. 11, pp. 2281.

Ringold, G., Dieckmann, B. & Lee, F. 1981, "Co-expression and amplification of dihydrofolate reductase cDNA and the Escherichia coli XGPRT gene in Chinese hamster ovary cells", *Journal of Molecular and Applied Genetics,* vol. 1, no. 3, pp. 165-175.

Sanchez, N., Kelly, P., Gallagher, C., Lao, N.T., Clarke, C., Clynes, M. & Barron, N. 2014, "CHO cell culture longevity and recombinant protein yield are enhanced by depletion of miR-7 activity via sponge decoy vectors", *Biotechnology Journal,* vol. 9, no. 3, pp. 396-404.

Sarró, E., Lecina, M., Fontova, A., Gòdia, F., Bragós, R. & Cairó, J.J. 2015, "Real-time and on-line monitoring of morphological cell parameters using electrical impedance spectroscopy measurements", *Journal of Chemical Technology and Biotechnology,* .

Schultz, D.C., Ayyanathan, K., Negorev, D., Maul, G.G. & Rauscher, F.J.,3rd 2002, "SETDB1: a novel KAP-1-associated histone H3, lysine 9-specific methyltransferase that contributes to HP1-mediated silencing of euchromatic genes by KRAB zinc-finger proteins", *Genes & Development,* vol. 16, no. 8, pp. 919-932.

Sellick, C.A., Croxford, A.S., Maqsood, A.R., Stephens, G.M., Westerhoff, H.V., Goodacre, R. & Dickson, A.J. 2015, "Metabolite profiling of CHO cells: Molecular reflections of bioprocessing effectiveness", *Biotechnology Journal,* vol. 10, no. 9, pp. 1434-1445.

Shah, P., Ding, Y., Niemczyk, M., Kudla, G. & Plotkin, J.B. 2013, "Rate-limiting steps in yeast protein translation", *Cell,* vol. 153, no. 7, pp. 1589-1601.

Steel, L.F. & Sanghvi, V.R. 2012, "Polycistronic expression of interfering RNAs from RNA polymerase III promoters" in *Functional Genomics* Springer, , pp. 347-359.

Strotbek, M., Florin, L., Koenitzer, J., Tolstrup, A., Kaufmann, H., Hausser, A. & Olayioye, M.A. 2013, "Stable microRNA expression enhances therapeutic antibody productivity of Chinese hamster ovary cells", *Metabolic Engineering,* vol. 20, pp. 157-166.

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,* vol. 110, no. 8, pp. 2208-2215.

Tabuchi, H., Sugiyama, T., Tanaka, S. & Tainaka, S. 2010, "Overexpression of taurine transporter in Chinese hamster ovary cells can enhance cell viability and product yield, while promoting glutamine consumption", *Biotechnology and Bioengineering,* vol. 107, no. 6, pp. 998-1003.

Tan, J.G., Lee, Y.Y., Wang, T., Yap, M.G., Tan, T.W. & Ng, S.K. 2015, "Heat shock protein 27 overexpression in CHO cells modulates apoptosis pathways and delays activation of caspases to improve recombinant monoclonal antibody titre in fed-batch bioreactors", *Biotechnology Journal,* vol. 10, no. 5, pp. 790-800.

Teixeira, A., Oliveira, R., Alves, P. & Carrondo, M. 2009, "Advances in on-line monitoring and control of mammalian cell cultures: supporting the PAT initiative", *Biotechnology Advances,* vol. 27, no. 6, pp. 726-732.

Trummer, E., Fauland, K., Seidinger, S., Schriebl, K., Lattenmayer, C., Kunert, R., Vorauer-Uhl, K., Weik, R., Borth, N. & Katinger, H. 2006, "Process parameter shifting: Part I. Effect of DOT, pH, and temperature on the performance of Epo-Fc expressing CHO cells cultivated in controlled batch bioreactors", *Biotechnology and Bioengineering,* vol. 94, no. 6, pp. 1033-1044.

Urlaub, G., Käs, E., Carothers, A.M. & Chasin, L.A. 1983, "Deletion of the diploid dihydrofolate reductase locus from cultured mammalian cells", *Cell,* vol. 33, no. 2, pp. 405-412.

Vazquez-Lombardi, R., Phan, T.G., Zimmermann, C., Lowe, D., Jermutus, L. & Christ, D. 2015, "Challenges and opportunities for non-antibody scaffold drugs", *Drug Discovery Today,* vol. 20, no. 10, pp. 1271-1283.

Vcelar, S., Jadhav, V., Melcher, M., Auer, N., Hrdina, A., Sagmeister, R., Heffner, K., Puklowski, A., Betenbaugh, M. & Wenger, T. 2018, "Karyotype variation of CHO host cell lines over time in culture characterized by chromosome counting and chromosome painting", *Biotechnology and Bioengineering,* vol. 115, no. 1, pp. 165-173.

Weenink, T., van der Hilst, J., McKiernan, R.M. & Ellis, T. 2018, "Design of RNA Hairpin Modules that Predictably Tune Translation in Yeast", *Synthetic Biology,* .

Weikert, S., Papac, D., Briggs, J., Cowfer, D., Tom, S., Gawlitzek, M., Lofgren, J., Mehta, S., Chisholm, V. & Modi, N. 1999, "Engineering Chinese hamster ovary cells to maximize sialic acid content of recombinant glycoproteins", *Nature biotechnology,* vol. 17, no. 11, pp. 1116.

Wesselhoeft, R.A., Kowalski, P.S. & Anderson, D.G. 2018, "Engineering circular RNA for potent and stable translation in eukaryotic cells", *Nature Communications,* vol. 9, no. 1, pp. 2629.

Wilson, T.J. & Kola, I. 2001, "The LoxP/CRE system and genome modification", *Gene Knockout Protocols,* , pp. 83-94.

Wlaschin, K.F. & Hu, W. 2007, "Engineering cell metabolism for high-density cell culture via manipulation of sugar transport", *Journal of Biotechnology,* vol. 131, no. 2, pp. 168-176.

Wroblewska, L., Kitada, T., Endo, K., Siciliano, V., Stillo, B., Saito, H. & Weiss, R. 2015, "Mammalian synthetic circuits with RNA binding proteins for RNA-only delivery", *Nature Biotechnology,* vol. 33, no. 8, pp. 839.

Wurm, F.M. 2004, "Production of recombinant protein therapeutics in cultivated mammalian cells", *Nature Biotechnology,* vol. 22, no. 11, pp. 1393-1398.

Wurm, F.M. & Hacker, D. 2011, "First CHO genome", *Nature biotechnology,* vol. 29, no. 8, pp. 718-720.

Wurm, F. & Bernard, A. 1999, "Large-scale transient expression in mammalian cells for recombinant protein production", *Current Opinion in Biotechnology,* vol. 10, no. 2, pp. 156-159.

Xu, X., Nagarajan, H., Lewis, N.E., Pan, S., Cai, Z., Liu, X., Chen, W., Xie, M., Wang, W. & Hammond, S. 2011, "The genomic sequence of the Chinese hamster ovary (CHO)-K1 cell line", *Nature Biotechnology,* vol. 29, no. 8, pp. 735-741.

Yang, Y., Fan, X., Mao, M., Song, X., Wu, P., Zhang, Y., Jin, Y., Yang, Y., Chen, L. & Wang, Y. 2017, "Extensive translation of circular RNAs driven by N6-methyladenosine", *Cell Research,* vol. 27, no. 5, pp. 626-641.

Yang, Z., Wang, S., Halim, A., Schulz, M.A., Frodin, M., Rahman, S.H., Vester-Christensen, M.B., Behrens, C., Kristensen, C. & Vakhrushev, S.Y. 2015, "Engineered CHO cells for production of diverse, homogeneous glycoproteins", *Nature Biotechnology,* vol. 33, no. 8, pp. 842.

Chapter 1

Ultra-deep next generation mitochondrial genome sequencing reveals widespread heteroplasmy in Chinese hamster ovary cells

Kelly, P.S., Clarke, C., **Costello, A.**, Monger, C., Meiller, J., Dhiman, H., Borth, N., Betenbaugh, M.J., Clynes, M. and Barron, N.
Metabolic Engineering, DOI: 10.1016/j.ymben.2017.02.001, PMID: 28188893
Impact Factor (2017): 7.674

My contributions to this study were strictly limited to the initial wet-lab work. I was responsible for the sub-culture and mitochondrial DNA (mtDNA) enrichment of ten of the twenty-two cell lines evaluated. I worked closely with Dr Paul Kelly to optimise the mtDNA enrichment and to eliminate contaminating nuclear DNA noise by PCR amplification of mtDNA with multiple overlapping primer sets.

ABSTRACT

Recent sequencing of the Chinese hamster ovary (CHO) cell and Chinese hamster genomes has dramatically advanced our ability to understand the biology of these mammalian cell factories. In this study, we focus on the powerhouse of the CHO cell, the mitochondrion. Utilizing a high-resolution next generation sequencing approach we sequenced the Chinese hamster mitochondrial genome for the first time and surveyed the mutational landscape of CHO cell mitochondrial DNA (mtDNA). Depths of coverage ranging from ~3,319X to 8,056X enabled accurate identification of low frequency mutations (>1%), revealing that mtDNA heteroplasmy is widespread in CHO cells. A total of 197 variants at 130 individual nucleotide positions were identified across a panel of 22 cell lines with 81% of variants occurring at an allele frequency of between 1% and 99%. 89% of the heteroplasmic mutations identified were cell line specific with the majority of shared heteroplasmic SNPs and INDELs detected in clones from 2 cell line development projects originating from the same host cell line. The frequency of common predicted loss of function mutations varied significantly amongst the clones indicating that heteroplasmic mtDNA variation could lead to a continuous range of phenotypes and play a role in cell to cell, production run to production run and indeed clone to clone variation in CHO cell metabolism. Experiments that integrate mtDNA sequencing with metabolic flux analysis and metabolomics have the potential to improve cell line selection and enhance CHO cell metabolic phenotypes for biopharmaceutical manufacturing through rational mitochondrial genome engineering.

1.1 INTRODUCTION

The continual improvement of bioprocesses over the last 20 years has enabled the production of g/L quantities of complex therapeutic proteins (e.g. monoclonal antibodies) from industrial scale Chinese hamster ovary (CHO) cell culture (Wurm, 2004, Li et al., 2010). These dramatic improvements in performance have been achieved, in part, through understanding the nutrient requirements of CHO cells to optimise media formulations. Industrial scale cell culture processes, where possible, now utilise chemically defined media that maintain growth rate and increase titre as well as eliminate the batch variation associated with biological components such as serum (Butler, 2005). The development of fed-batch cell culture strategies have also been central to achieving high product titres, counteracting the production of cellular waste products during cell culture and extending production runtimes. During exponential growth, CHO cells channel glucose and glutamine through the glycolytic pathway even in cases of high oxygen availability (aerobic glycolysis), a metabolic phenotype similar to the Warburg effect observed in cancer cells (Vander Heiden et al., 2009). The resulting secretion of lactate and ammonium inhibit cell growth and productivity as well as initiate apoptosis and decrease product quality (Butler, 2005). In a fed-batch process, cells are initially grown to a high cell density before the bioreactor environment is altered through e.g. reducing the temperature or altering the pH of cell culture. A stationary phase of cell growth is induced to shift CHO cells from a lactate production to consumption phenotype extending viability and maximising protein production (Butler, 2005). Metabolic flux analysis (MFA) and metabolomics approaches have proven to be powerful tools for understanding the molecular basis of CHO cell metabolic phenotypes to enable predictable, rapid and inexpensive optimisation of industrial bioprocesses (Quek et al., 2010). These techniques have enabled the development of complex closed loop feeding strategies to limit glucose and glutamine concentrations as well as identify genetic engineering targets to drive CHO cells toward metabolically desirable phenotypes (Young, 2013).

Recent studies of CHO cell metabolism have indicated that mitochondrial function is central to lactate production/consumption (Zagari et al., 2013) and indeed the variability observed in CHO cell metabolic phenotypes (Gilbert et al., 2013). Mitochondria play a central role in eukaryotic cellular energy metabolism via oxidative phosphorylation (OXPHOS) and have important functions in biological processes such as intracellular calcium signalling (Rizzuto et al., 2012) and apoptosis (Wang and Youle, 2009). While the overwhelming majority of proteins required to carry out these functions are transcribed from nuclear DNA (nDNA) (Calvo et al., 2006), mitochondria also contain a distinct, double stranded circular genome. Eukaryotic cells can contain more than 1,000 copies of mitochondrial DNA (mtDNA) packaged within DNA-protein structures known as mitochondrial nucleoids (each nucleoid contains 2–10 mtDNA molecules). mtDNA copy number varies according to cell type. For instance, individual myocardial muscle

cells contain an average of 6,000 copies while leukocytes may have as few as 350 per cell (Miller et al., 2003, Liou et al., 2010). In humans, a significant degree of variation in mtDNA copy number has been observed between the same tissues of different individuals as well as across multiple tissues from the same individual (Wachsmuth et al., 2016).

The mitochondrial genome is between 15 and 17 kb in length and contains 37 genes (28 on the guanine rich "heavy" or H-strand and 9 on the cytosine rich "light" or L-strand). mtDNA encodes 13 polypeptide subunits of OXPHOS complexes I, III, IV and V along with 2 ribosomal RNA subunits and 22 tRNAs required for intra-mitochondrial protein synthesis. The mtDNA genome is extremely compact, genes lack introns, intergenic regions are limited to 1 or 2 nucleotides and in some cases genes can overlap (e.g. ATP6 and ATP8). The only significant non-coding regulatory region is called the displacement loop (D-loop) and contains the origin of replication for the H-strand. Transcription is initiated from one of two H-strand promoters or a single promoter on the L-strand resulting in polycistronic RNA that is subsequently processed to produce mRNAs, tRNAs and rRNAs. Mitochondria utilise a distinct genetic code for mRNA translation allowing translation of all codons using only 22 tRNAs (Anderson et al., 1981, Barrell et al., 1980). mtDNA encodes for only two of the four nDNA stop codons. AGA and AGG are stop codons in mtDNA (as opposed to arginine in nDNA), with the UAA stop codon added to transcribed mRNAs via polyadenylation. The nDNA stop codon UGA encodes tryptophan in mtDNA while the AUA codon that encodes for isoleucine in nDNA encodes methionine in mtDNA.

Relatively inefficient DNA repair mechanisms and close proximity to reactive oxygen species (ROS) contribute to a mitochondrial genome mutation rate at least 10-fold higher than that of the nuclear genome. This high mutation rate has seen widespread application of mtDNA sequencing for studies in evolutionary biology, population genetics and forensic science. The first pathogenic human mitochondrial mutations were identified nearly 30 years ago and since then more than 250 polymorphisms, insertions and deletions have been implicated in metabolic disorders as well as cancer and diabetes. Mitochondrial genome polyploidy can give rise to two cellular states: 1) all mtDNA copies are identical, known as homoplasmy or 2) a mixture of wild-type and mutated mtDNA copies are present, known as heteroplasmy. In healthy cells, wild-type and mutated mtDNA copies can co-exist; mitochondrial dysfunction occurs when the ratio reaches a particular level known as the mitochondrial threshold effect, and in some cases, the frequency of heteroplasmy correlates with the severity of a clinical phenotype (Payne et al., 2013). In recent years, next generation sequencing (NGS) technologies have been utilized extensively for the study of heteroplasmy due to increased specificity, sensitivity and throughput in comparison to traditional Sanger sequencing. Although initially thought to be a rare phenomenon, NGS has revealed the prevalence of mitochondrial heteroplasmy in the human population as well as the

age-related increase of heteroplasmic variants (Li et al., 2015). Studies utilizing ultra-deep sequencing to identify very low frequency variants have indicated that heteroplasmy is universal with each cell containing a complex mixture of mitochondrial genotypes (Payne et al., 2013).

Efforts to understand the biology of CHO cell factories and improve industrial scale biopharmaceutical manufacturing have been dramatically enhanced since the landmark publication of the CHO-K1 genome (Xu et al., 2011). A wealth of sequence data is now freely available for several CHO cell lines as well as the Chinese hamster (Xu et al., 2011, Lewis et al., 2013, Brinkrolf et al., 2013, Kaas et al., 2015). Direct analyses of these data have permitted the first studies of CHO cell genome instability (Kaas et al., 2015), chromosomal rearrangement (Cao et al., 2012), and copy number variation (Lewis et al., 2013). Methods for expression profiling have also seen marked improvement in the CHO cell post-genomic era and overcome the reliance on homology with model species that limited early studies in the field. CHO-cell-specific sequence databases have increased the number of identifications from mass spectrometry based proteomic analysis (Meleady et al., 2012). The combination of genome sequence and NGS technology to study RNA has been employed to analyse mRNA and small RNA expression patterns (Gerstl et al., 2013, Diendorfer et al., 2015, Birzele et al., 2010) as well as to annotate transcripts (Rupp et al., 2014) and identify promoter regions (Wippermann et al., 2015).

While the availability of nuclear genome sequences has undoubtedly advanced CHO cell biology, we know little about the mtDNA and the impact of mutations on cell metabolism and bioprocess performance. Here, we present the first comprehensive survey of the CHO cell mitochondrial genome spanning a panel of cell lines originating from industry, the ATCC and our laboratory. The utilisation of next generation sequencing technology enabled high-resolution detection of mtDNA mutations including those occurring at low frequency. Our results indicate that heteroplasmy is widespread in CHO cell lines, tends to be cell line specific and that these mutations could play a role in metabolic phenotype variability.

1.2 MATERIALS AND METHODS

1.2.1 Extraction of DNA from Chinese hamster and mouse liver tissue

Genomic DNA was extracted from 30 mg of liver tissue from an outbred Chinese hamster and a CB17/lcr-Prkdcscid/Crl mouse (Supplementary Table 1) using a DNeasy Blood and Tissue kit (QIAGEN, 69581). Tissue samples were sheared using a Dounce homogenizer in 180 µl of ALT buffer. The purity and integrity of extracted tissue-derived genomic DNA was determined on a nano-drop and via a DNA-agarose gel stained with ethidium bromide (Supplementary Fig. 1A).

1.2.2 CHO cell culture and mtDNA extraction

22 CHO cell lines (Supplementary Table 1) were grown in suspension unless indicated otherwise and harvested at 72 h. All suspension cultured cell lines were seeded initially at 2×105 cells/mL in 5 mL of culture media. The 4 Biogen cell lines were cultured in suspension in proprietary chemically defined media supplied by the industry partner and cultured in-house. The 8 clones from Pfizer originating from 2 cell line development projects (CLD1 and CLD2), were grown in attached culture in DMEM supplemented with 5% serum. All remaining CHO cell lines were cultured in suspension in 5 mL of serum-free media at an initial density of 2×105 cells/mL. 17×106 cells were acquired at 72 h for mitochondrial DNA extraction. A bacterial mini-prep kit (QIAGEN, 27104) was used to reduce contaminating nuclear DNA and enrich for double stranded mtDNA in the CHO cell line samples as previously described (Quispe-Tintaya et al., 2013). Purity and integrity of isolated CHO mitochondrial DNA was determined using a NanoDrop and an agarose DNA gel (Supplementary Fig. 1B).

1.2.3 Amplification of mitochondrial DNA

To further eliminate nuclear DNA from the tissue and cell line samples, we amplified mtDNA fragments using a high-fidelity PCR kit (Life Technologies, 11304-011) (Supplementary Fig. 1C). CHO mtDNA primers were designed using the CHO cell mtDNA sequence available on GenBank (NC_007936.1). Two overlapping ~8.5 kb mtDNA fragments were designed to span the ~16.5 kb mitochondrial genome sequence (Supplementary Table 2). Another set of overlapping primers was designed based on the Mus musculus mitochondrial genome sequence (NC_005089.1). PCR amplification of each paired mitochondrial genome fragment was performed using the following thermo-cycler conditions: 94 °C for 2 min, 12 cycles at 94 °C for 30 s, 55 °C for 30 s and 68 °C for 8.5 min. The resulting PCR products for each sample was cleaned using Ampure® XP DNA-binding magnetic beads (Agencourt, A63880).

1.2.4 Nextera XT Mitochondrial DNA library preparation and sequencing

For each mtDNA sample, a serial dilution was performed in nuclease-free water and quantified using the Qubit dsDNA HS assay kit (Invitrogen, Q32851) to obtain a 0.2 ng/µl stock. 1 ng of mitochondrial DNA library was prepared using the Nextera XT DNA Sample Preparation Kit (FC-131–1024) in accordance with the manufacturer's specifications. After each library was fragmented, adapters were added followed by the incorporation of sample indexes by PCR. Each of the 24 uniquely indexed samples was passed through a PCR clean-up using Ampure® XP DNA-binding magnetic beads (Agencourt, A63880). Libraries were quantified using the Qubit dsDNA HS assay kit and fragment size distribution was determined using the High Sensitivity DNA Bioanalyzer kit (Agilent, 5067-4626) to confirm the recommended 500–600 bp range. Libraries were normalised to 4 nM in resuspension buffer prior to sequencing. Each 4 nM library was then pooled into a single sample and sequenced on an Illumina MiSeq (San Diego, CA) configured to produce 151 bp paired-end reads. Following sequencing, base calls were converted to 24 individual FASTQ format files for bioinformatics analysis. These data have been deposited in the NCBI SRA repository (SRP100361).

1.2.5 Reconstruction and annotation of the Cricetulus griseus mitochondrial genome sequence

To reconstruct the Chinese hamster mtDNA sequence the MITOBIM algorithm (Hahn et al., 2013) was utilised in combination with the CHO cell mtDNA sequence available on GenBank (NC_007936). Paired-end reads were merged using FLASH (Magoč and Salzberg, 2011) prior to assembly. MITOBIM assembles a mitochondrial genome by mapping sequencing reads (from Chinese hamster liver tissue) to a closely related sequence, in this case CHO cell mtDNA. The newly assembled Chinese hamster mtDNA was initially annotated using the MITOS (Bernt et al., 2013) and ARWEN (Laslett and Canbäck, 2008) webservers. Annotations were verified and if necessary, refined via BLAST and comparison to human, mouse and rat mtDNA.

1.2.6 CHO cell line mapping and data pre-processing

Reads corresponding to each of the 22 CHO cell lines sequenced were subjected to quality control assessment followed by the removal of adapter sequences and reads <50 bp using trimmomatic (Bolger et al., 2014). The remaining sequence data was mapped to the Chinese hamster mitochondrial genome reference using the BWA-MEM algorithm (Li and Durbin, 2009). Representation of a circular mitochondrial genome as a linear sequence (i.e. beginning at position 1 and ending at position 16283) can give rise to incomplete read mapping due to the introduction of an artificial sequence break. Paired-end reads that span the sequence break will be designated as unmapped and eliminated by algorithms such as BWA, decreasing depth at the start and end

of the mtDNA reference and affecting the ability to accurately detect variants in these regions. In this study the "double" alignment mapping strategy described by Ding et al. (2015) was utilised to ensure optimum alignment of reads to the Chinese hamster mitochondrial genome. Using this approach, CHO cell line reads were mapped to the original or "unshifted" reference sequence beginning at position 1 and ending at position 16283. For the second mapping run a new reference mtDNA sequence was created by joining the start and ends of the original sequence and introducing a new break point so that the sequence began at position 8000 and ended at position 7999 on the original reference (Supplementary Fig. 2).

1.2.7 Variant pre-processing

Following alignment against the "unshifted" and "shifted" reference sequence, reads with a MAPQ <20 were designated as "unmapped" and discarded. To pre-process the "unshifted" and "shifted" mapped data for variant calling, we followed the Genome Analysis ToolKit (GATK) (Van der Auwera et al., 2002) best practice guidelines. PCR duplicates can arise during library preparation following the amplification of the multiple copies of identical DNA fragments. Duplicates propagate errors in sample and library preparation across the dataset, violate the assumption of independence during variant calling and potentially result in the identification of false positive variants. Reads corresponding to PCR duplicates were identified using Picard (http://broadinstitute.github.io/picard/) and eliminated from further analyses. The remaining reads were realigned around INDELs accounting for mapping artefacts that arise from the independent read by read alignment process. Base quality score recalibration reduces the effect of systematic sequencing biases by first determining covariation between factors including nucleotide context (e.g. AC dinucleotides are often lower quality than TG) and base position within the read (bases at the ends of the reads generally have more mismatches). The Phred-scaled Q values are then adjusted accordingly to reduce false positives during variant calling.

1.2.8 Variant discovery and annotation

To identify CHO cell line mtDNA mutations using the dual mapping strategy, variant calls were made within specific regions of the "unshifted" and "shifted" Chinese hamster reference sequences (Supplementary Fig. 2). For the original "unshifted" reference, variants are called between positions 4000 and 12000, while sequence variants on the "shifted" reference sequence are called between positions 4000 and 12283, translating to the regions spanning 1–3999 and 12001–16283 on the original reference sequence (encompassing the joined start and ends). The bioinformatics pipeline incorporated both VarScan (Koboldt et al., 2012) and LoFreq (Wilm et al., 2012) to detect SNPs and INDELs. Both algorithms also output supporting information for

each identified mutation such as the number of times a nucleotide position was sequenced (i.e. depth) and the number of reads that support the variant allele. Only those SNPs and INDELs identified by both VarScan and LoFreq with a minor allele frequency (MAF) ≥1%, minimum sequencing depth >1,500X at the variant position and an average Phred-scaled base quality (≥ Q25) for the alternate allele were reported. Variants were eliminated if overrepresentation of reads supporting the mutation was observed in either forward or reverse direction (i.e. strand bias) (Guo et al., 2012). An additional threshold was employed for VarScan calls and only variants with $p<0.01$ were retained. Each identified INDEL was inspected manually to confirm potential false positives. Upon completion of the mutation detection pipeline, the coordinates of "shifted" sequence variants were transformed back to the original reference coordinates and combined with those variants identified following "unshifted" sequence analysis. To determine the putative effect of each mutation we first utilised snpEff (Cingolani et al., 2012) to annotate each mutation (i.e. frameshift, stop codon gained, start codon mutation, missense or non-synonymous). The PROVEAN algorithm (Choi et al., 2012) was utilised to predict the functional impact of missense variants on protein function. Those missense variants with a Provean score ≤−2.5 were classified as deleterious.

1.2.9 Estimation of contamination from nuclear mitochondrial sequences

Heteroplasmy detection can be confounded, particularly from whole genome sequencing data, by the presence of nuclear encoded copies of mitochondrial sequences (NumtS) (Just et al., 2015). In this study, potential contamination was reduced through long range PCR to enrich for mitochondrial DNA and in the case of cell lines, the utilisation of a bacterial mini-prep kit to eliminate non-circular DNA prior to amplification. To confirm the effectiveness of the enrichment strategy for mitochondrial DNA, we utilised the mouse mitochondrial sequencing data to assess potential contamination from NumtS. The variant identification pipeline was first used to identify SNPs and INDELs against the mouse reference sequence (NC_005089.1). To estimate the influence of NumtS, processed reads from mouse were separately aligned to the Mus musculus nuclear genome (mm9 assembly) using the BWA-MEM algorithm (the "–L" and "-T" parameters were set to "9,9" and "145" respectively). Those reads which mapped to known mm9 NumtS regions (Calabrese et al., 2012) with a MAPQ >20 were extracted, reconverted to FASTQ files and remapped against the mouse mtDNA reference sequence using the variant discovery pipeline described above. The thresholds of the LoFreq and VarScan algorithms were modified to account for the lower depth of coverage in order to determine if NumtS were influencing variant detection and heteroplasmy levels.

1.3 RESULTS

1.3.1 Reconstruction of the Cricetulus griseus mitochondrial DNA sequence

Examination of the NC_007936 mtDNA sequence and the publication describing its acquisition (Partridge et al., 2007) revealed that the sequence does not originate from Cricetulus griseus but the mitochondrial genome of either a CHO-K1 or CHO AL cell line. There are also a number of discrepancies between the annotation described in the publication and the GenBank entry. For example, the GenBank annotation states that tRNAAsn is encoded on mtDNA H-strand while the original publication places tRNAAsn on the L-strand. In this study, we sequenced the Chinese hamster mitochondrial genome to produce an accurate reference sequence for comparability of CHO cell line mtDNA as well as resolving any ambiguities in annotation (Fig. 1). Chinese hamster mtDNA was isolated from liver tissue and sequenced on the Illumina MiSeq platform yielding 640,142 paired-end reads. We reconstructed the mitochondrial genome from these data using the MITOBIM (Hahn et al., 2013) algorithm with the NC_007936 mtDNA sequence used as the "backbone" sequence. The assembled C. griseus mitochondrial genome is 16,283 bp in length (A=33.7%, C=22.8%, G=13.0%, T=30.5%) with an overall GC content of 35.7%.

Annotation was initially performed using MITOS (Bernt et al., 2013) and ARWEN (Laslett and Canbäck, 2008) and further refined through comparison of the Chinese hamster mtDNA with human, mouse and rat reference mtDNA sequences (Supplementary Table 3). The Chinese hamster mtDNA has conserved synteny with mammalian mitochondrial genomes with 13 protein-coding genes, 22 tRNAs and 2 ribosomal RNAs as well as a non-coding control region (D-loop). Nine genes are encoded on the mtDNA light strand (ND6, tRNAGln, tRNAAla, tRNAAsn, tRNACys, tRNATyr, tRNASer, tRNAGlu and tRNAPro) with the remaining 28 genes encoded by the H-strand. 9 protein-coding genes start with ATG initiation codon (COX1, COX2, ATP8, ATP6, COX3, ND4L, ND4, ND6, CYTB), 2 with an ATT codon (ND2 and ND5), 1 with a GTG (ND1) and 1 with an ATA codon (ND3). 8 genes terminated with the TAA codon with the ND1, ND2, COX3, ND6, and ND4 stop codons predicted to be completed via transcript polyadenylation. Comparison of the Chinese hamster mtDNA sequence to the currently available CHO cell line mitochondrial genome identified 7 variants, 5 in protein coding sequences with 2 mutations identified in the mtDNA D-loop comprising 4 SNPs, 2 deletions and an insertion (Table 1). The C. griseus mitochondrial genome sequence and corresponding annotation have been submitted to GenBank (accession no: KX576660).

1.3.2 Mapping of CHO cell line next generation sequencing data to the Cricetulus griseus mitochondrial genome

A total of 22 CHO cell lines sourced from industry partners, the ATCC and from our laboratory were sequenced (Supplementary Table 1). Adapter sequences were trimmed from each of the CHO cell line datasets followed by removal of low-quality reads and those less than 50 bp. Upon completion of this initial pre-processing stage the number of reads remaining in each sample ranged from 1,547,006 (DCU CHO-K1 SEAP) to 464,546 (Biogen DG44 #1) (Supplementary Table 4A). The BWA-MEM algorithm was used to align reads against the C. gresius mtDNA reference sequence as well as a modified version of the reference sequence where the original start and ends were joined and a new breakpoint introduced at 8,000 bp (Supplementary Fig. 2). This dual mapping strategy was utilised to account for the circularity of the mitochondrial genome and remove bias arising from the use of a linear reference sequence (e.g. discarding reads that spanned the artificial breakpoint) and therefore improve our ability to detect mutations around the start and ends of the mtDNA sequence.

Following alignment to both the shifted and unshifted reference sequence, reads with a MAPQ <20 were eliminated from further analysis. The DCU CHO-K1 SEAP sample was found to have the largest number of reads mapping with MAPQ <20, yet this represented only ~1% of the total reads in that dataset. On average >99% of reads mapped to the unshifted and shifted reference sequences with a MAPQ ≥20 (Fig. 2A; Supplementary Table 4B & 4C) demonstrating the effectiveness of the mtDNA isolation and amplification method utilised in this study. To ensure PCR duplicates were ignored in downstream stages of the bioinformatics analysis, duplicates were "marked" using the Picard tool (http://broadinstitute.github.io/picard/). The DCU CHO-K1 SEAP sample had the highest proportion of duplicates identified (~44%) in the sample set while the Biogen DG44 #1 sample had the lowest proportion of duplicates (~21%) (Fig. 2A). The duplicate marked data was further pre-processed for variant calling by INDEL realignment and base recalibration in line with the best practice guidelines (Van der Auwera et al., 2013, McKenna et al., 2010).

1.3.3 Identification of CHO cell mitochondrial genome variants

The average depth of coverage and perbase coverage for each cell line sample was calculated using samtools (Li et al., 2009). The lowest average depth of coverage across the unshifted reference sequence was observed for the Biogen DG44 #1 cell line (~3,319X) while the deepest coverage was observed for the DCU control sponge #1 cell line (~8,056X) (Fig. 2B, Supplementary Table 4D & 4E). While a negligible difference in average coverage (~4X) across the entire shifted and unshifted reference sequences was observed, the effectiveness of the dual mapping strategy is illustrated by an average increase in coverage across the first and last 100 bp

of the mtDNA reference sequence of 1,128X and 1,079X respectively. The coverage at each individual nucleotide position of the reference sequence was found to be extremely deep (Fig. 2C), permitting high resolution analysis of the CHO cell mitochondrial genome and confident identification of low frequency heteroplasmic variants across the 22 cell lines. The Lofreq and VarScan algorithms were utilised in parallel to identify mutations in CHO cell lines when compared with the Chinese hamster reference sequence. Only those SNPs and INDELs that met the following criteria were retained for further analysis: (1) identified by both algorithms, (2) a minor allele frequency >1%, (3) >1,500X sequencing depth at the mutant position, (4) there was no strand bias observed, (5) the average Phred-scaled base quality score for the alternative allele was ≥ Q25. For INDELs, the Q scores of the 10 flanking bases surrounding the variant position were examined and inspected manually. This procedure was carried out for both the unshifted and shifted reference sequences, and mutations within the defined calling regions were combined upon completion to produce the final variant set (Supplementary Fig. 2).

In total, 197 mutations (175 SNPs, 21 deletions and an insertion) were identified across the 22 CHO cell lines (Fig. 3 and Supplementary Tables 5–7). The SNPs identified corresponded to 99 nucleotide transitions (A↔G or C↔T) and 22 nucleotide transversions (A↔C, A↔T, G↔C or G↔T), yielding a 4.5:1 ts/tv ratio, similar to previous estimates for mutations in mammalian mtDNA (Belle et al., 2005). 130 (121 SNPs, 9 deletions and an insertion) individual variant nucleotide positions were detected in one or more of the CHO cell lines. The largest number of variants was identified in the ATCC DG44 and Biogen DG44 #1 samples with 30 and 21 mutations identified respectively (Fig. 4A & Supplementary Table 7) while the Biogen DG44 #2 cell line contained 9 mutated mtDNA positions. The least mutated cell line in the panel was the ATCC CHO-S cell line with only a single SNP identified while 6 SNPs were identified in the mitochondrial genome of the Biogen CHO-S cell line. Cell lines from the CHO-K1 lineages varied from 3 to as many as 12 variants. The number of variant positions in the Pfizer CHO-K1 cell lines derived from the same parental host ranged from 7 to 12. The remaining CHO-K1 cell lines all had less than 10 variant positions.

Of the 37 genes in the mitochondrial genome, 23 were found to have at least one variant position in one of the CHO cell lines sequenced (Fig. 4B). A SNP and an insertion were identified in the mtDNA D-loop in the ATCC DG44 and Pfizer CHO-K1 CLD1 #4 cell lines. The 16 S rRNA gene had the largest number of mutated positions with variants identified at 17 nucleotides. All of the 13 protein coding genes had at least one variant while 8 of 22 tRNAs harboured a mutation. CYTB and COX1 were found to have the largest number of mutations amongst protein coding genes while the ND3 gene had only one variant. A homoplasmic (MAF >99%) SNP in tRNAVal (m.1074 C>T) and the 16 S rRNA gene (m.2235 C>T) was identified in each of the 17 cell lines from the CHO-K1 lineage (Supplementary Table 7). The CHO cell lines from the S and DG44

lineages are identical to the Chinese hamster mtDNA reference sequence at these positions. The two Biogen DG44 cell lines had a shared homoplasmic mutation in tRNAVal (m1092A>G) yet this mutation was not detected in the ATCC DG44 cell line. The Biogen CHO S cell line contains the only homoplasmic mutation in a protein-coding gene, a SNP identified in CYTB (m.14311 C>T).

1.3.4 Identification of CHO cell line heteroplasmy

81% of all CHO cell line mutations identified in this study were heteroplasmic (i.e. the MAF>1% & <99%) with a minor allele frequency spanning from 1% to 96.2% (Fig. 4C). While the majority of heteroplasmic variants were identified in a single cell line, 11 of these mutations were shared in two or more cell lines (Table 2). The effectiveness of the dual mapping strategy is further demonstrated by the identification of two variants in the first 100 bp of the mitochondrial genome within tRNAPhe (m.62 C>T) and the 12 S rRNA gene (m.75 G>A). These low frequency heteroplasmies were not identified by the variant calling pipeline using the unshifted linear reference sequence. Next generation sequencing data from the Chinese hamster liver sample was also analysed using the variant detection pipeline. No heteroplasmic variants were identified in the Chinese hamster mitochondrial genome sequence. We did, however, identify a previously reported heteroplasmic insertion (Fan et al., 2012) and deletion (Bayona-Bafaluy et al., 2003) from the mouse liver mtDNA sequencing data using the mouse mtDNA reference sequence (GenBank accession: NC_005089) and the variant identification pipeline (Supplementary Table 8).

The mouse mtDNA sequencing data was utilised to demonstrate the effectiveness of the bioinformatics pipeline and estimate the potential influence of NumtS contamination on variant calling and heteroplasmy levels. Reads originating from mouse mtDNA sequencing were first analysed using an identical bioinformatics pipeline to the Chinese hamster and CHO cell line analyses, incorporating the shifted and unshifted reference sequence mapping. From this analysis, 3 known (Fan et al., 2012, Bayona-Bafaluy et al., 2003) mouse mtDNA variants were identified - a homoplasmic SNP (m.9461 T>C), a heteroplasmic deletion (m.5171delA) and an insertion (m.9820insAA) (Supplementary Table 8). To determine if these variants and their corresponding allele frequencies were influenced by NumtS, we stringently mapped reads against the mouse nuclear genome and extracted reads that aligned to known NumtS (Calabrese et al., 2012). The NumtS aligned reads were remapped to the mouse mtDNA reference sequence using the same alignment and variant pre-processing approach to that of Chinese hamster and CHO cell line data with the exception of variant calling thresholds that were modified to account for the lower depth of coverage. Reads were found to align predominantly to 5 regions in the mouse mtDNA reference sequence (Supplementary Fig. 3), which did not overlap with the 3 variants called on

the full dataset. Furthermore, no SNPs or INDELs were identified in the 5 regions where NumtS reads aligned indicating that NumtS contamination was not a contributing factor in either variant detection or heteroplasmy measurement.

1.3.5 Prediction of the effect of mitochondrial genome variations

Of the variants identified in this study, 62% (81/130) lie within protein coding regions of the mitochondrial genome. To determine the putative effects of these variants we utilised the snpEff (Cingolani et al., 2012) tool to annotate each mutation. Missense variants were the most common, accounting for 60.5% (49/81) of mutations in protein coding regions in comparison to 25% (18/81) of mutations predicted to be synonymous. PROVEAN predicted that 55% (27/49) of the amino acid substitutions arising from missense mutations would affect protein function (Supplementary Table 7). The remaining variants were predicted to result in frameshift mutations (7.5%) or premature stop codons (5%), with 1 start codon mutation identified. Of the 22 cell lines sequenced, 20 contained mutations that resulted in alteration of the protein coding sequence of at least 1 gene in the mitochondrial genome (Fig. 5A). Each protein coding gene harboured at least 1 mutation that altered the amino acid sequence in at least one sample (Fig. 5B). Frameshift mutations with a high probability of functional consequences were identified in ND1, COX1, ND4, ND5 and CYTB. Mutations resulting in a premature stop codon were observed in COX1, ND4L, ND6 and CYTB.

1.4 DISCUSSION

Analytical methods including metabolic flux analysis and metabolomics have shed light on CHO cell metabolic phenotypes, informing media design, feeding strategies and cell line engineering to increase the efficiency of industrial cell culture for biopharmaceutical production (Quek et al., 2010). Despite the clear utility of these methods, our understanding of the origins of variation in cellular metabolism in different CHO cell lines, clones and in some cases, between and over the course of production runs have not yet been completely unravelled (Young, 2013). Comparison of nuclear genome sequences has shown that CHO cell lineages harbour distinct mutations, and that millions of SNPs and INDELs arise during the development of a new cell line (Lewis et al., 2013). This plasticity of the CHO cell nuclear genome undoubtedly plays an underlying role in the range of bioprocess phenotype variation observed amongst CHO cell clones during cell line development. Eukaryotic cells also contain a separate non-nuclear polyploid genome within each mitochondrion. mtDNA sequence variants in the mitochondrial genome have been widely studied in human biology, and the association of specific mutations with a number of metabolic disorders is well established. In recent years, the development of massively parallel sequencing technology

has dramatically expanded our ability to study mitochondrial genomics and permitted analysis of low frequency mtDNA heteroplasmy. Here, we present the first survey of CHO cell mtDNA heterogeneity.

Considering the genomic instability of CHO cells, Chinese hamster mtDNA is an ideal common mitochondrial reference genome to compare sequence variants across cell lines. Assessment of the suitability of an existing Chinese hamster sequence available on GenBank (NC_007936) for this purpose revealed that the sequence is from either a CHOAL or CHO-K1 cell line (Partridge et al., 2007), not the Chinese hamster, and that a number of sequence features are incorrectly annotated. In order to ensure the accuracy of a reference sequence for comparison of cell lines, we sequenced mtDNA from the Chinese hamster. Comparative analysis of Chinese hamster mtDNA with the GenBank NC_007936 sequence identified 7 mutations comprising 4 SNPs and 3 INDELs within 4 protein-coding genes and the mitochondrial D-loop control region. The 16,283 bp sequence acquired in this study represents the first accurate reference for the analysis of the mitochondrial genome, providing an essential resource for future studies of CHO cell mtDNA.

To determine the prevalence of mutations in the CHO cell mitochondrial genome, 22 CHO cell lines derived from the CHO-K1, CHO-S and DG44 lineages, including industrial cell lines, engineered cell lines and clones generated from 2 cell line development projects were selected for mtDNA sequencing. The high depth of coverage achieved through massively parallel sequencing of the relatively small mitochondrial genome permits accurate identification of homoplasmic and heteroplasmic variants with minor allele frequencies as low as 1%. Established best practices were utilised to pre-process alignments before a conservative variant calling pipeline that required agreement between two algorithms was used for SNP and INDEL identification.

The heterogeneity of the CHO cell mitochondrial genome across the CHO cell lines analysed here is remarkable; cell lines were found to contain at least one to as many as 30 mutations in their mtDNA. We discovered both homoplasmic and heteroplasmic mutations at 130 nucleotide positions distributed across the entire mtDNA sequence with a total of 197 variants detected across the 22 CHO cell lines. A SNP or INDEL was identified in all 13 protein-coding genes, 8 tRNA genes, and both rRNA genes as well as within the D-loop region. Each of the protein coding genes were found to contain at least one heteroplasmic mutation in at least one cell line with CYTB, COX1 and ND5 each harbouring at least 10 distinct heteroplasmic variants. Of the 130 variant positions identified, only 4 (37 mutations across the 22 cell lines) were homoplasmic (MAF>99%). All 17 cell lines from the CHO-K1 lineage harboured a homoplasmic SNP within both tRNAVal (m.1074 C>T) and the 16 S rRNA (m.2235 C>T) gene. We also identified a

homoplasmic variant within tRNAVal (m.1092 A>G) in the two CHO-DG44 cell lines provided by Biogen, yet this mutation was not identified in the CHO-DG44 cell line sourced from the ATCC. The fourth homoplasmic variant was identified in the CYTB gene (m.14311 C>T) in the Biogen CHO-S cell line. CHO-K1 and CHO-S were both derived from the original Chinese hamster ovary tissue isolate before being sent to two different laboratories (Wurm, 2013). We did not detect the tRNAVal (m.1074 C>T) or 16 S rRNA (m.2235 C>T) mutations common to CHO-K1 cell lines in the CHO-S cell lines or the cell line sequenced in the Partridge et al. study (CHO-K1 cell or derived from CHOAL). In addition, the homoplasmic CYTB mutation (m.14311 C>T) observed in the Biogen CHO-S cell line was not identified in ATCC CHO-K1 or the ATCC CHO-S cell line. These findings indicate that CHO cell mtDNA mutations have arisen independently between cell lines following isolation from the original Chinese hamster tissue and over time, have become fixed in the mitochondrial genome. A homoplasmic tRNAVal (m.1092 A>G) mutation was identified in both CHO-DG44 cell lines provided by Biogen, and once again, this mutation was not present in the ATCC CHO-DG44. These mutations in CHO-DG44, which originated from a non-CHO-K1 or CHO-S lineage (Wurm, 2013), imply that homoplasmic mutations in CHO cell mtDNA can occur spontaneously and differ between CHO cell lines of the same lineage.

The overwhelming majority of mutations identified in this study were heteroplasmic and were detected at 126 individual nucleotide positions. Heteroplasmic mutations were identified in tRNA, rRNA and protein coding regions as well as the D-loop control region. Considering the widespread heteroplasmy identified in the mitochondrial genome of CHO cells, it was somewhat surprising that no heteroplasmy was detected the in Chinese hamster mitochondrial genome. A recent study by Li et al., (2015) reported tissue specific patterns of heteroplasmy in more than 150 humans across 12 tissue types and demonstrated that heteroplasmy is tissue-specific and, in some cases, tissues can be free of heteroplasmy while other tissues in the same individual can contain heteroplasmic mtDNA variants. It is also possible that very low frequency (MAF<1%) variations are present in the Chinese hamster liver tissue sample. While Li et al. utilised a >0.5% MAF threshold, we felt a MAF>1% threshold was an appropriate choice for this study, given the depth of mtDNA coverage obtained.

In comparison to homoplasmic mutations which tended to be lineage specific, 89% of heteroplasmic mutations were identified in a single cell line and ranged from 1% to 96%. For example, the 2 Biogen DG44 CHO cell lines shared a common homoplasmic SNP in tRNAVal (m.1092 A>G) yet no shared heteroplasmic variants were common to both cell lines. There were also no shared heteroplasmic mutations identified between several of the CHO-K1 cell lines (i.e. Biogen CHO-K1, Pfizer CHO-K1 2B6, Pfizer CHO-K1 114 cell lines, DCU CHO-K1 SEAP or the 4 DCU sponge transfected SEAP cell lines). Of those heteroplasmic mtDNA mutations that

were identified in more than one cell line, the majority were shared amongst clones originating from 2 distinct cell line development projects (CLD1 and CLD2).

Seven heteroplasmic variants were identified in 2 or more of the CLD1 clones while 4 heteroplasmies were found in 2 or more of the 3 CLD2 clones. Four shared heteroplasmic variants were identified in at least one clone from CLD1 and CLD 2. A SNP in tRNALys (m.7721 A>G) was identified in 1 clone from CLD1 (MAF=47%) and 2 clones from CLD2 (MAF=13% and MAF =9.5%). A missense variant (not predicted to be deleterious by PROVEAN) in CYTB (m.14849 G>A) was found in a single CLD1 clone and in 3 clones from CLD2. Three of the clones had an average mutation frequency of ~45% yet the MAF of the fourth clone was 1.5%, a marked difference from its counterpart clones within CLD2. The only mutation present in all 8 clones from CLD1 and CLD2 is a frameshift mutation in the CYTB gene (m.14136delA) with a MAF ranging from 7.7% to 52%. The m.14136delA mutation also appears in the ATCC CHO-K1 cell line (derived from an isolate of the original Chinese hamster ovary tissue) and it would seem that this mutation has been retained in the CLD project clones yet has been lost in the other CHO-K1 cell lines while the homoplasmic m.1074 C>T and m.2235 C>T mutations have become fixed in all CHO-K1 cell lines. The variation of mutation frequencies of heteroplasmic variants both within and across the two cell line development projects, following single cell cloning, is in line with the model of random assortment of mtDNA upon cell division (Jayaprakash et al., 2015).

It is not possible at this point to determine the impact of rRNA or tRNA mutations without further experimentation. Our results do, however, indicate that mtDNA mutations are not only widespread but also likely to influence mitochondrial function. Each of the protein coding genes in the mitochondrial genome were found to contain at least one SNP or INDEL in one cell line and CYTB, COX1 and ND5 each harboured at least 10 distinct mutations. PROVEAN analysis revealed that 55% of missense variants were predicted to affect biological function. A number of premature stop and frameshift mutations were also identified. For instance, CYTB was found to harbour a SNP (m.14378 G>A) predicted to result in a premature stop codon as well as a frameshift mutation (m.14136delA). The presence of these mtDNA mutations could play a role in CHO cell metabolic phenotype variation. The gain of stop and frameshift mutation in the CYTB gene are predicted to result in a loss of function and could lead to a diminished efficiency in mitochondrial aerobic respiration in the electron transport system centred on complex III. Weakened oxidative phosphorylation could signal reprogramming of cellular metabolism to rely more heavily on glycolysis (Zheng, 2012) and maintain cellular energy balance. Exclusive reliance on glycolytic metabolism, despite being energetically inefficient, has been shown to be associated with elevated cell growth in both cancer and CHO cells due to the intermediate metabolites of glycolysis that contribute to biomass accumulation (Templeton et al., 2013). All

but one protein coding variant identified were heteroplasmic and in some cases, loss of function mutations spanned a wide range. Mutation frequencies were found to vary by as much as 1.5–50% (m.3205delCT), indicating that the potential effects of these mutations could lead to a continuous distribution of phenotypes.

While the results of this study demonstrate the heterogeneity of CHO cell mtDNA, the polyploid nature of the mitochondrial genome presents considerable challenges to understanding the impact of mitochondrial mutations. A mtDNA variant might be spread across the entire population or confined to a subpopulation of cells. At the subcellular level, the mutation might be distributed across multiple mitochondria or indeed confined to a limited number of mitochondria. The emergence of new methods to sequence mtDNA at the single cell level (Jayaprakash et al., 2015) will play a valuable role in future studies and increase our understanding of the implications of a particular mutation in a CHO cell population. It will also be important to integrate MFA and metabolomics analysis with mtDNA sequencing to understand the biochemical threshold at which individual mutations affect CHO cell behaviour. The knowledge gained in doing so has the potential to enable precise cell line selection and ultimately rational genetic engineering to improve CHO cell phenotypes. For example, a recent study by Pieper et al., (in press) demonstrated that a better understanding of the role of a non-coding RNA (mitosRNA-1978) encoded in mitochondrial genome identified an genetic engineering route that improved CHO cell productivity. The development of mtDNA CRISPR-Cas9 based methods (Jo et al., 2015) and mitoTALENS (Bacman et al., 2013) for site specific mitochondrial genome editing will also provide routes to rational engineering of CHO cell mitochondria to enhance the metabolic performance of CHO cells for biopharmaceutical manufacture.

1.5 CONCLUSIONS

Widespread heteroplasmy in the CHO cell mitochondrial genome raises intriguing questions about the genetics and selection of mitochondrial mutations in CHO cells during cell culture and cell line development for biopharmaceutical production. Closely related clones derived from the same parental host and even originating from the same cell line development project can harbour distinct heteroplasmic variations. These variations in the mitochondrial genome are likely to affect mitochondrial function and could play a role in cell to cell, production run to production run and indeed clone to clone variation observed in CHO cell culture and cell line development. The combination of mtDNA sequencing with established techniques in metabolic flux analysis and metabolomics will be necessary to associate these mutations with desirable or undesirable CHO cell metabolism. The understanding of mtDNA variation could lead to new approaches to cell line screening and ultimately engineering of CHO cell mtDNA for more productive metabolic phenotypes.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge funding from Science Foundation Ireland (grant refs: 13/SIRG/2084, 13/IA/1841and 13/IA/1963) and the eCHO systems Marie Curie ITN programme (grant ref: 642663). The authors would also to acknowledge Lin Zhang (Pfizer Inc.), Scott Estes, Chapman Wright and Brian St. Germaine (Biogen Inc.) for access to cell lines and comments on the manuscript.

1.6 REFERENCES

S. Anderson, A.T. Bankier, B.G. Barrell, M.H.L. de Bruijn, et al. Sequence and organization of the human mitochondrial genome Nature, 290 (1981), pp. 457-465

S.R. Bacman, S.L. Williams, M. Pinto, S. Peralta, C.T. Moraes Specific elimination of mutant mitochondrial genomes in patient-derived cells by mitoTALENs Nat. Med., 19 (2013), pp. 1111-1113

B.G. Barrell, S. Anderson, A.T. Bankier, M.H. de Bruijn, et al. Different pattern of codon recognition by mammalian mitochondrial tRNAs Proc. Natl. Acad. Sci. USA, 77 (1980), pp. 3164-3166

M.P. Bayona-Bafaluy, R. Acín-Pérez, J.C. Mullikin, J.S. Park, et al. Revisiting the mouse mitochondrial DNA sequence Nucleic Acids Res., 31 (2003), pp. 5349-5355

E.M.S. Belle, G. Piganeau, M. Gardner, A. Eyre-Walker An investigation of the variation in the transition bias among various animal mitochondrial DNA Gene, 355 (2005), pp. 58-66

M. Bernt, A. Donath, F. Jühling, F. Externbrink, *et al.* MITOS: improved de novo metazoan mitochondrial genome annotation Mol. Phylogenet. Evol., 69 (2013), pp. 313-319

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

A.M. Bolger, M. Lohse, B. UsadelTrimmomatic: a flexible trimmer for Illumina sequence data Bioinformatics, 30 (2014), pp. 2114-2120

K. Brinkrolf, O. Rupp, H. Laux, F. Kollin, *et al.* Chinese hamster genome sequenced from sorted chromosomes Nat. Biotech., 31 (2013), pp. 694-695

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

F.M. Calabrese, D. Simone, M. Attimonelli Primates and mouse NumtS in the UCSC Genome Browser BMC Bioinforma., 13 (Suppl 4) (2012), p. S15

S. Calvo, M. Jain, X. Xie, S.A. Sheth, *et al.* Systematic identification of human mitochondrial disease genes through integrative genomics Nat. Genet., 38 (2006), pp. 576-582

Y. Cao, S. Kimura, T. Itoi, K. Honda, *et al.* Construction of BAC-based physical map and analysis of chromosome rearrangement in chinese hamster ovary cell lines Biotechnol. Bioeng., 109 (2012), pp. 1357-1367

Y. Choi, G.E. Sims, S. Murphy, J.R. Miller, A.P. Chan Predicting the functional effect of amino acid substitutions and indels PLoS One, 7 (2012), p. e46688

P. Cingolani, A. Platts, L.L. Wang, M. Coon, *et al.* A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3 Fly. (Austin), 6 (2012), pp. 80-92

A.B. Diendorfer, M. Hackl, G. Klanert, V. Jadhav, *et al.* Annotation of additional evolutionary conserved microRNAs in CHO cells from updated genomic data Biotechnol. Bioeng. (2015)

J. Ding, C. Sidore, T.J. Butler, M.K. Wing, *et al.* Assessing mitochondrial DNA variation and copy number in lymphocytes of ~2000 sardinians using tailored sequencing analysis tools PLOS Genet, 11 (2015), p. e1005306

W. Fan, C.S. Lin, P. Potluri, V. Procaccio, D.C. Wallacemt DNA lineage analysis of mouse L-cell lines reveals the accumulation of multiple mtDNA mutants and intermolecular recombination Genes Dev., 26 (2012), pp. 384-394

M.P. Gerstl, M. Hackl, A.B. Graf, N. Borth, J. Grillari Prediction of transcribed PIWI-interacting RNAs from CHO RNAseq data J. Biotechnol., 166 (2013), pp. 51-57

A. Gilbert, K. McElearney, R. Kshirsagar, M.S. Sinacore, T. Ryll Investigation of metabolic variability observed in extended fed batch cell culture Biotechnol. Prog., 29 (2013), pp. 1519-1527

Y. Guo, J. Li, C.-I. Li, J. Long, *et al.* The effect of strand bias in Illumina short-read sequencing data BMC Genom., 13 (2012), p. 666

C. Hahn, L. Bachmann, B. Chevreux Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads—a baiting and iterative mapping approach Nucleic Acids Res., 41 (2013), p. e129

A.D. Jayaprakash, E.K. Benson, S. Gone, R. Liang, *et al.* Stable heteroplasmy at the single-cell level is facilitated by intercellular exchange of mtDNA Nucleic Acids Res. (2015), p. gkv052

A. Jo, S. Ham, G.H. Lee, Y.-I. Lee, *et al.* Efficient mitochondrial genome editing by CRISPR/Cas9, efficient mitochondrial genome editing by CRISPR/Cas9 BioMed. Res. Int. BioMed. Res. Int. (2015), p. e305716

R.S. Just, J.A. Irwin, W. Parson Mitochondrial DNA heteroplasmy in the emerging field of massively parallel sequencing Forensic Sci. Int. Genet., 18 (2015), pp. 131-139

C.S. Kaas, C. Kristensen, M.J. Betenbaugh, M.R. AndersenSequencing the CHO DCB11 genome reveals regional variations in genomic stability and haploidy BMC Genom., 16 (2015), p. 160

D.C. Koboldt, Q. Zhang, D.E. Larson, D. Shen, *et al.* VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing Genome Res., 22 (2012), pp. 568-576

M.I. Krzywinski, J.E. Schein, I. Birol, J. Connors, *et al.* Circos: an information aesthetic for comparative genomics Genome Res. (2009)

D. Laslett, B. Canbäck ARWEN: a program to detect tRNA genes in metazoan mitochondrial nucleotide sequences Bioinformatics, 24 (2008), pp. 172-175

N.E. Lewis, X. Liu, Y. Li, H. Nagarajan, *et al.* Genomic landscapes of Chinese hamster ovary cell lines as revealed by the Cricetulus griseus draft genome Nat. Biotech., 31 (2013), pp. 759-765

F. Li, N. Vijayasankaran, A. (Yijuan) Shen, R. Kiss, A. Amanullah Cell culture processes for monoclonal antibody production mAbs, 2 (2010), pp. 466-477

H. Li, R. Durbin Fast and accurate short read alignment with Burrows-Wheeler transform Bioinforma. Oxf. Engl., 25 (2009), pp. 1754-1760

H. Li, B. Handsaker, A. Wysoker, T. Fennell, *et al.* The sequence alignment/Map format and SAMtools Bioinformatics, 25 (2009), pp. 2078-2079

M. Li, R. Schröder, S. Ni, B. Madea, M. Stoneking Extensive tissue-related and allele-related mtDNA heteroplasmy suggests positive selection for somatic mutations Proc. Natl. Acad. Sci., 112 (2015), pp. 2491-2496

C.-W. Liou, T.-K. Lin, J.-B. Chen, M.-M. Tiao, *et al.* Association between a common mitochondrial DNA D-loop polycytosine variant and alteration of mitochondrial copy number in human peripheral blood cells J. Med. Genet., 47 (2010), pp. 723-728

T. Magoč, S.L. Salzberg FLASH: fast length adjustment of short reads to improve genome assemblies Bioinforma. Oxf. Engl., 27 (2011), pp. 2957-2963

A. McKenna, M. Hanna, E. Banks, A. Sivachenko, *et al.* The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data Genome Res., 20 (2010), pp. 1297-1303

P. Meleady, R. Hoffrogge, M. Henry, O. Rupp, *et al.* Utilization and evaluation of CHO-specific sequence databases for mass spectrometry-based proteomics Biotechnol. Bioeng., 109 (2012), pp. 1386-1394

F.J. Miller, F.L. Rosenfeldt, C. Zhang, A.W. Linnane, P. Nagley Precise determination of mitochondrial DNA copy number in human skeletal and cardiac muscle by a PCR-based assay: lack of change of copy number with age Nucleic Acids Res., 31 (2003), p. e61

M.A. Partridge, M.M. Davidson, T.K. Hei The complete nucleotide sequence of Chinese hamster (Cricetulus griseus) mitochondrial DNA DNA Seq. J. DNA Seq. Mapp., 18 (2007), pp. 341-346

B.A.I. Payne, I.J. Wilson, P. Yu-Wai-Man, J. Coxhead, *et al.* Universal heteroplasmy of human mitochondrial DNA Hum. Mol. Genet., 22 (2013), pp. 384-390

Pieper, L.A., Strotbek, M., Wenger, T., Gamer, M., et al., Secretory pathway optimization of CHO producer cells by co-engineering of the mitosRNA-1978 target genes CerS2 and Tbc1D20. Metab. Eng., in press.

L.-E. Quek, S. Dietmair, J.O. Krömer, L.K. Nielsen Metabolic flux analysis in mammalian cell culture Metab. Eng., 12 (2010), pp. 161-171

W. Quispe-Tintaya, R.R. White, V.N. Popov, J. Vijg, A.Y. Maslov Fast mitochondrial DNA isolation from mammalian cells for next-generation sequencing BioTechniques, 55 (2013), pp. 133-136

R. Rizzuto, D. De Stefani, A. Raffaello, C. Mammucari Mitochondria as sensors and regulators of calcium signalling Nat. Rev. Mol. Cell Biol., 13 (2012), pp. 566-578

O. Rupp, J. Becker, K. Brinkrolf, C. Timmermann, *et al.* Construction of a Public CHO Cell Line Transcript Database Using Versatile Bioinformatics Analysis Pipelines PLoS One, 9 (2014), p. e85568

N. Templeton, J. Dean, P. Reddy, J.D. Young Peak antibody production is associated with increased oxidative metabolism in an industrially relevant fed-batch CHO cell culture Biotechnol. Bioeng., 110 (2013), pp. 2013-2024

Van der Auwera, G.A., Carneiro, M.O., Hartl, C., Poplin, R., et al., 2002 Curr. Protoc. Bioinforma., John Wiley & Sons, Inc.

G.A. Van der Auwera, M.O. Carneiro, C. Hartl, R. Poplin, *et al.* From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline Curr. Protoc. Bioinforma., 43 (2013), pp. 11.10.1-11.10.33

M.G. Vander Heiden, L.C. Cantley, C.B. Thompson Understanding the warburg effect: the metabolic requirements of cell proliferation Science, 324 (2009), pp. 1029-1033

M. Wachsmuth, A. Hübner, M. Li, B. Madea, M. Stoneking Age-Related and Heteroplasmy-Related Variation in Human mtdna Copy Number PLOS Genet., 12 (2016), p. e1005939

C. Wang, R.J. Youle The role of mitochondria in apoptosis Annu. Rev. Genet., 43 (2009), pp. 95-118

A. Wilm, P.P.K. Aw, D. Bertrand, G.H.T. Yeo, *et al.* LoFreq: a sequence-quality aware, ultra-sensitive variant caller for uncovering cell-population heterogeneity from high-throughput sequencing datasets Nucleic Acids Res., 40 (2012), pp. 11189-11201

A. Wippermann, O. Rupp, K. Brinkrolf, R. Hoffrogge, T. Noll The DNA methylation landscape of Chinese hamster ovary (CHO) DP-12 cells J. Biotechnol. (2015)

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

F.M. Wurm CHO Quasispecies—Implications for manufacturing processes Processes, 1 (2013), pp. 296-311

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

D. Young Metabolic flux rewiring in mammalian cell cultures Curr. Opin. Biotechnol. (2013), p. 24

F. Zagari, M. Jordan, M. Stettler, H. Broly, F.M. Wurm Lactate metabolism shift in CHO cell culture: the role of mitochondrial oxidative activity New Biotechnol., 30 (2013), pp. 238-245

J. Zheng Energy metabolism of cancer: glycolysis versus oxidative phosphorylation (Review) Oncol. Lett., 4 (2012), pp. 1151-1157

Table 1. Comparison of the Chinese hamster mitochondrial genome to the previously sequenced CHO cell line. 7 variants were identified within 4 protein coding genes as well as the D-loop control region. The position, mtDNA feature, variant type and nucleotide change observed are shown for each mutation. Chinese hamster reads were mapped against the CHO cell line mtDNA sequence (GenBank accession no. NC_007936) to determine the number of reads supporting each variant. All variants were homoplasmic (allele frequency >99%).

Position	Feature	Variant type	CHO Cell allele	C.griseusallele	# reads supporting variant
9306	COX3	SNP	A	G	5,201
11399	ND4	SNP	C	A	8,765
12683	ND5	SNP	G	A	11,464
13436	ND6	Deletion	AT	A	8,815
13456	ND6	Insertion	C	CA	8,461
15717	D-loop	Deletion	GC	C	7,522
15888	D-loop	SNP	G	T	8,056

Table 2. Heteroplasmic mitochondrial variants identified in 2 or more CHO cell lines. In total 11 individual heteroplasmic mutations were identified in two or more cell lines. A number of heteroplasmic sites were found in cells originating from 2 cell line development projects. A heteroplasmic frameshift mutation in Cytb was identified in all of the clones analysed.

Gene	Position	Alelle		Type	Minor Alelle Freq. (%)		Effect	CHO Cell lines
		Ref	Alt		Min	Max		
s-rRNA	105	G	A	SNP	1.7	3.0	Non-coding	Pfizer CHO-K1 CLD1 #1 Pfizer CHO-K1 CLD1 #2 Pfizer CHO-K1 CLD1 #4
l-rRNA	1575	G	A	SNP	1.2	50.0	Non-coding	Pfizer CHO-K1 CLD2 #3 ATCC DGG44
l-rRNA	2151	C	T	SNP	1.2	9.9	Non-coding	Pfizer CHO-K1 CLD1 #1 Pfizer CHO-K1 CLD1 #2 Pfizer CHO-K1 CLD1 #3 Pfizer CHO-K1 CLD1 #4
ND1	3205	GCT	G	Deletion	1.5	50.0	Frameshift variant	Pfizer CHO-K1 CLD1 #5 Pfizer CHO-K1 CLD2 #2 Pfizer CHO-K1 CLD2 #3
COX2	6996	G	A	SNP	14.0	26.0	Initiation codon variation	Pfizer CHO-K1 CLD1 #1 Pfizer CHO-K1 CLD1 #2 Pfizer CHO-K1 CLD1 #3

								Pfizer CHO-K1 CLD1 #4
tRNA^{Lys}	7721	A	G	SNP	9.5	47.0	Non-coding	Pfizer CHO-K1 CLD1 #5
								Pfizer CHO-K1 CLD2 #1
								Pfizer CHO-K1 CLD2 #3
ATP6	8067	C	T	SNP	1.8	6.8	Missense variation	Pfizer CHO-K1 CLD1 #1
								Pfizer CHO-K1 CLD1 #2
								Pfizer CHO-K1 CLD1 #3
								Pfizer CHO-K1 CLD1 #4
ND4	11431	TCA	T	Deletion	1.7	2.0	Frameshift variant	Biogen CHO-K1
								ATCC DGG44
								Biogen DG44 #2
tRNA^{Leu}	11659	AG	A	Deletion	1.2	2.8	Non-coding	Pfizer CHO-K1 CLD1 #1
								Pfizer CHO-K1 CLD1 #2
CYTB	14136	GA	G	Deletion	7.7	52.0	Frameshift variant	ATCC CHO-K1
								Pfizer CHO-K1 CLD1 #1
								Pfizer CHO-K1 CLD1 #2
								Pfizer CHO-K1 CLD1 #3
								Pfizer CHO-K1 CLD1 #4
								Pfizer CHO-K1 CLD1 #5
								Pfizer CHO-K1 CLD2 #1
								Pfizer CHO-K1 CLD2 #2

							Pfizer CHO-K1 CLD2 #3
14378	G	A	SNP	11.0	36.0	Premature stop codon	Pfizer CHO-K1 CLD1 #1
							Pfizer CHO-K1 CLD1 #2
							Pfizer CHO-K1 CLD1 #3
							Pfizer CHO-K1 CLD1 #4
14849	G	A	SNP	1.5	53.0	Missense variation	Pfizer CHO-K1 CLD1 #5
							Pfizer CHO-K1 CLD2 #1
							Pfizer CHO-K1 CLD2 #2
							Pfizer CHO-K1 CLD2 #3
15136	C	A	SNP	1.7	19.0	Missense variation	Pfizer CHO-K1 CLD1 #3
							Biogen DG44 #1

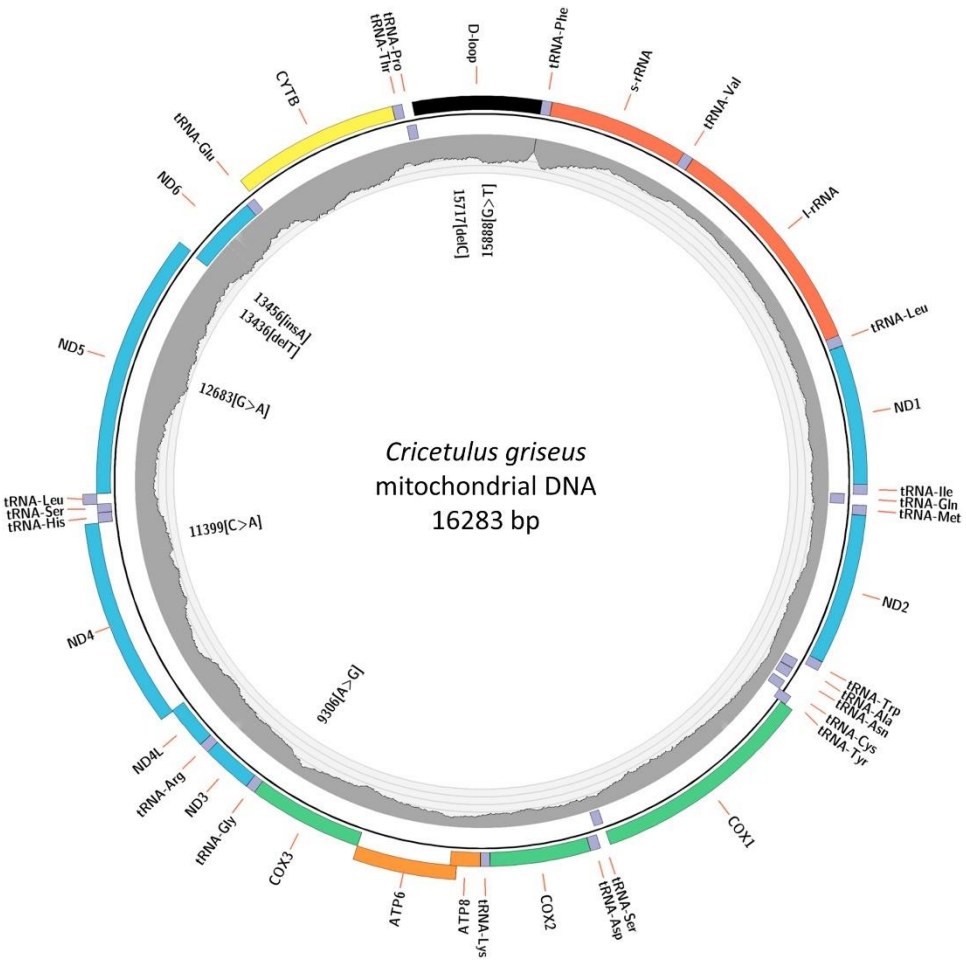


Fig. 1. **TheCricetulus griseusmitochondrial genome.** mtDNA was extracted from Chinese hamster liver tissue, deep sequenced and reconstructed using the MITOBIM algorithm. The resulting 16,283 bp mitochondrial DNA sequence had an average depth of coverage of ~6,417X (the depth of coverage at each base is shown as a grey histogram within the inner circle of the plot). 13 proteins, 22 tRNAs, 2 ribosomal RNAs and the non-coding D-loop control region are encoded by the Chinese hamster mitochondrial genome. 28 genes are present on the H-strand while 9 genes are on the L-strand of the mtDNA. Comparison of the Chinese hamster mtDNA sequence to the previously sequenced CHO cell line sequence on GenBank (accession no. NC_007936) revealed 7 mutations (4 SNPs and 3 INDELs) in protein coding genes (COX3, ND4, ND5, ND6) and the D-loop. The plot was generated using circos v.0.67 (Krzywinski et al., 2009).

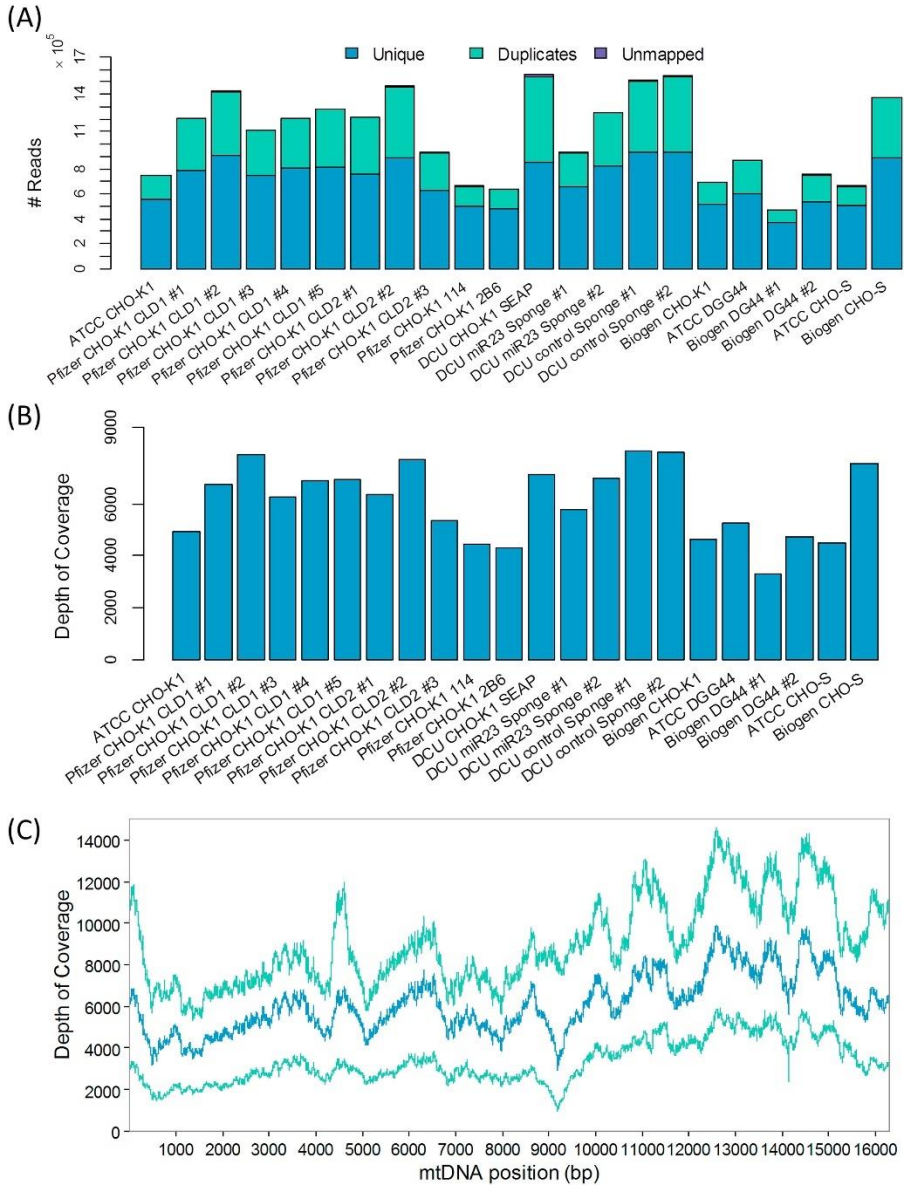


Fig. 2. High resolution sequencing of the CHO cell mitochondrial genome. (A) Number of mapped reads for the 22 CHO cell lines illustrating the number of unmapped reads (MAPQ <20), number of PCR duplicates and number of uniquely mapped reads. Note: a negligible difference in the total numbers of reads mapped against the shifted reference sequence were observed (Supplementary Table 4B & 4C). (B) Average sequencing depth for each CHO cell line sequenced. The average depth of coverage combines both variant calling regions within the unshifted and shifted reference sequences used for variant detection. (C) The mean (blue), minimum and maximum depth of coverage at each base position across the 22 CHO cell mitochondrial genomes sequenced. The utilisation of the dual mapping approach resulted in a 1,128X and 1,079X increase in coverage for the first and last 100 bp of the mtDNA reference sequence respectively.

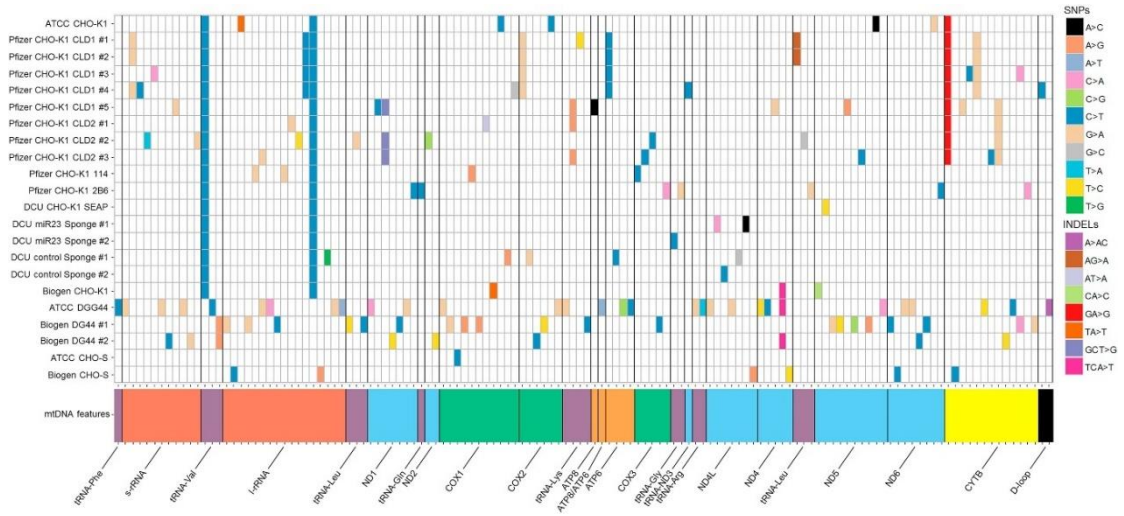


Fig. 3. The mutational landscape of CHO cell line mitochondrial genomes. A total of 197 mutations (175 SNPs and 22 INDELs) were identified for 22 CHO cell line samples at 130 positions in the mtDNA sequence. The nucleotide alternations detected for each cell line are shown along with the mtDNA feature harbouring the mutation. ATP8/ATP6 corresponds to the region of overlap between these genes. All protein coding genes along with the D-loop and 8 tRNA genes had a least 1 variant in 1 CHO cell line.

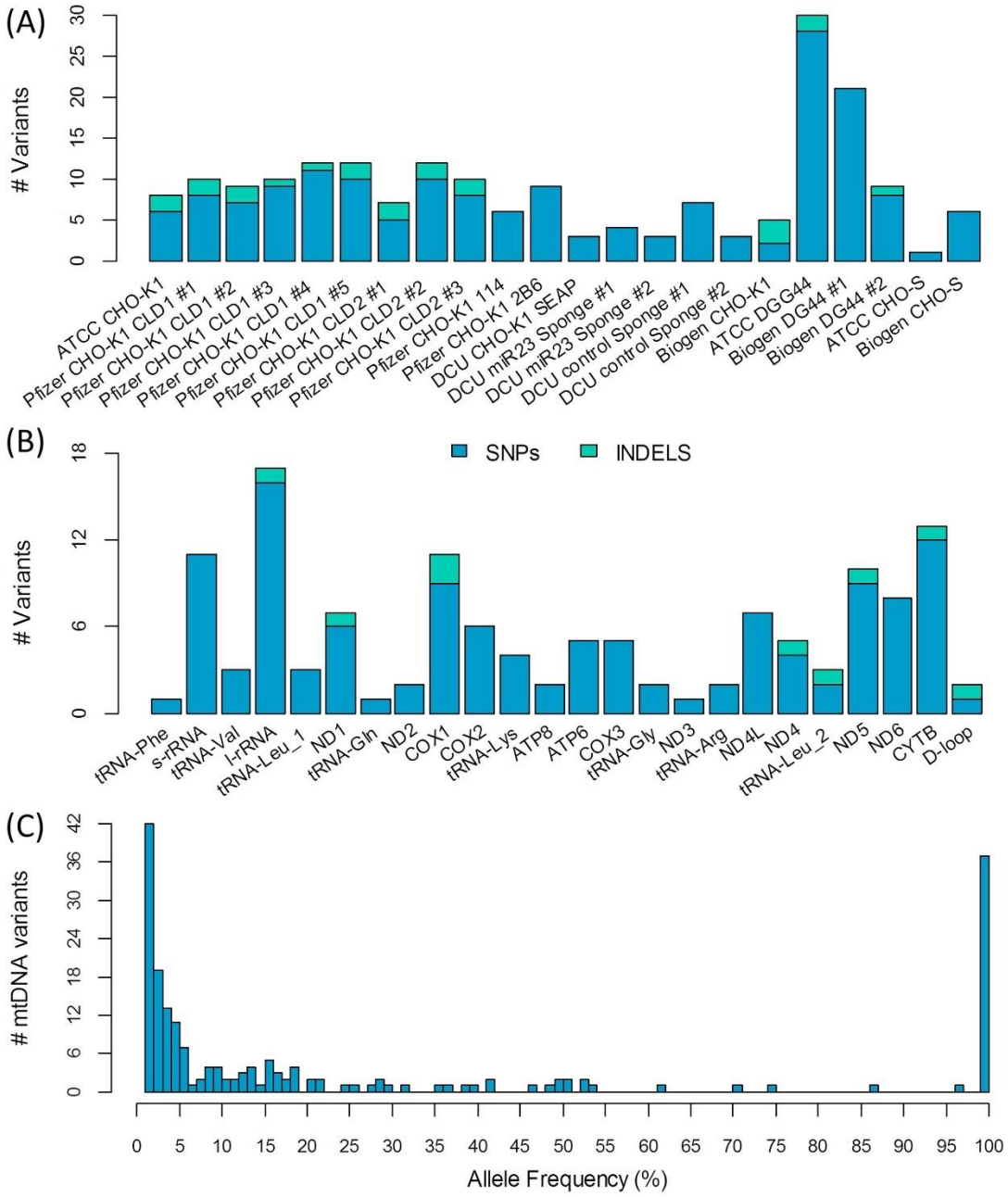


Fig. 4. Mitochondrial heteroplasmy is widespread in CHO cell lines. (A) Number of SNPs and INDELs identified for the 22 CHO cell line mitochondrial genomes sequenced. The mtDNA of 2 of the 3 DG44 cell lines sequenced were found to have the most variant positions while the ATCC CHO-S cell line had only a single mutated position. (B) Number of variant positions for the 24 genes found to harbour at least one mutation across the 22 CHO cell lines. (C) The majority of variants identified were heteroplasmic −160 out of 197 mtDNA mutations had an allele frequency <99%.

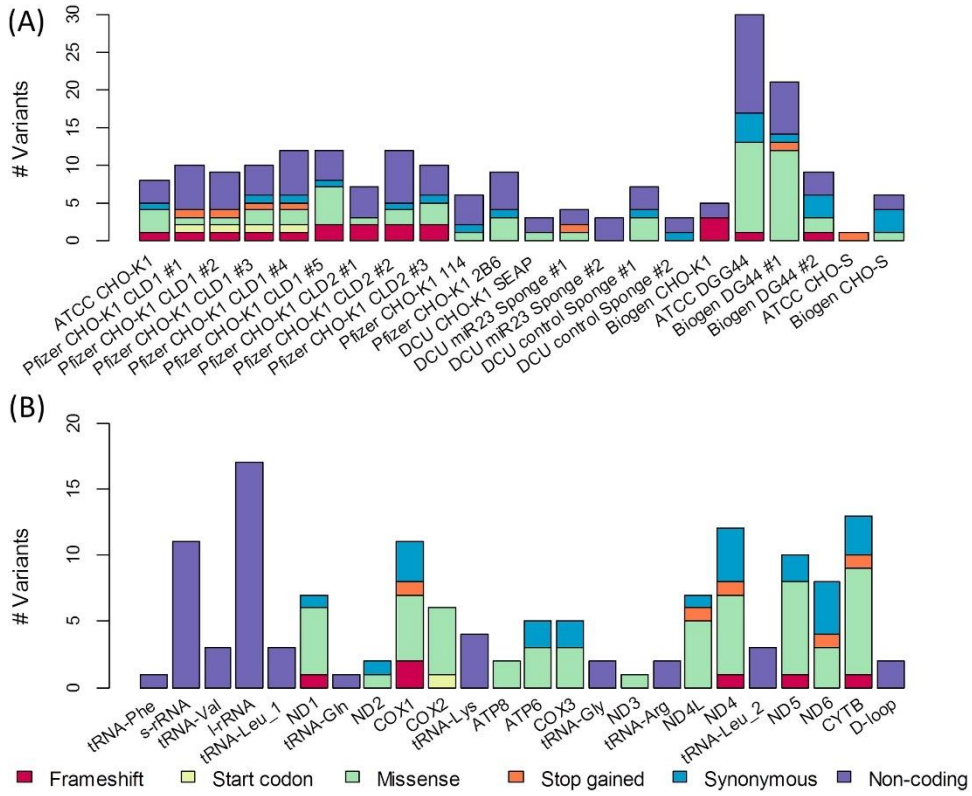


Fig. 5. Prediction of the effects of CHO cell mitochondrial genome mutations. The majority of mutations identified in this study occurred in protein coding genes. (A) 22 of the CHO cell lines sequenced contained at least one mutation that altered the protein sequence. (B) Each protein coding gene harboured a mutation that altered the amino acid sequence in at least one of the cell lines.

1.7 APPENDIX A

Supporting information associated with Chapter 1.

Supplementary Table 1: Sample Information. The origin of each sample sequenced in this study along with the respective file names corresponding to mtDNA sequencing data.

Sample ID	Lineage/Strain	Product	Sample Type
ATCC CHO-K1	CHO-K1	Non-producer	Cell line
Pfizer CHO-K1 CLD1 #1	CHO-K1	mAb	Cell line
Pfizer CHO-K1 CLD1 #2	CHO-K1	mAb	Cell line
Pfizer CHO-K1 CLD1 #3	CHO-K1	mAb	Cell line
Pfizer CHO-K1 CLD1 #4	CHO-K1	mAb	Cell line
Pfizer CHO-K1 CLD1 #5	CHO-K1	mAb	Cell line
Pfizer CHO-K1 CLD2 #1	CHO-K1	mAb	Cell line
Pfizer CHO-K1 CLD2 #2	CHO-K1	mAb	Cell line
Pfizer CHO-K1 CLD2 #3	CHO-K1	mAb	Cell line
Pfizer CHO-K1 2B6	CHO-K1	mAb	Cell line
Pfizer CHO-K1 114	CHO-K1	mAb	Cell line
DCU CHO-K1 SEAP	CHO-K1	SEAP	Cell line
DCU miR23 Sponge #1	CHO-K1	SEAP	Cell line
DCU miR23 Sponge #2	CHO-K1	SEAP	Cell line
DCU control Sponge #1	CHO-K1	SEAP	Cell line
DCU control Sponge #2	CHO-K1	SEAP	Cell line
Biogen CHO-K1	CHO-K1	Fc-fusion	Cell line
ATCC DGG44	CHO-DG44	Non-producer	Cell line
Biogen DG44 #1	CHO-DG44	Fc-fusion	Cell line
Biogen DG44 #2	CHO-DG44	mAb	Cell line
ATCC CHO-S	CHO-S	mAb	Cell line
Biogen CHO-S	CHO-S	Fc-fusion	Cell line
Chinese hamster	Outbred Chinese hamter	NA	Liver Tissue
Mouse	CB17/Icr-Prkdcscid/Crl	NA	Liver Tissue

Supplementary Table 2: Primer sequences for fragment amplification of the CHO cell mitochondrial genome as well as mtDNA isolated from Chinese hamster and mouse mtDNA. Primers were designed for PCR amplification of genomic CHO DNA using the CHOgenome.org database. A beta-actin control was designed against the whole CHO genome sequence with the reverse primers spanning an intron/exon junction. A positive control for the CHO mitochondrial genome was designed using the available CHO-K1 mitochondrial genome sequence on CHOgenome.org. Primer pairs (#1 and #2) were also designed using the available CHO-K1 mitochondrial genome sequence for high-fidelity PCR of the mitochondrial genomic DNA fragments. Finally, to account for possible sequence variation between the CHO mt-DNA sequence and the Chinese hamster mt-DNA sequence, primer sets (CH1-6/8 and CH2-86/7) were designed used the available CHO-K1 mt-DNA sequence available and matched for 100% sequence similarity with both the mouse (mmu) and rat (rno) mt-DNA sequences.

Genomic/Transcriptomic	Primer ID	Gene ID	Direction
Genomic PC	gB-act CHO-F	β-actin	Forward
	gB-act CHO-R	β-actin	Reverse
Genomic Mito PC	gMt posCHO-F	st-470	Forward
	gMt posCHO-R	st-598	Reverse
Mt genome fragment #1	pMt CHO #1-F	st-491	Forward
	pMt CHO #1-R	st-9304	Reverse
Mt genome fragment #2	pMt CHO #2-F	st-9180	Forward
	pMt CHO #2-R	st-598	Reverse
Conserved primer CH1-6/8	pMt cgr/mmu/rra For-1	st-577	Forward
	pMt cgr/mmu/rra Rev-1	st-8584	Reverse
Conserved primer CH2-86/7	pMt cgr/mmu/rra For-2	st-8582	Forward
	pMt cgr/mmu/rra Re2	st-686	Reverse

Supplementary Table 3: Chinese hamster mtDNA annotation. Coordinates for each of the 38 annotated features in the *C.griseus* mitochondrial genome.

Feature	Type	Start	Stop	Strand
tRNA-Phe	tRNA	1	69	+
s-rRNA	rRNA	71	1023	+
tRNA-Val	tRNA	1024	1095	+
l-rRNA	rRNA	1096	2657	+
tRNA-Leu_1	tRNA	2658	2732	+
ND1	CDS	2733	3687	+
tRNA-Ile	tRNA	3688	3756	+
tRNA-Gln	tRNA	3754	3824	-
tRNA-Met	tRNA	3829	3897	+
ND2	CDS	3898	4930	+
tRNA-Trp	tRNA	4931	4997	+
tRNA-Ala	tRNA	5000	5069	-
tRNA-Asn	tRNA	5072	5142	-
tRNA-Cys	tRNA	5174	5241	-
tRNA-Tyr	tRNA	5241	5308	-
COX1	CDS	5310	6854	+
tRNA-Ser_1	tRNA	6852	6920	-
tRNA-Asp	tRNA	6924	6992	+
COX2	CDS	6994	7677	+
tRNA-Lys	tRNA	7681	7744	+
ATP8	CDS	7746	7949	+
ATP6	CDS	7907	8587	+
COX3	CDS	8587	9370	+
tRNA-Gly	tRNA	9371	9438	+
ND3	CDS	9439	9786	+
tRNA-Arg	tRNA	9788	9855	+
ND4L	CDS	9857	10153	+
ND4	CDS	10147	11524	+
tRNA-His	tRNA	11525	11592	+
tRNA-Ser_2	tRNA	11593	11651	+
tRNA-Leu_2	tRNA	11651	11720	+
ND5	CDS	11721	13541	+
ND6	CDS	13525	14049	-
tRNA-Glu	tRNA	14050	14118	-
CYTB	CDS	14123	15265	+
tRNA-Thr	tRNA	15267	15333	+
tRNA-Pro	tRNA	15337	15403	-
D-loop	Non-coding	15404	16283	+

Supplementary Table 4: Pre-processed read counts, mapping rates and average depth of coverage for CHO cell line sequencing. The number of reads remaining following pre-processing along mapping and average coverage rate for each sample following shifted and unshifted sequence alignment.

Tables	
Table 4A	Preprocessed Read Counts
Table 4B	Unshifted reference mtDNA
Table 4C	Shifted reference mtDNA
Table 4D	Unshifted reference mtDNA
Table 4E	Shifted reference mtDNA

Supplementary Table 5: SNP calling outputs. Variant calling outs for 22 individual CHO cell lines including the depth at coverage at each SNP, number of forward and reverse reads supporting the reference and alternative allele, average base quality (Q) for each nucleotide at the SNP position and snpEff annotation.

Column label	Description
Position	Variant location on the reference mtDNA
Reference Allele	Nucleotide at reference location
Alternative Allele	Mutant nucleotide
Depth	Depth at the SNP position
Allele Frequency	Mutant nucleotide frequency
Forward Ref Allele	Number of reads in the forward direction supporting the reference n
Reverse Ref Allele	Number of reads in the reverse direction supporting the reference nu
Forward Alternative Allele	Number of reads in the forward direction supporting the mutant nuc
Reverse Alternative Allele	Number of reads in the reverse direction supporting the mutant nucl
Strand Bias	Overpresentation of mutation in either forward or reverse direction
A(Q)	Average Phred scaled quality score for A nucleotides
C(Q)	Average Phred scaled quality score for C nucleotides
G(Q)	Average Phred scaled quality score for G nucleotides
T(Q)	Average Phred scaled quality score for T nucleotides
snpEff Annotation	Putative effect of SNP predicted by snpEff

Supplementary Table 6: INDEL detection data. Variant calling outputs for 22 individual CHO cell lines including the depth at coverage at each INDEL, number of forward and reverse reads supporting the reference and alternative allele, average base quality (Q) at the INDEL position as well 5 upstream and downstream flanking regions nucleotide and the snpEff annotation.

Column label
Position
Reference Allele
Alternative Allele
Depth
Allele Frequency
Forward Ref Allele
Reverse Ref Allele
Forward Alternative Allele
Reverse Alternative Allele
Strand Bias
-5(Q)
-4(Q)
-3(Q)
-2(Q)
-1(Q)
INDEL(Q)
+1(Q)
+2(Q)
+3(Q)
+4(Q)
+5(Q)
snpEff Annotation

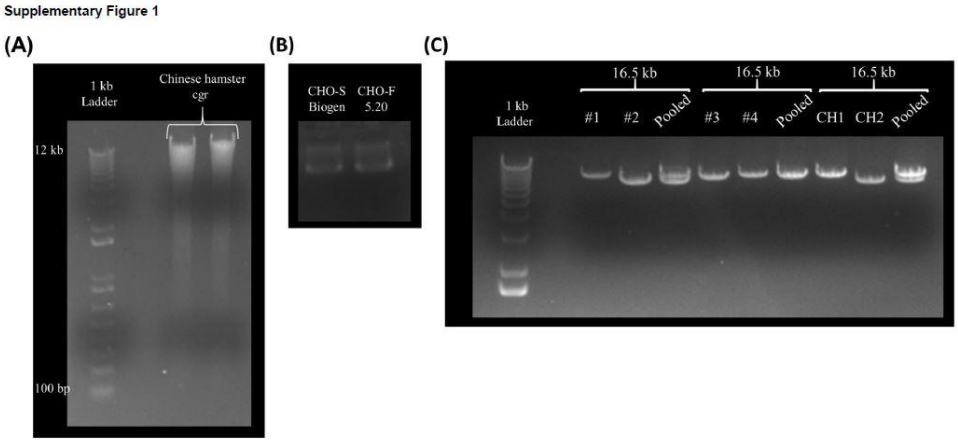
Supplementary Table 7: mtDNA SNP and INDEL summary. Base changes along with allele frequency at each variant position for the 22 CHO cell lines. Mutation annotations outputted by snpEff tool are provided for each SNP and INDEL. For missense variants, the PROVEAN classification and score are included. Missense variants with a PROVEAN score ≤−2.5 were predicted to be deleterious.

	Feature	*tRNA-Phe*	
Lineage	Nucleotide position	62	75
CHO-K1	ATCC CHO-K1		
	Pfizer CHO-K1 CLD1 #1		
	Pfizer CHO-K1 CLD1 #2		
	Pfizer CHO-K1 CLD1 #3		
	Pfizer CHO-K1 CLD1 #4		
	Pfizer CHO-K1 CLD1 #5		
	Pfizer CHO-K1 CLD2 #1		
	Pfizer CHO-K1 CLD2 #2		
	Pfizer CHO-K1 CLD2 #3		
	Pfizer CHO-K1 2B6		
	Pfizer CHO-K1 114		
	DCU CHO-K1 SEAP		
	DCU miR23 Sponge #1		
	DCU miR23 Sponge #2		
	DCU control Sponge #1		
	DCU control Sponge #2		
	Biogen CHO-K1		
DG44	ATCC DGG44	C>T [40%]	G>A [1.9%]
	Biogen DG44 #1		
	Biogen DG44 #2		
CHO-S	ATCC CHO-S		
	Biogen CHO-S		
	snpEff annotation	NC	NC
	Amino acid substitution		
	Provean score		
	Provean prediction*		
	Present in mitoMap		

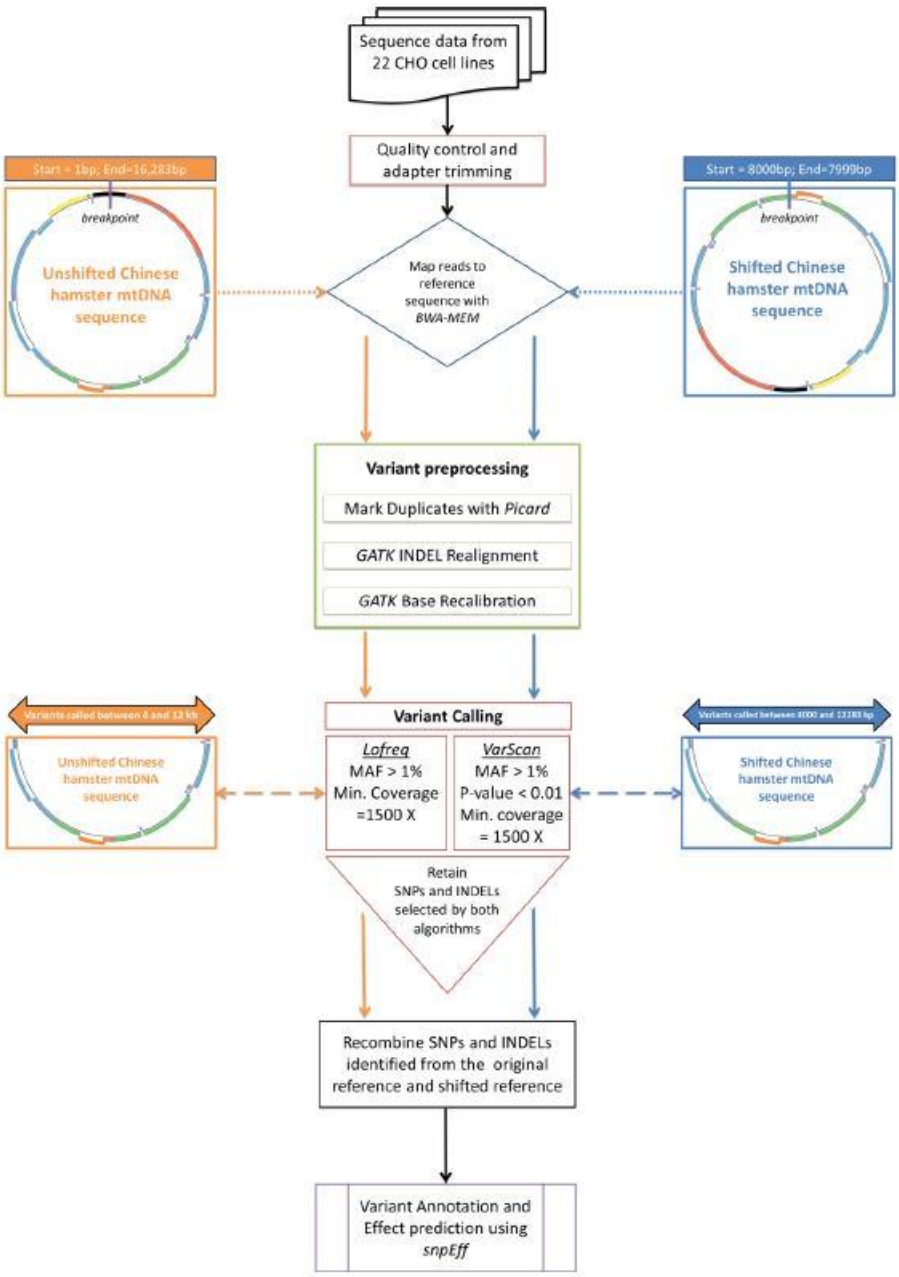
Ref	snpEff Annotation
NC	non_coding_transcript_variant
MS	missense_variant
FS	frameshift_variant
SYN	synonymous_variant
STOP	stop_gained

Supplementary Table 8: Detection of mutations in mtDNA isolated from the mouse liver sample. Using an identical dual mapping approach aligning reads to the unshifted and shifted mouse mitochondrial genome reference sequence we identified 3 mutations, two of which were heteroplasmic.

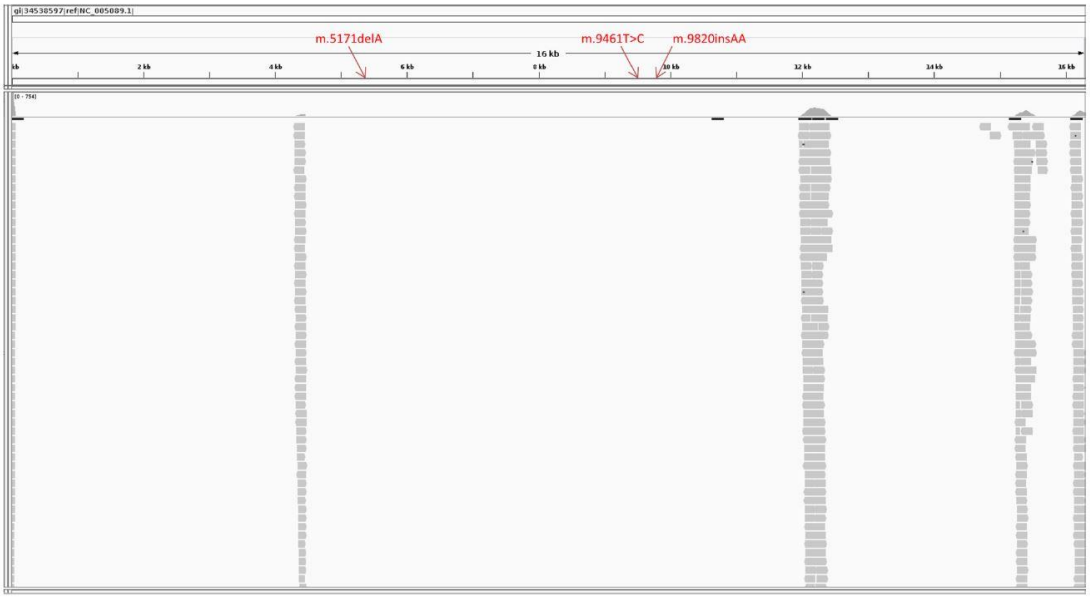
Position	Reference Allele	Alternative Allele	Depth	Allele Frequency	Forward Ref Allele	Reverse Ref Allele
9461	T	C	3147	0.998729	2	0



Supplementary Figure 1: mtDNA extraction, isolation and amplification. Ethidium bromide stained agarose gels were ran to assess the quality of **(A)** Genomic DNA isolated from liver samples of Chinese Hamster using a DNeasy Kit with high integrity DNA running at ~100 kb. **(B)**mitochondrial plasmid DNA isolated from CHO cell lines using a modified plasmid mini-prep kit with plasmid DNA running out as two bands (coiled and super-coiled) and **(C)** Mitochondrial genomic DNA fragments amplified by high-fidelity PCR and visualised individually and pooled.



Supplementary Figure 2: Dual mapping strategy and variant calling. In this study we utilised a dual mapping strategy to account for the mitochondrial genome circularity to maximize depth of coverage at the beginning and end of the mtDNA sequence. We also required agreement between two different variant calling algorithms for a SNP or INDEL to be reported.



Supplementary Figure 3: Alignment of potential NumtS to mtDNA. The mouse mtDNA genome was utilised as a control to determine if NumtS were influencing heteroplasmy measurements. Reads were first stringently aligned to know NumtS region the mm9 genome. Those reads that aligned to NumtS were extracted and remapped against the mouse mitochondrial region. The IGV diagram illustrates that potential NumtS reads aligned to 5 regions. The grey bars illustrate the location of reads that align to both the mouse nuclear and mitochondrial genomes demonstrating that NumtS contamination did not affect the 3 variants identified.

Chapter 2

Depletion of miR-378-3p Enhances CHO DP12 cell Growth via Up-regulation of the Ubiquitin-specific protease 14 (USP14)

Costello, A. *, Coleman, O. *, Lao, T. N., Henry, M., Meleady, P., Barron, N., Clynes, M.
* authors contributed equally,
Journal of Biotechnology, DOI: 10.1016/j.jbiotec.2018.10.008, PMID: 30389639
Impact Factor (2017): 2.533

I share joint primary authorship of this work with Ms. Orla Coleman. Within Chapter 4 I was responsible for miRNA sponge design and generation of all stable mixed pools used in the study. I conducted the initial phenotypic analysis of the miRNA depleted cells. Together with Ms. Coleman I aided in the sampling and sample preparation for quantitative label-free LC-MS/ MS analysis. Ms. Coleman conducted the LC-MS/ MS run and was responsible for analysing the raw data. Together we refined a list of candidate proteins found to be differential in abundance between the miR-378 depleted cells and the control for further functional validation. I completed the subsequent functional wet-lab work and Ms. Coleman prepared the accompanying "Data in Brief" article to make LC-MS/ MS files open source. Dr Nga T. Lao was responsible for constructing the miRNA-sponge vector backbone. Mr. Michael Henry optimised the LC-MS/ MS protocol used in the study. Dr Paula Meleady, Prof. Niall Barron and Prof. Martin Clynes contributed to the experimental design and in manuscript preparation.

ABSTRACT

miRNAs are potent molecular regulators of cellular behaviour. The manipulation of these small non-coding RNAs has been used to enhance industrially relevant phenotypes in Chinese Hamster Ovary (CHO) cells. We investigated the stable depletion of six miRNAs; miR-204-5p, 338-3p, 378-3p, 409-3p, 455-3p and 505-3p, robustly associated with cell growth rate from a previous profiling study. Inhibition of endogenous miR-378-3p function by miRNA-sponge-decoy improved peak cell density by 59%. Quantitative label free LC-MS/MS proteomic analysis of the fractionated cell cultures at day 4 and 8 of batch culture found 216 cytosolic and 114 membrane-associated proteins differentially expressed with stable miR-378-3p depletion. qRT-PCR of 8 genes; Clic4, Hnrnpa1, Prdx1, Actn4, Usp14, Srxn1, Canx and Gnb1, with unidirectional differential protein expression over the two time points of analysis was carried out. In-silico predictive algorithms; TargetScan and miRDB, were used to decipher possible direct targets of miR-378-3p. The Ubiquitin carboxyl-terminal hydrolase 14 (Usp14) protein was identified in the cytosolic fractions at both timepoints as differentially expressed with an increased abundance of 1.58-fold in the miR-378-3p depleted cells on day 8. Usp14 is a deubiquitinase (DUB) with previous reports of its up-regulation leading to increased proliferation of cancer cells. Overexpression of Usp14 in CHO cells had significant effects on cell growth supporting a role of Usp14 in the increased peak cell density seen with miR-378-3p depletion. This study highlights miR-378-3p as a novel engineering candidate for improving CHO cell growth. The use of sub-cellular fractionation also improved proteome coverage in the identification of novel miRNA targets.

Key words: Chinese hamster ovary, miRNA, Genetic engineering, Proteomics, Mass spectrometry

2.1 INTRODUCTION

Targeted genetic engineering of Chinese Hamster Ovary (CHO) cells has become more attainable since publication of the Chinese Hamster and CHO-K1 genomes (Xu et al., 2011; Brinkrolf et al., 2013; Lewis et al., 2014). The use of micro-RNA (miRNA) manipulation is one approach which has been promising. The potency of miRNAs comes from their small ~7-8nt recognition sequence (Bartel et al., 2009) enabling them to post-transcriptionally regulate the expression of multiple protein coding genes or entire pathways (Hackl et al., 2012). Constitutive miRNA manipulations have come in the form of exogenous over-expression of endogenous pri-miRNA hairpins (Klanert et al., 2014) or depletion of endogenous species by miRNA sponge-decoy (Ebert et al., 2007). miRNA engineering of CHO cells has been seen to alter; cell cycle (Sanchez et al., 2014), metabolism (Kelly et al., 2015), productivity (Sanchez et al., 2014; Jadhav et al., 2014; Loh et al., 2014; Fischer et al., 2015; Emmerling et al., 2015; Klanert et al., 2016; Scheollhorn et al., 2017), and inhibit apoptosis (Druz et al., 2013; Griffith et al., 2018). There are currently only a few examples of stable miRNA engineering resulting in enhanced CHO cell growth (Druz et al., 2013; Fischer et al., 2014), and only two miRNA profiling studies have focused on cell growth rate (Clarke et al., 2012; Klanert et al., 2016).

The present study aimed to validate whether prioritised miRNAs from a previous profiling study would impact cell specific growth rate (Clarke et al., 2012) and subsequently explore the biological function of miRNAs where depletion improves cell growth. Numerous challenges still exist in prediction of miRNA targets, chief among which is the lack of computational resources supporting the Chinese Hamster and CHO-K1 genomes (Fischer et al., 2015). Those that do exist fail to consider non-canonical binding events (Helwak et al., 2013). miRNA function to post-transcriptionally regulate the translation of protein coding mRNA. Unlike plant miRNA, which function to promote endonucleolytic cleavage of target mRNA through near perfect ~22nt base pairing (Filipowicz et al., 2008; Jones-Rhoades et al., 2006), miRNA mediated translational inhibition in mammalian cells is rarely through Argonaut 2 (AGO2) cleavage (Liu et al., 2004; Meister et al., 2004) but rather imperfect miRNA – mRNA binding resulting in translational repression or deadenylation (Wu et al., 2006) via AGO1 – 4 complexing (Pillai et al., 2004). The unpredictable nature of mammalian miRNA – mRNA interactions therefore limits the use of transcriptomic approaches. To overcome this, subcellular fractionation of miRNA depleted cultures in combination with quantitative label-free liquid chromatography mass spectrometry (LC-MS/MS) proteomic analysis was used to identify significantly differentially expressed proteins. Subcellular fractionation prior to high-resolution mass spectrometry allowed us to improve the limit of detection and thus identify changes between low abundant proteins. To our

knowledge this is the first report of subcellular proteomic analysis to investigate miRNA function in CHO cells.

2.2 MATERIALS AND METHODS

2.2.1 Cell culture and transfection

CHO DP12 (ATCC) cells were cultured in chemically defined Balan CD (Irvine Scientific) supplemented with 2 % (v/v) polyvinyl alcohol (PVA) and 4 mM L-glutamine (25030081 Gibco). Cells were seeded at $2x10^5$ cells/mL and sub-cultured in a routine 3-4-3-day manner in 50 mL spin tubes (87050T Helena-BioSciences). Cultures were maintained in an ISF1-XC climo-shaker (Kühner), at 37°C, 80 % humidity and 5 % CO_2. For parental cells, the viable cell density (VCD/ mL) was monitored using the ViaCount™ assay on a Guava® easyCyte benchtop cytometer (Merck Millipore). Selective pressure was applied periodically to maintain recombinant protein expression in DP12 with 200 nM Methotrexate (MTX) (Sigma-Aldrich). The formulae used to calculate growth rate, IVCD, AIVCD and Qp have been reported previously (Sanchez et al., 2014). Transient transfection of siRNA was carried out using MIRUS TransIT X2 (MIRUS) by complexing 30 nM of siRNA with 1 µL of reagent for 30 minutes in 100 µL of CHO S SFM II (Gibco). siRNAs in this study included a negative control (siNC - Cat # 51-01-14-03 – Integrated DNA Technologies), a positive control targeting Vasolin-containing protein (siVCP) (Doolan et al., 2010) and a Usp14 specific siRNA (Cat. No. 4390824 – Thermo Fisher Scientific). Transient over-expression of human USP14 was carried out with USP14 - NM_005151. (GenScript).

2.2.2 Vector construction and cloning

Construction of the sponge vectors and the stable mixed pools generated for this study has been described previously (Costello et al., 2017). The miRNA sequences sponges used in this study can be found in Supplementary Table 1. miRNA sponge specificity was evaluated in-silico using the http://genie.weizmann.ac.il/pubs/mir07/mir07_prediction.html (Kertesz et al., 2007) algorithm, with results for the NC and miR-378-3p sponges found in Supplementary Table 2. The 3' UTRs of CHO-K1 Actn4 and Usp14 were aligned with human, mouse and rat variants using Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo). The target region of each UTR was synthesised by (Integrated DNA Technologies), hybridized and cloned downstream of a d2eGFP reporter gene. This as previously described in detail (Costello et al., 2017).

2.2.3 Quantitative reverse-transcription polymerase chain reaction (qRT-PCR)

Cells were harvested by centrifugation at 1000 rpm for 5 minutes. Total RNA was isolated from 1-5x10^6 cells using Tri-reagent (Ambion) following the manufacturer's protocol. RNA quantification and quality were evaluated by NanoDrop (Thermo Fisher Scientific). RNA samples were treated with DNase I (Sigma Aldrich) prior to cDNA synthesis and stored at -80°C. Reverse transcription of total RNA was performed with the High Capacity cDNA Reverse Transcription Kit (Applied Biosystems), on a (G-Storm) thermocycler, in accordance with the manufacturer's protocol. qRT-PCR was run using a 7500 (Applied Biosystems) with Fast SYBR Green Master Mix (Applied Biosystems). The 2X SYBR master mix was combined with 20 ng of cDNA, 200 nM Forward and Reverse primers and water, and made up to a 20 µL final reaction volume. Each biological replicate sample was run in technical triplicate wells. Primer sequences can be found in Supplementary Table 3. Relative quantification was measured by the ddCt method with Gapdh as an endogenous control. For miRNA analysis, the TaqMan miRNA Assay® system (Applied Biosystems) was used. Reverse transcription of specific mature miRNA from total RNA was done using the Taqman® miRNA Reverse Transcription Kit (Applied Biosystems). Relative miRNA abundance was determined by qRT-PCR of three biological replicates run in technical triplicates using the ddCt method with U6 snRNA as an endogenous control.

2.2.4 IgG quantification

CHO DP12 secrete an IgG (anti-IL8) which is quantified by ELISA. Supernatant is harvested simultaneously with RNA and protein sampling, by centrifugation of the cell suspension at 91 x g for 5 minutes. Nunc-Immuno™ MicroWell™ 96 well solid plates (M9410 Sigma-Aldrich) were coated with capture antibody (80-104), (BETHYL Laboratories), diluted 1:100 in coating buffer (C3041-50CAP) (Sigma-Aldrich) overnight at 4°C. Plate washing consisted of 3 x 100 µL /well of ELISA wash buffer (T9039) (Sigma-Aldrich). Blocking was carried out for 1 hr at room temperature with Blocking Buffer (T6789) (Sigma-Aldrich). Standards of Human serum standard (RS10-110) were diluted to range from 0-1000 ng/mL. Samples were diluted to fall within the range of the standard curve and incubated on the plate at room temperature for 1 hr. Detection antibody, HRP antibody (A80-104P) (BETHYL Laboratories), was diluted 1: 100,000 and 100 µL was added to each well and incubated for 1hr at room temperature. To each well, 100 µL of TMB substrate was added and incubated in the dark for 15 minutes. The reaction was stopped using 0.18 M H_2SO_4 and the plate was analysed on a MultiSKAN GO plate reader (Thermo Fisher Scientific) at 450nm.

2.2.5 Protein extraction and in-solution protein digestion

Triplicate biological replicates for day 4 and day 8 of culture following stable mir-378 depletion and NC-spg controls were collected for proteomic analysis. Cells were harvested by centrifugation at 91xg for 5 minutes, at 4°C. Cell pellets were subjected to fractionation using the Mem-PER Plus membrane protein extraction kit (89842 - Thermo Fisher Scientific). The extraction was carried out according to the manufacturer's instructions to yield membrane and cytosolic enriched fractions. Protein concentrations were determined using the Quick Start Bradford protein assay kit (BioRad). Protein samples were purified and digested for mass spectrometry analysis as previously described (Coleman et al., 2017). In short, 100 µg of protein from each sample was purified using filter-aided sample preparation and digested using sequencing grade modified trypsin (Promega) at a ratio of 1:50 (protease: protein) overnight at 37°C. The peptides were purified using Pierce C18 spin columns (Thermo Fisher Scientific), then dried using vacuum centrifugation and re-suspended in an appropriate volume of loading buffer (2% acetonitrile (ACN), 0.1% trifluoroacetic acid (TFA) in LC-MS grade water).

2.2.6 LC-MS/MS and Quantitative label-free data analysis

LC–MS/MS analysis of the membrane and cytosolic fractions of mir-378-spg and NC-spg for day 4 and 8 was carried out as previously described (Henry and Meleady, 2017). In short, peptides were separated using an Ultimate 3000 nanoHPLC over 180 min which was coupled in-line to a hybrid linear LTQ Orbitrap (Thermo Fisher Scientific).

2.2.7 Statistical Analysis

The values reported represent the mean of three independent experiments with error displayed as standard deviation. Statistical analysis was carried out by a two-tailed homoscedastic student t-test, ($p \leq 0.05$ *, $p \leq 0.01$**, $p \leq 0.001$ ***) of three independent experimental replicates. Quantitative label-free data analysis was performed using Progenesis QI for Proteomics as previously described (Henry et al., 2018). The MS files were searched against the *Cricetulus griseus* CRIGR database from NCBI containing 44,065 sequences (fasta file downloaded November 2015). Differentially expressed (DE) proteins between the NC-spg and mir-378-spg have a statistical ANOVA p-value <0.05, a minimum of two peptides contributing to each protein identification and a minimum fold change of 1.25 in relative abundance between the NC-spg and mir-378-spg. Data visualization was achieved as before (Henry et al., 2017) using the updated ggplot2 package in R. Relative quantification of qRT-PCR products was based on three biological replicates run in technical triplicates using the ddCt method with Gapdh as endogenous control for mRNA and U6 snRNA as an endogenous control for miRNA.

2.3 RESULTS

2.3.1 Functional validation of miRNAs associated with cellular growth rate

A previous profiling study conducted by our group identified genes, proteins and miRNAs whose expression was robustly associated with cell specific growth rate (Clarke et al., 2012). Sister clones from the same original transfected pool were evaluated for growth phenotypes over 40 passages. Clones with similar cell specific productivity [24 (± 3) pg/ cell/ day] were characterized by growth rate as "fast" (≥ 0.025 cells/ mL/ hr) or "slow" (≤ 0.023 cells/ mL/ hr). The study found 93 miRNAs of the 667 analysed to be differentially expressed between the two subgroups, 17 of which were anticorrelated with growth. miRNAs whose expression was consistently high across CHO-K1 derived cell lines used by Hackl et al. (Hackl et al. 2011) were prioritized for miRNA sponge mediated knockdown to functionally validate their role in regulating cell specific growth rate. This resulted in a panel of six miRNAs to be tested for phenotypic impact; miR-204-5p, miR-338-3p, miR-378-3p, miR-409-3p, miR-455-3p and miR-505-3p. For each candidate miRNA a specific miRNA-sponge-decoy (Fig. 1A) was designed as carried out in previous studies (Ebert et al., 2007; Kluvier et al., 2012). Each sponge sequence contained 10 miRNA responsive elements (MRE) in the 3' untranslated region (UTR) of a reporter gene. A destabilized GFP (d2eGFP) was used as a reporter due to its short half-life. To assess the effect of candidate miRNA depletion on cell growth the sponge vectors were used to generate stable mixed pools in a model IgG producer line, CHO DP12. The relative quantification (RQ) of mature target miRNAs in stable mixed pools were assessed with respect to a non-specific negative control sponge (NC-spg) (Sanchez et al., 2014) (Fig. 1B). Statistically significant miRNA knockdown was observed for five of the six stable mixed pools; miR-204-5p ($p = 0.003107$), miR-338-3p ($p = 0.00227$), miR-378-3p ($p = 0.00202$), miR-455-3p ($p = 0.00284$) and miR-505-3p (0.00075). There was no significant knockdown of miR-409-3p ($p = 0.08679$). The assessment of cell growth following stable depletion of specific candidate miRNAs revealed three of the six prioritized candidates to have a significant effect on cell growth in this cell line (Fig. 1C). Depletion of miR-378-3p, miR-455-3p and miR-505-3p caused significant increases in peak viable cell density on day 8 of culture of 60%, 25% and 24% respectively. From this initial screen we identified miR-378-3p depletion as having a biologically significant impact on CHO cell growth and prioritized this candidate for subsequent investigation.

The batch culture performance of the 378-spg cells was evaluated in a direct comparison to the NC-spg cells. Phenotypic assessment was conducted at a 5 mL scale in 50 mL spin tube reactors. Stable depletion of miR-378 in CHO DP12 cells achieved a 59% ($p=4.33x10^{-5}$) increase in viable cell density [VCD /mL] over the NC-spg on day 8 of culture with no compromise of culture viability (Fig. 2A). Stable mixed pools of NC-spg and miR-378-spg cells were scaled to 60 mL cultures and the strong growth phenotype was stable with this scale up (Fig. 2B). The cellular

growth rate of the 60 mL cultures during exponential growth was improved from 0.0177 cells/ mL/ hr in the NC-spg cells to 0.0215 cells/ mL/ hr in the miR-378-3p depleted cells (p=0.02256) (Fig. 2C). The integrated viable cell density (IVCD) (Fig. 2D) and accumulated IVCD (AIVCD) (Fig. 2E) were both increased significantly in the 378-spg cells on day eight by 61% ($p = 1.92 \times 10^{-5}$) and 44% ($p = 1.16 \times 10^{-5}$) respectively. There was no difference in volumetric titre between the NC-spg and 378-spg cells in the 60 mL scale (Fig. 2F). The cell specific productivity (Qp) was significantly reduced in the 378-spg cells on days 4, 6 and 8 of culture by; 32% ($p = 0.025$), 23% ($p = 0.0139$) and 35% ($p = 0.00289$) respectively (Fig. 2G).

2.3.2 Differential proteomic analysis of CHO DP12 cells following miR-378 depletion

Proteomic analysis for this study was carried out using a subcellular enrichment strategy to yield two fractions, membrane and cytosolic (Fig. 3A), to improve proteome coverage. Sample complexity negatively affects the ability to detect, identify and quantify low-abundance proteins by MS because low-abundance peptides are beyond the limit of detection by MS in whole proteome analyses. By fractionating the complex cell proteome, the dynamic range of each subcellular fraction is increased in comparison to a total proteome analysis thus proteins at low abundance have a better chance of being detected. By fractionating our mir-378-spg and NC-spg stable mixed pools we improve signal-to-noise and proteome coverage of our samples within the LTQ-OrbiTrap XL mass spectrometer.

Proteomic analysis identified an average of 700 proteins per cytosolic fraction and 590 proteins per membrane fraction, (Supplementary, Associated Data in Brief). Label-free quantitative data analysis was carried out using Progenesis QI for Proteomics to identify proteins whose abundance significantly changed in comparison to a non-specific negative control as a result of sponge depletion of miRNA-378. Differentially expressed (DE) proteins with significantly increased or decreased expression between samples were determined by using a fold-change cut-off of ±1.25-fold and an ANOVA *p*-value of <0.05 between the two stable mixed pools for each fraction and the timepoint analysed. Sequestering miR-378-3p using sponge decoy engineering allows the endogenous targets of miR-378-3p to have repealed regulation in the cells thus exhibiting an increased protein abundance level in the miR-378-spg sample sets. To achieve a better understanding of the effect of miR-378-3p on the proteome, samples were analysed at day 4 and day 8 of culture to capture the mid-exponential and stationary phases of growth.

Unsupervised Pearson hierarchical clustering of the differentially expressed proteins shows clear separation of the cell lines, NC-spg and miR-378-spg, into distinct sample groups based on protein abundance differences resulting from miR-378 depletion as visualised by heatmaps for

both fractions and timepoints (Fig. 3). Quantitative analysis of day 4 of culture identified 27 (Fig. 3B) and 81 (Fig. 3D) differentially expressed proteins for the membrane and cytosolic fractions, respectively, with a minimum of 1.25-fold change between the NC-spg line and the miR-378-spg line (Supplementary, Associated Data in Brief). Of those significant proteins 7 from the membrane fraction and 28 from the cytosolic showed higher abundance in the mir-378-spg sample set which suggests they are potential targets of miR-378. Late stage culture at day 8 identified 95 (Fig. 3C) and 151 (Fig. 3E) significantly differentially expressed proteins for the membrane and cytosolic fractions respectively, (Supplementary, Associated Data in Brief). Of those proteins, 72 from the membrane fraction and 73 from the cytosolic fractions exhibited a higher abundance level in the mir-378-spg stable cell line.

A cohort of differentially expressed proteins were found to have consistently higher abundance in the miR-378-spg cell line at both stages of culture analysed; 6 proteins in the cytosolic fraction and 2 proteins in the membrane fraction. All eight of these proteins; Chloride intracellular channel protein 4 (Clic4), Heterogenous nuclear ribonucleoprotein A1 (Hnrnpa1), Peroxiredoxin 1 (Prdx1), Alpha-actinin 4 (Actn4), Ubiquitin carboxyl-terminal hydrolase 14 (Usp14), Sulfiredoxin 1 (Srxn1), Calnexin (Canx) and G protein subunit beta 1 (Gnb1), were investigated at the transcript level by qRT-PCR for day 4 (Fig. 3F) and day 8 (Fig. 3G). Relative quantification (RQ) of mRNA for 3 of the 8 genes showed a statistically significant directional correlation between protein and mRNA levels for; Actn4, Usp14 and Gnb1 on both days 4 and 8. The RQ of the remaining genes was unchanged on day 4 and 8, apart from Clic4, whose expression increased on day 8. The molecular function and biological process associated with the 8 genes is shown in Table 1. In silico miRNA target prediction software was used to identify potential miR-378 – mRNA interactions of proteins from the differentially expressed list. Hnrnpa1 is a predicted target of miR-378-3p by both TargetScan 7.1 and miRDB algorithms and broadly conserved in vertebrates while Actn4, Usp14 and Srxn1 have conserved target sites in rodents (Table 2). The remaining genes had no predicted miR-378-3p target sites in their 3'UTR, potentially indicative of secondary effects as the result of stable miR-378-3p depletion.

2.3.3 Determining direct targets of miR-378-3p from LC-MS/ MS data

Actn4 and Usp14 were up-regulated at both mRNA and protein levels in the miR-378-spg cell line. These two genes also contain a single predicted miR-378-3p target site in their 3' UTR (Fig. 4A). To investigate direct interaction between miR-378-3p and the two genes transient over-expression of miR-378-3p was performed using a miRNA mimic in the original DP12 cell line (Fig. 4B). Exogenous expression of miR-378-3p significantly reduced the abundance of Actn4 ($p = 0.00083$) and Usp14 ($p = 0.04309$) mRNA. The miR-378-3p target region of each 3' UTR was cloned with flanking sequence downstream of a d2eGFP reporter gene (Fig. 4C). The effects

of these UTR motifs on reporter gene translation were monitored with respect to a negative control (NC). Each construct showed no significant difference in the total cells GFP positive post transfection (Fig. 4D). There was no significant impact on the % of total cells GFP positive (Fig. 4E), yet there was a significant reduction in the median fluorescent intensity of the constructs containing UTR elements from Actn4 ($p = 3.056 \times 10^{-5}$) and Usp14 ($p = 0.0353$) with respect to the NC (Fig. 4F).

2.3.4 Functional analysis of Usp14 validates LC-MS/ MS data

Of the 4 differentially expressed proteins with predicted miR-378 binding sites in their UTR, only two have been reported to effect cell proliferation; Hnrnpa1 (He et al., 2005) and Usp14 (Wang et al., 2014). However, He et al., (2005) demonstrated that RNAi mediated knockdown of single heterogenous nuclear ribonucleoproteins had no effect on cell proliferation. Only by targeting a combination of Hnrnpa1 and Hnrnpa2 or Hnrnpa3 cellular proliferation rate was reduced. As neither of these proteins were up-regulated in the 378-spg cell line we postulated that this may not be the source of enhanced growth. The Ubiquitin carboxyl-terminal hydrolase 14 (Usp14) had the strongest p-value of the differentially expressed proteins which is clearly visible when analysed by volcano plot (Fig. 5A). Usp14 has also previously been associated with cellular proliferation in ovarian cancer (Wang et al., 2014).

To functionally validate the differential expression of Usp14 by depletion of miR-378 via miRNA-sponge decoy, as a source of enhanced cell growth, exogenous expression and repression of Usp14 in CHO DP12 cells was evaluated (Fig. 5). Usp14 is highly conserved between human and Chinese hamster, with 96% homology (Sup.Fig. 1). The human USP14 open reading frame (ORF) was transiently over-expressed in CHO DP12 cells. Transient expression of the human USP14 in DP12 cells was assessed with respect to an empty vector (EV) control. To evaluate whether there was a correlation between the level of USP14 up-regulation and phenotype, the USP14 plasmid was used at different concentrations, 0.25, 0.5, 1 and 2 µg, made up to a total of 2 µg with empty vector in each case. After 48 h transfected cells were lysed and over expression (OE) of USP14 protein levels assessed by Western blot (Fig. 5B). The level of up-regulation increased with over-expression plasmid concentration as anticipated. Over-expression of USP14 was also monitored at the transcript level by qRT-PCR (Fig. 5C). Cells were monitored for VCD (Fig. 5D) and viability (Fig. 5E) 48h post transfection. There was a significant ($p=0.016$) increase in cell growth of 20% at the lowest level of Usp14-OE evaluated (Fig. 5D). At high levels of Usp14-OE (1µg) cells showed a significant reduction in VCD of 70% (Fig. 5D) and viability (55%) (Fig. 5E). Increased growth with 0.25 µg of USP14-OE plasmid did not significantly improve product titre (Fig. 5F).

To elucidate the effect on cell growth of reduced Usp14 activity siRNA mediated knockdown of endogenous CHO DP12 Usp14 was conducted. CHO DP12 cells were transiently transfected with 30 nM siRNA. Knockdown at the protein level was evaluated by Western blot 72 h after transfection (Fig. 5G). There was a significant (89%) knockdown of Usp14 mRNA 24 h post transfection (Fig. 5H), with no significant alteration in Usp14 mRNA levels between all controls. Knockdown of Usp14 resulted in a significant (p= 0.0315) 15% reduction in viable cell density compared to the cells only control, and a 10% reduction versus the negative control (p= 0.0455) (Fig. 5I). Knockdown of Usp14 had no significant effect on cell viability (Fig. 5J) or volumetric titer (Fig. 5K).

Given the role of Usp14 in the un-folded protein (UPR) response we investigated the expression of UPR marker genes with stable miR-378-3p depletion. Six genes, three involved in PERK signalling; Atf4, Atf6 and Chop, and three involved in protein folding; Grp78, Grp94 and Xbp1 (spliced variant) were evaluated at transcript level on days four and eight of culture (Fig. 6). There was a significant reduction in Chop (p = 0.0237), Grp78 (p = 0.0454) and Xbp1 (p = 0.01118) mRNA on day four of culture. These genes were not found to be DE in the LC-MS/ MS data suggesting this lower mRNA abundance did not significantly alter them at the protein level.

2.4 DISCUSSION

The notion of multi-miRNA engineering of CHO cells is not new, yet the catalogue of miRNAs with known robust phenotypes required to implement such strategies does not yet exist (Costello et al., 2018). As it currently stands there are reports of several promising miRNA candidates to stably enhance CHO cell specific productivity (Sanchez et al., 2014; Jadhav et al., 2014; Loh et al., 2014; Fischer et al., 2015; Emmerling et al., 2015; Klanert et al., 2016; Scheollhorn et al., 2017) and inhibit apoptosis (Druz et al., 2013; Griffith and Kelly et al., 2018). Currently there are very few studies showing significantly enhanced CHO cell growth through miRNA manipulation (Druz et al., 2013; Fischer et al., 2014). The present study evaluated the effect of stable depletion of miRNAs previously associated with CHO cell growth rate (Clarke et al., 2012). Of the six miRNAs evaluated, 50% significantly improved peak viable cell density of CHO DP12 cells in batch culture (Fig. 1C). Perhaps unsurprisingly the three miRNA that evoked a growth phenotype had higher basal expression than those that did not show any effect (Sup. Fig. 2). Target miRNA knockdown was observed for five of the candidates evaluated (Fig. 1B). The theory of target mediated miRNA protection (TMMP) (Chatterjee, et al. 2011), a proposed phenomenon wherein the miRNA is protected from exoribonuclease activity by target association, could explain the apparent increase in miR-338 abundance with stable sponge expression (Fig. 1B). TMMP has been reported with a previous miRNA knockdown in CHO cells (Sanchez et al., 2014). From this screen we identified the 59% improvement in peak viable

cell density via miR-378-3p depletion as a biologically significant impact. The improved growth phenotype remained statistically significant in the scale up from 5 to 60 mL (Fig. 2B). The increase in VCD did not lead to any improvement in volumetric titer (Fig. 2D). This reduction in cell specific productivity could be the cause of the improved growth. miR-378-3p has been described as an onco-miR and is associated with tumorigenesis and proliferation in nasopharyngeal carcinoma (Yu et al., 2014), prostate (Nguyen et al., 2013; Avgeris et al., 2014; Chen et al., 2015), colorectal (Wang et al., 2014; Zhang et al., 2014) and gastric (Fei et al., 2013) cancers.

Eight proteins; Clic4, Hnrnpa1, Prdx1, Actn4, Usp14, Srxn1, Canx and Gnb1, had the highest abundance in the miR-378-spg cells for both time points of LC-MS/MS analysis. These 8 proteins were assessed at the transcript level by qRT-PCR, 3 of which; Actn4, Usp14 and Gnb1, showed a significant directional correlation between mRNA and protein levels at both points of culture analysed. This cohort of proteins are potentially direct targets of miR-378 as our data proves their protein expression levels significantly increase upon depletion of miR-378-3p. However due to the complexity of eukaryotic miRNA – mRNA interactions, translational repression associated with miRNA targeting may not involve mRNA turnover. Four of the eight proteins; Hnrnpa1, Actn4, Usp14 and Srxn1, contained a predicted miR-378-3p binding site in the 3'UTR. Transient over expression of miR-378-3p in DP12 cells resulted in significant knockdown of Actn4 and Usp14 (Fig. 4B) and presence of their predicted binding region in the 3' UTR of a GFP reporter caused significant reduction in median fluorescence intensity (Fig. 4D & E). Direct targets or secondary effects, these proteins are very likely to contribute to the growth phenotype seen with miR-378-3p depletion. Clic4 has previous associations with cell proliferation (Tung et al., 2009), as does Prdx1 (Cai et al., 2018), USP14 (Wang et al., 2015) and Hnrnpa1 (He et al., 2005). Two proteins from the list are also associated with cellular un-folded protein response (UPR); Usp14 and Canx. Calr was identified as significantly up-regulated on day 8 of culture only. Usp14 (Lee et al, 2010), Canx and Calr (Pieren et al., 2005) are all involved in cellular UPR and may contribute to the reduced specific productivity of the miR-378-3p depleted cells.

Ubiquitin carboxyl-terminal hydrolase 14 (Usp14) had the strongest statistically differentially expressed protein abundance between the NC-spg and miR-378-spg cells (ANOVA, $p = 5.51x10^{-4}$). Elevated levels of cellular Usp14 has previously been linked to proliferation and tumorigenesis in breast (Zhu et al., 2015), leukemic (Ishiwata et al., 2001), colorectal (Shinji et al., 2006), lung (Wu et al., 2013) and ovarian (Wang et al., 2014) cancers. Usp14 plays a major role in the regulation of the proteasome (Lee et al, 2010) and is one of three proteasome-associated deubiquitinating (DUB) enzymes (Hanna et al, 2006; Bashore et al, 2015; Lee et al, 2016). The proteasome is an essential mechanism in eukaryotes, responsible for non-lysosomal intracellular protein degradation, regulating many aspects of cell physiology and responses including; cell

division, proliferation and apoptosis (Finley, 2009; Schrader et al., 2009; Wang et al., 2014). Proteins are tagged for proteasome mediated degradation by the covalent attachment of ubiquitin or ubiquitin-like proteins (Ub or UBLs). Three classes of proteins (E1, E2 and E3) act in a sequential manner to ubiquitinate proteins destined for proteasome mediated turn-over. The process starts with the activation of the highly conserved 78-amino acid protein, Ub, by the E1 (activator) class of enzymes in an ATP dependant reaction. Followed by transfer of the activated Ub to the E2 (conjugase) class and finally covalent tagging of target proteins by E3 (ligase) enzymes. Targeted proteins may be mono or poly-ubiquitinated, (Jin et al, 2009). The tagged protein is finally degraded by the proteasome. Cellular ubiquitination is balanced by deubiquitination (Zhu et al, 2016), this is achieved by DUBs. Mammalian proteasomes are associated with three predominant DUBs; RPN11, UCH37 and USP14 (Verma et al, 2002; Yao et al, 2002; Koulich et al, 2008; Anderson et al, 2005). Unlike RPN11, which is a stoichiometric subunit of the proteasome, UCH37 and USP14 can work independently to disassemble or trim but not degrade Ub chains, thus impeding the tagged protein's commitment to proteasome degradation (Lam et al, 1997; Hanna et al, 2006; Hedge et al, 2007).

Over expression of human USP14 in the original CHO DP12 cell line showed a dynamic response to different levels of up-regulation. With a minor increase in USP14 protein abundance VCD was improved by 20%, yet at high levels of protein expression there was a 70% reduction in VCD and 55% reduction in viability. Knockdown of endogenous Usp14 by siRNA significantly reduced VCD. Exogenous expression and repression of Usp14 had anticipated positive and negative effects on CHO DP12 cell growth respectively, providing confidence in the LC-MS/ MS differential expression analysis. However, the effect of this single gene manipulation was less impactful on cell growth than knockdown of miR-378-3p.

2.5 CONCLUSIONS

This data represents the first subcellular proteomic approach to understanding the mechanism of action of microRNAs in CHO cells. Forced expression and repression of the strongest statistically differentially expressed protein from the LC-MS/ MS data gave anticipated inverse phenotypes. Stable depletion of miR-378-3p enhanced CHO DP12 cell density by 59%, a much greater impact that transient over-expression of the top candidate from the LC-MS/ MS data. This confirms the known benefit of miRNA engineering strategies, in that a single miRNA manipulation can affect multiple genes simultaneously. Finally, while the cell specific productivity was reduced with miR-378-3p depletion, its beneficial impact on cell growth could make miR-378-3p a candidate for multi-miRNA engineering strategies.

ACKNOWLEDGMENTS

This work was conducted under the financial support of Scientific Foundation of Ireland (SFI) grant numbers [13/IA/1963] and [13/IA/1841]. The authors would also like to thank Dr. Colin Clarke of the National Institute for Bioprocessing Research and training miRNA candidates used in this study.

CONFLICT OF INTEREST

The authors declare no financial or commercial conflict of interest.

2.6 REFERENCES

Anderson, C., Crimmins, S., Wilson, J. A., Korbel, G. A., Ploegh, H. L., & Wilson, S. M. (2005) Loss of Usp14 results in reduced levels of ubiquitin in ataxia mice. *Journal of Neurochemistry. 95*, 724-731.

Avgeris, M., Stravodimos, K., & Scorilas, A. (2014) Loss of miR-378 in prostate cancer, a common regulator of KLK2 and KLK4, correlates with aggressive disease phenotype and predicts the short-term relapse of the patients. *Biological Chemistry. 395*, 1095-1104.

Bartel, D. P. (2009) MicroRNAs: Target recognition and regulatory functions. *Cell. 136*, 215-233.

Bashore, C., Dambacher, C. M., Goodall, E. A., Matyskiela, M. E., Lander, G. C., & Martin, A. (2015) Ubp6 deubiquitinase controls conformational dynamics and substrate degradation of the 26S proteasome. *Nature Structural & Molecular Biology. 22*, 712-719.

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

Cai, A., Zeng, W., Cai, W., Liu, J., Zheng, X., Liu, Y., Yang, X., Long, Y., & Li, J. (2018) Peroxiredoxin-1 promotes cell proliferation and metastasis through enhancing Akt/mTOR in human osteosarcoma cells. *Oncotarget. 9*, 8290.

Chen, Q., Zhou, W., Han, T., Du, S., Li, Z., Zhang, Z., Shan, G., & Kong, C. (2016) MiR-378 suppresses prostate cancer cell growth through downregulation of MAPK1 *in vitro* and *in vivo. Tumor Biology. 37*, 2095-2103.

Clarke, C., Henry, M., Doolan, P., Kelly, S., Aherne, S., Sanchez, N., Kelly, P., Kinsella, P., Breen, L., Madden, S. F., Zhang, L., Leonard, M., Clynes, M., Meleady, P., & 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*, 656-2164-13-656.

Coleman, O., Henry, M., Clynes, M., & Meleady, P. (2017) Filter-aided sample preparation (FASP) for improved proteome analysis of recombinant chinese hamster ovary cells. *Heterologous Protein Production in CHO Cells: Methods and Protocols*, vol. 1603, 187-194.

Costello, A., Lao, N., Clynes, M., & Barron, N. (2017) Conditional knockdown of endogenous MicroRNAs in CHO cells using TET-ON-SanDI sponge vectors. *Heterologous Protein Production in CHO Cells: Methods and Protocols*, vol. 1603, 87-100.

Costello, A., Lao, N. T., Gallagher, C., Capella Roca, B., Julius, L. A., Suda, S., Ducrée, J., King, D., Wagner, R., & Barron, N. (2018) Leaky expression of the TET-On system hinders control of endogenous miRNA abundance. *Biotechnology Journal*, 1800219.

Doolan, P., Meleady, P., Barron, N., Henry, M., Gallagher, R., Gammell, P., Melville, M., Sinacore, M., McCarthy, K., & Leonard, M. (2010) Microarray and proteomics expression profiling identifies several candidates, including the valosin-containing protein (VCP), involved in regulating high cellular growth rate in production CHO cell lines. Biotechnology and Bioengineering. 106, 42-56.

Druz, A., Son, Y., Betenbaugh, M., & Shiloach, J. (2013) Stable inhibition of mmu-miR-466h-5p improves apoptosis resistance and protein production in CHO cells. *Metabolic Engineering. 16*, 87-94.

Ebert, M. S., Neilson, J. R., & Sharp, P. A. (2007) MicroRNA sponges: Competitive inhibitors of small RNAs in mammalian cells. *Nature Methods. 4*, 721-726.

Emmerling, V. V., Fischer, S., Stiefel, F., Holzmann, K., Handrick, R., Hesse, F., Hörer, M., Kochanek, S., & Otte, K. (2015) Temperature-sensitive miR-483 is a conserved regulator of recombinant protein and viral vector production in mammalian cells. *Biotechnology and Bioengineering*

Fei, B., & Wu, H. (2013) MiR-378 inhibits progression of human gastric cancer MGC-803 cells by targeting MAPK1 *in vitro. Oncology Research Featuring Preclinical and Clinical Cancer Therapeutics. 20*, 557-564.

Filipowicz, W., Bhattacharyya, S. N., & Sonenberg, N. (2008) Mechanisms of post-transcriptional regulation by microRNAs: Are the answers in sight?. *Nature Reviews.Genetics. 9*, 102.

Finley, D. (2009) Recognition and processing of ubiquitin-protein conjugates by the proteasome. *Annual Review of Biochemistry. 78*, 477-513.

Fischer, S., Buck, T., Wagner, A., Ehrhart, C., Giancaterino, J., Mang, S., Schad, M., Mathias, S., Aschrafi, A., Handrick, R., & Otte, K. (2014) A functional high-content miRNA screen identifies miR-30 family to boost recombinant protein production in CHO cells. *Biotechnology Journal. 9*, 1279-1292.

Fischer, S., Handrick, R., & Otte, K. (2015) The art of CHO cell engineering: A comprehensive retrospect and future perspectives. *Biotechnology Advances. 33*, 1878-1896.

Griffith, A., Kelly, P. S., Vencken, S., Lao, N. T., Greene, C. M., Clynes, M., & Barron, N. (2017) miR-CATCH identifies biologically active miRNA regulators of the Pro-Survival gene XIAP, in chinese hamster ovary cells. *Biotechnology Journal.* 13(3), p.1700299.

Hackl, M., Jakobi, T., Blom, J., Doppmeier, D., Brinkrolf, K., Szczepanowski, R., Bernhart, S. H., zu Siederdissen, C. H., Bort, J. A. H., & Wieser, M. (2011) Next-generation sequencing of the chinese hamster ovary microRNA transcriptome: Identification, annotation and profiling of microRNAs as targets for cellular engineering. *Journal of Biotechnology*. 153, 62-75.

Hanna, J., Hathaway, N. A., Tone, Y., Crosas, B., Elsasser, S., Kirkpatrick, D. S., Leggett, D. S., Gygi, S. P., King, R. W., & Finley, D. (2006) Deubiquitinating enzyme Ubp6 functions noncatalytically to delay proteasomal degradation. *Cell. 127*, 99-111.

He, Y., Brown, M. A., Rothnagel, J. A., Saunders, N. A., & Smith, R. (2005) Roles of heterogeneous nuclear ribonucleoproteins A and B in cell proliferation. *Journal of Cell Science*. 118, 3173-3183.

Hegde, A. N., & Upadhya, S. C. (2007) The ubiquitin–proteasome pathway in health and disease of the nervous system. *Trends in Neurosciences. 30*, 587-595.

Helwak, A., Kudla, G., Dudnakova, T., & Tollervey, D. (2013) Mapping the human miRNA interactome by CLASH reveals frequent noncanonical binding. *Cell*. 153, 654-665.

Henry, M., Gallagher, C., Kelly, R. M., Frye, C. C., Osborne, M. D., Brady, C. P., Barron, N., Clynes, M., & Meleady, P. (2018) Clonal variation in productivity and proteolytic clipping of an fc-fusion protein in CHO cells: Proteomic analysis suggests a role for defective protein folding and the UPR. *Journal of Biotechnology.*

Henry, M., & Meleady, P. (2017) Clinical proteomics: Liquid Chromatography–Mass spectrometry (LC–MS) purification systems. In: Anonymous , Protein Chromatography. Springer, pp. 375-388

Henry, M., Power, M., Kaushik, P., Coleman, O., Clynes, M., & Meleady, P. (2017) Differential phosphoproteomic analysis of recombinant chinese hamster ovary cells following temperature shift. *Journal of Proteome Research. 16*, 2339-2358.

Ishiwata, S., Katayama, J., Shindo, H., Ozawa, Y., Itoh, K., & Mizugaki, M. (2001) Increased expression of queuosine synthesizing enzyme, tRNA-guanine transglycosylase, and queuosine levels in tRNA of leukemic cells. *The Journal of Biochemistry. 129*, 13-17.

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

Jones-Rhoades, M. W., Bartel, D. P., & Bartel, B. (2006) MicroRNAs and their regulatory roles in plants. *Annual Reviews in Plant Biology. 57*, 19-53.

Kelly, P. S., Breen, L., Gallagher, C., Kelly, S., Henry, M., Lao, N. T., Meleady, P., O'Gorman, D., Clynes, M., & Barron, N. (2015) Re-programming CHO cell metabolism using miR-23 tips the balance towards a highly productive phenotype. *Biotechnology Journal. 10*, 1029-1040.

Kertesz, M., Iovino, N., Unnerstall, U., Gaul, U., & Segal, E. (2007) The role of site accessibility in microRNA target recognition. *Nature Genetics. 39*, 1278.

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, 538-544.

Klanert, G., Jadhav, V., Shanmukam, V., Diendorfer, A., Karbiener, M., Scheideler, M., Bort, J. H., Grillari, J., Hackl, 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, pp.150-161.

Kluiver, J., Gibcus, J. H., Hettinga, C., Adema, A., Richter, M. K., Halsema, N., Slezak-Prochazka, I., Ding, Y., Kroesen, B. J., & van den Berg, A. (2012) Rapid generation of microRNA sponges for microRNA inhibition. *PloS One. 7*, e29275.

Koulich, E., Li, X., & DeMartino, G. N. (2008) Relative structural and functional roles of multiple deubiquitylating proteins associated with mammalian 26S proteasome. *Molecular Biology of the Cell. 19*, 1072-1082.

Lam, Y. A., Xu, W., DeMartino, G. N., & Cohen, R. E. (1997) Editing of ubiquitin conjugates by an isopeptidase in the 26S proteasome. *Nature. 385*, 737.

Lee, B., Lee, M. J., Park, S., Oh, D., Elsasser, S., Chen, P., Gartner, C., Dimova, N., Hanna, J., & Gygi, S. P. (2010) Enhancement of proteasome activity by a small-molecule inhibitor of USP14. *Nature. 467*, 179-184.

Lee, B. H., Lu, Y., Prado, M. A., Shi, Y., Tian, G., Sun, S., Elsasser, S., Gygi, S. P., King, R. W., & Finley, D. (2016) USP14 deubiquitinates proteasome-bound substrates that are ubiquitinated at multiple sites. *Nature. 532*, 398-401.

Lewis, N. E., Liu, X., Li, Y., Nagarajan, H., Yerganian, G., O'Brien, E., Bordbar, A., Roth, A. M., Rosenbloom, J., & Bian, C. (2013) Genomic landscapes of chinese hamster ovary cell lines as revealed by the cricetulus griseus draft genome. *Nature Biotechnology. 31*, 759-765.

Liu, J., Carmell, M. A., Rivas, F. V., Marsden, C. G., Thomson, J. M., Song, J. J., Hammond, S. M., Joshua-Tor, L., & Hannon, G. J. (2004) Argonaute2 is the catalytic engine of mammalian RNAi. *Science (New York, N.Y.). 305*, 1437-1441.

Meister, G., Landthaler, M., Patkaniowska, A., Dorsett, Y., Teng, G., & Tuschl, T. (2004) Human Argonaute2 mediates RNA cleavage targeted by miRNAs and siRNAs. *Molecular Cell. 15*, 185-197.

Nguyen, H. C. N., Xie, W., Yang, M., Hsieh, C., Drouin, S., Lee, G. M., & Kantoff, P. W. (2013) Expression differences of circulating microRNAs in metastatic castration resistant prostate cancer and low-risk, localized prostate cancer. *The Prostate. 73*, 346-354.

Pieren, M., Galli, C., Denzel, A., & Molinari, M. (2005) The use of calnexin and calreticulin by cellular and viral glycoproteins. *The Journal of Biological Chemistry. 280*, 28265-28271.

Pillai, R. S., Artus, C. G., & Filipowicz, W. (2004) Tethering of human ago proteins to mRNA mimics the miRNA-mediated repression of protein synthesis. *RNA (New York, N.Y.). 10*, 1518-1525.

Sanchez, N., Kelly, P., Gallagher, C., Lao, N. T., Clarke, C., Clynes, M., & Barron, N. (2014) CHO cell culture longevity and recombinant protein yield are enhanced by depletion of miR-7 activity via sponge decoy vectors. *Biotechnology Journal. 9*, 396-404.

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), pp.1046-1058.

Schrader, E. K., Harstad, K. G., & Matouschek, A. (2009) Targeting proteins for degradation. *Nature Chemical Biology. 5*, 815-822.

Shinji, S., Naito, Z., Ishiwata, S., Ishiwata, T., Tanaka, N., Furukawa, K., Suzuki, H., Seya, T., Matsuda, A., & Katsuta, M. (2006) Ubiquitin-specific protease 14 expression in colorectal cancer is associated with liver and lymph node metastases. *Oncology Reports. 15*, 539-543.

Tung, J. J., Hobert, O., Berryman, M., & Kitajewski, J. (2009) Chloride intracellular channel 4 is involved in endothelial proliferation and morphogenesis *in vitro. Angiogenesis. 12*, 209-220.

Verma, R., Aravind, L., Oania, R., McDonald, W. H., Yates, J. R.,3rd, Koonin, E. V., & Deshaies, R. J. (2002) Role of Rpn11 metalloprotease in deubiquitination and degradation by the 26S proteasome. *Science (New York, N.Y.). 298*, 611-615.

Wang, K., Ma, J., Zhang, F., Yu, M., Xue, J., & Zhao, J. (2014) MicroRNA-378 inhibits cell growth and enhances l-OHP-induced apoptosis in human colorectal cancer. *IUBMB Life. 66*, 645-654.

Wang, Y., Wang, J., Zhong, J., Deng, Y., Xi, Q., He, S., Yang, S., Jiang, L., Huang, M., & Tang, C. (2015) Ubiquitin-specific protease 14 (USP14) regulates cellular proliferation and apoptosis in epithelial ovarian cancer. *Medical Oncology. 32*, 379.

Wang, Y., Wang, J., Zhong, J., Deng, Y., Xi, Q., He, S., Yang, S., Jiang, L., Huang, M., & Tang, C. (2015) Ubiquitin-specific protease 14 (USP14) regulates cellular proliferation and apoptosis in epithelial ovarian cancer. *Medical Oncology. 32*, 379.

Wu, N., Liu, C., Bai, C., Han, Y., Cho, W., & Li, Q. (2013) Over-expression of deubiquitinating enzyme USP14 in lung adenocarcinoma promotes proliferation through the accumulation of β-catenin. *International Journal of Molecular Sciences. 14*, 10749-10760.

Wu, N., Liu, C., Bai, C., Han, Y., Cho, W., & Li, Q. (2013) Over-expression of deubiquitinating enzyme USP14 in lung adenocarcinoma promotes proliferation through the accumulation of β-catenin. *International journal of molecular sciences. 14*, 10749-10760.

Wu, L., Fan, J., & Belasco, J. G. (2006) MicroRNAs direct rapid deadenylation of mRNA. *Proceedings of the National Academy of Sciences of the United States of America. 103*, 4034-4039.

Xu, X., Nagarajan, H., Lewis, N. E., Pan, S., Cai, Z., Liu, X., Chen, W., Xie, M., Wang, W., & Hammond, S. (2011) The genomic sequence of the chinese hamster ovary (CHO)-K1 cell line. *Nature Biotechnology. 29*, 735-741.

Yao, T., & Cohen, R. E. (2002) A cryptic protease couples deubiquitination and degradation by the proteasome. *Nature. 419*, 403.

Yu, B., Peng, X., Zhao, F., Liu, X., Lu, J., Wang, L., Li, G., Chen, H., & Li, X. (2014) MicroRNA-378 functions as an onco-miR in nasopharyngeal carcinoma by repressing TOB2 expression. *International Journal of Oncology. 44*, 1215-1222.

Zhang, G., Zhou, H., Xiao, H., Li, Y., & Zhou, T. (2014) MiR-378 is an independent prognostic factor and inhibits cell growth and invasion in colorectal cancer. *BMC Cancer. 14*, 109.

Zhu, L., Yang, S., He, S., Qiang, F., Cai, J., Liu, R., Gu, C., Guo, Z., Wang, C., & Zhang, W. (2016) Downregulation of ubiquitin-specific protease 14 (USP14) inhibits breast cancer cell proliferation and metastasis, but promotes apoptosis. *Journal of Molecular Histology. 47*, 69-80.

Table 1: Proteins with Increased Peptide Abundance on Day 4 and Day 8

Gene ID	Fraction	FC Day 4	FC Day 8	Anova Day 4	Anova Day 8	Molecular Function	Biological Process
Clic4	Cytosolic	1.37	1.55	2.23E-04	2.26E-02	Anion Channel Activity	Anion Transport
Hnrnpa1	Cytosolic	1.32	1.85	5.72E-04	4.53E-02	RNA Binding	Cellular Response to Glucose Starvation
Prdx1	Cytosolic	1.33	1.85	2.38E-03	2.26E-02	Oxioreductase Activity	Cell Proliferation
Actn4	Cytosolic	1.26	1.28	3.94E-03	3.71E-03	Actin Binding	Protein Transport
Usp14	Cytosolic	1.45	1.58	5.27E-03	5.51E-04	Hydrolase	Ubl Conjugation Pathway
Srxn1	Cytosolic	1.35	1.37	9.24E-03	3.04E-02	Antioxidant	Cellular Response to Oxidative Stress
Canx	Membrane	1.52	1.60	2.35E-02	1.78E-03	Calcium Ion Binding	Exocytosis
Gnb1	Membrane	1.30	1.28	1.83E-02	1.26E-03	GTPase Activity	G-protein Coupled Receptor Signalling Pathway

Table 2: TargetScan 7.1 Analysis of miR-378-3p and UTR binding of DE genes

Gene ID	Position in 3'UTR /miRNA	Predicted consequential pairing of target region (top) and miRNA (bottom)	Site type	Context++ score	Context++ score percentile	Weighted context++ score	Conserved branch length
Hnrnpa1	Position 1766-1772 of HNRNPA1 3' UTR mmu-miR-378a-3p	5' ...UAUUGAGCCAAAACUAGUCCAGU... 3' GGAAGACUGAGGUUCAGGUCA	7mer-m8	-0.22	88	-0.21	3.822
Hnrnpa1	Position 850-857 of HNRNPA1 3' UTR mmu-miR-378a-3p	5' ...UUAAUGCCACCUAUAAGUCCAGA... 3' GGAAGACUGAGGUUCAGGUCA	8mer	-0.44	98	-0.42	0
Actn4	Position 368-374 of ACTN4 3' UTR mmu-miR-378a-3p	5' ...CCGCCUGCCCUAAGAAGUCCAGC... 3' GGAAGACUGAGGUUCAGGUCA	7mer-m8	-0.13	74	-0.13	0.403
Usp14	Position 444-450 of USP14 3' UTR mmu-miR-378a-3p	5' ...GUGCAAUCAAGUAUUGUCCAGAC... 3' GGAAGACUGAGGUUCAGGUCA	7mer-A1	-0.16	79	-0.12	0.403
Srxn1	Position 1428-1434 of SRXN1 3' UTR mmu-miR-378a-3p	5' ...CUUCUGCAAACCUAGAGUCCAGG... 3' GGAAGACUGAGGU---UCAGGUCA	7mer-m8	-0.14	75	-0.07	0

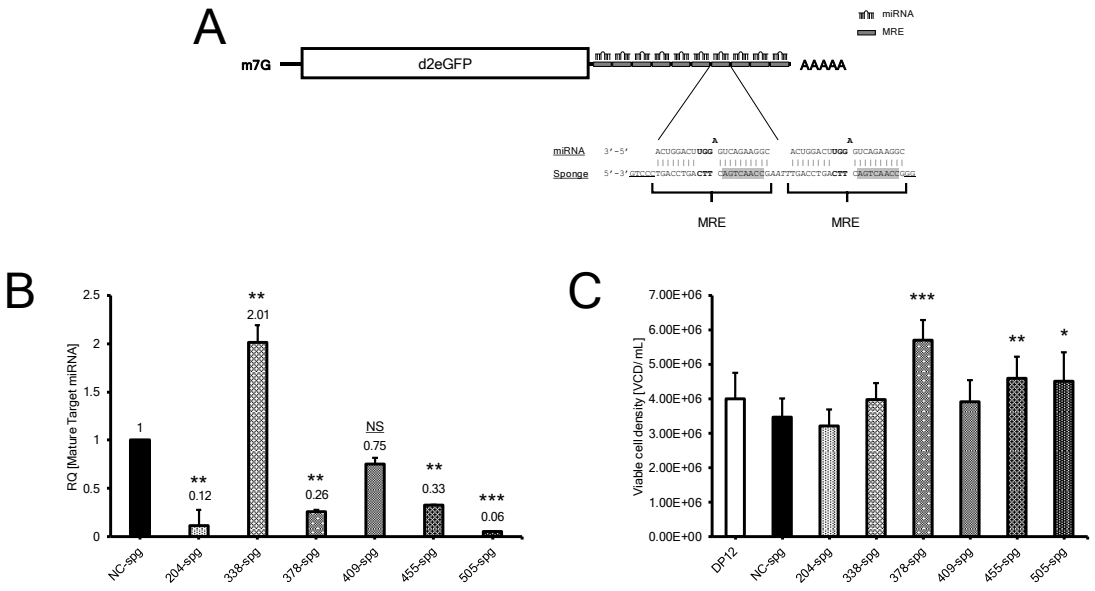


Fig 1. Functional assessment of growth associated miRNAs in stable mixed pools. (A) A schematic representation of the miRNA sponge design used in this study. miRNA sponge sequences were cloned downstream of a destabilized GFP (d2eGFP) reporter gene. miRNA seed regions are highlighted and a 4-nucleotide mismatch or "bulge" sequence in (bold). (B) Stable mixed pools of cells expressing a specific miRNA sponge decoy were generated in CHO DP12 cells for miRNAs; 204-5p, 338-3p, 378-3p, 409-3p, 455-3p, 505-3p and a non-specific negative control (NC-spg). miRNA sponge efficacy was evaluated for all candidates by analysing the relative quantification of specific mature miRNA in sponge cell lines compared to a non-specific negative control sponge. (C) Viable cell density [VCD/ mL] of CHO DP12 and stable specific miRNA sponge expressing mixed pools for miRNAs with respect to the NC-spg mixed pool. Statistics were carried out by a two-tailed homoscedastic student t-test, ($p \leq 0.05$ *, $p \leq 0.01$**, $p \leq 0.001$ ***), values based on the mean of three biological replicates with error as standard deviation of replicates.

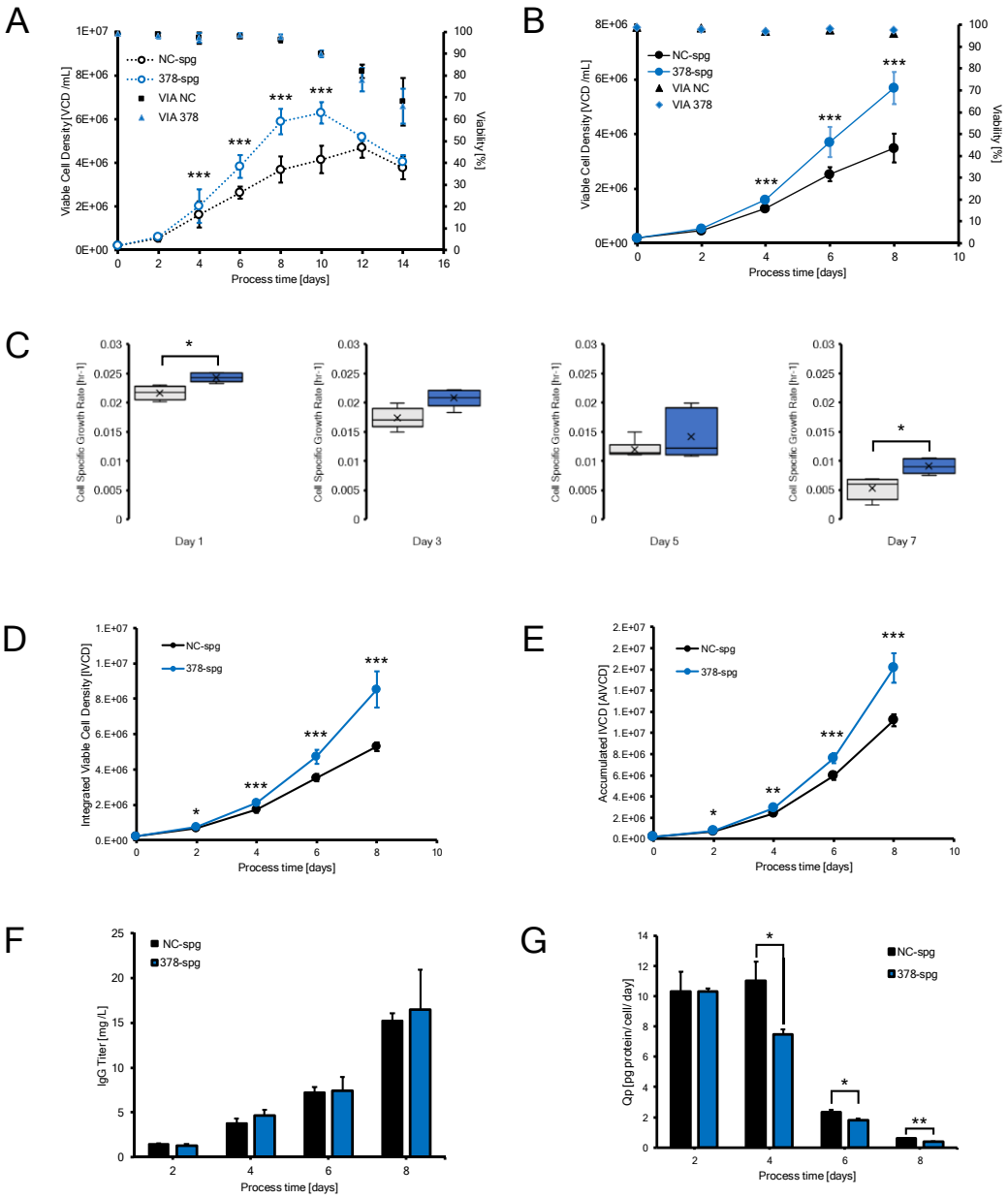


Fig 2. Batch analysis of stable miR-378-3p depletion in CHO DP12 cells. (A) The Negative control sponge (NC-spg) was grown in parallel with the miR-378-3p-sponge cell line (378-spg). Viable cell density (VCD) /mL is seen on the y-axis (left), viability (right), and culture duration (Days) on the x-axis. (B) [VCD/ mL] and culture viability of mixed pools scaled to 60mL scale. (C) Cell specific growth rate of NC-spg and 378-spg cells at 60mL scale. (D) Integrated VCD (IVCD) of NC-spg and 378-spg cells at 60mL. (E) Accumulated IVCD of NC-spg and 378-spg cells at 60mL scale. (F) IgG production (mg /L) for NC-spg and 378-spg cells grown at 60mL scale. (G) Cell specific productivity (Qp) (pg/ cell/ day) for NC-spg and 378-spg cells grown at 60mL scale. Statistics were carried out by a two-tailed homoscedastic student t-test, ($p \leq 0.05$ *, $p \leq 0.01$**, $p \leq 0.001$ ***), values based on the mean of three biological replicates with error as standard deviation of replicates.

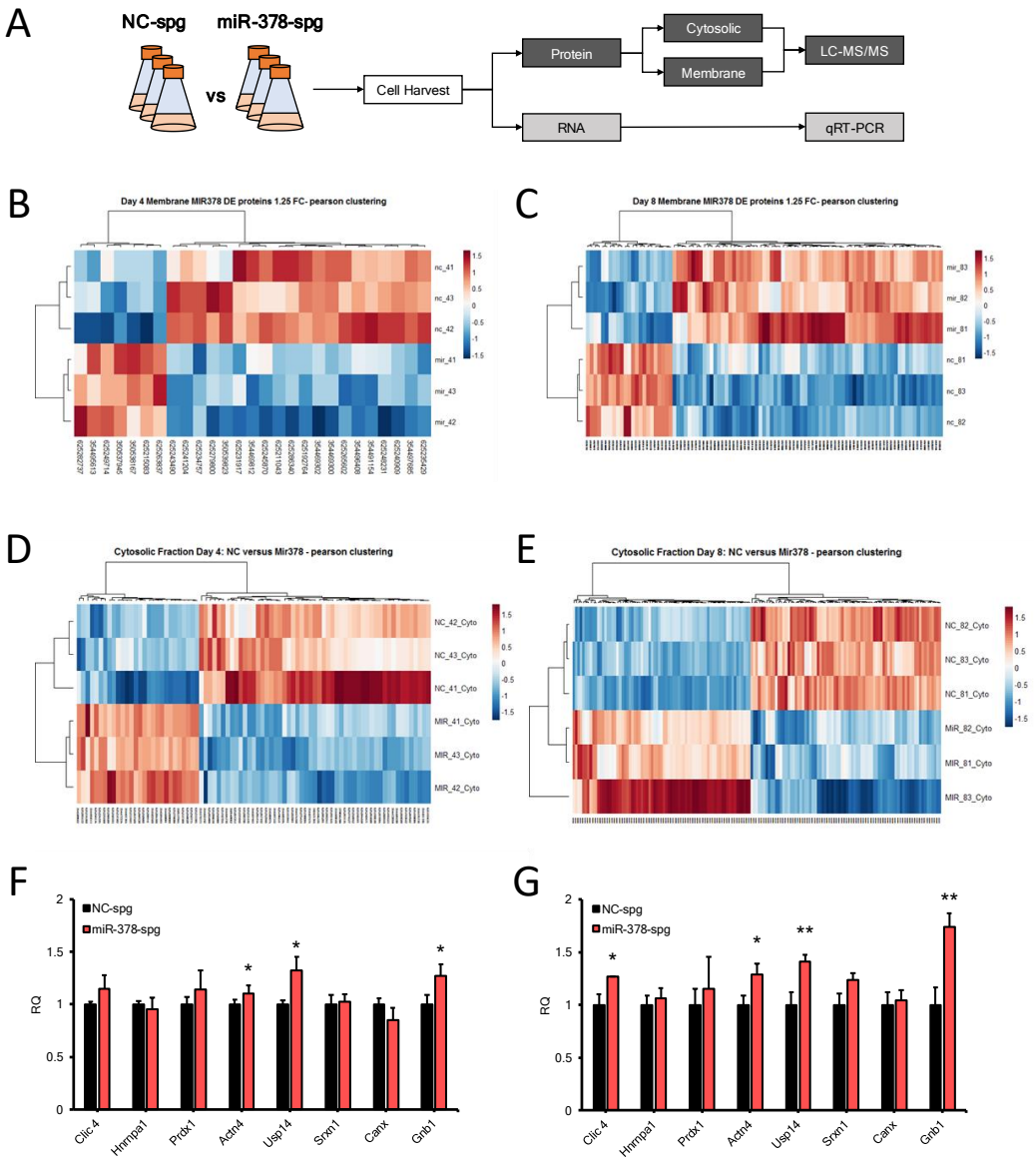


Fig 3. Differential expression analysis of NC-spg and 378-spg cells. (A) Schematic representation of experimental set up. Three biological replicates of NC-spg and 378-spg cells were grown at 60mL with cells harvested on days 4 and 8 of culture. Cells were used for proteomic and transcript analysis. Relative protein abundance visualized by heat map for differentially expressed membrane proteins on day 4 (B) and day 8 (C), and cytosolic proteins day 4 (D) and day 8 (E). Transcript analysis was undertaken for matched differentially expressed proteins on both days of analysis. qRT-PCR was carried out in technical triplicates for each gene on day 4 (F) and day 8 (G). Relative quantification was done using Gapdh as an endogenous control. Statistical analysis consisted of a two-tailed homoscedastic student t-test, ($p \leq 0.05$ *, $p \leq 0.01$**, $p \leq 0.001$ ***), values based on triplicate biological RNA samples.

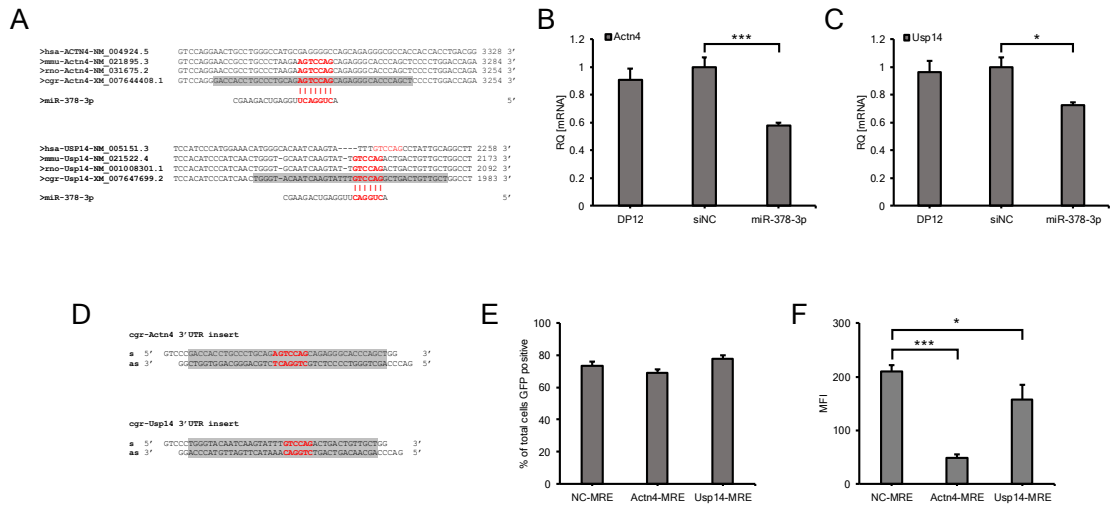


Fig 4. Functional validation of Actn4 and Usp14 mRNA interaction with miR-378-3p. (A) A multiple alignment Actn4 and Usp14 mRNAs from human, mouse, rat and CHO-K1 revealed conservation of miR-378-3p binding sites in rodents but not human. The predicted miRNA responsive element (MRE) is highlighted in bold. miR-378-3p was transiently over-expressed and the relative quantification (RQ) of Actn4 (B) and Usp14 (C) mRNA was compared to a negative control siRNA (siNC). (D) The predicted MRE for Actn4 and Usp14 with 30bp up and downstream flanking sequence was cloned downstream of a GFP reporter with one copy of the MRE in each case. DP12 cells were transfected with one of an; negative control (NC-MRE), Actn4-MRE or Usp14-MRE plasmids. The % of total cells GFP positive (E) and corresponding median fluorescence intensity (MFI) (F) for each construct. Statistics were carried out by a two-tailed homoscedastic student t-test, ($p \leq 0.05$ *, $p \leq 0.001$ ***), values based on three biological replicates.

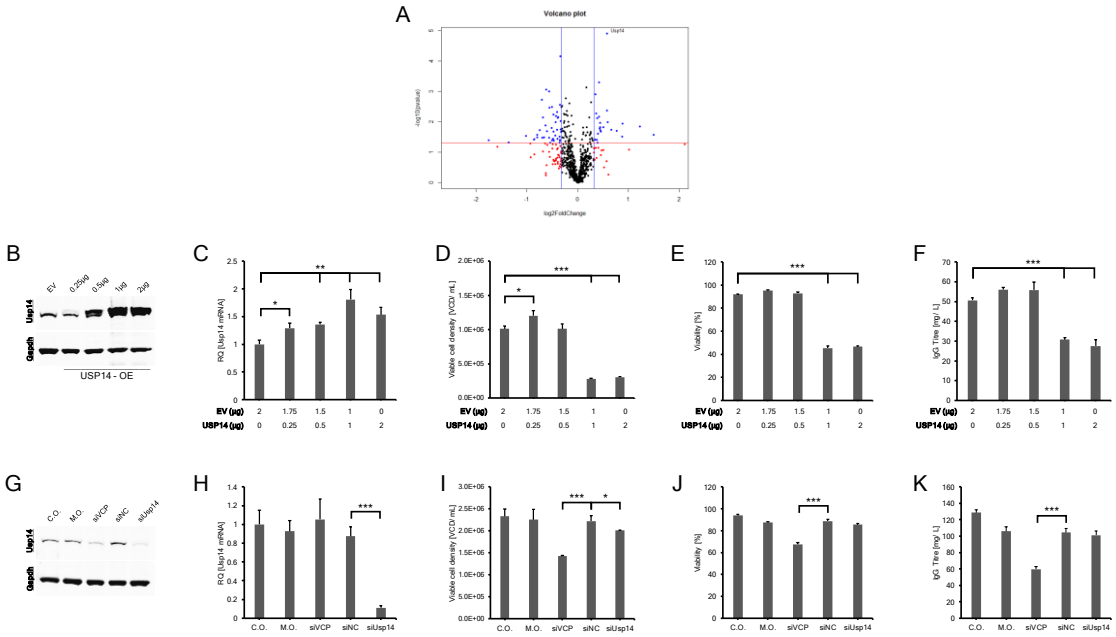


Fig 5. Functional validation of LC-MS/ MS and in silico predicted miR-378-3p target Usp14 in CHO DP12 cells. (A) The distribution of differentially expressed proteins based on *p*-value and fold-change between mir-378-spg and NC-spg from the cytosolic fraction of day 8 of culture is shown by volcano plot. USP14 was transiently over-expressed (OE) in CHO DP12 cells at different levels. To achieve this the USP14 expression vector was diluted with an empty vector (EV). (B) Protein expression 48hrs post transfection was analysed by Western blot. (C) USP14 over expression was evaluated at transcript level by qRT-PCR. Cells transfected with USP14 over-expression vector were monitored for viable cell density (D), viability (E) and volumetric titre (F). Transient knockdown of endogenous Usp14 was done with siRNA. Knockdown was monitored with respect to cells only (C.O.) and vehicle only (M.O.) controls. An siRNA targeting VCP was used as a positive control. (G) The expression of Usp14 was assessed at protein level 72hrs after transfection by Western blot. (H) Endogenous transcript levels of Usp14 was measured by qRT-PCR. The effects of Usp14 knockdown on viable cell density (I), viability (J) and volumetric productivity (K) were monitored 72hrs post transfection. Statistics were carried out by a two-tailed homoscedastic student t-test, ($p \leq 0.05$ *, $p \leq 0.01$**, $p \leq 0.001$ ***), values based on the mean of three biological replicates with error as standard deviation of replicates.

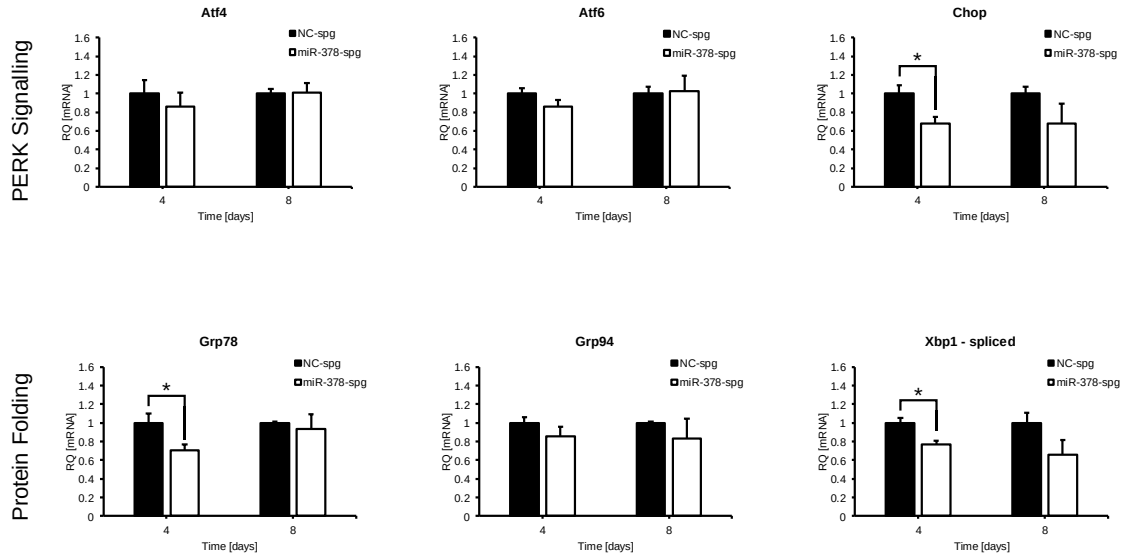


Fig 6. Evaluation of UPR marker genes with miR-378-3p depletion. Marker genes of the un-folded protein response were evaluated by qRT-PCR. The relative quantification (RQ) of gene mRNA was compared on days four and eight of culture with miR-378-3p depletion to the NC-spg line. Three genes; Aft4, Atf6 and Chop are involved in the PERK signalling pathway and the three involved in protein folding; Grp78, Grp94 and Xbp1. Statistics were carried out by a two-tailed homoscedastic student t-test, ($p \leq 0.05$ *, $p \leq 0.01$**, $p \leq 0.001$ ***), values based on three biological replicates.

2.7 APPENDIX B

Supporting information associated with Chapter 2.

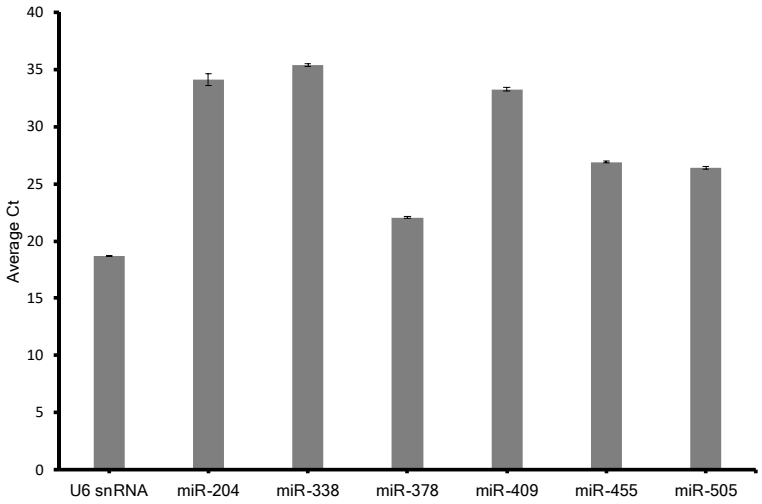
```
CLUSTAL O(1.2.4) multiple sequence alignment (96.36% alignment)


cgr-Usp14 MFLFTVTVKWGKEKFEGVELNTDESPMVFKAQLFALTGVQPARQKVMVKGGTLKDDDWGN 60
hsa-USP14 MPLYSVTVKWGKEKFEGVELNTDEPPMVFKAQLFALTGVQPARQKVMVKGGTLKDDDWGN 60
          * *::****************** ***********************************
cgr-Usp14 IKMKNGMTILMMGSADALPEEPSAKTVFVEDMTEEQLATAMELPCGLTNLGNTCYMNATV 120
hsa-USP14 IKIKNGMTLLMMGSADALPEEPSAKTVFVEDMTEEQLASAMELPCGLTNLGNTCYMNATV 120
          **:*****:****************************:********************
cgr-Usp14 QCIRSVPELKDALKRYAGALRASGEMASAQYITAALRDLFDSMDKTSSSIPPIILLQFLH 180
hsa-USP14 QCIRSVPELKDALKRYAGALRASGEMASAQYITAALRDLFDSMDKTSSSIPPIILLQFLH 180
          ************************************************************
cgr-Usp14 MAFPQFAEKGEQGQYLQQDANECWVQMMRVLQQKLEAIEDDSATEVSETDSSASAVTPSK 240
hsa-USP14 MAFPQFAEKGEQGQYLQQDANECWIQMMRVLQQKLEAIEDDSVKETD--SSSASAATPSK 238
          ************************:*****************..*.. .*****.****
cgr-Usp14 KKSLIDQFFGVEFETTMKCTESEEEEVTKGKENQLQLSCFINQEVKYLFTGLKLRLQEEI 300
hsa-USP14 KKSLIDQFFGVEFETTMKCTESEEEEVTKGKENQLQLSCFINQEVKYLFTGLKLRLQEEI 298
          ************************************************************
cgr-Usp14 TKQSPTLQRNALYIKSSKISRLPAYLTIQMVRFFYKEKESVNAKVLKDVKFPLMLDVYEL 360
hsa-USP14 TKQSPTLQRNALYIKSSKISRLPAYLTIQMVRFFYKEKESVNAKVLKDVKFPLMLDMYEL 358
          ************************************************************:***
cgr-Usp14 CTPELQEKMVSFRSKFKDLEDKKVNQQPNANDKKSSPPKEVKYEPFSFADDIGSNNCGYY 420
hsa-USP14 CTPELQEKMVSFRSKFKDLEDKKVNQQPNTSDKKSSPQKEVKYEPFSFADDIGSNNCGYY 418
          ****************************:.****** **********************
cgr-Usp14 DLQAVLTHQGRSSSSGHYVSWVKRKQDEWIKFDDDKVSIVTPEDILRLSGGGDWHIAYVL 480
hsa-USP14 DLQAVLTHQGRSSSSGHYVSWVKRKQDEWIKFDDDKVSIVTPEDILRLSGGGDWHIAYVL 478
          ************************************************************
cgr-Usp14 LYGPRRVEIMEEESEQ 496
hsa-USP14 LYGPRRVEIMEEESEQ 494
          ****************
```

An * (asterisk) indicates positions which have a single, fully conserved residue.

A : (colon) indicates conservation between groups of strongly similar properties as below - roughly equivalent to scoring > 0.5 in the Gonnet PAM 250 matrix:
A . (period) indicates conservation between groups of weakly similar properties as below - roughly equivalent to scoring =< 0.5 and > 0 in the Gonnet PAM 250 matrix.

Supplementary Figure 1: Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo) alignment of the amino acid sequence for CHO-K1 (cgr) and human (hsa) Usp14.



Target	Avg Ct	Ct Std err
U6 snRNA	18.672	0.064
miR-204	34.1	0.514
miR-338	35.41	0.092
miR-378	22.065	0.089
miR-409	33.268	0.147
miR-455	26.921	0.052
miR-505	26.421	0.132

Supplementary Figure 2: Average Ct values for candidate miRNAs; miR-204-5p, miR-338-3p, miR-378-3p, miR-455-3p and miR-505-3p in CHO DP12 cells during exponential growth.

Supplementary Table 1: miRNA sponge sequences

Sequence Name	miRNA Target	miRBase ID	miRNA Sponge Sequence (5'-3')
NC-spg s	-	-	GTCCCAAGTTTTCAGAAAGCTAACACCGGAAGTTTTCAGAAAGCTAACAGG
NC-spg as	-	-	GACCCTGTTAGCTTTCTGAAAACTTCCGGTGTTAGCTTTCTGAAAACTTGG
miR-204-5p-spg s	miR-204-5p	MI0020454	GTCCCAGGCATAGGACTGAAAGGGAAAATTAGGCATAGGACTGAAAGGGAAGG
miR-204-5p-spg as	miR-204-5p		GACCCTTCCCTTTCAGTCCTATGCCTAATTTTCCCTTTCAGTCCTATGCCTGG
miR-338-spg s	miR-338	MI0020499	GTCCCTCAACAAAATCGAGATGCTGGAAATTTCAACAAAATCGAGATGCTGGAGG
miR-338-spg as	miR-338		GACCCTCCAGCATCTCGATTTTGTTGAAATTTCCAGCATCTCGATTTTGTTGAGG
miR-378-spg s	miR-378-3p	MI0020515	GTCCCGCCTTCTGACGTTAGTCCGATAATTGCCTTCTGACGTTAGTCCGATGG
miR-378-spg as	miR-378-3p		GACCCATCGGACTAACGTCAGAAGGCAATTATCGGACTAACGTCAGAAGGCGG
miR-409-3p-spg s	miR-409-3p	MI0020520	GTCCCAGGGGTTCACTTTCAACATTCAATTAGGGGTTCACTTTCAACATTCGG
miR-409-3p-spg as	miR-409-3p		GACCCGAATGTTGAAAGTGAACCCCTAATTGAATGTTGAAAGTGAACCCCTGG
miR-455-3p-spg s	miR-455-3p	MI0020528	GTCCCAGTGTATATGAAATGGACTGCAATTAGTGTATATGAAATGGACTGCGG
miR-455-3p-spg as	miR-455-3p		GACCCGCAGTCCATTTCATATACACTAATTGCAGTCCATTTCATATACACTGG
miR-505-3p-spg s	miR-505-3p	MI0020538	GTCCCAGAGGAAACCATTTGTGTTGACAATTAGAGGAAACCATTTGTGTTGACGG
miR-505-3p-spg as	miR-505-3p		GACCCGTCAACACAAATGGTTTCCTCTAATTGTCAACACAAATGGTTTCCTCTGG

Supplementary Table 2: Evaluating Sponge – miRNA Interaction and Specificity

Prediction Rank	Sponge	miRNA Target	MRE	Score
1	NC-spg	hsa-miR-769-3p	2	-8.3
2	NC-spg	hsa-miR-920	2	-7.05
3	NC-spg	hsa-miR-1306	10	-6.18
4	NC-spg	hsa-miR-138	5	-6.14
5	NC-spg	hsa-miR-637	4	-5.73
6	NC-spg	hsa-miR-146b-5p	9	-5.34
7	NC-spg	hsa-miR-626	9	-5.17
8	NC-spg	hsa-miR-7	9	-5.14
9	NC-spg	hsa-miR-571	5	-4.87
10	NC-spg	hsa-miR-199b-5p	5	-4.46
11	NC-spg	hsa-miR-146a	9	-4.38
12	NC-spg	hsa-miR-492	2	-4.3
13	NC-spg	hsa-miR-630	9	-4.28
14	NC-spg	hsa-miR-1263	2	-4.27
15	NC-spg	hsa-miR-1299	9	-4.18

Prediction Rank	Sponge	miRNA Target	MRE	Score
1	miR-378-spg	hsa-miR-378	10	-30.48
2	miR-378-spg	hsa-miR-422a	10	-26.51
3	miR-378-spg	hsa-miR-432	7	-16.91
4	miR-378-spg	hsa-miR-383	7	-16.34
5	miR-378-spg	hsa-miR-1268	8	-15.81
6	miR-378-spg	hsa-miR-663	3	-13.08
7	miR-378-spg	hsa-miR-214	7	-12.72
8	miR-378-spg	hsa-miR-665	7	-11.58
9	miR-378-spg	hsa-miR-890	8	-11.21
10	miR-378-spg	hsa-miR-1287	3	-10.06
11	miR-378-spg	hsa-miR-650	10	-9.95
12	miR-378-spg	hsa-miR-524-3p	4	-9.23
13	miR-378-spg	hsa-miR-944	4	-8.83
14	miR-378-spg	hsa-miR-548j	4	-7.73
15	miR-378-spg	hsa-miR-510	7	-7.57

hsa-miR-378-3p: 5’ A**CUGGACUU**GGAGUCAGAAGGC 3’
hsa-miR-422-5p: 5’ A**CUGGACUU**AGGGUCAGAAGGC 3’

Supplementary Table 3: Primer Information (qRT-PCR)

DE - 378	Gene Symbol	Primer Name	Sequence (5'-3')	Tm (°C)	Amplicon Size (bp)
UP	**Clic4**	Clic4 FP	GAGGGGGCTCTTGAAGACAC	60	145
UP	**Clic4**	Clic4 RP	GCAGTCTGCCAGTGTCATCT	60	145
UP	**Hnrnpa1**	Hnrnpa1 FP	GGTCGAGGGGGAAACTTCAG	60	195
UP	**Hnrnpa1**	Hnrnpa1 RP	ATAGCCACTGCCCTGGTTTC	60	195
UP	**Prdx1**	Prdx1 FP	ACTTTTGTGTGCCCCACTGA	60	157
UP	**Prdx1**	Prdx1 RP	TCATGGGTCCCAATCCTCCT	60	157
UP	**Actn4**	Actn4 FP	CAGATCCTCACCCGAGATGC	60	151
UP	**Actn4**	Actn4 RP	TCTCCACATCGTAGCCCAGA	60	151
UP	**Usp14**	Usp14 FP	GAGCTTCAGGGGAAATGGCT	60	193
UP	**Usp14**	Usp14 RP	ACCCAGCACTCGTTAGCATC	60.1	193
UP	**Srxn1**	Srxn1 FP	GTCCTCTGGATCAAAGGGGC	60.1	109
UP	**Srxn1**	Srxn1 RP	GCTTGGCAGGAATGGTCTCT	60	109
UP	**Canx**	Canx FP	TCTGGCAGCGACCTTTGATT	60	182
UP	**Canx**	Canx RP	GGACCAGAGCTCCAAACCAA	59.9	182
UP	**Gnb1**	Gnb1 FP	TGGGATGTCCGAGAAGGGAT	60	114
UP	**Gnb1**	Gnb1 RP	AGCATCGTCTGAACCAGTGG	60	114
DOWN	**S100a4**	S100a4 FP	CTGAAGGAGCTGCTGACCAG	60.4	152
DOWN	**S100a4**	S100a4 RP	ATGGCAATGCAGGACAGGAA	60	152
DOWN	**Calr**	Clar FP	CTCTGGCAGGTCAAGTCTGG	60	199
DOWN	**Calr**	Clar RP	CCTCAGCCTCTTCCTCCTCT	60	199
DOWN	**Hsp90b1**	Hsp90b1 FP	ACCGGGAAGCAACAGAGAAG	60	105
DOWN	**Hsp90b1**	Hsp90b1 RP	CCGTGAGACGCTGAGATACC	60	105
DOWN	**Rpl26**	Rpl26 FP	AGGACACTACAAAGGCCAGC	60	161
DOWN	**Rpl26**	Rpl26 RP	TGTCCAGCTTTAGCCTGGTG	60	161
-	**Gapdh**	Gapdh FP	TGGCTACAGCAACAGAGTGG	59.9	144
-	**Gapdh**	Gapdh RP	GTGAGGGAGATGATCGGTGT	60	144
-	**Xbp1**	Xbp1 FP	CTCCAGAGACGGAGTCCAAG	60	181
-	**Xbp1**	Xbp1 FP	ACAGGGTCCAACTTGTCCAG	60	181
-	**Grp78**	Grp78 FP	GTGCAGAAACTTCGTCGTGA	60	162
-	**Grp78**	Grp78 RP	GGTAGATCGGAACAGGTCCA	59.9	162
-	**Grp94**	Grp94 FP	ACCGGGAAGCAACAGAGAAG	60	105
-	**Grp94**	Grp94 RP	CCGTGAGACGCTGAGATACC	60	105
-	**Atf6**	ATF6 FP	GCCACTGAAGGAAGACAAGC	60	173
-	**Atf6**	ATF6 RP	TGATGGTTTTTGCTGGAACA	60.1	173
-	**Atf4**	Atf4 FP	TTGCAACCTTTTCCCTGTTC	60.1	167
-	**Atf4**	Atf4 RP	TCTGAGGGGGTGTCTTCATC	60	167
-	**Chop**	Chop FP	TTTTGCCTTGGAGACGGTGT	60	160
-	**Chop**	Chop RP	TAGGGATGCGGAGTCTAGGG	59.9	160
-	-	Usp14 3'UTR MRE s	GTCCCTGGGTACAATCAAGTATTTGTCCAGACTGACTGTTGCTGG	-	-
-	-	Usp14 3'UTR MRE as	GACCCAGCAACAGTCAGTCTGGACAAATACTTGATTGTACCCAGG	-	-
-	-	Actn4 3'UTR MRE s	GTCCCGACCACCTGCCCTGCAGAGTCCAGCAGAGGGCACCCAGCTGG	-	-
-	-	Actn4 3'UTR MRE as	GACCCAGCTGGGTCCCCTCTGCTGGACTCTGCAGGGCAGGTGGTCGG	-	-

Data-in-brief article

Title: A proteomic profiling dataset of recombinant Chinese hamster ovary cells showing enhanced cellular growth following miR-378 depletion

Authors: Orla Coleman[1]*, Alan Costello[1]*, Michael Henry[1], Nga T. Lao[1], Niall Barron[2,3], Martin Clynes[1]* and Paula Meleady[1]*
* These authors contributed equally

Affiliations:

[1]National Institute for Cellular Biotechnology, Dublin City University, Dublin 9, Ireland
[2]National Institute for Bioprocess Research and Training, Blackrock, Co. Dublin, Ireland
[3]University College Dublin, Belfield, Dublin 4, Ireland
Contact email: orla.coleman2@mail.dcu.ie

Abstract
The proteomic data presented in this article provide supporting information to the related research article "Depletion of endogenous miRNA-378-3p increases peak cell density of CHO DP12 cells and is correlated with elevated levels of Ubiquitin Carboxyl-Terminal Hydrolase 14" [1]. Control and microRNA-378 depleted CHO DP12 cells were profiled using label-free quantitative proteomic profiling. CHO DP12 cells were collected on day 4 and 8 of batch culture, subcellular proteomic enrichment was performed, and subsequent fractions were analyzed by liquid chromatography tandem mass spectrometry (LC-MS/MS). Here we provide the complete proteomic dataset of proteins significantly differentially expressed by greater than 1.25-fold change in abundance between control and miR-378 depleted CHO DP12 cells, and the lists of all identified proteins for each condition.

Specifications Table

Subject area	Biotechnology
More specific subject area	Proteomics
Type of data	Excel Tables and Figure
How data was acquired	LTQ-Orbitrap XL mass spectrometer (Thermo Scientific), Progenesis QI for Proteomics (Non-linear Dynamics, Waters) and Proteome Discoverer software (Thermo Scientific)
Data format	Relative quantitation calculated and qualitative analysis
Experimental factors	Subcellular enrichment for of CHO DP12 control cells and cells depleted of microRNA-378
Experimental features	Quantitative proteomic profiling of CHO DP12 cells following microRNA-378 depletion
Data source location	Dublin, Ireland
Data accessibility	Data available in this article
Related research article	Costello, A., Coleman, O., Lao, N.T., Henry, M., Meleady, M., Barron, N. & Clynes, M. Depletion of Endogenous miRNA-378-3p Increases Peak Cell Density of CHO DP12 Cells and is Correlated with Elevated Levels of Ubiquitin Carboxyl-Terminal Hydrolase 14. Journal of Biotechnology, In Press, [1]

Value of the data

- This data reveals protein expression patterns associated with microRNA-378
- Differentially expressed proteins between control and miR-378 depleted CHO cells may serve as indicators of CHO cell growth
- This dataset reports enriched proteins from the cytosolic and membrane subcellular fractions of CHO DP12 cells
- This data provides proteomic profiles for two time-points of CHO DP12 batch culture; exponential and stationary phase

Data

The data presents a quantitative proteomic profiling of subcellular-enriched protein fractions from day 4 and day 8 cultures of CHO DP12 cells following microRNA-378 stable depletion. Both the cytosolic and membrane protein enriched fractions were analysed to identify significantly differentially expressed proteins between control and miR-378 depleted CHO cells (miR-378-spg) for each timepoint. Differentially expressed proteins between control and miR-378-spg cells are required to have (i) a p-value ≤ 0.05 on the peptide and the protein level and (ii) a minimum of 1.25-fold change in normalized abundance levels.

Tables 1-4 list the differentially expressed proteins with an increased abundance in miR-378 depleted cells when compared to control cells. Proteins with an increased abundance in miR-378-spg cells represent potential direct targets of miR-378 in CHO cells and are of most interest. Tables 1-4 report the accession number, peptide count, number of unique peptides, ANOVA p-value, q-value, maximum fold-change and protein name. Supplementary table S1 presents the complete list of all differentially overexpressed and under expressed proteins for each subcellular fraction and time-point. Supplementary table S2 presents the qualitative list of all identified proteins for each condition (control and miR-378-spg), subcellular enriched fraction (cytosolic and membrane protein enriched) and time-point (day 4 and day 8 of culture). Heat maps are shown in figure 1 that outlines the clustering of significantly increased versus decreased proteins in miR-378-spg cells, as compared to control cells.

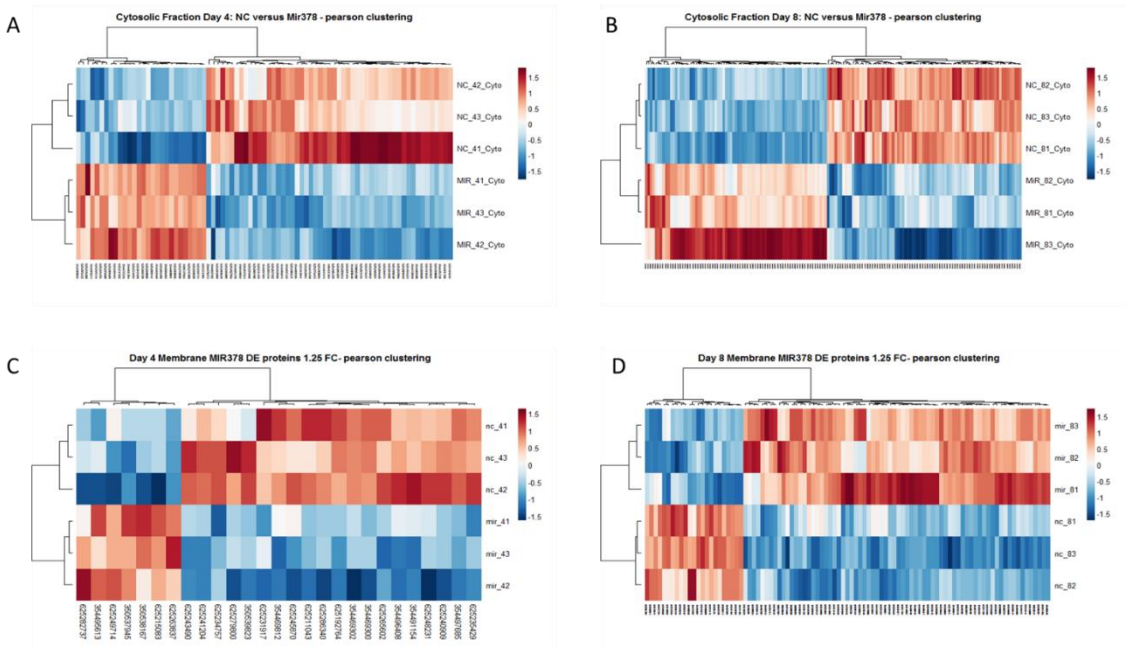


Figure 1: Heat maps of differentially expressed proteins in miR-378-spg CHO cells. 1A and 1B show the clustering of significantly increased and decreased proteins identified in the cytosolic enriched fraction of miR-378-spg cells for day 4 and day 8, respectively. 1C and 1D show the clustering of differentially expressed proteins identified in the membrane enriched fraction of miR-378-spg when compared to control on day 4 and day 8 of culture, respectively. The normalised abundance values of differentially expressed proteins were log2 transformed and hierarchical Pearson clustering was performed on Z-score normalised intensity values.

Table 1: Mass spectrometric identification of 28 proteins from the cytosolic enriched protein fraction with ≥1.25-fold increase in the miR-378 depleted CHO cells on day 4 of cell culture

Accession	Peptides	Unique peptides	Anova (p)	Q value	Max fold change	Protein name
625285532	1	1	1.08E-02	2.96E-02	1.51	splicing factor 3B subunit 3
625233305	2	2	5.27E-03	1.98E-02	1.45	ubiquitin carboxyl-terminal hydrolase 14 isoform X4
354504493	9	9	1.24E-03	1.13E-02	1.44	6-phosphogluconate dehydrogenase, decarboxylating isoform X1
625250820	4	4	7.81E-03	2.69E-02	1.43	copine-1 isoform X3
354500682	1	1	2.00E-02	3.88E-02	1.41	cytochrome b5
625231502	1	1	3.25E-02	4.62E-02	1.40	leucine-rich repeat-containing protein 47 isoform X2, partial
625204380	3	3	2.23E-04	7.06E-03	1.37	chloride intracellular channel protein 4 isoform X1
625279800	1	1	5.08E-03	1.95E-02	1.37	caveolin-1 isoform X1
354481364	1	1	3.55E-02	4.65E-02	1.36	crk-like protein isoform X1
625258134	1	1	9.24E-03	2.76E-02	1.35	sulfiredoxin-1 isoform X2
625290509	1	1	3.70E-02	4.65E-02	1.34	T-complex protein 1 subunit beta isoform X2
350537945	9	9	2.38E-03	1.35E-02	1.33	peroxiredoxin-1
625225560	2	2	5.72E-04	7.39E-03	1.32	heterogeneous nuclear ribonucleoprotein A1 isoform X1
625260720	1	1	1.75E-02	3.77E-02	1.31	TAR DNA-binding protein 43 isoform X1
354477234	2	2	2.29E-02	3.97E-02	1.30	F-actin-capping protein subunit alpha-2 isoform X1
354502560	2	2	1.95E-03	1.35E-02	1.30	protein DJ-1
354480001	1	1	1.29E-02	3.06E-02	1.30	T-complex protein 1 subunit delta
354495613	1	1	9.68E-03	2.82E-02	1.30	thrombomodulin
625250988	1	1	2.32E-02	3.97E-02	1.29	inositol-3-phosphate synthase 1 isoform X2
625224185	1	1	1.14E-02	2.99E-02	1.29	spermidine synthase
625280088	10	10	4.60E-03	1.85E-02	1.28	alpha-enolase isoform X3
625234360	1	1	4.51E-02	4.89E-02	1.28	glutaredoxin-3 isoform X2
625280141	2	2	2.09E-03	1.35E-02	1.28	cytosolic acyl coenzyme A thioester hydrolase isoform X2
625258715	1	1	4.84E-03	1.90E-02	1.26	branched-chain-amino-acid aminotransferase, cytosolic isoform X3
625267589	1	1	3.94E-03	1.70E-02	1.26	alpha-actinin-4 isoform X2
625233493	1	1	4.49E-02	4.89E-02	1.26	26S proteasome non-ATPase regulatory subunit 13 isoform X3
625237309	2	2	1.62E-03	1.30E-02	1.25	adenosylhomocysteinase
625240103	2	2	1.88E-02	3.81E-02	1.25	T-complex protein 1 subunit epsilon

Table 2: Mass spectrometric identification of 73 proteins from the cytosolic enriched protein fraction with ≥1.25-fold increase in the miR-378 depleted CHO cells on day 8 of cell culture.

Accession	Peptides	Unique peptides	Anova (p)	Q value	Max fold change	Protein Name
625185523	1	1	3.14E-02	5.71E-03	18.24	60S ribosomal protein L26 isoform X1
625188420	1	1	3.92E-02	5.85E-03	10.57	heterogeneous nuclear ribonucleoprotein A3 isoform X1
350539695	1	1	1.13E-02	5.34E-03	7.67	protein disulfide-isomerase precursor
346421364	1	1	4.09E-02	5.85E-03	7.35	calreticulin precursor
625242946	1	1	4.88E-02	6.22E-03	4.95	protein S100-A4
625223066	1	1	3.69E-02	5.85E-03	4.78	60S ribosomal protein L22 isoform X1
625195560	1	1	4.30E-02	6.00E-03	4.45	protein disulfide-isomerase A6
625229196	2	2	4.03E-02	5.85E-03	4.13	40S ribosomal protein S6
354478978	1	1	2.80E-02	5.71E-03	3.70	protein S100-A6
625203562	1	1	2.20E-02	5.51E-03	2.84	14-3-3 protein beta/alpha
354495666	1	1	3.30E-02	5.71E-03	2.74	60S ribosomal protein L27a isoform X1
625237172	1	1	3.93E-02	5.85E-03	2.67	cathepsin D
625272649	1	1	4.01E-02	5.85E-03	2.66	nucleophosmin isoform X4
625191956	1	1	3.70E-02	5.85E-03	2.64	glucosidase 2 subunit beta isoform X1
625221706	5	5	2.41E-02	5.51E-03	2.60	alpha-enolase isoform X2
354499455	2	2	3.84E-02	5.85E-03	2.34	60S ribosomal protein L29
625193837	1	1	4.78E-02	6.22E-03	2.31	60S ribosomal protein L4 isoform X1
346986359	2	2	3.15E-02	5.71E-03	2.28	calreticulin precursor
350539629	1	1	4.93E-02	6.22E-03	2.21	40S ribosomal protein S4
625278207	1	1	2.91E-02	5.71E-03	2.09	transcription elongation factor B polypeptide 2 isoform X2
625234436	3	3	3.74E-02	5.85E-03	2.04	acyl-CoA-binding protein
625290232	1	1	2.66E-02	5.51E-03	1.89	60S ribosomal protein L18 isoform X2
625194917	1	1	1.47E-02	5.34E-03	1.88	14-3-3 protein gamma
350537945	7	7	2.26E-02	5.51E-03	1.85	peroxiredoxin-1
625225560	3	3	4.53E-02	6.15E-03	1.85	heterogeneous nuclear ribonucleoprotein A1 isoform X1
625265794	1	1	1.13E-02	5.34E-03	1.84	60S ribosomal protein L35a
350537423	4	4	3.03E-02	5.71E-03	1.79	78 kDa glucose-regulated protein precursor
354484084	2	2	3.84E-02	5.85E-03	1.77	40S ribosomal protein S3a
354487474	5	5	3.26E-02	5.71E-03	1.75	endoplasmin
625242866	3	3	3.26E-02	5.71E-03	1.74	tropomyosin alpha-3 chain isoform X7
625271377	1	1	3.10E-02	5.71E-03	1.73	peptidyl-prolyl cis-trans isomerase FKBP4 isoform X2
350538733	1	1	3.20E-02	5.71E-03	1.71	60S ribosomal protein L13
625218325	1	1	1.98E-02	5.45E-03	1.71	Y-box-binding protein 3, partial
625219233	2	2	1.46E-02	5.34E-03	1.71	heterogeneous nuclear ribonucleoprotein D0, partial
354497356	1	1	3.08E-02	5.71E-03	1.69	ADP-ribosylation factor 3
625286340	1	1	4.66E-02	6.16E-03	1.66	annexin A5
354507332	1	1	1.72E-02	5.34E-03	1.63	60S ribosomal protein L8
346227155	3	3	2.10E-02	5.46E-03	1.62	elongation factor 2
625223526	1	1	1.81E-02	5.39E-03	1.59	40S ribosomal protein S3 isoform X1
625233305	3	3	5.51E-04	2.52E-03	1.58	ubiquitin carboxyl-terminal hydrolase 14 isoform X4
625263837	1	1	1.23E-02	5.34E-03	1.55	reticulocalbin-3 isoform X2
625204380	1	1	2.26E-02	5.51E-03	1.55	chloride intracellular channel protein 4 isoform X1
354506476	2	2	1.53E-02	5.34E-03	1.54	glutathione S-transferase Mu 7
625190571	3	3	2.60E-02	5.51E-03	1.47	tropomyosin alpha-4 chain
354497863	1	1	2.57E-02	5.51E-03	1.45	RNA-binding protein FUS isoform X1
625203986	2	2	2.64E-02	5.51E-03	1.44	peptidyl-prolyl cis-trans isomerase A
354475571	1	1	2.89E-02	5.71E-03	1.42	NSFL1 cofactor p47 isoform X1
625198438	1	1	2.48E-02	5.51E-03	1.42	ran-specific GTPase-activating protein
625282303	1	1	1.94E-02	5.45E-03	1.40	serine/threonine-protein kinase SMG1 isoform X3
625223520	2	2	1.48E-02	5.34E-03	1.39	serpin H1 isoform X1
625190791	1	1	2.90E-03	4.00E-03	1.39	60S ribosomal protein L7a-like

354471594	1	1	2.06E-02	5.45E-03	1.38	cathepsin B
625227859	3	3	1.48E-02	5.34E-03	1.37	glutathione S-transferase Mu 6
625258134	1	1	3.04E-02	5.71E-03	1.37	sulfiredoxin-1 isoform X2
625225201	1	1	1.69E-02	5.34E-03	1.35	annexin A2 isoform X1
625262042	2	2	1.34E-02	5.34E-03	1.35	heat shock protein beta-1 isoform X2
625180993	2	2	1.11E-02	5.34E-03	1.34	eukaryotic initiation factor 4A-I-like
350540646	1	1	4.13E-02	5.85E-03	1.32	phosphoglycerate kinase 1
625222844	3	3	2.44E-04	1.80E-03	1.32	septin-11 isoform X1
625240830	1	1	2.66E-02	5.51E-03	1.31	nucleoside diphosphate kinase B
625222011	1	1	4.63E-02	6.16E-03	1.30	inosine-5'-monophosphate dehydrogenase 2 isoform X1
625289462	1	1	3.93E-02	5.85E-03	1.30	calcium-regulated heat stable protein 1
625199022	1	1	1.31E-02	5.34E-03	1.30	m7GpppX diphosphatase
354489619	1	1	2.61E-02	5.51E-03	1.29	isocitrate dehydrogenase [NADP] cytoplasmic
625249460	2	2	1.10E-02	5.34E-03	1.29	src substrate cortactin
625202098	1	1	4.66E-02	6.16E-03	1.29	myosin light polypeptide 6-like
625256794	2	2	1.36E-04	1.60E-03	1.29	fatty acid-binding protein, adipocyte
350538479	2	2	3.11E-02	5.71E-03	1.28	tubulin beta-6 chain
625267589	5	5	3.71E-03	4.07E-03	1.28	alpha-actinin-4 isoform X2
625206697	1	1	2.49E-02	5.51E-03	1.26	ATP-binding cassette sub-family F member 1 isoform X1
354483012	2	2	1.40E-02	5.34E-03	1.25	heterogeneous nuclear ribonucleoprotein R
625249889	1	1	4.29E-03	4.53E-03	1.25	caldesmon isoform X3
354465044	1	1	2.33E-02	5.51E-03	1.25	rab GDP dissociation inhibitor beta

Table 3: Mass spectrometric identification of 7 proteins from the membrane protein enriched fraction with ≥1.25-fold increase in the miR-378 depleted CHO cells on day 4 of cell culture.

Accession	Peptides	Unique peptides	Anova (p)	Q value	Max fold change	Protein name
350538167	1	1	2.35E-02	3.75E-02	1.52	calnexin precursor
354495613	2	2	2.29E-02	3.71E-02	1.37	thrombomodulin
625263837	3	3	2.63E-03	2.00E-02	1.34	reticulocalbin-3 isoform X2
350537945	3	3	3.71E-03	2.34E-02	1.33	peroxiredoxin-1
625249714	1	1	3.11E-02	3.99E-02	1.31	perilipin-4 isoform X14
625282737	1	1	4.67E-02	4.88E-02	1.31	protein dpy-30 homolog
625215083	3	3	1.83E-02	3.38E-02	1.30	guanine nucleotide-binding protein G(I)/G(S)/G(T) subunit beta-1 isoform X1

Table 4: Mass spectrometric identification of 72 proteins from the membrane protein enriched fraction with ≥1.25-fold increase in the miR-378 depleted CHO cells on day 8 of cell culture.

Accession	Peptides	Unique peptides	Anova (p)	Q value	Max fold change	Protein Name
625244585	2	2	1.10E-02	2.45E-02	2.52	histone H2A.V isoform X2
354496412	1	1	6.56E-03	2.14E-02	2.45	histone H1.0
354480100	5	5	2.14E-02	3.33E-02	2.17	histone H2B type 1
354494381	1	1	1.63E-04	1.65E-02	2.12	fibronectin isoform X1
354494231	1	1	9.56E-03	2.38E-02	2.08	high mobility group nucleosome-binding domain-containing protein 5 isoform X1
345842361	1	1	3.64E-02	4.43E-02	2.08	high mobility group protein HMG-I/HMG-Y
625206001	7	7	1.31E-02	2.76E-02	2.06	histone H3.1-like
625285909	3	3	1.01E-02	2.38E-02	1.84	histone H2A type 1-H-like isoform X1
625229196	1	1	4.46E-02	4.84E-02	1.81	40S ribosomal protein S6
625205207	1	1	2.86E-03	1.65E-02	1.77	rRNA 2'-O-methyltransferase fibrillarin, partial
354480104	6	6	1.53E-02	2.89E-02	1.74	histone H1.4 isoform X1
625289934	1	1	3.81E-02	4.46E-02	1.73	calumenin isoform X2
350537403	1	1	2.63E-02	3.77E-02	1.68	DNA topoisomerase 2-alpha
625262546	1	1	3.48E-02	4.36E-02	1.68	replication protein A 14 kDa subunit
625209863	1	1	1.44E-02	2.87E-02	1.67	alpha-parvin
625234125	4	4	1.10E-02	2.45E-02	1.63	elongation factor 1-gamma
350538167	3	3	1.78E-03	1.65E-02	1.60	calnexin precursor
625284147	1	1	8.52E-03	2.30E-02	1.50	legumain
350539823	1	1	4.09E-04	1.65E-02	1.50	heat shock cognate 71 kDa protein
625204124	1	1	4.42E-02	4.82E-02	1.47	platelet glycoprotein 4
625256908	1	1	3.23E-03	1.65E-02	1.47	septin-2
625211254	2	2	2.82E-02	3.86E-02	1.47	plectin isoform X1
625260069	1	1	7.33E-03	2.14E-02	1.44	14-3-3 protein epsilon isoform X2
354504493	2	2	1.54E-03	1.65E-02	1.44	6-phosphogluconate dehydrogenase, decarboxylating isoform X1
625231575	2	2	7.14E-03	2.14E-02	1.43	eukaryotic initiation factor 4A-II isoform X1
625274484	1	1	4.37E-02	4.80E-02	1.42	serum albumin isoform X3
625262669	1	1	3.56E-02	4.39E-02	1.42	cellular nucleic acid-binding protein isoform X2
625243141	1	1	2.10E-02	3.32E-02	1.41	ATP-dependent RNA helicase DDX39A
625216841	1	1	6.75E-03	2.14E-02	1.41	coronin-1B
625292335	1	1	1.81E-02	3.04E-02	1.40	high mobility group protein B2 isoform X2
354489619	1	1	7.40E-03	2.14E-02	1.40	isocitrate dehydrogenase [NADP] cytoplasmic
625215758	1	1	2.88E-03	1.65E-02	1.39	enoyl-CoA delta isomerase 1, mitochondrial isoform X1
354483223	1	1	2.25E-02	3.44E-02	1.39	prolyl 4-hydroxylase subunit alpha-1 isoform X1
354467247	1	1	2.41E-03	1.65E-02	1.39	NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 9, mitochondrial
625284339	3	3	1.46E-02	2.87E-02	1.38	succinate dehydrogenase ubiquinone] iron-sulfur subunit, mitochondrial isoform X2, partial
625231917	1	1	2.18E-02	3.36E-02	1.38	guanine nucleotide-binding protein subunit beta-4
625238921	1	1	4.00E-02	4.53E-02	1.38	EH domain-containing protein 4 isoform X2
354500398	1	1	1.25E-03	1.65E-02	1.37	ubiquitin-like modifier-activating enzyme 1 isoform X1
625190571	1	1	2.57E-02	3.74E-02	1.35	tropomyosin alpha-4 chain

354485048	1	1	3.51E-02	4.37E-02	1.35	polymerase I and transcript release factor
354485701	1	1	4.91E-04	1.65E-02	1.35	stomatin-like protein 2, mitochondrial
354492573	1	1	4.19E-02	4.72E-02	1.35	actin-related protein 3B isoform X1
354465900	2	2	1.57E-02	2.93E-02	1.35	ATP-dependent RNA helicase DDX3X isoform X1
354484391	1	1	2.57E-02	3.74E-02	1.35	14-3-3 protein zeta/delta
625190862	1	1	1.66E-02	2.97E-02	1.34	NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 10, mitochondrial isoform X1
625240103	1	1	4.16E-03	1.83E-02	1.34	T-complex protein 1 subunit epsilon
625188420	1	1	4.22E-02	4.73E-02	1.33	heterogeneous nuclear ribonucleoprotein A3 isoform X1
625211596	1	1	6.89E-04	1.65E-02	1.33	60S ribosomal protein L7 isoform X1
625248231	1	1	2.96E-03	1.65E-02	1.33	ADP/ATP translocase 1 isoform X2
625251833	1	1	2.86E-02	3.86E-02	1.32	hydroxymethylglutaryl-CoA lyase, mitochondrial isoform X3
350540646	2	2	1.59E-02	2.93E-02	1.32	phosphoglycerate kinase 1
625254434	1	1	6.32E-03	2.14E-02	1.32	superoxide dismutase [Mn], mitochondrial isoform X2
625249635	1	1	5.33E-03	2.06E-02	1.31	lon protease homolog, mitochondrial
625208910	1	1	4.97E-02	5.03E-02	1.30	septin-7 isoform X1
354486011	1	1	4.29E-02	4.74E-02	1.30	acyl-coenzyme A thioesterase 1 isoform X1
625213146	1	1	9.70E-03	2.38E-02	1.29	integrin beta-1 isoform X1
625232358	3	3	4.03E-04	1.65E-02	1.29	lipoprotein lipase isoform X1
625279800	1	1	4.90E-02	5.00E-02	1.28	caveolin-1 isoform X1
625235290	1	1	4.73E-02	4.89E-02	1.28	peroxiredoxin-5, mitochondrial
625215083	2	2	1.26E-03	1.65E-02	1.28	guanine nucleotide-binding protein G(I)/G(S)/G(T) subunit beta-1 isoform X1
346986359	3	3	2.62E-03	1.65E-02	1.27	elongation factor 1-alpha 1
625201330	1	1	1.81E-02	3.04E-02	1.27	cell division control protein 42 homolog
625288359	2	2	1.73E-02	3.01E-02	1.27	dephospho-CoA kinase domain-containing protein
552953713	1	1	1.43E-02	2.87E-02	1.26	40S ribosomal protein S7
354482483	1	1	2.73E-02	3.82E-02	1.26	vimentin
625243995	2	2	3.76E-02	4.46E-02	1.26	leucine-rich PPR motif-containing protein, mitochondrial isoform X2
625184898	1	1	3.26E-02	4.24E-02	1.25	39S ribosomal protein L12, mitochondrial isoform X1
625236680	1	1	1.11E-02	2.45E-02	1.25	60 kDa heat shock protein, mitochondrial
625183009	2	2	1.37E-02	2.85E-02	1.25	triosephosphate isomerase isoform X1
354486540	2	2	2.91E-03	1.65E-02	1.25	hydroxymethylglutaryl-CoA synthase, mitochondrial
625224152	1	1	2.50E-02	3.72E-02	1.25	nuclear body protein SP140-like isoform X1
625291524	1	1	3.79E-02	4.46E-02	1.25	mitochondrial import inner membrane translocase subunit Tim13 isoform X3, partial

Experimental Design, Materials and Methods

1. Subcellular protein extraction and in-solution protein digestion

Triplicate biological samples for control and miR-378 depleted cells were collected on day 4 and day 8 of batch cultures. Subcellular protein enrichment was achieved using the Mem-Per Plus Membrane protein extraction kit (#89842, Thermo Fisher Scientific) which yielded a cytosolic and membrane protein enriched fraction. Protein concentration was determined using the QuickStart Bradford assay (Bio-rad). Equal concentrations (100 µg) of protein from each sample were purified and trypsin digested for mass spectrometry using the filter-aided sample preparation method as previously described [2]. The resulting peptide samples were purified using Pierce C18 spin columns then dried using vacuum centrifugation and suspended in 2% acetonitrile and 0.1% trifluoracetic acid in LC grade water prior to LC-MS/MS analysis.

2. Label-free liquid chromatography mass spectrometry

Quantitative label-free liquid-chromatography mass spectrometry (LC–MS/MS) analysis of mir-378-spg and NC-spg membrane and cytosolic fractions from day 4 and day 8 was carried out using a Dionex UltiMate™ 3000 RSLCnano system (Thermo Fisher Scientific) coupled to a hybrid linear LTQ Orbitrap XL mass spectrometer (Thermo Fisher Scientific). LC-MS/MS methods were applied as previously described [3]. A 5 µL injection of each sample was loaded onto a C18 trapping column (PepMap100, C18, 300 µm × 5 mm; Thermo Fisher Scientific). Each sample was desalted for 5 min using a flow rate of 25 µL/min with 2% ACN, 0.1% TFA before being switched online with the analytical column (PepMap C18, 75 µm ID × 250 mm, 3 µm particle and 100 Å pore size; (Thermo Fisher Scientific)). Peptides were eluted using a binary gradient of Solvent A (2% ACN and 0.1% formic acid in LC grade water) and Solvent B (80% ACN and 0.08% formic acid in LC grade water). The following gradient was applied; 6–25% solvent B for 120 min and 25–50% solvent B in a further 60 min at a column flow rate of 300 nL/min. Data was acquired with Xcalibur software, version 2.0.7 (Thermo Fisher Scientific). The LTQ Orbitrap XL was operated in data-dependent mode with full MS scans in the 400–1200 m/z range using the Orbitrap mass analyser with a resolution of 30,000 (at m/z 400). Up to three of the most intense ions (+1, +2, and +3) per scan were fragmented using collision-induced dissociation (CID) in the linear ion trap. Dynamic exclusion was enabled with a repeat count of 1, repeat duration of 20 s, and exclusion duration of 40 s. All tandem mass spectra were collected using a normalized collision energy of 32%, and an isolation window of 2 m/z with an activation time of 30 ms.

3. Quantitative Label-free LC-MS/MS Data Analysis

Protein identification was achieved using Proteome Discoverer 2.1 with the Sequest HT and MASCOT algorithm followed by Percolator validation [4] to apply a false-discovery rate <0.01. Data was searched against the NCBI Chinese Hamster (*Cricetulus griseus*) protein database containing 44,065 sequences (fasta file downloaded November 2015). The following search parameters were used for protein identification: (1) precursor mass tolerance set to 20 ppm, (2) fragment mass tolerance set to 0.6 Da, (3) up to two missed cleavages were allowed, (4) carbamidomethylation of cysteine set as a static modification and (5) methionine oxidation set as a dynamic modification. The complete lists of all identified proteins from the cytosolic and membrane enriched fractions of day 4 and day 8 cell cultures of the control (NC378-spg) and miR-378-spg are provided in supplementary table S2.

Quantitative label-free data analysis was performed using Progenesis QI for Proteomics (version 2.0; Nonlinear Dynamics, a Waters company) as described by the manufacturer (www.nonlinear.com). To counteract potential drifts in retention time a reference run was assigned to which all MS data files were aligned. The triplicate samples from the two experimental groups (NC-378-spg and miR-378-spg) were set up for differential analysis and label-free relative quantitation was carried out after peak detection, automatic retention time calibration and normalisation to account for experimental variation. The experimental analyses performed compared the three biological replicates for control cells to miR-378-spg triplicates for each timepoint and subcellular fraction collected. The following settings were applied to filter peptide features (1) peptide features with a one-way ANOVA p-value <0.05 between experimental groups, (2) mass peaks with charge states from +1 to +3 and (3) greater than one isotope per peptide. The normalised data is transformed prior to statistical analysis, using an arcsinh transformation to meet the assumptions of the one-way ANOVA test. A mascot generic file (mgf) was generated from all exported MS/MS spectra which satisfied the peptide filters, the mgf was used for peptide and protein identification in Proteome Discoverer. Protein identifications were imported into Progenesis and considered differentially expressed if they passed the following criteria: (i) a protein one-way ANOVA p-value <0.05 and (ii) a ≥1.25-fold change in relative abundance between the two experimental groups. All differentially expressed proteins identified between NC378-spg and miR-378-spg cells are reported in supplementary table S1.

Heatmaps illustrating protein abundances for statistically significant and differentially expressed proteins were designed using ggplot2 in R-studio. The normalised abundance values of differentially expressed proteins were determined using Progenesis QI for Proteomics and were loaded as a txt file into R-studio and the data was log2 transformed. Hierarchical Pearson

clustering was then performed on Z-score normalised intensity values by clustering both samples and proteins.

Acknowledgements

This research was carried out with financial support from Science Foundation Ireland, grant number 13/IA/1841.

References

[1] Costello, A., Coleman, O., Lao, N.T., Henry, M., Meleady, M., Barron, N. & Clynes, M. Depletion of Endogenous miRNA-378-3p Increases Peak Cell Density of CHO DP12 Cells and is Correlated with Elevated Levels of Ubiquitin Carboxyl-Terminal Hydrolase 14. Journal of Biotechnology [In Press]

[2] Coleman, O., Henry, M., Clynes, M., & Meleady, P. (2017) Filter-aided sample preparation (FASP) for improved proteome analysis of recombinant chinese hamster ovary cells. Heterologous Protein Production in CHO Cells: Methods and Protocols, 187-194.

[3] Henry, M., & Meleady, P. (2017) Clinical proteomics: Liquid Chromatography–Mass spectrometry (LC–MS) purification systems. Methods in Mol. Bol. Springer, pp. 375-388

[4] Käll, L., Canterbury, J.D., Weston, J., Noble, W.S. & MacCoss, M.J. (2007) Semi-supervised learning for peptide identification from shotgun proteomics datasets. *Nature methods.* Vol. 4(11), pp 923-925.

Chapter 3

Conditional Knockdown of Endogenous MicroRNAs in CHO Cells Using TET-ON-SanDI Sponge Vectors

Costello, A., Lao, T.N., Clynes, M. & Barron, N.
Methods in Molecular Biology, DOI: 10.1007/978-1-4939-6972-2_6, PMID: 28493125
Impact Factor (2017): 1.200

I was the primary contributor to the experimental work in this chapter. Dr Nga Lao and I worked closely to draft the manuscript. Prof. Martin Clynes and Prof. Niall Barron contributed to experimental design and redaction of the manuscript.

ABSTRACT

MicroRNA (miRNA) are small, noncoding RNAs of about 22 nucleotides in length and have proven to be targets of genetic modifications for desirable phenotype in the biotech industry. The use of a constitutively expressed sponge vector in Chinese hamster ovary (CHO) cells for down regulating the levels of endogenous microRNAs has shown to be more advantageous than that of the synthetic antagomirs. The application of miRNA sponges in biotechnological processes, however, could be more efficient, if expression of miRNA sponges could be timely tuned. In this chapter, we present a method for generation of stable CHO cell lines expressing TET-ON-SanDI-miRNA-sponge which only turns on in the presence of an inducer.

3.1 INTRODUCTION

Chinese Hamster Ovary (CHO) cells have been the workhorse of the biopharmaceutical industry since their development in the 1950s ***(1)***. Process optimization is a major area of interest in the CHO community. This includes numerous factors of the fed batch process, such as vessel design, media composition, temperature shift, online monitoring of process variables (pO_2, CO_2, nutrient and glucose concentration) to name a few. On a cellular level the standard industry approach is to screen a large panel of clones for desired traits. However, the process of single cell cloning is both time consuming and gives no information on why one clone is outperforming the others.

MiRNAs with their ability to impact, regulate complex cellular pathways and alter phenotypes have become a popular genetic engineering tool ***(2)***. Traditionally, miRNA loss of function studies would be carried- out using chemically modified oligonucleotides. These are chemically synthesized antisense molecules (antagomirs or anti-miRs) that inhibit endogenous miRNAs by directly binding miRNAs, hence rendering them inactive. However, application of anti-miRs in biotechnological processes is unpractical, as they are transiently expressed. MiRNA sponges, developed by Ebert et al., ***(3)*** as an effective method for scavenging endogenous miRNA *in vivo* have been shown to be as effective as the present antisense technology, and their activities are specific to miRNA seed families. Kluvier et al., ***(4)*** developed a simple and effective method to generate miRNA sponges from short oligo duplexes with SanDI overhangs. Constitutive stable expression of miRNA sponges in CHO cells has been successful and suggests that they can be used as potential tools to generate high producer CHO cells of desirable phenotypes ***(5, 6)***. However, the levels of specific miRNAs are tightly regulated in the cell in response to the cell's needs. MiRNAs that positively impact on specific productivity or product quality but have a negative phenotype with regard to growth in a constitutive expression system are of no real benefit in biotech processes. Therefore, by mimicking nature and timely tuning the level of specific miRNAs at specific stages of culture, analogous to the use of inducible promoter, one could potentially improve a product yield or growth beneficial phenotype, more efficiently respond to biotechnological processes.

The need for the tight regulation of specific genes has existed for decades. Gene function studies would in principle benefit greatly from not only a controllable on/off expression but also controlled expression at a defined level. The inducible control of gene activity by means of heat shock**,** heavy metal ions ***(7, 8)*** or hormones ***(9)*** have been described, however, these systems are limited by their "leakiness" ***(8)*** and pleiotropic effects ***(10)***. In relation to CHO cell fed batch culture, temperature shift is routinely used to boost cellular productivity.

A tetracycline-responsive promoter was introduced by Gossen and Bujard in 1992 ***(13)***. This system based on the Tn10-specified tetracycline-resistance operon of *Escherichia coli (E.coli)*

(12) where in the transcription of resistance genes are negatively regulated by tetracycline repressor (tetR). In the presence of tetracycline, tetR is unable to bind to the promoter region of the operon, allowing transcription of the resistance genes. By attaching the tetR to the C-terminal of a VP16 from herpes simplex virus (HSV), a hybrid trans-activator was created that stimulates minimal promoters fused with a tetracycline operator (tetO), a promoter was generated that is silent in the presence of even very low levels of tetracycline, due to the prevention of the tetracycline-controlled trans-activator (tTA) binding to the (tetO), but can be induced ~1000 fold. This is referred to as the TET-off or tTA dependent system. In 1995, Gossen *et al*, ***(13)*** explored the possibility of reversing the effect of tetracycline and its derivatives on the tetO. Random mutagenesis of the Tn10 tetR gene was used to identify the sequence responsible for tetR binding to tetO in the presence of tetracycline which was named reverse tTA (rtTA) ***(14).*** This resulted in the TET-ON system, an inducible promoter, tune able with respect to doxycycline (Dox) concentration.

For the conditional knockdown of endogenous miRNAs, we take advantage of the "TET-ON" system described above. In this chapter, we outline methods to generate TET-ON-*SanDI*-miR sponge constructs containing a desired number of miRNA binding sites and to characterize stable CHO cells expressing the inducible miR-sponges which are expressed in the presence of doxycycline inducer, a tetracycline derivative.

3.2 MATERIALS

3.2.1 CLONING AND PCR SCREENING

1. Oligos: purchase from Integrated DNA Technology (IDT) or MWG Eurofins.
2. *E. coli* DH5α for routine sub cloning kit (Invitrogen).
3. SOC medium (Invitrogen).
4. Ampicillin, sodium salt: Prepare 100mg/ml stock by dissolve 1g of ampicillin in sterilized deionized water to a final volume of 10mL. Store aliquots at -20^{o}C. One mL of 100mg/mL stock is used for 1 L of medium to achieve a final concentration 100ug/mL.
5. Luria Bertani (LB): Dissolve 25g of powder LB in deionized water to a final volume of 1 L. Autoclave.
6. Luria Bertani-Agar (LB): Dissolve 25g of powder LB in deionized water to a final volume of 1 L. Add 15g agar. Autoclave. Allow to cool before adding antibiotic.
7. Restriction enzymes: Fisher Scientific or New England Bio labs, store at -20^{o}C.

8. Fast Alkaline Phosphatase (Fisher Scientific).
9. T4 Polynucleotide Kinase (NEB).
10. T4 ligase (Roche).
11. 10 X T4 Ligation Buffer (NEB)
12. MyTaq Red DNA Polymerase (BIOLINE)
13. Petri Dishes
14. Plasmid Mini-prep kit and Midi-prep kit (Qiagen or Invitrogen).
15. QiaQuick Gel Extraction Kit (Qiagen).
16. PCR cleaning kit (Qiagen or Invitrogen)
17. 50 X Tris-Acetate-EDTA (TEA) (Fisher Scientific): For routine electrophoresis, dilute 10 mL with 490 mL water (1 X TEA).
18. Agarose Gel; for a 0.8% (w/v) gel mix 0.8 g of agarose powder with 100 mL of 1X TEA buffer. Microwave for 1-2 minutes. Cool down the solution for 5 minutes and add 8 µL Safe View Nucleic Stain (NBS Biologicals). Pour solution into the gel casting tray and insert comb. Allow the gel to solidify.
19. DNA Gel Loading Dye (6X).
20. 100bp DNA Ladder (New England BioLabs).
21. Heat block.
22. Water bath.
23. PCR machine.
24. Orbital shaker.
25. Incubator.
26. Gel running unit.
27. Trans-illuminator.

3.2.2 STABLE CHO CELL LINE DEVELOPMENT

1. TET-ON-*SanDI*-sponge vector (in house made) or basic vectors can be purchased from Addgene and modified.
2. CHO cells: order from ATCC (CRL-12445).
3. Medium for transfection: Dulbecco's Modified Eagle's Medium/Nutrient Mixture F12-Ham (SIGMA ALDRICH) and CHO-S-SFM (Gibco)
4. Standard Fetal Bovine Serum (PAA-The Cell Culture Company)
5. 0.5M EDTA (Ambion) to make 100 X EDTA: add 5.3mL 0.5M EDTA to 44.7mL water
6. PBS buffer: Dissolve 8g NaCL, 0.2g KCl, 1.44g Na2HPO and 0.24g KH2PO4 in 800mL water, adjust pH to 7.4. Adjust volume to 1L with water. Autoclave.
7. 10 X Trypsin (GIBCO).
8. Trypsin-EDTA solution: For 500mL: add 50mL 10 X Trypsin, and 5mL 100 X EDTA into 445mL PBS.
9. Doxycycline hyclate (DOX) (SIGMA). To make 50mg/mL solution: Dissolve 1g in 20mL sterilized water. Store at -80oC. To make 100ug/mL solution, add 1mL 50mg/mL solution to 49mL sterilized water. Store 1mL aliquots at -20oC.
10. Transfection reagent: TransIT-X2 Dynamic Delivery System (Mirus).
11. Guava easyCyte (EMD Millipore)

3.3 METHODS

3.3.1 METHOD FOR CLONING AND SCREENING OF E. COLI TRANSFORMANT CLONES

3.3.1.1 Design SanDI sponge oligos for cloning into the backbone vector

The *SanDI* sponge oligos contain two miRNA binding sites (MBS) with *SanDI* overhangs for cloning and a 4-5nt spacer between (Figure 1). The nucleotide composition of the spacer can be altered as required. The MBS is a sequence complementary to the mature miRNA of interest with a 3 base pair mismatch and 1 nucleotide deletion starting at base 9-12 from the 5' end of the miRNA. This creates a bulge, inhibiting AGOII, a component of the RISC complex from degrading the transcript. Mature miRNA sequences can be found on miRNA databases such as miRBase (http://www.mirbase.org/) or in literature ***(15)***.

1. Obtain the mature miRNA from miRBase (http://www.mirbase.org/) or in literature.
2. Reverse complement to obtain one MBS.

3. Manually modify to add the bulged nucleotides, avoid the 8 nucleotides of seed region.

4. Add 4-5 spacer and a second MBS unit. This is the sense sponge oligo.

5. Reverse complement the sense sponge oligo to obtain the antisense sponge oligo.

6. Add overhang for cloning into *SanDI* site.

7. Input the sense bulged sponge oligo into the online miRNA prediction tool http://genie.weizmann.ac.il/pubs/mir07/mir07_prediction.html or STarMir ohttp://sfold.wadsworth.org/cgi-bin/starmir.pl. In both web sites, the lower ddG or ΔG total is expected for the perfectly matched MBS than the bulged MBS.

8. Oligos can be ordered from any companies that provide oligos for standard polymerase chain reaction (pcr). Input the sequence in the 5'-3' orientation.

9. Similar procedure is done to design negative sponge oligos using scramble sequence.

3.3.1.2 Cloning SanDI-sponge oligos into backbone TET-ON-SanDI-vector

3.3.1.2.1 Oligo Annealing and Phosphorylation (see Note 1)

1. Re-suspend the oligo's to give a 100µM stock.

2. In a clean 0.2ml PCR tube make up the following reaction:

1µL 100µM Oligo 1

1µL 100µM Oligo 2

1µL 10 X T4 Ligation buffer (NEB)

6.5µL Nuclease free water

0.5µL T4PNK (NEB)

Incubate at 37°C for 30 minutes.

3. On a pre-heated heating block, incubate tubes at 100°C for 5 minutes. Spin down tubes to bring condensation in the lids to the bottom of the tube.

4. Return tubes to the heat block, turn off heat block and allow the reaction to cool slowly in the block for at least 2 hours or until the temperature in the block drops below 40°C.

5. Store at -20°C.

6. The oligo duplex should be diluted 1:3, to a concentration of 300-400ng/µL, with nuclease free water prior to its use in a ligation reaction.

3.3.1.2.2 Vector digestion and De-phosphorylation

1. In a 0.2ml PCR tube make up the following reaction

5uL TET-ON-*SanDI*-HYG vector (5ug)

10uL Fast Digest buffer (Thermo Scientific)

5uL Fast Digest *SanDI* (Thermo Scientific)

5uL Fast Alkaline phosphatase

75uL Nuclease free water

2. Incubate for 1 hour at 37°C.
3. Clean digested vector using PCR cleaning kit, according to the supplier instruction. Store at -20ºC.
4. Verify the concentration of the digested vector using a nanodrop. Store at -20ºC.

3.3.1.2.3 Ligation

1. In a 0.2ml PCR tube make up the following reaction

1 µL TET-ON-*SanDI*-HYG digested above (section 3.1.2.2) (50ng)

1µL Oligo Duplex (section 3.1.2.1 above)

1µL 10 X T4 Ligase Buffer (Roche Diagnostics)

1µL T4 Ligase (Roche Diagnostics)

7 µL Nuclease free water

2. Incubate overnight at 16°C in a water bath.

Store at -20°C or use immediately for transformation in the following step.

3.3.1.2.4 Transformation

1. To a 1.7ml Eppendorf tube add 50µL of competent DH5α sub cloning efficiency (Invitrogen) and 5µL of the ligation mix from section 3.1.2.3.

2. Incubate for 30 minutes on ice.
3. Transformation is done by heat shocking the cells for 30 seconds at 42°C (water bath warmed in advance).
4. Incubate on ice for 2 minutes.
5. Add 500µL of S.O.C media (Invitrogen).
6. Revive the cells by incubating at 37°C for 1 hour with shaking (220rpm)
7. Pellet the cells by centrifuging at 4,000 rpm for 3 minutes.
8. Decant all but 50µL of supernatant.
9. Re-suspend the pellet and plate the total 50µL of cells on an LB agar plate containing ampicillin.
10. Incubate plates upside down overnight at 37°C.

3.3.1.3 Identify and verify the transformant clones

To identify colonies positive for oligo inserts a PCR screening method is used. This provides fast method of screening a large number of clones.

1. On a LB agar plate (100 µg/ mL, ampicillin) draw a grid of 8-16 squares and number each section

1-8/16.

2. Add 3µL of nuclease free water to 250µL PCR tubes.
3. Pick single colonies form the LB plates (section 3.1.2.4). Suspend the single colony in 3µL nuclease free water. Take 2µL of this and spread it in one square of the gridded plate. Each PCR reaction corresponds to a square on the gridded plate. Incubate plates overnight at 37°C.
4. Perform PCR reaction with 1µL cells as described in the following section.

   4.1 Make a PCR master mix with the following components per reaction;

5µL 2X MyTag reaction mix

0.5μL 10 μM forward primer

0.5 μL 10 μM reverse primer

3μL Nuclease free water

Mix well by gently pipetting

4.2 Add 9μL of master mix to each reaction tube that contain 1μL cell.

4.3 Perform PCR as follows:

Initial denaturation	94°C, 1min
25 cycles :	94°C, 15sec; 55°C, 15sec; and 72°C, 1min
Hold:	4°C, ∞

5. Run all 10μL of pcr reactions on a 0.8% agarose gel.
6. Miniprep. of the correspondent clones containing the interest number of MBS, according to the kit provider.
7. Quantify the concentration of the plasmid DNA by a nanodrop.
8. Verify the plasmid miniprep. by sequencing (*see* **Note 2** & **Note 3**).
9. Store plasmid at -20°C.

3.3.2 METHOD FOR GENERATION OF STABLE CHO CELL LINES EXPRESSING THE TET-ON-SANDI-MIR SPONGE

The plasmid is now ready for transfection into the CHO cell lines of interest. The choice of transfection method and transfection reagents depends on CHO strains. Pools of stable CHO cell lines are first selected based on the antibiotic marker on the backbone vector (*see* **Note 4**). Preliminary experiments should be carried out on the pools before single stable CHO cell lines are isolated for further characterization.

3.3.2.1 Induction testing (see Note 5)

This should be done first with the CHO cell pools expressing the negative control sponge (NC). A range of concentration of DOX should be tested to determine the concentration of DOX

suitable for the CHO strain of interest. Induction of TET-ON systems can be achieved with as little as 10ng /ml DOX. In addition, the induction regime should also be determined (Figure 2). As seen in Figure 2, induction of over 60% of the total mixed population was achieved, Flow Cytometry (FACS) can be used to obtain sub-pools for different levels of miR-sponge expression if necessary.

3.3.2.2 Induction for phenotyping and molecular analysis

A wide range of bioprocess-related phenotypes of CHO cell pools or sub-pools expressing the inducible sponge vectors can be further characterized such as cell proliferation, life span of the CHO culture, pharmaceutical product quality or stress response. These characterizations should be carried out in parallel for both the miR-sponge and negative control (NC) sponge. Experiments should be carefully designed to obtain molecular data such as the levels of endogenous miRNA, putative target transcripts and proteins in the absence and presence of inducer in order to gain insight into the biological role of the interest miRNA. Functional characterization of single stable clone expressing the miRNA sponge is carried in similar way.

3.3.2.3 Statistical Analysis

The values reported represent the mean of three independent experiments with error displayed as standard deviation. Statistical analysis was carried out by a two-tailed homoscedastic student t-test, ($p \leq 0.05$ *, $p \leq 0.01$**, $p \leq 0.001$ ***) of three independent experimental replicates.

3.4 NOTES

1. Molecular cloning and bacteria-related work are done on the bench. Standard care should be taken including cleaning the bench and pipettes with 70% industrial methanol solution. Molecular reagents are stored at -20oC freezer, kept on ice when carrying out the procedure and return immediately to the freezer after use.

2. It is essential that the number of MBS of the transformant clones is verified by sequencing before use to generate stable transfected CHO cell lines.

3. The number of MBS inserts for specific miRNA and scramble sequence should be the same for accurate assessment.

We found the optimal concentration of oligo to use to range between 300-400ng, this gives a ratio of ~1:1000 vector: insert, this correlated with the work seen in (4).

4. The optimal concentration of antibiotic used for selection of stable CHO cell lines should be determined in advance using antibiotic killing curve approach.

5. Induction should be tested first with the backbone vector before use for cloning. We found that DOX is quite toxic to CHO cells, however at the low concentration at 10ng/mL (which is needed to be determined based on the CHO strain used) the negative effect of DOX on CHO cell viability is negligible, and at the same time sufficient to use for isolation of induced population for further characterization. The leaky expression of TET inducible systems is well known (11). Efforts to limit the extent of basal expression have included; titration of vector, multi-vector approaches, rtTA promoter optimization (16) and incorporation of a transcriptional repressor/silencer (17).The addition of DOX to growth media in this study did not initiate transcription of the miRNA sponge, it merely enhanced it in a conditional fashion.

ACKNOWLEDGEMENTS

This work was supported by the Scientific Foundation of Ireland (SFI) grant number 13/IA/1963 13/IA/1841

3.5 REFERENCES

1. Lewis NE, Liu X, Li Y, Nagarajan H, Yerganian G, O'Brien E, Bordbar A, Roth AM, Rosenbloom J, Bian C, Xie M, Chen W, Li N, Baycin-Hizal D, Latif H, Forster J, Betenbaugh MJ, Famili I, Xu X, Wang J, Palsson BO (2013) Genomic landscapes of Chinese hamster ovary cell lines as revealed by the Cricetulus griseus draft genome. Nat Biotechnol 31:759-765. doi: 10.1038/nbt.2624

2. Jadhav V, Hackl M, Druz A, Shridhar S, Chung CY, Heffner KM, Kreil DP, Betenbaugh M, Shiloach J, Barron N, Grillari J, Borth N (2013) CHO microRNA engineering is growing up: recent successes and future challenges. Biotechnol Adv 31:1501-1513. doi: 10.1016/j.biotechadv.2013.07.007

3. Ebert MS, Neilson JR, Sharp PA (2007) MicroRNA sponges: competitive inhibitors of small RNAs in mammalian cells. Nat Methods 4:721-726. doi: nmeth1079

4. Kluiver J, Gibcus JH, Hettinga C, Adema A, Richter MK, Halsema N, Slezak-Prochazka I, Ding Y, Kroesen BJ, van den Berg A (2012) Rapid generation of microRNA sponges for microRNA inhibition. PLoS One 7:e29275. doi: 10.1371/journal.pone.0029275

5. Sanchez N, Kelly P, Gallagher C, Lao NT, Clarke C, Clynes M, Barron N (2014) CHO cell culture longevity and recombinant protein yield are enhanced by depletion of miR-7 activity via sponge decoy vectors. Biotechnol J 9:396-404. doi: 10.1002/biot.201300325

6. Kelly PS, Breen L, Gallagher C, Kelly S, Henry M, Lao NT, Meleady P, O'Gorman D, Clynes M, Barron N (2015) Re-programming CHO cell metabolism using miR-23 tips the balance towards a highly productive phenotype. Biotechnol J 10:1029-1040. doi: 10.1002/biot.201500101

7. Brinster RL, Chen HY, Warren R, Sarthy A, Palmiter RD (1982) Regulation of metallothionein--thymidine kinase fusion plasmids injected into mouse eggs. Nature 296:39-42

8. Mayo KE, Warren R, Palmiter RD (1982) The mouse metallothionein-I gene is transcriptionally regulated by cadmium following transfection into human or mouse cells. Cell 29:99-108. doi: 0092-8674(82)90094-0

9. Lee F, Mulligan R, Berg P, Ringold G (1981) Glucocorticoids regulate expression of dihydrofolate reductase cDNA in mouse mammary tumour virus chimaeric plasmids. Nature 294:228-232

10. Lee SW, Tsou AP, Chan H, Thomas J, Petrie K, Eugui EM, Allison AC (1988) Glucocorticoids selectively inhibit the transcription of the interleukin 1 beta gene and decrease the stability of interleukin 1 beta mRNA. Proc Natl Acad Sci U S A 85:1204-1208

11. Gossen M, Bujard H (1992) Tight control of gene expression in mammalian cells by tetracycline-responsive promoters. Proc Natl Acad Sci U S A. 89(12):5547-5551

12. Tovar K, Hillen W (1989) Tet repressor binding induced curvature of tet operator DNA. Nucleic Acids Res 17:6515-6522

13. Gossen M, Freundlieb S, Bender G, Müller G, Hillen W, Bujard H (1995) Transcriptional activation by tetracyclines in mammalian cells. Science 268(5218):1766-1769

14. Hecht B, Muller G, Hillen W (1993) Noninducible Tet repressor mutations map from the operator binding motif to the C terminus. J Bacteriol 175:1206-1210

15. Hackl, M, Jakobi, T, Blom, J., Doppmeier, D, Brinkrolf, K, Szczepanowski, R, Bernhart, S.H, Honer Zu Siederdissen, C, Bort, J.A., Wieser, M, Kunert, R, Jeffs, S, Hofacker, IL, Goesmann, A, Puhler, A, Borth, N. & Grillari, J (2011)Next-generation sequencing of the Chinese hamster ovary microRNA transcriptome: Identification, annotation and profiling of microRNAs as targets for cellular engineering. Journal of Biotechnology 153(1-2): 62-75.

16. Zabala M, Wang L, Hernandez-Alcoceba R, Hillen W, Qian C, Prieto J, Kramer MG (2004) Optimization of the Tet-on system to regulate interleukin 12 expression in the liver for the treatment of hepatic tumors. Cancer Res 64:2799-2804

17. McGee Sanftner LH, Rendahl KG, Quiroz D, Coyne M, Ladner M, Manning WC, Flannery JG (2001) Recombinant AAV-mediated delivery of a tet-inducible reporter gene to the rat retina. Mol Ther 3:688-696. doi: 10.1006/mthe.2001.0308 [doi].

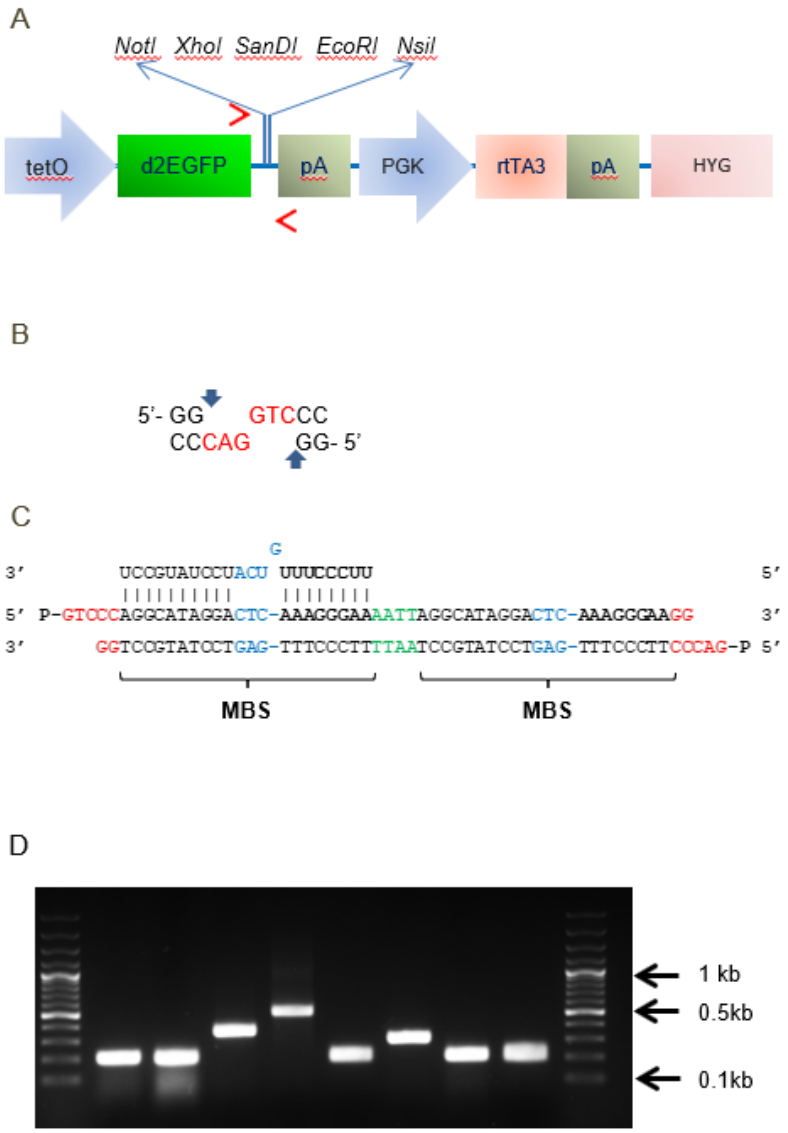


Figure 1. Cloning oligos into TET-ON-*SanDI*-sponge vector **(A)** Illustrates the Tetracycline inducible sponge vector with a Tetracycline inducible promoter (tetO) transcriptionally regulates expression of an unstable green fluorescent protein (d2GFP) reporter gene. The transcriptional factor rtTA3 presents in the same vector and constitutively expressed. An antibiotic marker (HYG) allows for the selection of stable CHO cell lines. Primers for screening of sponge insert are red arrow heads **(B)** *SanDI* site with its overhang bases are in red color. **(C)** An example of a sponge oligo duplex design with two binding sites (capital letters) for CHO miR-204 (cgr-miR-204). Overhang bases in the duplex for cloning into vector are in red color, and the spacer sequences between each binding site are in green. The blue turquoise colour letters indicate mismatch site. **(D)** PCR screening for sponge inserts of different size. Amplicons from different transformant *E. coli* clones (lanes 2 to 8), 100bp DNA ladder (lanes 1 and 9).

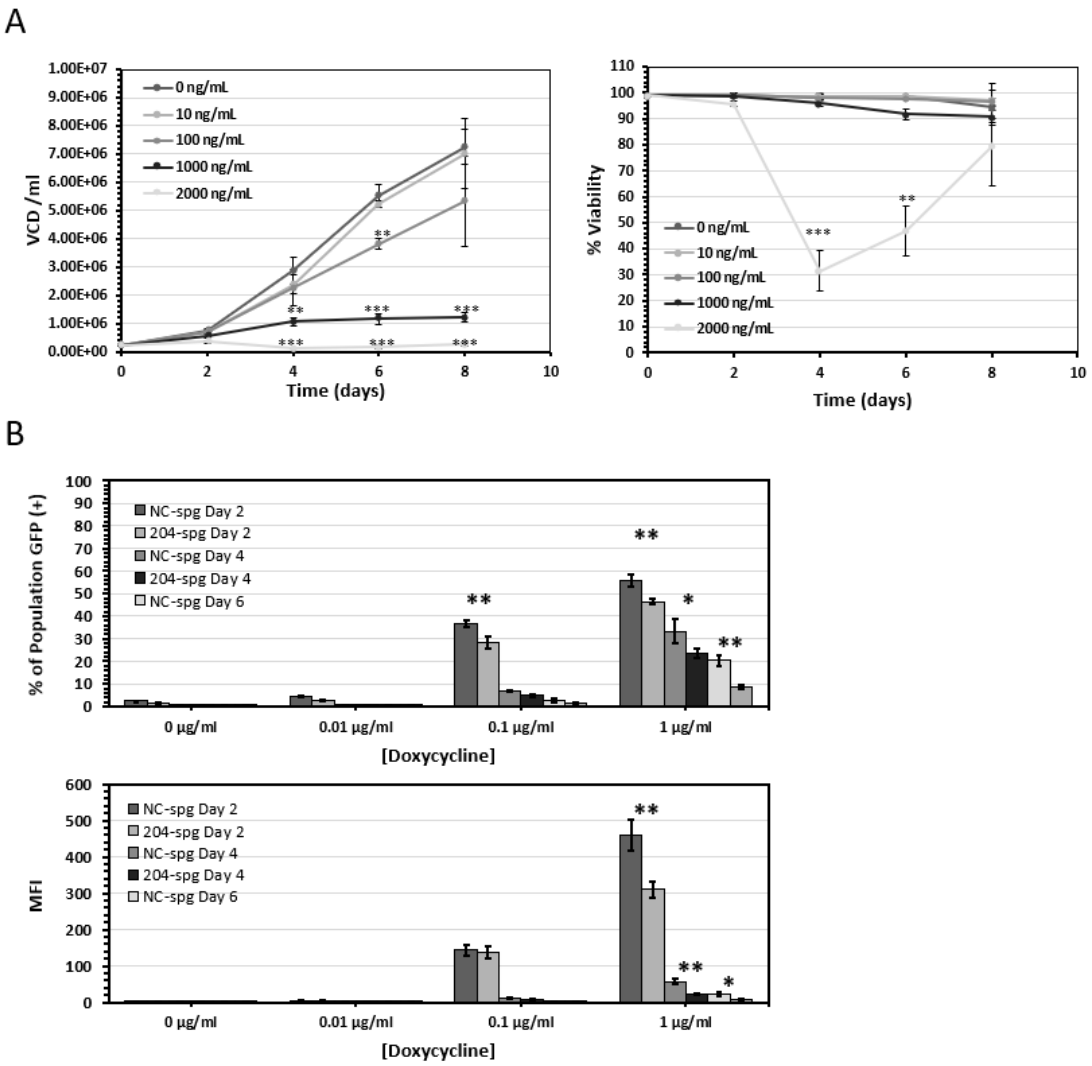


Figure 2. Testing of stable CHO cell lines expressing inducible miR-sponge. **(A)** Induction of stable CHO cells expressing a negative control (NC) sponge with different concentrations of DOX, a derivative of tetracycline **(B)** Characterisation of GFP shift in CHO cell lines expressing a NC sponge and a miRNA sponge. Percentage (%) of total population GFP positive(left), and mean fluorescence intensity (MFI) (right).

Chapter 4

Leaky Expression of the TET-On System Hinders Control of Endogenous miRNA Abundance

---

Costello, A., Lao, N.T., Gallagher, C., Capella Roca, B., Julius, L.A., Suda, S., Ducrée, J., King, D., Wagner, R., Barron, N. and Clynes, M.
Biotechnology Journal, DOI: 10.1002/biot.201800219, PMID: 29989353
Impact Factor (2018): 3.507

I was the primary contributor to all aspects of experimental work and manuscript preparation described in Chapter 4. Dr Nga T. Lao was responsible for constructing the miRNA-sponge vector backbone. Dr Clair Gallagher performed the cell sorting for the study. Mrs. Berta Capella Roca aided in the generation of the inducible stable mixed pools. Dr Albina L. Julius, Dr Damien King and Prof. Jens Ducrée were responsible for the single cell microfluidics study, I provided the cells. Dr Srinivas Suda provided the stable EPO producing clonal cell line and Prof. Roland Wagner kindly provided the industrial clone used. Prof. Martin Clynes and Prof. Niall Barron contributed to experimental design and redaction of the manuscript.

ABSTRACT

With the ability to affect multiple genes and fundamental pathways simultaneously, miRNA engineering of Chinese Hamster Ovary (CHO) cells has significant advantages over single gene expression or repression. Tight control of these molecular triggers is desirable as it could in theory allow on/ off or even tunable regulation of desirable cellular phenotypes. The present study investigated the potential of employing a tetracycline inducible (TET-On) system for conditional knockdown of specific miRNAs but encountered several challenges. We show a significant reduction in cell proliferation and culture viability when maintained in media supplemented with the TET-On induction agent Doxycycline at concentrations commonly reported. Calculation of a mature miRNA and miRNA sponge mRNA copy number demonstrated that leaky basal transgene expression in the un-induced state, is sufficient for significant miRNA knockdown. This work highlights challenges of the TET-On inducible expression system for controlled manipulation of endogenous miRNAs with two examples; miR-378 and miR-455. We suggest a solution involving isolation of highly inducible clones and use a single cell analysis platform to demonstrate the heterogeneity of basal expression and inducibility. Finally, we describe numerous strategies to minimize leaky transgene expression and alterations to current miRNA sponge design.

Key words: Chinese hamster ovary cells, TET-On, miRNA, Biotechnology

Abbreviations: CHO, Chinese hamster ovary, **miRNA**, microRNA, **miR**, microRNA, **multi-miRNA**, multiple microRNA, **TET-ON**, Tetracycline On inducible expression system, **tetO**, tetracycline operon, **Dox**, Doxycycline Hychlate, **rtTA3**, reverse tetracycline controlled trans activator, **FACS**, fluorescence activated cell sorting, **UTR**, untranslated region, **d2eGFP**, destabilized green fluorescent protein, **MRE**, microRNA responsive element, **VCD**, viable cell density, **MFI**, mean fluorescence intensity, **AU**, autofluorescence units **RQ**, relative quantification, **AQ**, absolute quantification, **PCC**, Pearson correlation coefficient, **t**, time, **min**, minutes, **hr**, hours

4.1 INTRODUCTION

The endogenous expression of microRNAs (miRNAs) is dynamic in nature, with diverse profiles between cell types and even stages of industrial cell culture [1]. Process control is paramount in the production of homogenous high-quality protein from mammalian cell factories and has been an area of major advancements in the biopharmaceutical industry. These developments have largely come in the form of culture media and feed formulation, along with on and off-line monitoring systems. Genetic engineering of CHO cells has seen increased focus since publication of the first CHO-K1 genome [2,3] including several reports of miRNA manipulation leading to desirable cell phenotypes [4-16] In all cases, the miRNA up or down regulation was through constitutive intervention. The next step logically is to engineer multiple miRNA with beneficial effects in the same cell. To avoid redundancy in multi-miRNA engineering, conditional regulation of each miRNA should ideally be achieved with timely induction. A multi-miRNA engineering strategy could be used to enhance growth, productivity and or anti-apoptotic behaviours at defined stages of culture.

The tetracycline inducible system [17] is the most widely used inducible gene expression system [18] and is broadly implemented to control gene expression in eukaryotic cells [19]. This system is based on the Tn10-specified tetracycline-resistance operon of *E. coli* [20] wherein the transcription of resistance genes is negatively regulated by the tetracycline repressor (tetR). In the presence of tetracycline, tetR is unable to bind to the promoter region of the operon, allowing transcription of the resistance genes. By attaching the tetR to the C-terminal of VP16 from herpes simplex virus (HSV), a hybrid trans-activator, the tetracycline-controlled trans-activator (tTA), was created that stimulates minimal promoters fused with a tetracycline operator (tetO). There are two types of tetracycline-controlled gene expression, TET-Off and TET-On. The TET-Off [21] system uses a tTA that binds to a tetO in the absence of tetracycline or its homologue doxycycline (Dox), to promote transcription. The addition of tetracycline or Dox to the system decouples the tTA from the tetO and transcription is silenced. The TET-On system [22] was created by random mutagenesis of the tetR gene, identifying the sequence responsible for tetO binding. The modified tTA was named reverse tTA or (rtTA) as it works in the opposite manner to tTA. The presence of Dox enables rtTA to bind to the tetO and up-regulate transcription in a dose dependant manner.

miRNAs are small, ~22 nucleotide non-coding RNAs that function in post-transcriptional regulation through pairing with partial complementarity to the 3' untranslated regions (UTRs) of protein coding genes [23]. The benefit of targeting miRNAs in genetic engineering of mammalian cells is two-fold; unlike gene expression and repression strategies, manipulation of a single miRNA can affect clusters of protein coding genes simultaneously, without increased translational burden on the cell. Stable constitutive suppression of specific endogenous miRNA

function in CHO cells has been reported previously [7, 11, 13]. This is achieved by introduction of synthetic decoy target sequences, referred to as miRNA sponges [24]. miRNA sponges act by "soaking up" specific miRNAs of interest, freeing their endogenous targets from translational repression. The levels of endogenous miRNAs are tightly regulated in response to the cell's needs, with diverse expression among cell types, stage and species [25]. A profiling study [26] of sister clones from the same original transfected pool with similar specific productivity but varying growth rates, revealed a reduction in miR-378 and miR-455 expression to correlate with higher growth rates in CHO cells. There will always be a trade-off between cellular metabolism being driven towards either protein expression or biomass accumulation [27]. Many constitutive cell engineering strategies aimed at improving protein production result in reduced cellular growth rate and vice-versa. To negate this, we aimed to investigate the potential of conditional regulation of endogenous miRNA species at defined stages of culture.

4.2 MATERIALS AND METHODS

4.2.1 Cell Culture

Three CHO producer cell lines were used in this study; DP12 (ATCC), SK15-EPO (in-house generated clone) and C#112 (Rentschler Biopharma). DP12 and C#112 are monoclonal antibody (mAb) producers and SK15-EPO produces Erythropoietin (EPO). CHO DP12 and SK15-EPO were cultured in chemically defined Balan CD (Irvine Scientific) supplemented with 2% (v/v) polyvinyl alcohol (PVA) and 4mM L-glutamine (25030081-Gibco). C#112 was cultured in ProCHO5 (Lonza), supplemented with 4mM L-glutamine (25030081-Gibco) and 1:200 anti-clumping agent (0010057AE-Gibco). Sub-culturing in a routine 3-4-3-day manner with seeding at $2x10^5$ cells/ mL in 50mL spin tube (87050T-Helena-BioSciences), 5mL working volume. Suspension cultures were performed in an orbital shaker (Kühner), 37°C, 80% humidity and 5% CO_2. Viable cell density was monitored using the ViaCount™ on a Guava® easyCyte benchtop cytometer (Merck Millipore, UK).

4.2.2 Vector Construction and Generation of Stable Mixed Pools

Construction of the TET-On sponge vectors and stable mixed generation was previously described elsewhere [28]. The miRNA sponge sequences used are as follows; miR-378-sense (5'–GTCCCGCCTTCTGACGTTAGTCCGATAATTGCCTTCTGACGTTAGTCCGATGG–3'), miR-378-antisense (5'–GACCCATCGGACTAACGTCAGAAGGCAATTATCGGACTAACGTCAGAAGGCGG–3'), miR-455-sense (5'–GTCCCAGTGTATATGAAATGGACTGCAATTAGTGTATATGAAATGGACTGCGG–3'),

miR-455-antisense (5'–GACCCGCAGTCCATTTCATATACACTAATTGCAGTCCATTTCAT ATACACTGG–3'), NC-spg-sense (5'- GTCCCAAGTTTTCAGAAAGCTAACACCGGAAGTTTTCAGAA AGCTAACAGG – 3') and NC-spg-antisense (5'–GACCCTGTTAGCTTTCTGAAAACTTCCGGTGTTAGC TTTCTGAAAACTTGG–3'). The miRNA sponge sequences were cloned downstream of a destabilized GFP (d2eGFP) [29] using SanDI restriction sites as previously described [30]. This method allows concatemeric integration of multiple MREs in a single step, thus enabling the generation of miRNA sponges of varying length. Enrichment of Dox sensitive subpopulations was done by sorting GFP positive cells post incubation with 10ng/ mL Dox for 24hrs. Cell sorting was performed on a FACS Aria (Becton Dickinson), gating for dark cells with the original DP12 cell line, and collection of $5x10^4$ cells for each subpopulation. Cells were collected in 6 well culture plates, 2mL per well. Collection media; Balan CD, 4mM L-glutamine, 400µg/ mL Hygromycin B (Roche), 200mM MTX (Sigma-Aldrich) and 1X Pen Strep (Gibco), supplemented with 20% pre-conditioned media.

4.2.3 Relative and Absolute Quantitative RT-PCR

Cells were harvested by centrifugation at 1000rpm for 5mins. Total RNA was isolated from 1-$5x10^6$ cells using Trizol Reagent (Ambion), following the manufacture's protocol. RNA quantification and quality were measured by NanoDrop (Thermo Fischer Scientific). Relative Quantification (RQ) of mature miRNA was done using the TaqMan® miRNA Assay system (Applied Biosystems), and expression calculated using the $2^{(-\Delta\Delta Ct)}$ method. Reverse transcription (RT) of specific mature miRNA or endogenous control U6 snRNA was done with specific TaqMan® miRNA Assays (Applied Biosystems) for 20ng total RNA, with no deviations in timing from manufacture's protocol. qRT-PCR was run for all miRNA analysis on an (Applied Biosystems) AB7500 using manufacturers specifications. All reactions were run in technical triplicate. The TaqMan® threshold cycle (Ct) values were converted into absolute copy number using a specific standard curve from synthetic miRNA of interest, this method was adopted from [31]. For gene expression, reverse transcription of total RNA was performed with the High Capacity cDNA Reverse Transcription Kit (Applied Biosystems), for 2000ng of total RNA, with no deviation in timing from the manufacturer's protocol. qRT-PCR was run using an AB7500 (Applied Biosystems) with Fast SYBR Green Master Mix (Applied Biosystems). 2X SYBR was combined with 20ng of cDNA, 200nM Forward and Reverse primers, and water made up to 20µL final reaction volume. Each biological replicate sample was run in technical triplicate wells. The Ct values were converted to absolute copy number using a standard curve of the d2eGFP amplicon.

4.2.4 Single Cell Microfluidic Analysis

Lab on a chip fabrication, device-priming and cell loading was performed as previously described [32]. Trenches were imaged with an Olympus IX81 motorized inverted microscope coupled to a Hamamatsu ORCA – RE digital camera. TIFF images were analyzed in ImageJ (version 1.46r). Fluorescence intensity was quantified by "region of interest" analysis.

4.2.5 Statistical Analysis

The values reported represent the mean of three independent experiments with error displayed as standard deviation. Statistical analysis was carried out by a two-tailed homoscedastic student t-test, ($p \leq 0.05$ *, $p \leq 0.01$**, $p \leq 0.001$ ***) of three independent experimental replicates.

4.3 RESULTS

4.3.1 TET-On inducible miRNA sponge vectors

Sponge decoy sequences were designed to bind endogenous miRNA targets at both 5' and 3' ends with a central bulge site (Figure 1A). Each sponge consisted of 10 miRNA responsive elements (MRE) and were inserted downstream of a destabilized GFP (d2eGFP) reporter gene under the control of a TET-On promoter (Figure 1B). The d2eGFP was employed to facilitate monitoring of induction through fluorescence. In the un-induced state, (Figure 1B), the rtTA, in this case rtTA3 [33] cannot bind to the tetO 5' of the promoter (pTET) and sponge expression is low. This allows for target miRNA to function naturally. Upon induction with Dox addition, (Figure 1C), miRNA sponge expression is up-regulated. This creates an abundance of decoy target sites to sequester the miRNA of interest, thus inhibiting endogenous function.

4.3.2 Evaluation of TET-ON induction variables

To assess the phenotypic impact of specific miRNAs at defined culture stages, it was imperative to first formulate a robust induction strategy with a control vector not targeting any miRNA. CHO DP12 cells were stably transfected with a TET-On vector containing a Negative Control sponge (NC-spg) [7]. Stable mixed pools of NC-spg cells were exposed to 0, 100 or 1000ng/ mL of Dox to test induction. In conjunction with the titration experiment, Dox-feeding was evaluated by exposing cultures to 100 or 1000ng/ mL repeatedly at; 0, 48 and 96hrs of culture. Cells exposed to Dox were monitored for induction at 24hr intervals. The percentage of total cells GFP positive with respect to the cells only, no Dox, control (Figure 2A) and corresponding mean fluorescence intensity (MFI) (Figure 2B) was evaluated over a 144hr time-period. There was a titration response to the induction agent with 1000ng /mL of Dox inducing 66 % of cells compared to 46% induced with 100ng/ mL, (Figure 2A). At 144hrs after a single Dox addition,

cells dosed with 100ng/ mL showed 5% GFP positive cells and those dosed with 1000ng/ mL 33% GFP positive (Figure 2A). Feeding of 1000ng/ mL Dox at 48hr intervals sustained induction of 50% of the total cells over the 144hr period of analysis. The NC-spg cells were then assessed for growth with induction on days 2 (100ng/ mL) and 4 (1000ng/ mL) (Figure 2C). A significant difference in the maximum viable cell density [VCD/ mL] reached by the NC-spg cell line was observed between the un-induced, "NC-spg (-Dox)", and induced, "NC-spg (+Dox)", cells. The NC-spg cell line reached a peak cell density of $8.09x10^6$ cells/ mL on day 8 of culture. The addition of Dox reduced this to $6.14x10^6$ cells/mL, ($p=0.016$), a ~24% decrease in maximum VCD. The significant reduction in cell growth for the NC-spg cells under induced conditions was a major concern. Two factors were identified as potential causes of reduced cell growth upon induction: toxicity of the induction agent Dox; and increased GFP expression. To examine Dox toxicity DP12 cells were cultured in media supplemented with Dox at the concentrations; 0, 10, 100, 1000 and 2000ng/ mL 48hrs post inoculation (Figure 2D). The presence of 100, 1000 and 2000ng/ mL Dox significantly reduced the VCD of DP12 cells by 40, 78 and 85% respectively.

4.3.3 Correlating absolute miRNA abundance with leaky miRNA-sponge expression

Previous profiling studies in our lab [26] identified genes, proteins and miRNAs whose expression was robustly associated with cell specific growth rate. The study used sister clones from the same original transfected pool characterized for growth behaviour over 40 passages. Differential expression analysis of 667 miRNAs revealed miR-378 and miR-455 to have low abundance in "fast" growing clones. To investigate the depletion of miRNA miR-378 and miR-455 as means of enhancing cell growth, we implemented a miRNA-sponge mediated knockdown strategy. Cell growth of miR-378 and miR-455 sponge expressing cells was compared to NC-spg cells in the absence (-Dox) or presence of 1000ng/ mL Dox (+Dox). There was a significant increase in cell growth with and without induction of the miR-378-spg mixed pool over the NC-spg cells (Figure 3A). Relative quantification (RQ) of mature miR-378 in the NC-spg and miR-378 mixed pools revealed significant knockdown of miR-378 in the absence of Dox (-Dox). Approximately (77%) knockdown was observed due to leaky expression alone, with no increased knockdown of miR-378 following induction (Figure 3B). A similar trend was witnessed in the stable miR-455-spg mixed pool. The un-induced miR-455-spg cells showed a significant increase in VCD over the NC-spg under the same conditions (Figure 3C) however there was no significant increase in VCD with induction. Again, there was significant knockdown (67%) of miR-455 without induction (Figure 3D). The leaky expression of inducible promoters is well documented and can vary greatly between cell types [18]. Leaky sponge expression could explain why considerable target miRNA knockdown was observed in the un-induced cells.

In the un-induced state, DP12 cells expressing the NC-spg show ~10% GFP positive cells at low intensity, this is indicative of low-level constitutive reporter-gene expression. To identify the potential limitations of leaky miRNA-sponge expression on our candidate miRNAs, absolute quantification (AQ) of mature miRNA and d2eGFP was performed. The AQ of d2eGFP mRNA in the original NC-spg stable mixed pool was 496 copies per cell, un-induced (-Dox) and 76,000 copies per cell, induced (+Dox) (Figure 3E). This equates to 153-fold increase in mRNA expression upon induction. As each mRNA transcript contains 10 miRNA responsive elements (MRE), there are 4,964 MREs per cell, without induction (Figure 3E). AQ of miR-378 was assessed for three CHO producer lines; DP12, SK15-EPO and C#112. The AQ of miR-378 was 12,452, 10,856 and 2,451 copies per cell for; DP12, SK15-EPO and C#112 respectively (Figure 3F). With target miRNA abundance ranging from ~2 – 12k copies per cell and basal sponge expression producing ~5k MREs per cell, there would be considerable miRNA knockdown without any induction.

4.3.4 Generating cells with high induction capacity

It has been reported that translational burden due to high transgene expression can reduce cell growth [15]. To evaluate potential translational burden on the cells due to the reporter gene d2eGFP, NC-spg cells were induced with 10ng/ mL of Dox for 24hrs. The induced cells were sorted into subpopulations based on medium (MED) or high (HIGH) MFI. Both the MED and HIGH populations achieved >90% GFP positive cells with 10ng/ mL Dox (Figure 4A) and had corresponding MFI of 513 and 2163 respectively, (Figure 4B). The un-induced cells of MED and HIGH populations showed leaky GFP expression of 11% and 22% respectively. There was no significant difference in growth between the MED population and the original DP12 cell line (Figure 4C). The HIGH cells grew significantly worse than the original DP12 cells both un-induced ($p=0.01298$) and induced ($p=0.00065$). There was a significant difference between the HIGH cells induced and uninduced ($p=0.01$). Culture viability was also reduced by 35%, for the HIGH population upon induction (Figure 4D). Through FACS enrichment of cells induced at non-toxic concentrations of Dox, it was possible to isolate subpopulations with predictable induction characteristics and determine an upper threshold of reporter gene expression.

4.3.5 Single cell analysis of clonal heterogeneity in basal expression and induction capacity

Leaky sponge expression was sufficient for significant knockdown of target miRNAs in stable mixed pools (Figure 5). As leaky expression was increased with final induction capacity (Figure 4A), this suggested a relationship between basal expression and induction capacity. These mixed pools were generated by random integration of the TET-On vector, creating a heterogenous

population. To investigate clonal heterogeneity as it relates to basal reporter gene expression and induction capacity, a novel single cell analysis platform was employed [33]. Limited dilution was used to load single "MED" cells into microfluidic trenches (Figure 5A). The intensity of d2eGFP expression was measured in auto-fluorescence units [AU] at time zero to identify basal expression for each clone. Unlike hormone inducible systems that respond in minutes, the TET-On system has a reported induction lag of ~4hrs [19]. To comprehensively test our TET-On system, media containing 10ng/ mL Dox was perfused into the trenches and AU monitored every 15 minutes for a 4hr time-period. The analysis of 80 single cells was grouped based on initial AU levels, for cells with AU <50 (Figure 5B), 50 – 100 (Figure 5C) and > 100 (Figure 5D). The heterogeneity of Dox sensitivity is evident with initial basal expression of d2eGFP being an order of magnitude in difference 24 – 303 AU. The fold increase of AU with respect to baseline is represented for cells with initial AU <50 (Figure 5E), 50 – 100 (Figure 5F) and > 100 (Figure 5G). Of the 80 clones tested, 73% induced >1.2 fold in the 4hr time-period. There was no correlation between the basal expression of d2eGFP and induction capacity (Figure 5H). Nor was there any correlation between basal expression and time to induce (Figure 5I). A Pearson correlation coefficient of (PCC = -0.855) indicated a strong correlation between time to first induction and final fold increase over the 4hr period (Figure 5J).

4.4 DISCUSSION

The data presented here highlights considerations and limitations of TET-On inducible express systems as it pertains to the conditional downregulation of specific miRNA function in CHO cells. The concentration of Dox required to induce >50% of cells had a significant effect on cell growth (Figure 2C). Addition of Dox to culture media at concentrations exceeding 100ng/ mL significantly reduced cell growth and viability (Figure 2D). Cytotoxic effects of the induction agent Dox on mammalian cells has been reported previously [34]. FACS enrichment of cells inducible with non-toxic levels of Dox was used to evaluate the effects of high reporter gene expression on cell growth and viability (Figure 4). It was possible to derive subpopulations of NC-spg cells with predictable distinct induction characteristics. Two subpopulations were derived based on medium or high reporter gene expression. Induction of the MED population had no significant effects on cell growth or viability, while high level reporter gene expression resulted in a significant reduction of both (Figure 4C-D). One key difference noticed between the MED and HIGH cells is the levels of "leaky" reporter gene expression is higher in the HIGH cells. This is likely due to high transgene copy number or a positional effect due to the random integration. To explore a relationship between basal expression and inducibility, a single cell analysis platform was utilized. Monitoring 80 single cells from the MED population over a 4hr period revealed no correlation between basal expression and the final fold increase in reporter

gene expression (Figure 5H). There was a strong link between the time taken to induce and the final fold induction (Figure 5J). To eliminate error from this analysis, the 20% of cells that showed no induction, <1.2 – fold increase, (Figure 5B-G) were removed. This link between time to induce and final induction capacity could serve as screening criteria in the identification of highly inducible clones.

Transient phenotypic screening of miRNA function has been reported previously [8, 15]. This was achieved using miRNA Mimics or AntagomiRs to evaluate the effects of miRNA over-expression or depletion respectively, providing a fast, high throughput means of identifying miRNA engineering candidates. For the reasons already discussed inducible expression systems do not lend themselves to use as miRNA screening tools. Yet a plethora of miRNA with known biological function in CHO cells already exists, with functions ranging from; cell growth [6, 8] cell cycle [4, 7] metabolism [11] productivity [7 - 9, 11 - 15] and inhibition of apoptosis [4, 16]. Stable manipulation in all above cases resulted in one specific phenotype, yet the inherent tradeoff in cell metabolism driving antagonistic desirable behaviors, growth and protein production, still exists as a major bottleneck. We generated stable mixed pools for the controlled downregulation of endogenous miR-378 and miR-455 in DP12 cells. The specific miRNA sponge decoy in each case was driven by a TET-On promoter. In both cases there was a significant phenotypic difference between the miRNA targeting and negative control sponges in the un-induced state (Figure 3A, C). Evaluation of mature target miRNA levels revealed there was significant knockdown (70%) without induction (Figure 3B, D). Through absolute quantification of sponge mRNA and miR-378 levels we found that leaky sponge expression generated enough MREs for ~50% knockdown. The knockdown seen in the stable mixed pools for miR-378 and miR-455 was greater than 50%. As depletion of free miRNA abundance in cases improved cell growth, it could be that cells exhibiting higher levels of leaky sponge expression had a competitive advantage. Previously reported AQ of 5 miRNAs across 7 tissues [31] report moderate miRNA abundance ~2000 copies per cell, with high >30,000 and low abundance of ~500 copies per cell. Further evaluation of absolute miR-378 levels in two other producer lines; SK15-EPO and C#112 (Figure 3F) shows the cell line specificity of these molecules as described by [1]. The levels of miR-378 are similar in the two CHO-K1 derived producers; DP12 and SK15, whereas C#112 of CHO DG44 lineage had significantly lower levels. We suggest that screening the absolute levels of miRNAs of interest should be done in conjunction with differential expression analysis. This could possibly aid in predicting which cell lines will respond to target miRNA manipulations.

We propose several possible strategies to resolve the issue of leaky knockdown of specific miRNAs. First, by reducing the efficacy of synthetic sponge transcripts. By reducing the MREs from 10 – 1 the number decoy target sites generated by leaky expression alone would reduce by

10-fold. The miRNA sponge sequences used in this study were designed to include a bulge site. The sponges are designed to bind both 5' and 3' ends of specific miRNAs, this differs somewhat from canonical eukaryotic miRNA seed only binding. By mimicking canonical eukaryotic miRNA binding and changing the sponge design to seed only binding, this may reduce efficacy of target miRNA depletion but could increase the events of non-targeted seed family binding [35, 36].

Leaky expression from TET promoters is well documented, provoking active repression strategies. The Kid-1 protein and approximately one-third of all other zinc finger proteins contain a highly conserved 75 amino acid domain at their NH2 terminus named Krüppel-associated box (KRAB) [37]. Tetracycline-Reversible silencing of eukaryotic promoters was investigated by fusion of the tTA from the TET-Off system [17] and KRAB [38] to produce TetR-KRAB [39]. TetR localizes the KRAB silencing domain to the tetO upstream of a promoter driving a gene of interest. Where low basal expression is paramount such as animal models, the TetR-KRAB system has been employed [40, 41]. Yet even with the presence of an acting transcriptional repressor domain, leaky expression persists, with the TetR-KRAB system being described as "less leaky" than the TET-On [42]. An alternative means to reduce leaky sponge abundance is active repression of the sponge mRNA itself. This could be achieved by RNAi mediated degradation of the sponge transcript. Active degradation of leaky MRE abundance could be forced by co-expression of a constitutively expressed shRNA targeting either the reporter gene or sponge sequence. In this context, induction of the sponge would need to meet a threshold by which sponge expression exceeds that of the targeting shRNA before target miRNA knockdown can occur.

4.5 CONCLUSION

In summation, we aimed to evaluate phenotypic impact of specific miRNA depletion at defined culture stages, encountering numerous limitations of the TET-On inducible system for this application. In stable mixed pools, the concentration of induction agent Doxycycline required for reporter gene expression had adverse effects on cell growth and viability. A cell sorting strategy was devised to isolate subpopulations of cells inducible with a non-toxic dose of Dox. These subpopulations showed great induction capacity with minimal Dox addition. As anticipated, high levels of reporter-gene expression also had significant detrimental growth effects, with induction of these cells leading to significantly reduced growth and compromised viability. Through the investigation of detrimental side effects of high transgene expression, we serendipitously devised a method to generate highly inducible and predictable gene expression in CHO cells, with no adverse Dox associated effects. We correlated the absolute levels of a specific miRNA, and MREs with mere leaky expression. Showing significant stable knockdown of target miRNAs

without need for induction, we recognize that the transcript levels from basal leaky expression are negligible pertaining to controlled gene expression. However, we suggest that with the potency of their function, and volatile expression patterns, miRNAs are more sensitive than protein coding genes and controlled expression requires a more refined approach.

ACKNOWLEDGMENTS

This work was conducted under the financial support of Scientific Foundation of Ireland (SFI) grant numbers [13/IA/1963] and [13/IA/1841]. The work in figure 5 was supported by Science Foundation Ireland (SFI) and Fraunhofer-Gesellschaft under the SFI Strategic Partnership Programme [16/SPP/3321], the EDRF and the MicroPAT project funded by Enterprise Ireland [CF-2017- 0686-P]. Also, a special thank you to Prof. Roland Wagner and Rentschler Biopharma SE, Laupheim, Germany for the C#112 cell line.

CONFLICT OF INTEREST

The authors declare no financial or commercial conflict of interest.

4.6 REFERENCES

[1] J. A. H. Bort, Hackl, M., Höflmayer, H., Jadhav, V., Harreither, E., Kumar, N., Ernst, W., Grillari, J., Borth, N., Dynamic mRNA and miRNA profiling of CHO-K1 suspension cell cultures. *Biotechnol. J.* 2012, *7*, 500-515.

[2] X. Xu, Nagarajan, H., Lewis, N. E., Pan, S., Cai, Z., Liu, X., Chen, W., Xie, M., Wang, W., Hammond, S., The genomic sequence of the Chinese hamster ovary (CHO)-K1 cell line. *Nat. Biotechnol.* 2011, *29*, 735.

[3] N. E. Lewis, Liu, X., Li, Y., Nagarajan, H., Yerganian, G., O'Brien, E., Bordbar, A., Roth, A. M., Rosenbloom, J., Bian, C., Xie, M., Chen, W., Li, N., Baycin-Hizal, D., Latif, H., Forster, J., Betenbaugh, M. J., Famili, I., Xu, X., Wang, J., Palsson, B. O., Genomic landscapes of Chinese hamster ovary cell lines as revealed by the Cricetulus griseus draft genome. *Nat. Biotechnol.* 2013, *31*, 759-765.

[4] N. Sanchez, Gallagher, M., Lao, N., Gallagher, C., Clarke, C., Doolan, P., Aherne, S., Blanco, A., Meleady, P., Clynes, M., MiR-7 triggers cell cycle arrest at the G1/S transition by targeting multiple genes including Skp2 and Psme3. *PloS One* 2013, *8*, e65671.

[5] A. Druz, Son, Y., Betenbaugh, M., Shiloach, J., Stable inhibition of mmu-miR-466h-5p improves apoptosis resistance and protein production in CHO cells. *Metab. Eng.* 2013, *16*, 87-94.

[6] M. Strotbek, Florin, L., Koenitzer, J., Tolstrup, A., Kaufmann, H., Hausser, A., Olayioye, M. A., Stable microRNA expression enhances therapeutic antibody productivity of Chinese hamster ovary cells. *Metab. Eng.* 2013, *20*, 157-166.

[7] N. Sanchez, Kelly, P., Gallagher, C., Lao, N. T., Clarke, C., Clynes, M., Barron, N., CHO cell culture longevity and recombinant protein yield are enhanced by depletion of miR-7 activity via sponge decoy vectors. *Biotechnol. J.* 2014, *9*, 396-404.

[8] S. Fischer, Buck, T., Wagner, A., Ehrhart, C., Giancaterino, J., Mang, S., Schad, M., Mathias, S., Aschrafi, A., Handrick, R., A functional high-content miRNA screen identifies miR-30 family to boost recombinant protein production in CHO cells. *Biotechnology Journal* 2014, *9*, 1279-1292.

[9] V. Jadhav, Hackl, M., Klanert, G., Bort, J. A. H., Kunert, R., Grillari, J., Borth, N., Stable overexpression of miR-17 enhances recombinant protein production of CHO cells. *J. Biotechnol.* 2014, *175*, 38-44.

[10] G. Klanert, 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.

[11] P. S. Kelly, Breen, L., Gallagher, C., Kelly, S., Henry, M., Lao, N. T., Meleady, P., O'Gorman, D., Clynes, M., Barron, N., Re-programming CHO cell metabolism using miR-23 tips the balance towards a highly productive phenotype. *Biotechnol. J.* 2015, *10*, 1029-1040.

[12] S. Fischer, Paul, A. J., Wagner, A., Mathias, S., Geiss, M., Schandock, F., Domnowski, M., Zimmermann, J., Handrick, R., Hesse, F., miR-2861 as novel HDAC5 inhibitor in CHO cells enhances productivity while maintaining product quality. *Biotechnol. Bioeng.* 2015, *112*, 2142-2153.

[13] V. V. Emmerling, Fischer, S., Stiefel, F., Holzmann, K., Handrick, R., Hesse, F., Hörer, M., Kochanek, S., Otte, K., Temperature-sensitive miR-483 is a conserved regulator of recombinant protein and viral vector production in mammalian cells. *Biotechnol. Bioeng.* 2016, *113*, 830-841.

[14] M. Schoellhorn, 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.

[15] S. Fischer, Marquart, K. F., Pieper, L. A., 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.

[16] A. Griffith, Kelly, P. S., Vencken, S., Lao, N. T., Greene, C. M., Clynes, M., Barron, N., miR-CATCH Identifies Biologically Active miRNA Regulators of the Pro-Survival Gene XIAP, in Chinese Hamster Ovary Cells. *Biotechnol. J.* 2017, vol. 13 issue 3, 1700299.

[17] M. Gossen, Bujard, H., Tight control of gene expression in mammalian cells by tetracycline-responsive promoters. *Proc. Natl. Acad. Sci. U. S. A.* 1992, *89*, 5547-5551.

[18] A. T. Das, Zhou, X., Metz, S. W., Vink, M. A., Berkhout, B., Selecting the optimal Tet-On system for doxycycline-inducible gene expression in transiently transfected and stably transduced mammalian cells. *Biotechnol. J.* 2016, *11*, 71-79.

[19] Gossen, M., Bujard, H., Tetracyclines in the control of gene expression in eukaryotes, *Tetracyclines in Biology, Chemistry and Medicine*, Springer, 2001, pp. 139-157. Birkhäuser, Basel.

[20] K. Tovar, Hillen, W., Tet repressor binding induced curvature of tet operator DNA. *Nucleic Acids Res.* 1989, *17*, 6515-6522.

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

[22] B. Hecht, Muller, G., Hillen, W., Noninducible Tet repressor mutations map from the operator binding motif to the C terminus. *J. Bacteriol.* 1993, *175*, 1206-1210.

[23] D. P. Bartel, MicroRNAs: target recognition and regulatory functions. *Cell* 2009, *136*, 215-233.

[24] M. S. Ebert, Neilson, J. R., Sharp, P. A., MicroRNA sponges: competitive inhibitors of small RNAs in mammalian cells. *Nat. Methods* 2007, *4*, 721-726.

[25] P., Landgraf, M., Rusu, R., Sheridan, A., Sewer, N., Iovino, A., Aravin, S., Pfeffer, A., Rice, A.O. & M. Landthaler, 2007, "A mammalian microRNA expression atlas based on small RNA library sequencing", *Cell*, vol. 129, no. 7, pp. 1401-1414.

[26] C. Clarke, Henry, M., Doolan, P., Kelly, S., Aherne, S., Sanchez, N., Kelly, P., Kinsella, P., Breen, L., Madden, S. F., Integrated miRNA, mRNA and protein expression analysis reveals the role of post-transcriptional regulation in controlling CHO cell growth rate. *BMC Genomics* 2012, *13*, 656.

[27] H., Hefzi, K.S., Ang, M., Hanscho, A., Bordbar, D., Ruckerbauer, M., Lakshmanan, C.A., Orellana, D., Baycin-Hizal, Y. Huang, & D. 2016, Ley, "A consensus genome-scale reconstruction of Chinese hamster ovary cell metabolism", *Cell Systems*, vol. 3, no. 5, pp. 434-443. e8.

[28] A. Costello, Lao, N., Clynes, M., Barron, N., Conditional knockdown of endogenous microRNAs in CHO Cells using TET-ON-Sandi sponge vectors. *Heterologous Protein Production in CHO Cells: Methods and Protocols, Methods in Molecular Biology, vol. 1603*, 2017, 87-100.

[29] X. Li, Zhao, X., Fang, Y., Jiang, X., Duong, T., Fan, C., Huang, C. C., Kain, S. R., Generation of destabilized green fluorescent protein as a transcription reporter. *J. Biol. Chem.* 1998, *273*, 34970-34975.

[30] Kluiver, J., Gibcus, J.H., Hettinga, C., Adema, A., Richter, M.K., Halsema, N., Slezak-Prochazka, I., Ding, Y., Kroesen, B.J. & van den Berg, A. 2012, "Rapid generation of microRNA sponges for microRNA inhibition", *PloS One*, vol. 7, no. 1, pp. e29275.

[31] C. Chen, Ridzon, D. A., Broomer, A. J., Zhou, Z., Lee, D. H., Nguyen, J. T., Barbisin, M., Xu, N. L., Mahuvakar, V. R., Andersen, M. R., Real-time quantification of microRNAs by stem–loop RT–PCR. *Nucleic Acids Res.* 2005, *33*, e179-e179.

[32] T. M. O'Connell, King, D., Dixit, C. K., O'Connor, B., Walls, D., Ducrée, J., Sequential glycan profiling at single cell level with the microfluidic lab-in-a-trench platform: a new era in experimental cell biology. *Lab on a Chip* 2014, *14*, 3629-3639.

[33] A. T. Das, Zhou, X., Vink, M., Klaver, B., Verhoef, K., Marzio, G., Berkhout, B., Viral evolution as a tool to improve the tetracycline-regulated gene expression system. *J. Biol. Chem.* 2004, *279*, 18776-18782.

[34] G. Ermak, V. J. Cancasci, K. J. A. Davies, Cytotoxic effect of doxycycline and its implications for tet-on gene expression systems. *Analytical Biochem.* 2003, *318*, 152-154.

[35] B. P. Lewis, Burge, C. B., Bartel, D. P., Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* 2005, *120*, 15-20.

[36] M. S. Ebert, Sharp, P. A., MicroRNA sponges: progress and possibilities. *RNA* 2010, *16*, 2043-2050.

[37] H. J. Thiesen, Bellefroid, E., Revelant, O., Martial, J. A., Conserved KRAB protein domain identified upstream from the zinc finger region of Kox 8. *Nucleic Acids Res.* 1991, *19*, 3996.

[38] R. Witzgall, O'Leary, E., Leaf, A., Onaldi, D., Bonventre, J. V., The Krüppel-associated box-A (KRAB-A) domain of zinc finger proteins mediates transcriptional repression. *Proc. Natl. Acad. Sci. U. S. A.* 1994, *91*, 4514-4518.

[39] U. Deuschle, Meyer, W., Thiesen, H., Tetracycline-reversible silencing of eukaryotic promoters. *Mol. Cell. Biol.* 1995, *15*, 1907-1914.

[40] L. H. McGee Sanftner, Rendahl, K. G., Quiroz, D., Coyne, M., Ladner, M., Manning, W. C., Flannery, J. G., Recombinant AAV-mediated delivery of a tet-inducible reporter gene to the rat retina. *Mol. Ther.* 2001, *3*, 688-696.

[41] J. Szulc, Wiznerowicz, M., Sauvain, M. O., Trono, D., Aebischer, P., A versatile tool for conditional gene expression and knockdown. *Nat. Methods* 2006, *3*, 109-116.

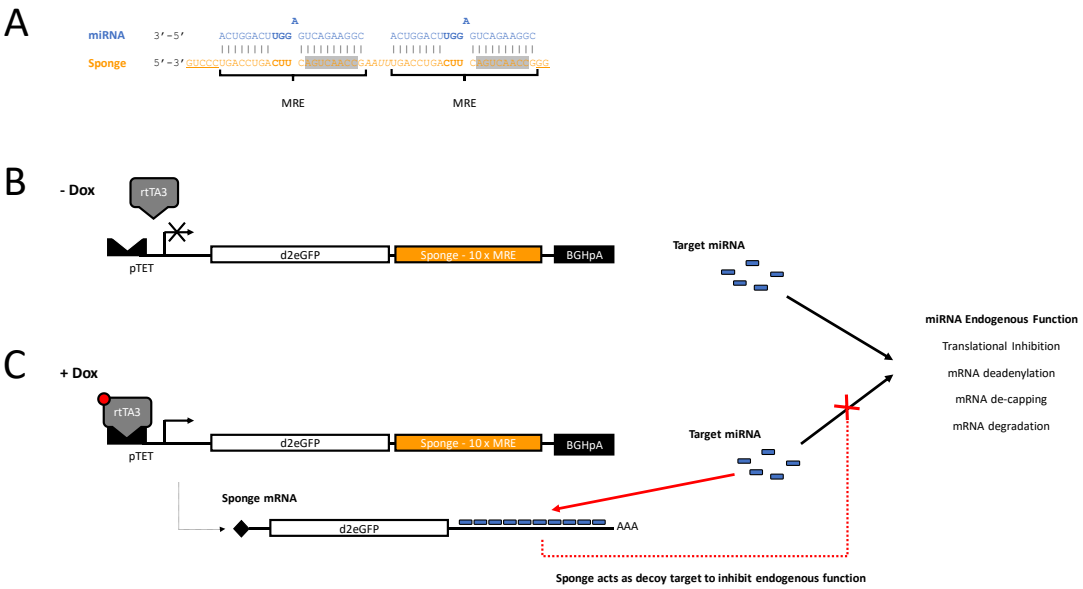


Figure 1. [Conditional miRNA knockdown with TET-ON controlled miRNA-sponge expression: (A) A schematic representation of the miRNA sponge design used in this study. The sponge consists of miRNA responsive elements (MRE) that bind the target miRNA of interest. The miRNA seed binding region is highlighted and a 4nt mismatch in bold. miRNA sponges were cloned downstream of a destabilized (d2eGFP) reporter gene and each sponge sequence contained 10 MREs. The reporter gene – sponge construct is driven by a TET-On inducible promoter (pTET). Transcription is enhanced by binding of the rtTA3 trans-activator to a tetO in the promoter. (B) In the un-induced state and the absence of the induction agent Dox, rtTA3 cannot bind to the tetO region of the promoter. This allows target miRNAs to function naturally. (C) Dox addition enables localization of rtTA3 to pTET and subsequent up-regulation of miRNA-sponge expression. Increased sponge expression generates large numbers of decoy target sites for the miRNAs of interest to bind, diverting them from endogenous functions.]

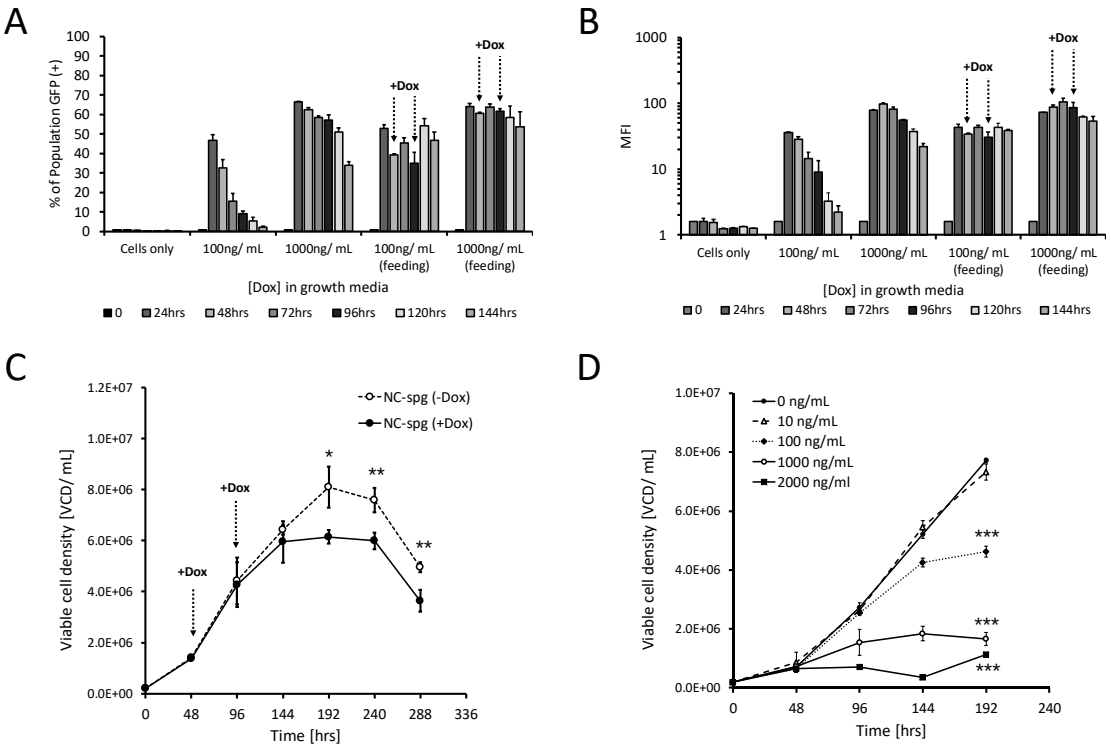


Figure 2. [Induction in stable mixed pools requires toxic concentrations of Dox: CHO DP12 cells stably expressing a negative control sponge (NC-spg) were evaluated for induction. Cells were grown in the presence of Doxycycline Hychlate (Dox) at 100 or 1000ng/ mL for 144 hrs with respect to a cells only control. To sustain induction over this time, Dox-feeding, indicated by arrows, was also assessed by the addition of either 100 or 1000ng/ mL Dox time 0, 48 and 96hrs of culture. miRNA sponge induction was monitored through fluorescence of the reporter gene (d2eGFP). (A) The percentage of total cells GFP positive at 24hr intervals post induction. (B) The corresponding mean fluorescence intensity (MFI) of induced cells. (C) The viable cell density [VCD/ mL] of NC-spg cells was monitored in the absence (-Dox) and presence (+Dox) of Dox. Dox addition consisted of 100ng/ mL at 48hrs and 1000ng/ mL at 96hrs after inoculation. (D) The effects of Dox on the growth of the original DP12 cell line was evaluated by culturing cells in media supplemented with; 0, 10, 100, 1000 and 2000ng/ mL of Dox. Values are the average of three biological replicate cultures. Error is the standard deviation of three biological replicates. Statistics were carried out by a two-tailed homoscedastic student t-test, ($p \leq 0.05$ *, $p \leq 0.01$**, $p \leq 0.001$ ***) of three biological replicates.]

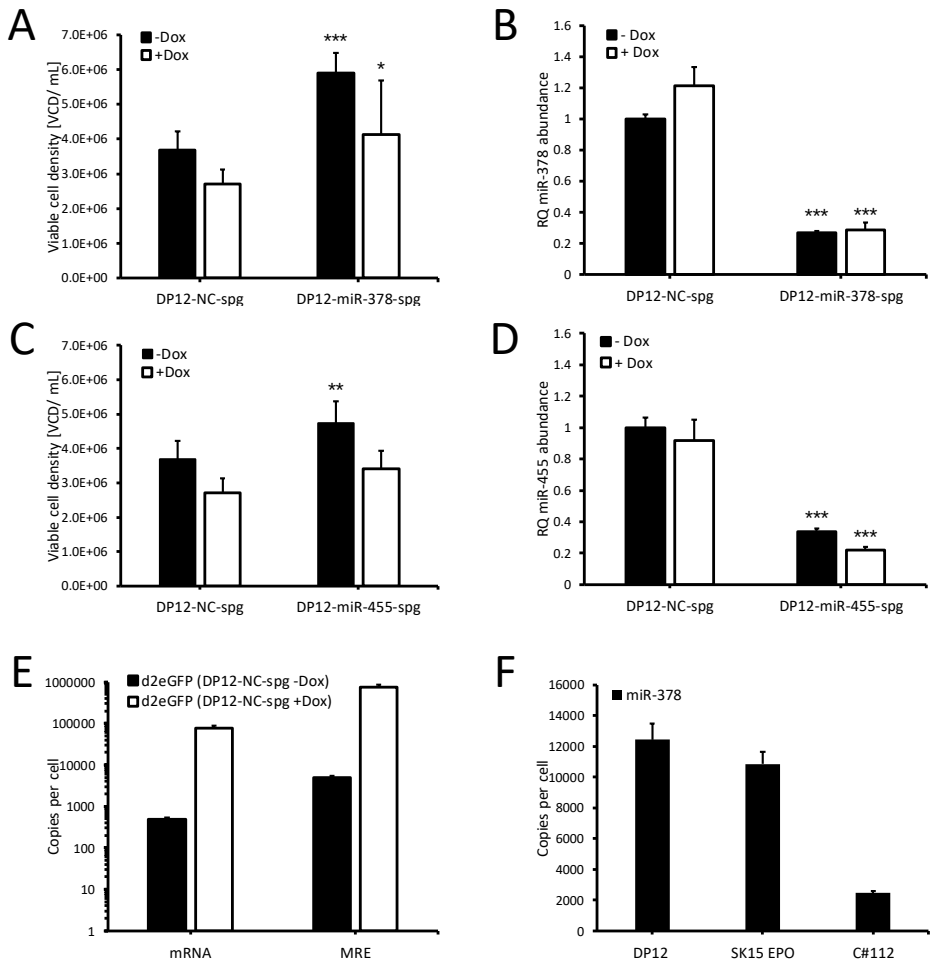


Figure 3. [Leaky sponge expression is sufficient for miRNA knockdown: (A) Maximum viable cell density (VCD/ mL) of NC-spg and miR-378-spg stable mixed pools in the absence (-Dox) or presence (+Dox) of 1000ng/ mL Dox for 48hrs. (B) Relative quantification (RQ) of miR-378 abundance in stable NC-spg and miR-378-spg stable mixed pools induced (+Dox) and un-induced (-Dox). (C) VCD of NC-spg and miR-455-spg stable mixed pools with (+Dox) and without (-Dox) induction. (D) RQ of miR-455 abundance in stable NC-spg and miR-378 specific sponge stable mixed pools induced (+Dox) and un-induced (-Dox). (E) Absolute quantification of d2eGFP-sponge mRNA and miRNA responsive elements (MRE) per cell. (F) Absolute quantification of miRNA-378-3p (miR-378) copies per cell, for three CHO producer lines; DP12, SK15 EPO and C#112. Values in all cases are based on the average of three biological replicate cultures. Error is the standard deviation between three biological replicate cultures. Statistics were carried out (A-D) by a two-tailed homoscedastic student t-test with respect to the un-treated (-Dox) DP12-NC-spg cells, ($p \leq 0.05$ *, $p \leq 0.01$**, $p \leq 0.001$ ***).]

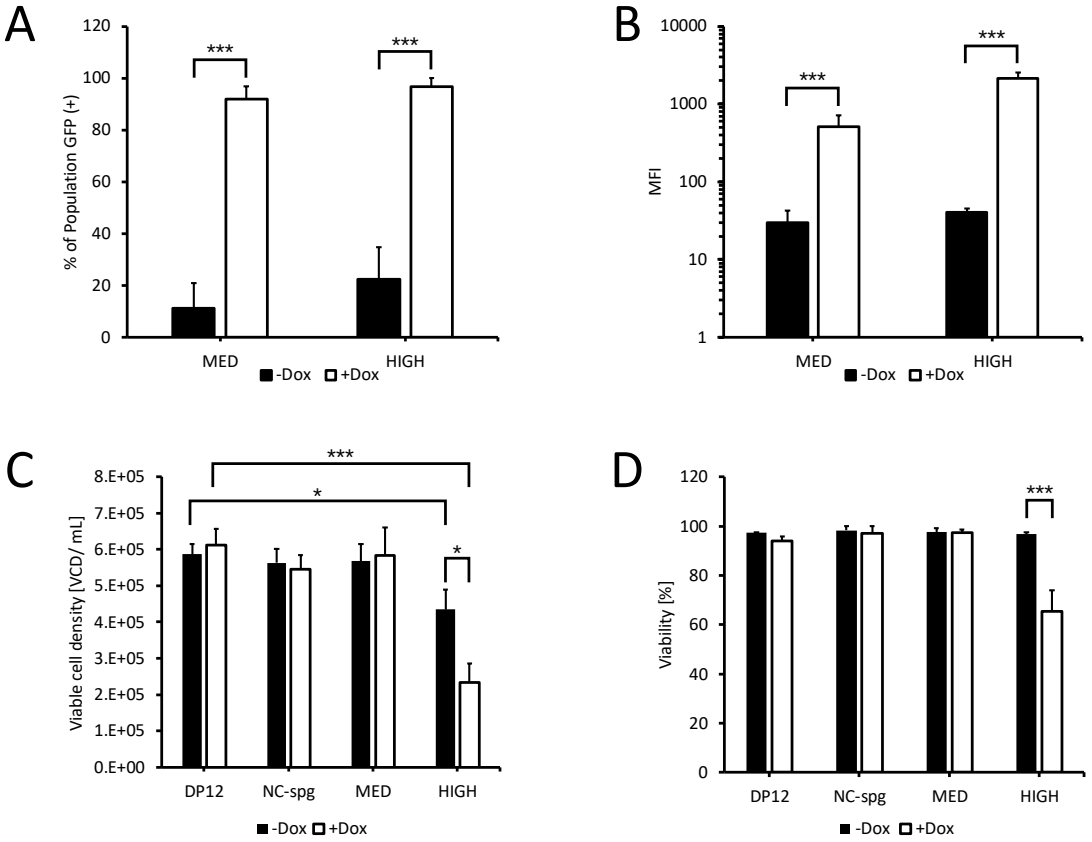


Figure 4. [Growth and induction characteristics of FACS enriched Dox sensitive sub-populations: The NC-spg cells were induced with 10ng/ mL Dox for 24hrs. Two sub-populations of induced cells were isolated by FACS sorting based on mean fluorescence intensity (MFI). (A) The % of total cells GFP positive for FACS enriched subpopulations with medium (MED) and high (HIGH) reporter-gene expression, cultured with or without 10ng/ mL Dox for 48hrs. (B) Corresponding MFI of MED and HIGH cells. (C) The viable cell density [VCD/ mL] 48hrs after Dox addition for DP12, the original NC-spg mixed pool, MED and HIGH cells. (D) The corresponding viability [%] of DP12, NC-spg, MED and HIGH cells. Values in all cases are based on the average of three biological replicate cultures. Error is the standard deviation between three biological replicate cultures. Statistics were carried out by a two-tailed homoscedastic student t-test, ($p \leq 0.05$ *, $p \leq 0.01$**, $p \leq 0.001$ ***).]

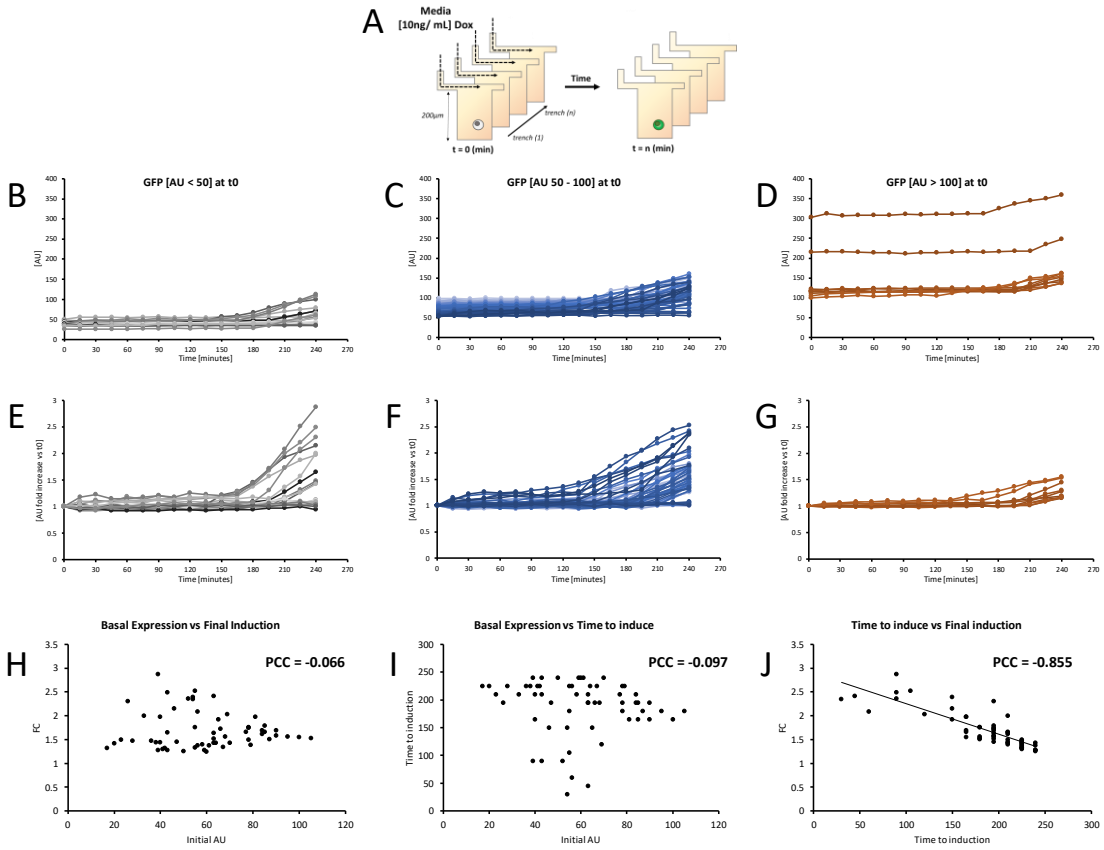


Figure 5. [Heterogeneity in basal sponge expression and induction capacity: In a stable mixed pool the basal sponge expression has a high degree of variability between cells. (A) Single cells from the NC-spg MED population were isolated in microfluidic trenches. Media supplemented with 10ng/ mL Dox was perfused into each trench at time zero, (t = 0 min). The fluorescence of each single cell was monitored in 15 minutes increments. GFP auto-fluorescence units [AU] for 18 single cells over a 4hr period for cells with basal AU > 50 (B), 50 – 100 (C) and > 100 (D). The GFP [AU] normalized to the value of each single cell at time zero for cells > 50 (E) 50 – 100 (F) and > 100 (G). To find a correlation between basal expression and inducibility, the 20% of cells that did not induce over the 240-minute time course were eliminated from this analysis. (H) A representation of final fold induction with respect to basal expression. (I) The cells time to first induction versus basal expression. (J) Correlation of final fold induction and time to induce.]

Chapter 5

Continuous Translation of Circularized mRNA Improves Recombinant Protein Titer

---

Costello, A., Lao, T.N., Clynes, M. * & Barron, N. *
* authors contributed equally
Metabolic Engineering, revised manuscript returned to journal.
Impact Factor (2017): 7.674

I was the primary contributor to all aspects of experimental work and manuscript preparation described in Chapter 5. Dr Nga Lao contributed with extensive discussion of appropriate controls and supported experimentally with optimised methods for immunoblotting and ELISA for the EPO product. Prof. Martin Clynes and Prof. Niall Barron contributed to experimental design and redaction of the manuscript.

ABSTRACT

Recent success in demonstrating the translation of circular RNA open reading frames or circular mRNA, may offer a new avenue for improving recombinant protein production from cell and cell-free expression platforms. Initiation and termination are two rate limiting steps of translation. Circular RNA as a class of RNA is defined by covalent joining of terminal ends to give a closed loop structure. By encoding a gene lacking a stop codon on a circular RNA molecule an infinite open reading frame is generated permitting continuously translating circular mRNA (CTC mRNA). CTC mRNAs have shown promise in enhancing the production of multimeric polyproteins in bacterial cell-free expression systems. Problems arise when homogenous, functional post-translationally modified protein is required. To produce post-translationally modified, secreted protein from an CTC mRNA we investigated co-translational cleavage of nascent polypeptide chains by incorporating a 2A "self-cleavage" peptide motif. Using a model recombinant human glycoprotein Erythropoietin (EPO) we demonstrate for the first time the ability to produce secreted protein from continuously translating circular mRNA in live mammalian cells. Both cell-specific and volumetric productivity were improved by using circular mRNAs. This study pioneers the potential of recombinant protein production from CTC mRNA in mammalian cells.

5.1 INTRODUCTION

Circular RNA (circRNA) is a recent example of how hybridization-free profiling techniques such as RNA-seq (1) have redefined our picture of the transcriptome with identification and quantification of novel, non-canonical transcripts. Eukaryotic circRNA was first discovered in the early 1990's (2, 3). Yet, until recently circRNA was disregarded, believed to be an artefact of splicing error. It has now been established by independent studies that circRNA is highly abundant and the presence of circular isoforms can exceed that of associated linear RNA (4-8). The expression of circRNA is said to be species and cell-type specific with no discernible correlation between linear transcript expression and circRNA biogenesis (9). Studies evaluating the function of endogenous circRNAs have mostly found non-coding species (8, 10). However, interest in the translation of exogenous circular mRNAs has seen recombinant protein expression evaluated in bacterial, cell-free, insect and mammalian cell expression systems (11-15). While protein derived from exogenous circular mRNA was achieved in all cases, the strategy for enhancing protein production over associated linear mRNA differed.

CircRNA is a covalently closed RNA structure composed of exons (8), introns (16) or both (10) and produced by non-colinear splicing or "back-splicing" of pre-mRNA (6). Both endogenous (6) and exogenous (14) circRNAs have shown greater stability over associated linear mRNAs. Lacking the free ends required for exonuclease-mediated degradation makes circRNAs resistant to several mechanisms of RNA turnover, extending their lifespan. A very recent publication evaluated this characteristic of circRNA in recombinant protein production (15). Early work by Perriman et al., (11) and more recently Abe et al., (12) evaluated the expression of recombinant protein from infinite circular mRNAs in *E. coli* and *E. coli* lysate respectively. By its very nature a circular RNA is infinite, and by removing the stop codon in a circular mRNA perfectly divisible by three, an infinite open reading frame is created. In the case of Abe et al., (12) production of a concatemeric FLAG-tag sequence saw >100-fold improvement in yield of total protein. The issue with infinite circular mRNAs is that rolling circle translation (12, 13) results in large multimeric polyproteins that are of no use if functional, homogenous product is required.

Self-cleavage peptides have been implemented in di or polycistronic cassettes in place of viral internal ribosome entry sites (IRES) to overcome the reduced translation of ORFs placed down-stream of an IRES. The cleavage of "2A" motifs is universal in eukaryotes, however the efficiency of 2A motifs can differ (17). First discovered in foot and mouth disease virus, 2A refers to the specific region of the picornavirus polyproteins wherein ribosome skipping of the glycyl-prolyl di-peptide at the C-terminus leads to cleavage (18, 19). T2A from the *Thosea asigma virus* has been reported to have the highest self-cleavage efficiency in CHO DG44 cells (20). Encoding a 2A self-cleavage peptide in a circularized infinite ORF could in principle allow

rolling circle translation (RCT) and yield homogenous protein. The findings of Donnelly et al., (21) indicating a non-proteolytic, but rather ribosomal skipping action of the 2A sequences lends itself to RCT-like action as translation is not re-initiated but continued. The present study investigated if co-translational self-cleavage of polyproteins derived from an infinite circular mRNA could improve the translation efficiency of a secreted therapeutic glycoprotein and improve recombinant protein yield.

5.2 MATERIALS AND METHODS

5.2.1 Circular RNA Constructs

Human Splicing Finder (HSF) (http://www.umd.be/HSF3) (22) was used to compare the confidence of exon splice signals from common house-keeping genes. A confidence cut-off of 90 was used and candidates were prioritized on minimal alternative splice signals, finally exons required a flanking intron >300bp. The human ACTB exon 4 was chosen for cloning. Genomic DNA was harvested from human BJ fibroblast cells using a Blood & Cell Culture DNA Mini Kit (Qiagen). The human ACTB exon-4 locus was amplified by PCR using My Taq 2X (Bioline). The locus was amplified in two fragments to facilitate the incorporation of a unique SanDI site in the exon sequence. The first fragment included ~450bp of the 5' flanking intron, with HindIII and SanDI in the forward and reverse primers respectfully. The second fragment included ~100bp of the endogenous 3' intron, with SanDI and EcoRV in the forward and reverse primers. pcDNA3.1 and PCR amplicons were digested using Fast Digest (FD) enzymes (Thermo Scientific) at 37°C for 30 minutes, and gel purified using the PureLink™ Quick Gel Extraction and PCR Purification Combo Kit (Invitrogen). Both fragments were cloned into pcDNA3.1 (+) HYG in a single reaction triple ligation, using T4 DNA Ligase (Roche). To clone the reverse complementary intron, 350bp of the 5' intron sequence was PCR amplified with XbaI and EcoRV sites on the forward and reverse primers, and subsequently ligated into the vector. The same intron flanks as those used to circularize ACTB exon 4 were used in the circEPO constructs with the exon sequence replaced by the rhEPO ORF. Multiple restriction enzyme sites were added to each end of the ORF. HSF was again used to optimize the splice sites, by single base pair changes in the splice acceptor and donor motifs. To allow for ribosomal entry into the circular mRNA constructs a short "eukaryotic IRES-like site" was encoded 5' of the start codon, as described previously (23). The T2A motif was taken from Kim et al., (17). All virtual cloning and vector analysis was performed in Vector NTI 10.3.1 (Invitrogen). For the circEPO expression vectors, the EPO and intron flanking sequence were synthesized (GenScript). The sequences of all circEPO variants were verified by Sanger sequencing (Eurofins Genomics) and can be found in (Supplementary Figure 1).

5.2.2 Cell Culture and treatments

1.1.1. CHO-K1 (ATCC) were grown Cell Culture and treatments

in CHO S SFM II (Gibco) with (2.52g/ L) polyvinyl alcohol (PVA), and sub-cultured in a 3-4-3-day manner. Cells were grown in a Climo-Shaker (Kühner), 37°C, 80% humidity and 5% CO_2 with routine seeding at $2x10^5$ cells/ mL in 50mL spin tube (TPP), 5mL working volume. For the batch cultures in (Figure 6) cells were grown in 250 mL Erlenmeyer flasks (Corning). Cell growth and viability was measured in technical triplicate using the Guava EasyCyte ViaCount programme (Merck Millipore). Prior to transfection, cells were washed twice with CHO S SFM II and seeded between $0.2\text{-}1x10^6$ cells/ mL in 2 mL of CHO S SFM II. Transfection was performed in accordance with the MIRUS Transit 2X protocol. In brief, a complex of a [1:2] ratio µg pDNA to µL transfection reagent was formed in 100µL of CHO S SFM II and pre-warmed to 37°C, for 30mins, then added to the cell suspension, and incubated for 24-96hrs. Actinomycin D (Sigma Aldrich) treatments were performed in 24 well suspension plates (Fisher Scientific). For RNA stability analysis 5µg/ mL of Actinomycin D was added to triplicate wells and cells subsequently harvested at various timepoints. A time zero control of cells harvested at time of Actinomycin D addition was used as a reference.

5.2.3 RNA isolation and RNase R treatment

Cells were harvested by centrifugation at 91xg for 5 minutes. Total RNA was isolated from $1\text{-}5x10^6$ cells using Tri-reagent (Ambion), in accordance with manufacture's protocol. RNA quantification and quality were evaluated by NanoDrop (Thermo Scientific). RNA samples were treated with DNase I (Sigma Aldrich) prior to cDNA synthesis. RNase R (24) treatment was used to validate RNA circularization, for this 2µg of total RNA was either treated with (15Units/µgRNA) of RNase R (Epicentre® Biotechnologies) or mock treated with RNase free water (Ambion), for 10 minutes at 37°C on a heat block.

5.2.4 Semi and Quantitative RT-PCR

Reverse transcription of total RNA was performed with the High Capacity cDNA Reverse Transcription Kit (Applied Biosystems), on a thermocycler (G-Storm), in accordance with the manufacture's protocol. RT-qPCR was performed on a 7500 (Applied Biosystems) using Fast SYBR Green Master Mix (Applied Biosystems). The 2X SYBR was mixed with 20ng of cDNA, 200nM forward and reverse primers, and water made up to 20µL final reaction volume. Each biological replicate sample was run in technical triplicate wells. Primer sequences can be found in (Supplementary Table 1). Relative quantification was calculated using the ddCt method with respect to endogenous Gapdh. Absolute values of RNA transcripts were calculated as previously

described (25). RNA decay rates were calculated using the method outlined by Jeck et al., (2013) (6).

5.2.5 Protein Isolation, quantification and PNGase F treatment

Cells were lysed with 100µL of lysis buffer; 7M Urea, 2M Thiourea, 4% CHAPS, 30mM Tris adjusted to pH8.5, per $1x10^6$ cells. Lysis was completed by passing the cell pellet through a syringe 10 times and rotating for 1hr. Bradford reagent (Bio-Rad) and pre-diluted bovine serum albumin (BSA) standards (Thermo Fischer Scientific) were used to quantify protein concentration. For evaluation of secreted protein, cell supernatant was harvested by centrifugation, 91xg for 5minutes. PNGase F (R&D Systems) was used to remove N-linked glycans from EPO products. Prior to PNGase F treatment, sample denaturation was conducted in a 30µL reaction volume using 0.5% SDS and 50mM DTT then incubated at 100°C for 10min. The reaction was cooled to room temperature and 3µL of 10% Triton-x100 added. The reaction was then split in two and either mock treated with 15µL of assay buffer; 0.1M Tris, pH 7.5, or 5ng of PNGase F in assay buffer. The PNGase F reaction was incubated at 37°C for 2hrs.

5.2.6 Immunoblotting

Preparation of protein samples; 5-30µL of concentrated supernatant or 20µg of lysate was made up to a total volume of 30µL, mixed with 10 µL of NuPAGE LDS Sample Buffer (4x) (Thermo Fischer Scientific), 4 µL 10X NuPAGE Sample Reducing Agent (Thermo Fischer) and incubated at 100°C for 10 minutes. Samples were loaded on a Bolt 4-12% Bis-Tris Plus Gel (Thermo Fischer Scientific). Each gel contained a lane with 5 µL of SeeBlue Plus 2 Protein Standard (Invitrogen). Recombinant human EPO (Merck Millipore) was used as a positive control. Gels were run at 180 V. The gel, Amersham Protan Nitrocellulose Membrane (Sigma-Aldrich), and filter paper were soaked in equilibration buffer for 20 minutes. Protein transfer was done on a Fast blotter system (Thermo Fischer Scientific) using the 7-minute pre-programmed method. Membranes were blocked for 1hr to overnight with Odyssey Blocking Buffer PBS (Li-Cor), at 4°C. Primary antibody, Rabbit polyclonal anti-human EPO AB-286-NA (R&D Systems), was diluted to a final concentration of 400ng/mL in PBS, was added and the membrane incubated at 4°C overnight. Primary antibody was decanted, and the membrane washed four times, five minutes each, in PBS containing 0.1% Tween 20. Secondary antibody 800 Green anti-rabbit (Li-Cor) was added to membrane and incubated at room temperature for 1 hour. Membranes were imaged using an Odyssey imager (Li-Cor). Detection of endogenous Gapdh for lysate samples was done using a primary mAb (60004-1-Ig) from (Proteintech) at 0.1ng/ mL for 1hr at room temperature. Secondary antibody 680 Red anti-mouse (Li-Cor) was used to visual the Gapdh control.

5.2.7 ELISA

EPO was quantified by enzyme linked immunosorbent assay (ELISA) of cell culture supernatant. Supernatant was harvested simultaneously with RNA and protein sampling, by centrifugation of cell suspension at 91xg for 5 minutes. Samples consist of pooled technical triplicates, measured in technical triplicate, with 3 biological replicates. Nunc-Immuno™ MicroWell™ 96 well solid plates (M9410 Sigma-Aldrich) were coated with 100µL of capture antibody MAB-287 (R&D Systems), diluted 1:500 in coating buffer (C3041-50CAP) (Sigma-Aldrich) for 1hr to overnight at 4°C. Plate washing consisted of 3 x 100µL /well washes of ELISA wash buffer (T9039) (Sigma-Aldrich). Blocking was done for 1 hr at room temperature with (T6789) Blocking Buffer (Sigma-Aldrich). Standards of recombinant human EPO (Cat no. 329871) (Merck Millipore) were diluted to range from 8000-0pg/mL in sample dilution buffer (T6789 + 0.05% Tween20) (Sigma-Aldrich). Samples were diluted to fall within the range of standard curve and 100µL incubated on the plate at 37°C for 1 hr. Detection was achieved using Rabbit Immunoglobulin-HRP (P0448) (Dako, Agilent) diluted 1:2000 with 100µL added to each well and incubated for 1hr. 100µL of development solution; 1.2mL of 3,3',5,5'-Tetramethylbenzidin at 1mg/ mL in dimethyl sulfoxide, 2.4µL 30% (v/v) hydrogen peroxide and 10.8mL Phosphate-Citrate buffer (P4809) (Merck), was added to each well and incubated for 30minutes. The reaction was stopped by adding 100µL 0.18M H_2SO_4 to each well. Protein concentration was determined by reading the plate on a Multi-scan Go (Thermo Fischer Scientific).

5.2.8 Statistical Analysis

Unless otherwise stated, values reported represent the mean of three independent experiments with error displayed as standard deviation. Statistical analysis was carried out by a two-tailed homoscedastic student t-test, ($p \leq 0.05$ *, $p \leq 0.01$**, $p \leq 0.001$ ***) of three independent experimental replicates.

5.3 RESULTS

5.3.1 Co-translational cleavage of rolling circle translation products

The general mechanism by which mRNA is translated to polypeptides can be broken down into three steps (Figure 1A). The translation initiation and termination steps are relatively slow compared to elongation (12). To overcome these rate limiting steps of translation we examined the use of continuously translating circular mRNA (Figure 1B). Continuously translating circular mRNAs (CTC mRNA) are circular RNAs encoding a protein with no stop codon. In theory if the CTC mRNA is a multiple of three nucleotides the ribosome should circle the molecule indefinitely producing a repeating polyprotein. A repeating polyprotein is of little use if homogenous, functional protein is required. To overcome this, we sought to investigate co-translational cleavage of the growing polyprotein with the inclusion of a sequence encoding a 2A

self-cleavage peptide (Figure 1C) in the CTC mRNA. First it was imperative that a robust means of producing circular RNA *in vivo* was obtained and to investigate if the mechanism of rolling circle translation was feasible in living Chinese hamster ovary cells.

5.3.2 Canonical splice sites produce circular RNA with high efficiency in CHO cells

To date, rolling circle translation has only been achieved with synthesized RNA artificially circularized with T4 RNA ligase (12, 13). Large scale RNA synthesis is not cost-effective in the context of recombinant protein production. It has been reported previously that exon circularization is dependent on canonical splice signals (26), is facilitated by complementary intron pairing (8) and that the intron length can affect the level of transcript circularization (10). We chose the human ACTB exon 4 as a model of RNA circularization based on the assumption that house-keeping genes, with their high levels of expression, have strong splicing efficiency. The ACTB exon 4 genomic locus was amplified with 450bp 5' and 100bp 3' flanking intron sequence. To facilitate exon circularization, approximately 350bp of the 5' intron was cloned in reverse orientation downstream of the endogenous 3' intron flank (Figure 2A). Upon transcription the 5' intron hybridizes to the reverse complementary intron (RCI) bringing the splice signals into close proximity. circRNA is formed harboring a unique back-splice junction (BSJ). To detect the presence of circularized ACTB exon 4 multiple primer sets were used. Three primer sets were common to both linear and circular isoforms of the transcript (Figure 2B). Another three primer sets were designed to amplify the BSJ; one traversing the junction, one with the reverse primer embedded in the BSJ and the final one with the forward primer embedded in the BSJ (Figure 2B). The efficiency of transcript circularization by complementary intron pairing was determined by distinguishing exogenous transcript isoforms from endogenous Actin using multiple primer sets (Figure 2C). Absolute quantification (Figure 2D) enabled the approximation of exogenous transcript circularization to be 92% of the total exon over-expression in CHO-K1 cells (Figure 2E).

5.3.3 Circularization of human EPO open reading frame

Having established that RNA can be circularized efficiently in Chinese hamster ovary cells, human Erythropoietin (EPO) was chosen as a model glycoprotein to evaluate the potential for recombinant protein therapeutic production from circular mRNA. EPO expression vectors were designed and synthesized for both linear and circular mRNA isoforms (Figure 3). The constructs included a control linear mRNA (Figure 3A). To verify that RNA circularization was occurring through intron pairing and not just the presence of splice signals, an un-structured intron control was used to express linear pre-mRNA (L pre-mRNA) (Figure 3B). In this case the EPO ORF was flanked 5' with a splice acceptor (SA) and 3' with a splice donor (SD) along with a 5' intron. To

generate circular mRNA a reverse complementary intron was cloned downstream of the 3' SD (Figure 3C). The circular EPO ORF was made infinite by the removal of the stop codon (Figure 3D). Finally, the 2A self-cleavage motif from *Thosea asigma virus* (T2A) (17) was encoded in this infinite circular mRNA (Figure 3E).

In designing constructs to circularize the EPO ORF numerous steps were undertaken to ensure highly efficient splicing with no variance between circRNAs harboring unique motifs. Human splice site finder 3.0 (HSF) (http://www.umd.be/HSF3/HSF.html) was used to identify splice signals and ensure the splice confidence score was not compromised with alterations to the EPO ORF (Figure 4A). A splice confidence threshold of 90 was used to distinguish between strong and weak signals. The HSF software provides a map of splice acceptor (SA) and splice donor (SD) motifs found in the input sequence. A linear vector map of the structured intron construct for the circularization of EPO aligned about the HSF output shows how the expected SA and SD align to that predicted. Because of differences at the 3' of the EPO ORF between C mRNA, CTC mRNA and 2ACTC mRNA the confidence of the SD prediction differed greatly. To overcome this the ORF was directly bordered with a short sequence of restriction enzyme cloning sites. This both normalized the splice signals and would enable the use of the backbone construct for circularization of other genes. RNA folding software (http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi) was used to qualitatively evaluate the structured intron pairing (Figure 4B). It was evident that the designed structured intron sequence should form the anticipated hairpin-like structure to bring the ends of the gene ORF into close proximity.

The utilization of multiple primer sets to distinguish transcript variants was applied to evaluate EPO ORF circularization; a common primer set amplifying the EPO (ORF), an intron-exon bridge primer pair to detect linear pre-mRNA (Intron bridge) and a set generating an amplicon traversing the back-splice junction (BSJ) which should only be detected in the presence of circRNA (Figure 4C). RNase R has been shown to digest all linear RNA and is used to enrich for circRNA species in RNA-Seq studies (6). Total RNA was harvested from CHO-K1 cells that had been transiently transfected with linear and circular mRNA plasmid constructs. The total RNA was then either treated or mock treated with RNase R prior to cDNA synthesis and subsequent PCR analysis (Figure 4D). RNase R digestion was verified by the inclusion of a Gapdh control for both treated and mock treated samples. The common primer set amplifying the ORF detected EPO expression in all mock treated cases. As anticipated the presence of BSJ was only detected in cells transfected with the C mRNA construct. The circular EPO ORF was resistant to RNase R digestion evidencing its circularity. To investigate if splicing was altered between EPO circular mRNAs with different elements, the back-splice junction (BSJ) of each construct was amplified (Figure 4E). To provide template for sequencing the entire circular RNA molecule in each case was also amplified using the forward primer from the ORF primer set with the reverse primer

from the BSJ specific primers (Figure 4E). Sequencing of the BSJ derived from each construct showed an identical splice junction in all cases (Figure 4F). Absolute quantification of each transcript variant was measured by RT-qPCR (Figure 4G). Circular RNA was only detected in the structured intron sample, with near complete circularization. The intron-bridge primer set was detected at relatively low levels with respect to linear and circular forms. In the case of C mRNA this can be explained by circular mRNA formation. In the case of the L pre-mRNA construct it is suggested that the lack of RNA circularization should not impede the loss of intron sequence by linear splicing. It has been reported that the lack of free ends makes circular RNA resistant to most cellular nucleases increasing stability of the molecule (6). The intracellular stability of linear and circular EPO RNA was evaluated by Actinomycin D treatment of transiently transfected cells. The circular EPO mRNA (circEPO) proved to be more stable than linear EPO mRNA over a period of 96hrs (Figure 4H). In the case of linear EPO (Lin EPO) and Gapdh, the transcripts decrease to near un-detectable levels, approximately 10 copies per cell. In contrast, the circular variant degraded at a much lower rate, with 1800 copies per cell still detectable after 96hrs.

5.3.4 Continuously translating circular mRNA produces glycosylated polyproteins

Rolling circle translation of the EPO ORF in CHO-K1 cells transiently transfected with pDNA expressing one of; L mRNA, C mRNA, CTC mRNA or 2ACTC mRNA, was evaluated by resolving whole cell lysate (WCL) in an SDS-PAGE gel (Figure 5A). EPO was detected at three molecular weights in the lysate for CTC mRNA transfected CHO-K1 cells. 2A self-cleavage motifs are reported to work with a ribosome skipping action wherein steric hindrance of Pro-Gly-Pro in the translating ribosome forces the nascent polypeptide to leave and continue translation with Pro. Inclusion of a 2A motif in the CTC mRNA abolished the formation of high molecular weight polyprotein variants (Figure 5A). Human EPO has a molecular weight of 21.3kDa however due to high levels of sialylation on its three N-linked glycans, the protein migrates to ~37kDa on an SDS-PAGE gel. To determine a more accurate estimation of the EPO variants found in the CTC mRNA transfected cells, PNGase F was used to remove all N-linked glycans (Figure 5B). Three bands were observed in the mock-treated lysates (- PNGase F). Removal of the glycans improved resolution as the uniformity of the protein condensed to strong bands (+ PNGase F). Three bands were found in the PNGase F treated CTC mRNA EPO sample. These bands migrated to approximately 23, 46 and 68kDa, indicative of mono, di and trimeric EPO polyproteins, respectively. The rolling circle translation (RCT) of CTC mRNA stopping at a trimeric form was surprising. Previous reports of RCT have shown very high molecular weight polyproteins. These studies were however conducted over short time periods in cell free expression systems. To address this CTC mRNA and a L pre-mRNA control lacking a stop codon (Supplementary Figure 2A) were transfected for 16hrs. A similar pattern of EPO variants was

seen in CTC mRNA transfected CHO-K1 lysate (Supplementary Figure 2B). Higher molecular weight variants were not formed, this gave confidence that the poly-EPO variants were just that and not the product of read-through translation of downstream sequence. Read-through translation of the control would result in the addition of 35 amino acid residues and increase the molecular weight to 24.8kDa, which is considerably smaller than the dimer with an apparent molecular weight of 46kDa. Secreted EPO was produced by all constructs (Figure 5C). EPO produced by L mRNA and C mRNA appeared similar in molecular weight, while that derived from CTC mRNA and 2ACTC mRNA had a higher molecular weight. In the case of 2ACTC mRNA this is due to the inclusion of the 2A motif on the C-terminal of the EPO protein. Secreted EPO was produced by the CTC mRNA construct. The product had a higher apparent molecular weight, we believe this to be the result of post-translational cleavage of the growing poly-protein chain. EPO has a natural signal peptide which is removed prior to its secretion. Post-translational processing of the poly-protein chain to release the signal peptide from the N-terminus of one product would result in that sequence remaining at the C-terminus of the proceeding product. This would explain the higher apparent molecular weight of this product.

5.3.5 Continuously translating circular mRNA improves cell productivity

To investigate the validity of circular mRNA as an alternative to linear mRNA for recombinant protein production, constructs were evaluated by transient transfection of CHO-K1 cells in 50 mL batch culture. The vectors used in this study also drive the expression of a resistance gene Hygromycin-B-phosphotransferase (HYG) under a PKG promoter. Relative quantification of HYG mRNA normalized to endogenous Gapdh was used to evaluate relative transfection efficiency. There was no significant difference in HYG mRNA levels between cultures (Figure 6A). The relative expression of linear and circular EPO mRNA isoforms was normalized to HYG mRNA (Figure 6B). At 48hrs post transfection there was a fold increase in EPO mRNA abundance of 11.99 ($p = 6.43x10^{-5}$), 9.29 ($p = 2.23x10^{-6}$) and 12.39 ($p = 3.47x10^{-7}$) for C mRNA, CTC mRNA and 2ACTC mRNAs respectively over that of the L mRNA control. The viable cell density (VCD/ mL) was monitored at 24hr intervals (Figure 6C). There was no significant reduction in VCD between cells expressing L mRNA and C mRNA ($p = 0.14039$). The VCD at 96hrs post transfection of cells expressing CTC mRNA and 2ACTC mRNA was significantly reduced with respect to that of L mRNA (Figure 6C). The culture viability (%) dropped following transfection in all cases with no significant difference between cells expressing linear and circular EPO mRNA variants (Figure 6D). There was no significant difference in volumetric titer between cells expressing L mRNA or C mRNA 96hrs post transfection ($p = 0.0977$) (Figure 6E). There was a significant increase in secreted EPO titer from L mRNA (1.677 mg/ L) to (3.64 mg/ L, $p = 9.031x10^{-5}$) with CTC mRNA and to (4.065 mg/ mL, $p = 0.00485$) with 2ACTC mRNA. This

equates to a 2.17-fold and 2.42-fold improvement in secreted EPO titer by rolling cycle translation of CTC mRNA and 2ACTC mRNA respectively. At 96hrs post transfection the cell specific productivity (Qp) of cultures expressing CTC mRNA and 2ACTC mRNA was (3.47-fold, $p = 1.545x10^{-5}$) and (3.25-fold, $p = 0.00133$) higher respectively over cells expressing L mRNA (Figure 6F). The cell specific productivity improvement from CTC mRNA is difficult to determine as quantifying polyproteins with multiple binding motifs on the same molecule by ELISA is likely to be inaccurate. As rolling circle translation produced largely di- and trimeric variants of EPO intracellularly, it was thought that perhaps the protein structure was hindering long polyprotein repeats. Several genes involved in the un-folded protein response were evaluated at transcript level (Supplementary Figure 4) to investigate if there was any response to exogenous circular mRNA expression and the resulting polyproteins. There was a small but insignificant down-regulation of all nine genes in cells expressing circular mRNA compared to linear mRNAs and non-transfected cell controls. It remains to be seen whether this would still be the case in a high producing stable cell line.

5.4 DISCUSSION

Improving cell specific productivity of recombinant proteins has been a focus of the biotechnology industry since the 1980's. Recent publication of the industry workhorse CHO-K1 genome (27) has seen efforts shift from traditional process engineering approaches towards host cell engineering. As we continue to assimilate the nuances of eukaryotic translation, engineering transgene mRNA may gain increased attention also. To date, there has been limited investigation of circular mRNAs for the production of recombinant protein in mammalian cells (13, 15). Abe et al., (13) showed rolling circle translation of a continuously translating circular mRNA Flag-tag in Hela cells. Wesselhoeft and colleagues investigated optimization of RNA circularization in HEK-293 cells using heavily modified nucleic acids achieving a high degree of circularization and reported on improved protein yield over a linear mRNA control due to the enhanced stability of circular RNA. Here we report for the first time a strategy for highly efficient production of circular mRNA from a structured intron expression vector taking inspiration form the strategy put forward by Hansen et al., (8) in the validation studies of endogenous circRNA species. Hansen demonstrated an exogenous mimic of the complementary intron pairing of the circRNA biogenesis pathway (8) whereas, previous studies have relied on the use of synthesized RNAs delivered directly to cells (12, 13, 15). We also demonstrate co-translational cleavage of polyproteins from continuously translating circular mRNA to their monomeric form by incorporation of a 2A peptide.

As exon circularization is dependent on canonical splice signals (25) it was thought to use a single exon as a model for circularization efficiency via complementary intron pairing. The choice of candidates for this model of circRNA biogenesis efficiency was prioritized from common house-keeping genes on the assumption that their expression is so abundant and uniform, the splice signals and surrounding enhancer motifs must work with very high efficiencies. Circularization of the human ACTB exon 4 via complementary intron pairing achieved >90% circularization in CHO-K1 cells. Advancing this strategy to circularization of transgene mRNA to create circular and continuously translating circular mRNAs, a number of steps were undertaken to ensure uniformity in splicing between constructs. Sequencing of the back-splice junctions from EPO; C mRNA, CTC mRNA and 2ACTC mRNA showed that all constructs contained an identical splice junction. This is imperative when using CTC mRNAs due the fact that rolling circle translation of a circular mRNA not perfectly divisible by three would read out of frame on the second (and third) rotation of the ribosome. Circularization of the EPO ORF was validated in three ways; amplification of the unique back-splice junction, RNase R resistance and sequencing confirmation of the back-splice junction. It has previously been demonstrated that while the length of flanking intron sequence positively correlates with the fold increase in circRNA products in exogenous expression of naturally occurring circRNAs, it is not always necessary for their formation (10). Our results suggest that complementary structured intron flanking is essential for the expression of exogenous circRNAs from pDNA. Complementary intron pairing facilitated the circularization of >90% of total exogenous transcript for both the model exon and EPO ORF constructs, producing a circular mRNA with 2-log-fold increased stability over its linear counter-part. We suggest that expression of circular mRNA by a structured intron vector is a viable means for scalable expression of transgene RNA.

Rolling circle translation of the EPO ORF produced multimeric di- and trimer variants of the protein. These higher molecular weight polyproteins were retained in the cell (Figure 5A). PNGase F treatment of purified-EPO from CHO-K1 cell lysate showed that mono, di and trimeric versions of the proteins all contained N-linked glycans. The most prominent form existing in the cell lysate was a dimer. Why the ribosome produces monomeric EPO from CTC mRNA and does not rotate the transcript more than three times is not yet clear. As the polyproteins appear to be glycosylated, it can be assumed that the polypeptide chain is reaching the ER and subsequently the golgi. EPO has a globular structure and the spacer produced by translation of the back-splice junction is very small compared to the protein. It may be that large concatemeric globular poly-EPO may be an unstable structure and lead to inhibition of further RCT action. This was not observed by Abe et al., (13) previously however their repeating FLAG-tag of only eight amino acids is a much simpler structure than the 193 residues that make up EPO. It could also be that the native signal peptide in the human EPO sequence is post-translationally cleaved as the poly-

protein chain is growing. This may explain why we don't see very high molecular weight variants and how protein is being secreted from the cells. De-glycosylation of the secreted products with PNGase F showed the product of CTC mRNA to be higher in molecular weight than that of L mRNA and C mRNA. This further supports the suspected retention of signal peptide sequence (Figure 5C). Another possibility is that very large poly-EPOs were not being resolved in the SDS-PAGE gel. Determination of polyprotein by ELISA is also difficult as the repeating motif possibly limits accurate quantification. The reduced growth of CHO-K1 cells expressing circular RNA variants could be the result of higher protein production diverting energy from growth to productivity or the cell's discrimination between linear and circular RNA. We postulated that long concatemeric proteins would stress the cell's post-translational machinery but found no significant impact on hallmark genes of the un-folded protein response (Supplementary Figure 4). The increase in EPO mRNA abundance with circularization (Figure 6B) could be the reason for the moderate reduction in viable cell density (28). This may be a limiting factor of recombinant protein production from circular mRNA in live mammalian cells.

This study raises some interesting questions. What is known is that we can robustly produce circular mRNA in live mammalian cells from pDNA using a structured intron expression vector. Alterations to the gene sequence do not affect circularization efficiency and circular mRNA is more stable than linear mRNA. Rolling circle translation of recombinant human EPO produced large multimeric variants and incorporation of a 2A self-cleavage motif abolished this resulting in the highest levels of secreted protein. The reduced cell growth found with expression of these constructs should not be an issue as volumetric productivity was enhanced. Use of synthetic biology to counter the cells response to these circular mRNAs may offer even greater improvements in titer. What can be concluded is that there's still largely un-tapped potential improvements in translational efficiency to be gained from optimization of transgene mRNA structure.

FUNDING

This work was conducted under the financial support of Scientific Foundation of Ireland (SFI) grant numbers [13/IA/1963] and [13/IA/1841].

5.5 REFERENCES

1. Monger, C., Kelly, P.S., Gallagher, C., Clynes, M., Barron, N. and Clarke, C. (2015) Towards next generation CHO cell biology: Bioinformatics methods for RNA-Seq-based expression profiling. *Biotechnology Journal*, **10,** 950-966.

2. Nigro, J.M., Cho, K.R., Fearon, E.R., Kern, S.E., Ruppert, J.M., Oliner, J.D., Kinzler, K.W. and Vogelstein, B. (1991) Scrambled exons. *Cell*, **64,** 607-613.

3. Cocquerelle, C., Mascrez, B., Hetuin, D. and Bailleul, B. (1993) Mis-splicing yields circular RNA molecules. *FASEB J.*, **7,** 155-160.

4. Burd,C.E., Jeck,W.R., Liu,Y., Sanoff,H.K., Wang,Z. and Sharpless,N.E. (2010) Expression of linear and novel circular forms of an INK4/ARF-associated non-coding RNA correlates with atherosclerosis risk. *PLoS Genetics*, **6,** e1001233.

5. Salzman, J., Gawad, C., Wang, P.L., Lacayo, N. and Brown, P.O. (2012) Circular RNAs are the predominant transcript isoform from hundreds of human genes in diverse cell types. *PloS One*, **7,** e30733.

6. Jeck, W.R., Sorrentino, J.A., Wang, K., Slevin, M.K., Burd, C.E., Liu, J., Marzluff, W.F. and Sharpless, N.E. (2013) Circular RNAs are abundant, conserved, and associated with ALU repeats. *RNA*, **19,** 141-157.

7. Memczak, S., Jens, M., Elefsinioti, A., Torti, F., Krueger, J., Rybak, A., Maier, L., Mackowiak, S.D., Gregersen, L.H. and Munschauer, M. (2013) Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature*, **495,** 333-338.

8. Hansen, T.B., Jensen, T.I., Clausen, B.H., Bramsen, J.B., Finsen, B., Damgaard, C.K. and Kjems, J. (2013) Natural RNA circles function as efficient microRNA sponges. *Nature*, **495,** 384-388.

9. Salzman, J., Chen, R.E., Olsen, M.N., Wang, P.L. and Brown, P.O. (2013) Cell-type specific features of circular RNA expression. *PLoS Genetics*, **9,** e1003777.

10. Li, Z., Huang, C., Bao, C., Chen, L., Lin, M., Wang, X., Zhong, G., Yu, B., Hu, W. and Dai, L. (2015) Exon-intron circular RNAs regulate transcription in the nucleus. *Nature Structural & Molecular Biology*, **22,** 256-264.

11. Perriman, R. and Ares, M. (1998) Circular mRNA can direct translation of extremely long repeating-sequence proteins in vivo. *RNA*, **4,** 1047-1054.

12. Abe, N., Hiroshima, M., Maruyama, H., Nakashima, Y., Nakano, Y., Matsuda, A., Sako, Y., Ito, Y. and Abe, H. (2013) Rolling circle amplification in a prokaryotic translation system using small circular RNA. *Angewandte Chemie*, **125,** 7142-7146.

13. Abe, N., Matsumoto, K., Nishihara, M., Nakano, Y., Shibata, A., Maruyama, H., Shuto, S., Matsuda, A., Yoshida, M., Ito, Y., et al. (2015) Rolling circle translation of circular RNA in living human cells. *Scientific Reports*, **5,** 16435.

14. Wang, Y. and Wang, Z. (2015) Efficient backsplicing produces translatable circular mRNAs. *RNA*, **21,** 172-179.

15. Wesselhoeft, R.A., Kowalski, P.S. and Anderson, D.G. (2018) Engineering circular RNA for potent and stable translation in eukaryotic cells. *Nature Communications,* **9,** 2629.

16. Talhouarne, G.J. and Gall, J.G. (2014) Lariat intronic RNAs in the cytoplasm of xenopus tropicalis oocytes. *RNA,* **20,** 1476-1487.

17. Kim, J.H., Lee, S., Li, L., Park, H., Park, J., Lee, K.Y., Kim, M., Shin, B.A. and Choi, S. (2011) High cleavage efficiency of a 2A peptide derived from porcine teschovirus-1 in human cell lines, zebrafish and mice. *PloS One,* **6,** e18556.

18. Donnelly, M.L., Hughes, L.E., Luke, G., Mendoza, H., ten Dam, E., Gani, D. and Ryan, M.D. (2001) The 'cleavage'activities of foot-and-mouth disease virus 2A site-directed mutants and naturally occurring '2A-like'sequences. *Journal of General Virology,* **82,** 1027-1041.

19. de Felipe, P., Luke, G.A., Brown, J.D. and Ryan, M.D. (2010) Inhibition of 2A-mediated 'cleavage'of certain artificial polyproteins bearing N-terminal signal sequences. *Biotechnology Journal,* **5,** 213-223.

20. Chng, J., Wang, T., Nian, R., Lau, A., Hoi, K.M., Ho, S.C., Gagnon, P., Bi, X. and Yang, Y. (2015) Cleavage efficient 2A peptides for high level monoclonal antibody expression in CHO cells. *MAbs,* **7,** 403-412.

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

22. Desmet, F., Hamroun, D., Lalande, M., Collod-Béroud, G., Claustres, M. and Béroud, C. (2009) Human splicing finder: An online bioinformatics tool to predict splicing signals. *Nucleic Acids Research,* **37,** e67-e67.

23. Yang, Y., Fan, X., Mao, M., Song, X., Wu, P., Zhang, Y., Jin, Y., Yang, Y., Chen, L. and Wang, Y. (2017) Extensive translation of circular RNAs driven by N6-methyladenosine. *Cell Res*earch, 27(5), p.626.

24. Suzuki, H., Zuo, Y., Wang, J., Zhang, M.Q., Malhotra, A. and Mayeda, A. (2006) Characterization of RNase R-digested cellular RNA source that consists of lariat and circular RNAs from pre-mRNA splicing. *Nucleic Acids Research,* **34,** e63-e63.

25. Costello, A., Lao, N.T., Gallagher, C., Roca, B.C., Julius, L.A., Suda, S., Ducrée, J., King, D., Wagner, R., Barron, N., 2018. Leaky Expression of the TET-On System Hinders Control of Endogenous miRNA Abundance. *Biotechnology Journal,* 1800219.

26. Starke, S., Jost, I., Rossbach, O., Schneider, T., Schreiner, S., Hung, L. and Bindereif, A. (2015) Exon circularization requires canonical splice signals. *Cell Reports,* **10,** 103-111.

27. Lewis, N.E., Liu, X., Li, Y., Nagarajan, H., Yerganian, G., O'Brien, E., Bordbar, A., Roth, A.M., Rosenbloom, J., Bian, C., et al. (2013) Genomic landscapes of chinese hamster ovary cell lines as revealed by the cricetulus griseus draft genome. *Nature Biotechnology,* **31,** 759-765.

28. Gorochowski, T.E., Ellis, T., 2018. Designing efficient translation. *Nature Biotechnology* 36, 934–935. https://doi.org/10.1038/nbt.4257

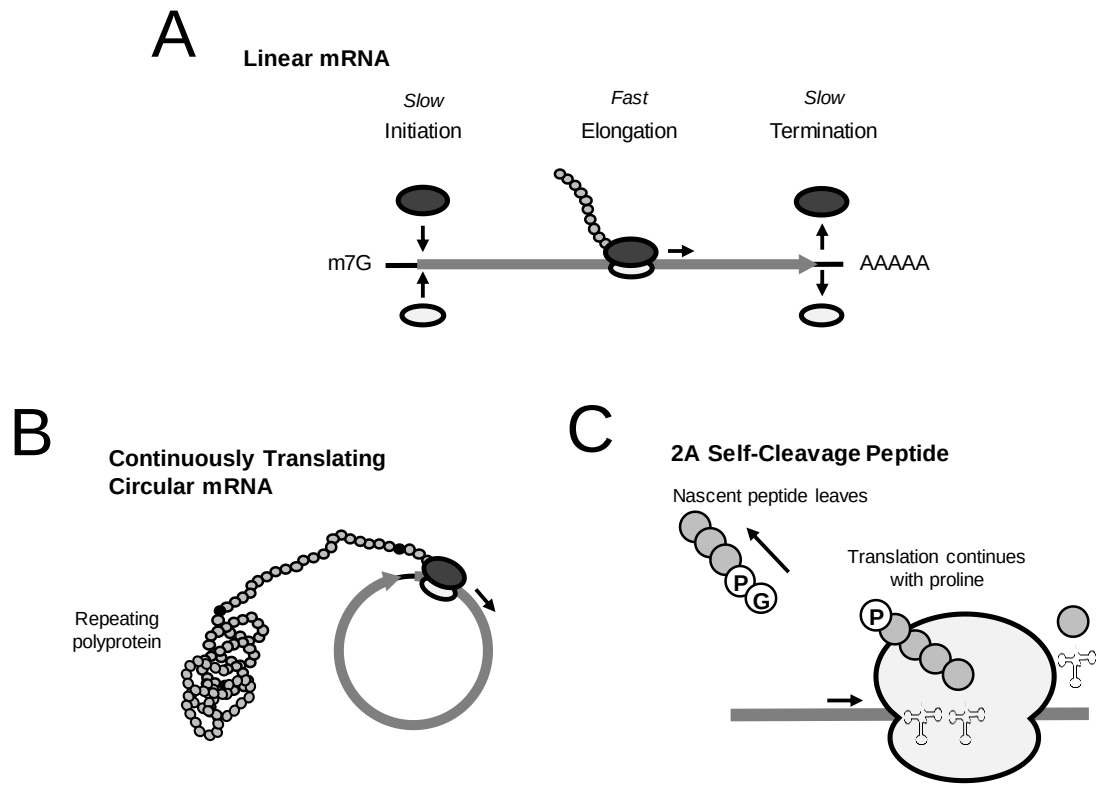


Figure 1. Co-translational cleavage of polyprotein chains

(A) Translation of linear mRNA is summarized in three steps; initiation, elongation and termination. Both the initiation and termination of eukaryotic translation are slow with respect to the elongation step. (B) Continuously translating circular mRNA can overcome the rate limiting step of translational termination by continual production of a repeating polyprotein. (C) Inclusion of a 2A self-cleavage motif in a continuously circular mRNA may enable the removal of translational termination while producing homogenous protein.

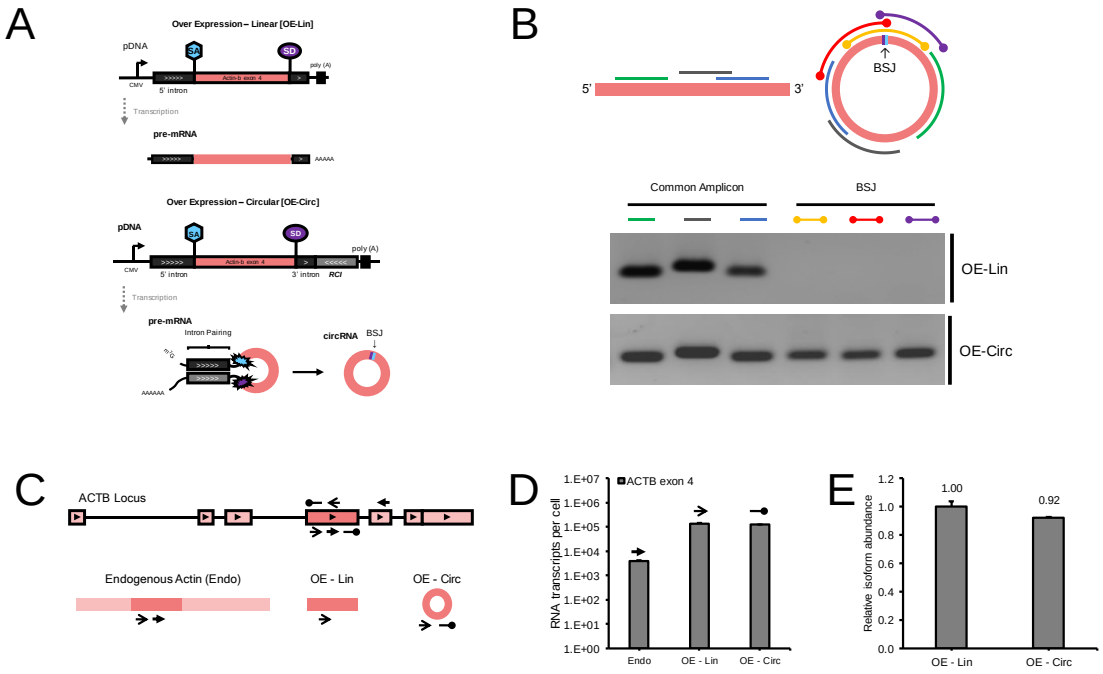


Figure 2. Efficient exon circularization with complementary flanking introns

(A) Schematic representation of expression vector design for exon circularization. The ACTB exon 4 with endogenous flanking introns, splice acceptor (SA) and slice donor (SD) were cloned downstream of a CMV promoter, this served as a linear control (OE-Lin). To facilitate exon circularization (OE-Circ) the 5' intron sequence was inserted in reverse orientation downstream of the 3' intron to create a reverse complementary intron (RCI). Upon transcription structed intron pairing brings splice signals in close proximity generating a back-splice junction (BSJ) in a circRNA. (B) Exon circularization was evaluated by multiple primer sets for both linear and circular transcript isoforms. (C) Three primer sets were used to determine the absolute quantification of exogenous ACTB exon 4 isoforms; closed arrow (endogenous), open arrow (exogenous linear) and oval arrow (exogenous circular). Endogenous CHO-K1 actin was discriminated from exogenous ACTB exon 4 by using a primer set including a (D) Absolute transcript levels for endogenous, exogenous linear and exogenous circular ACTB exon 4. (E) Relative exogenous isoform abundance.

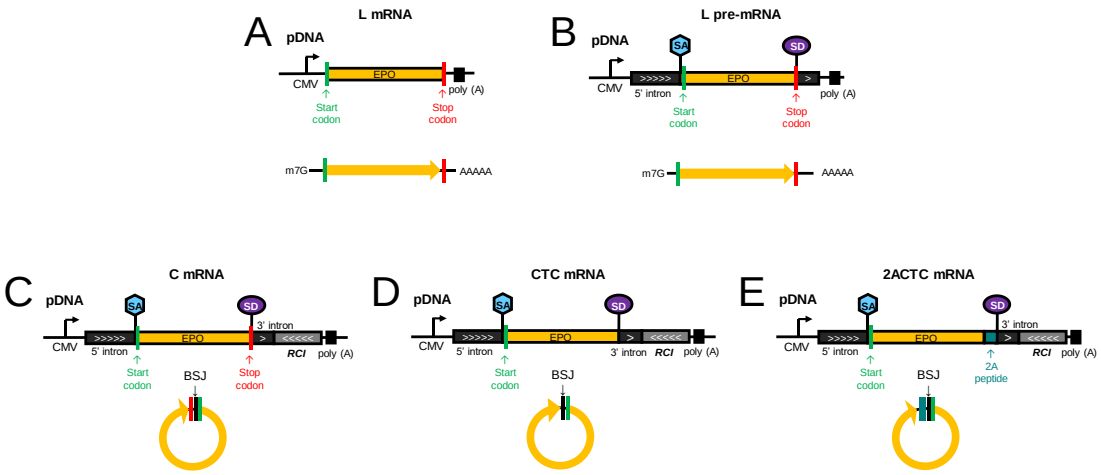


Figure 3. EPO expression from linear, circular and continuously translating circular mRNAs

(A) EPO open reading frame contains transcriptional start site (TSS) and stop codon. Transcript exists as a linear mRNA (L mRNA) with poly adenylation tail and 5' cap. (B) Un-structed 5' intron and flanking splice signals express a linear pre-mRNA (L pre-mRNA). (C) Circular mRNA (C mRNA) expression vector contains reverse complementary intron (RCI). C mRNA contains a TSS and stop codon as linear and linear pre-mRNAs. (D) Continuously translating circular mRNA (CTC mRNA) is created by removal of the stop codon from the EPO open reading frame. (E) T2A self-cleavage peptide motif encoded in a CTC mRNA (2ACTC mRNA).

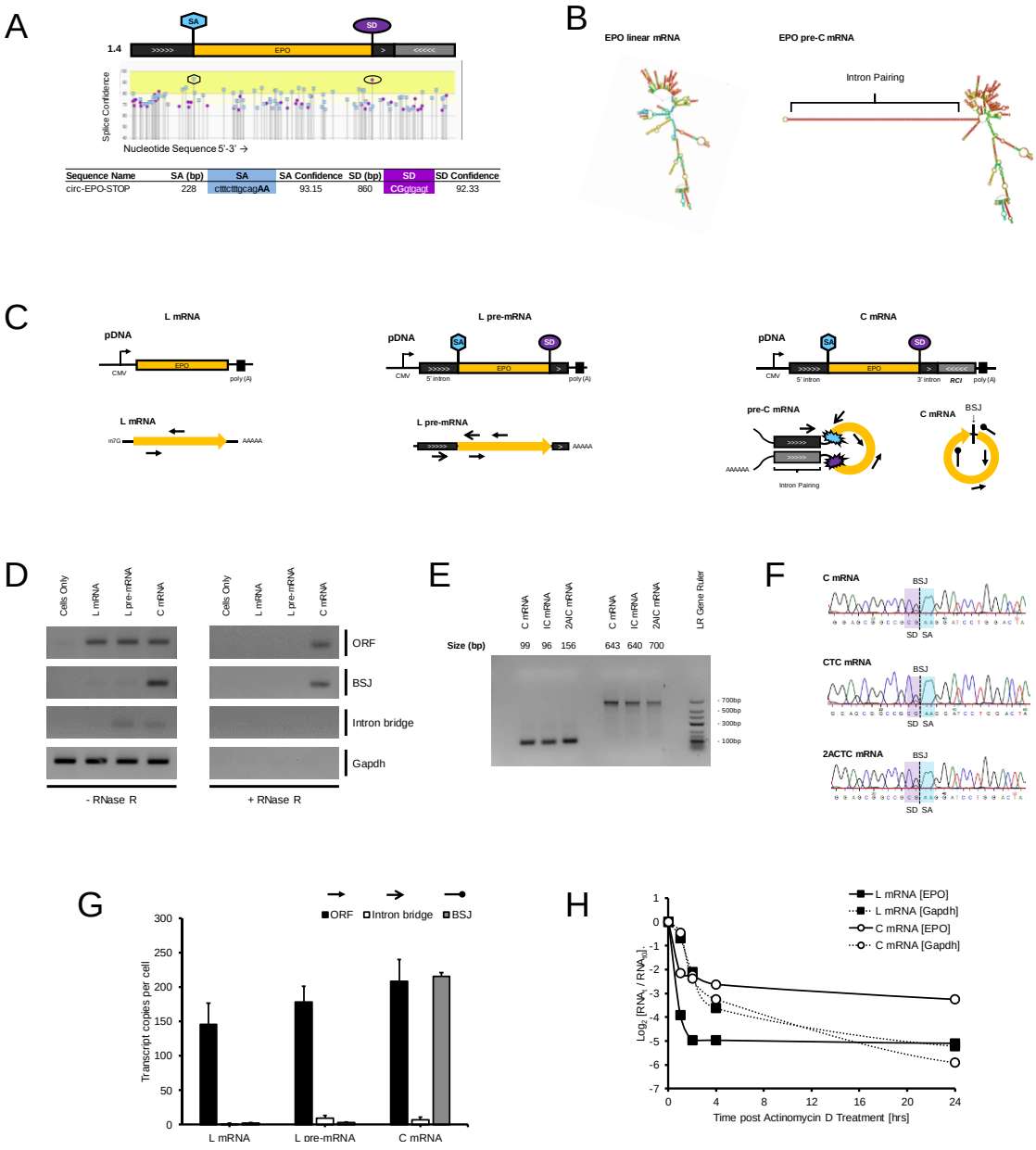


Figure 4. Validation of EPO RNA circularization

(A) Human splice site finder 3.0 (HSF) (http://www.umd.be/HSF3/HSF.html) analysis of EPO ORF and flanking intron sequence. A vector map of the structured intron flanked EPO open reading frame (ORF) is aligned to the splice signal analysis. Intron sequence is depicted as (Black/ grey box) with arrows showing sequence direction. Slice acceptor (SA blue) and splice donor (SD purple) motifs predicted by HSF software. (B) RNA secondary structure analysis of linear (left) and structured intron flanked (right) for the EPO ORF. (C) Discrimination of circular RNA from linear RNA is achieved by the utilization of multiple PCR primer sets. A common primer set amplifying the EPO ORF (bold arrow). Amplification of intron-exon bridge only present in the pre-mRNA expressed by the L pre-mRNA control and the C mRNA expression

vector (open arrow). A third primer set (oval arrow) amplifies the unique back-splice junction only found in the presence of C mRNA. (D) Semi-quantitative PCR analysis of EPO mRNA transcript variants using three primer sets. RNase R treatment was used to remove all linear RNA from total RNA prior to cDNA synthesis. Gapdh was used as a control to ensure complete digestion of linear RNA. (E) The back-splice junction (BSJ) of each construct was amplified. To provide template for sequencing the entire circular RNA molecule in each case was also amplified. (F) Sequencing of the BSJ derived from each construct showed an identical splice junction in all cases. (G) Absolute quantification of transgene RNA transcript variants. The absolute levels of transcripts were measured using standard curves for each specific amplicon. (H) RNA stability for EPO L mRNA and C mRNA following Actinomycin D treatment. Gapdh mRNA was also evaluated for cells transfected with either L mRNA or C mRNA expressing vectors. The rate of RNA decay was calculated by the Log_2 [RNA_t/ RNA_{t0}].

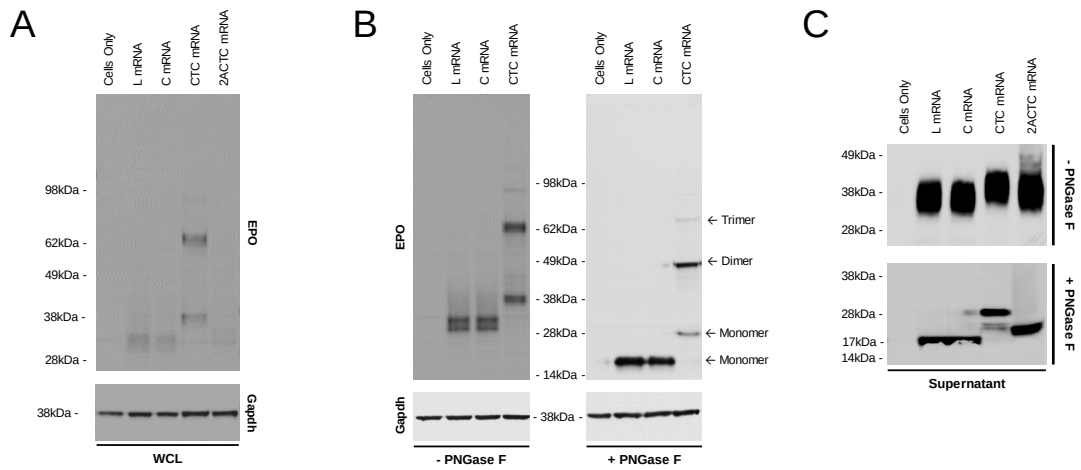


Figure 5. Rolling circle translation of EPO produces poly-proteins in CHO cells

(A) CHO-K1 cells were transiently transfected for 48hrs with EPO expression vectors encoding one of L mRNA, C mRNA CTC mRNA or 2ACTC mRNA transcript isoforms. Whole cell lysates (WCL) were resolved in an SDS-PAGE gel. (B) CHO-K1 lysates treated or mock treated with PNGase F to remove N-linked glycans. (C) Concentrated supernatant of CHO-K1 cells transiently transfected for 48hrs with linear and circular EPO constructs. Concentrated supernatants were also treated with PNGase F to remove N-linked glycans.

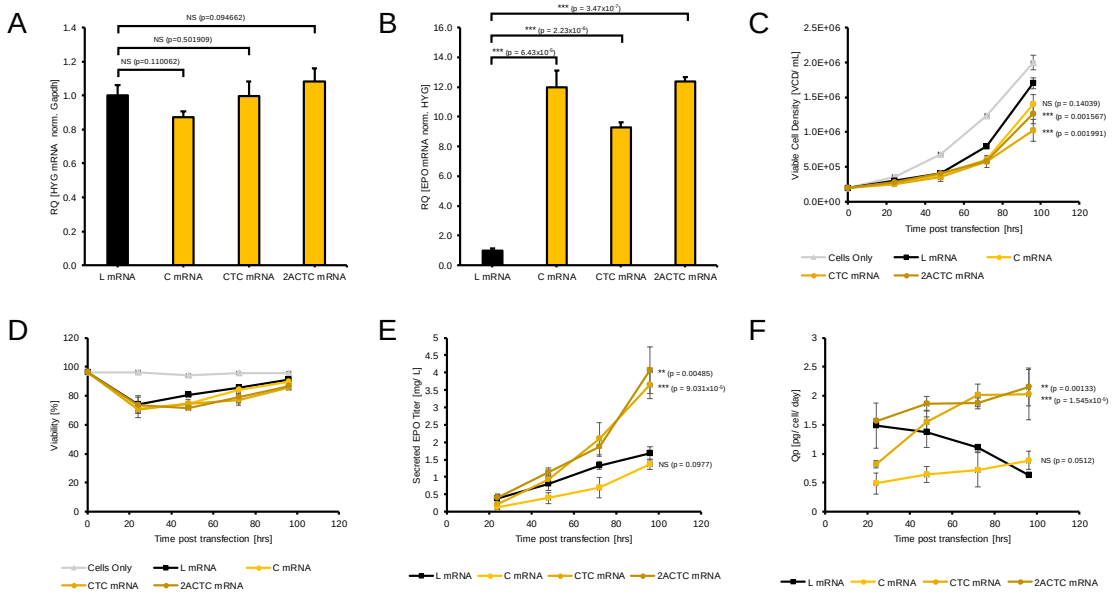
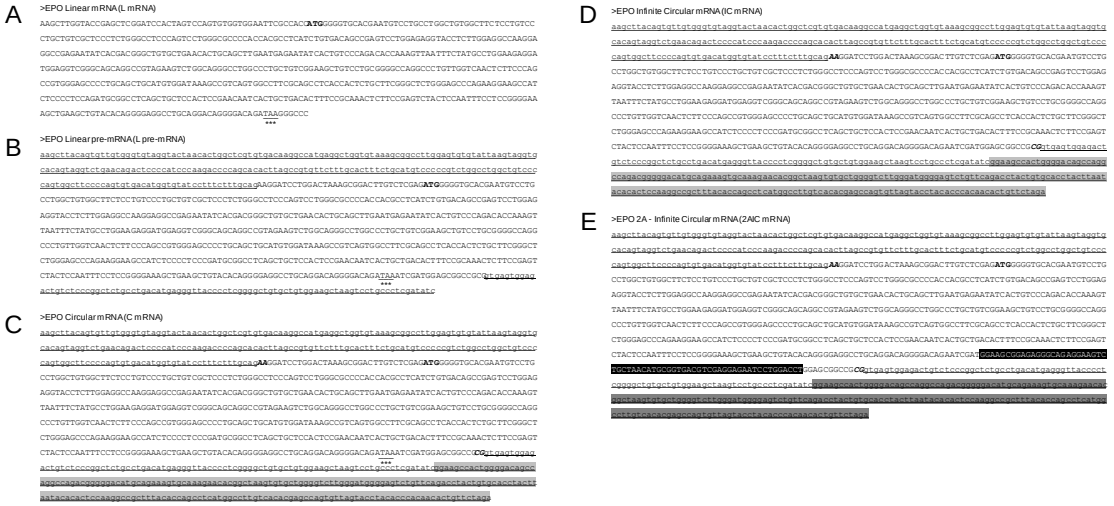


Figure 6. Evaluation of circular EPO constructs in transient batch culture

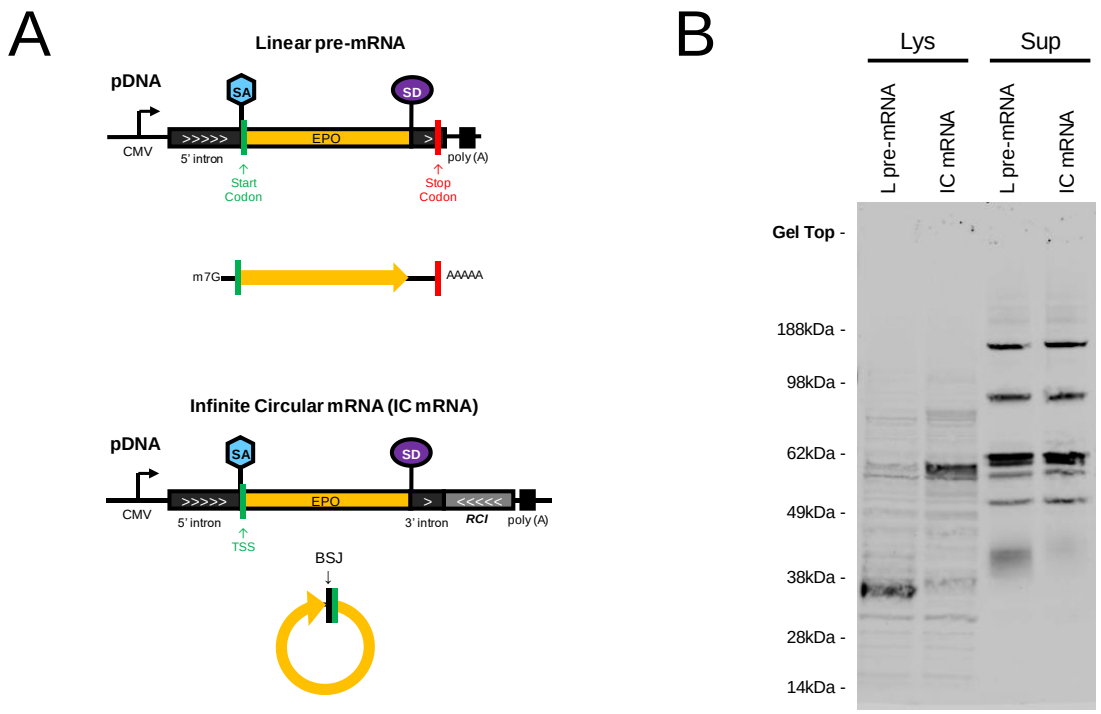
CHO-K1 cells were transiently transfected with EPO expression vectors encoding one of L mRNA, C mRNA CTC mRNA or 2ACTC mRNA transcript isoforms in 50 mL cultures. (A) The vectors used in this study also drive the expression of a resistance gene Hygromycin-B-phosphotransferase (HYG) under a PKG promoter. Relative quantification (RQ) of HYG mRNA was calculated using the ddCt method with respect to endogenous Gapdh 48hrs post transfection. (B) RQ of EPO mRNA was calculated using the ddCt method with respect to HYG mRNA 48hrs post transfection. (C) Viable cell density [VCD/ mL] was monitored at 24hr intervals. (D) Viability (%) of cultures following transfection measured at 24hr intervals. (E) Secreted EPO titer was quantified by ELISA of culture supernatant harvested every 24hrs. (F) Cell specific productivity (Qp) (pg/ cell/ day) was calculated for each timepoint of analysis. Values are based on biological replicates, (n=3). Statistical analysis consisted of a two-tail homoscedastic student t-test, ($p \leq 0.05$*, $p \leq 0.01$**, $p \leq 0.001$***).

5.6 APPENDIX C

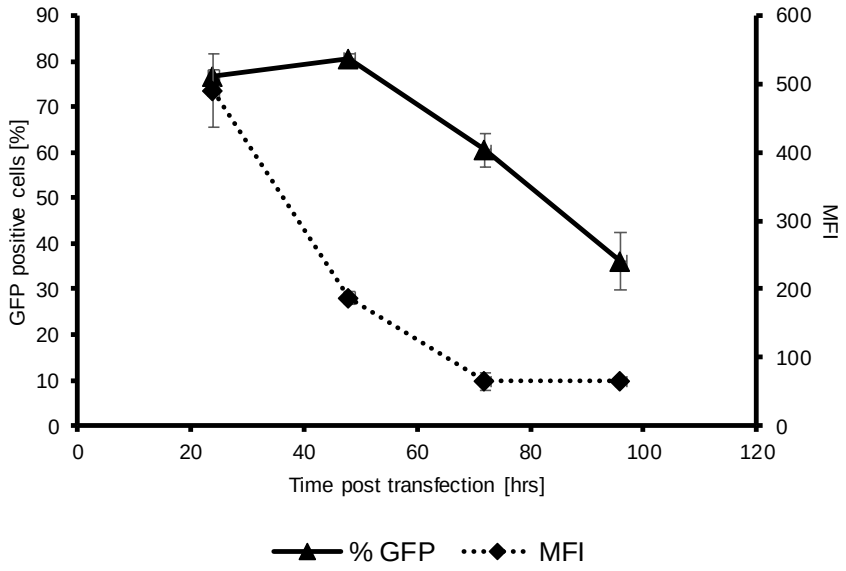
Supporting information associated with Chapter 5.



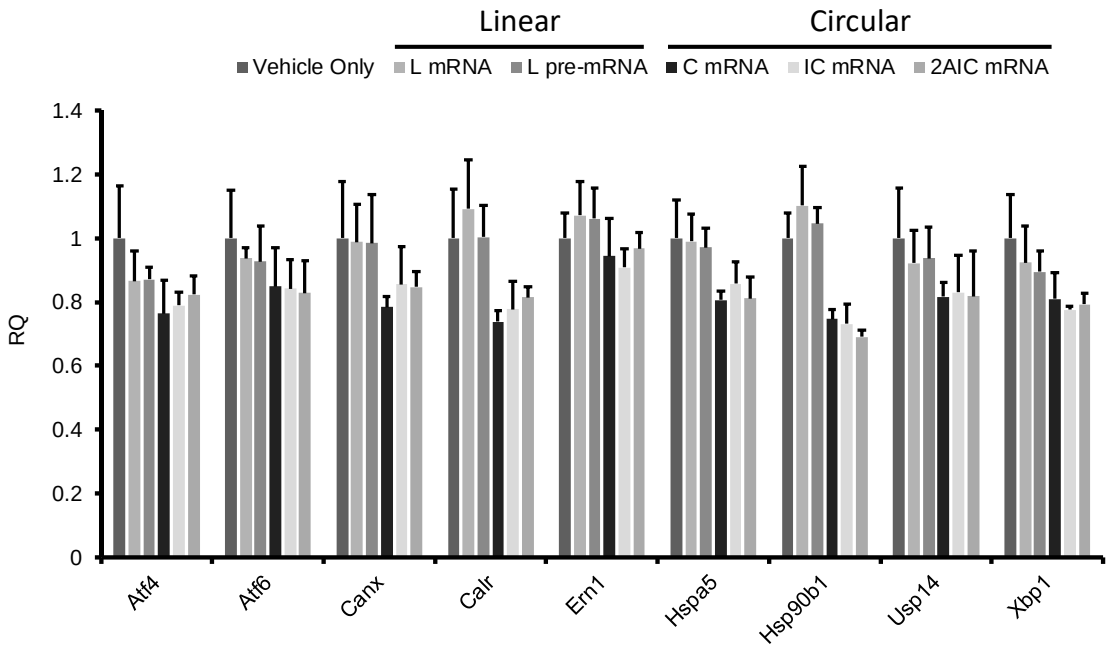
Supplementary Figure 1. Sequence Information of linear and circular RNA constructs. The sequence of EPO expression constructs used in this study; linear mRNA (A), linear pre-mRNA (B), circular mRNA (C), continuously translating circular mRNA (D) and 2A circular mRNA (E). Start codons are in (bold), stop codons underlined with (***), intronic sequence is underlined and in lower case, splice signals are in (bold italics), reverse complementary introns are highlighted in (grey) and 2A motif highlighted in (black, white lettering).



Supplementary Figure 2. Confirmation of RCT with linear pre-mRNA control. (A) To confirm the high molecular weight band seen in the cell lysates of IC mRNA expressing cells was the product of rolling circle translation and not read-through of the down-stream intron a linear pre-mRNA control lacking the normal stop codon was constructed. The downstream intron contains a stop codon. (B) Western blot analysis of cells lysates and supernatants from CHO-K1 transfected for 16hrs.



Supplementary Figure 3. Evaluation of transfection efficiency. A GFP reporter in the same pcDNA 3.1 (+) HYG backbone vector was used to evaluate transient transfection efficiency of CHO-K1 cells. The % of total cells GFP positive is depicted by the solid black line and the corresponding median fluorescence intensity by the dashed line.



Supplementary Figure 4. Evaluation of un-folded protein response. Relative quantification of nine genes; Atf4, Atf6, Canx, Calr, Ern1, Hspa5, Hsp90b1, Usp14, Xbp1, associated with un-folded protein response was evaluated for linear and circular RNA constructs with respect to a vehicle only control.

Supplementary Table 1. Sequence information of primers used in this study.

Primer Name	Sequence (5'-3')	Tm	Amplicon (bp)
LinEPO-ORF-FP	ACTTGTCTCGAGATGGGGGT	60	140
LinEPO-ORF-RP	AGGTACCTCTCCAGGACTCG	60	140
circEPO-FP	AGCTGAAGCTGTACACAGGG	59	99 (+ T2A 156)
circEPO-RP	TCTCGAGACAAGTCCGCTTT	59	99 (+T2A 156)
LinEPO-EB-FP	AGGTCTGAACAGACTCCCCA	60	171
LinEPO-EB-RP	CATCTCGAGACAAGTCCGCT	60	171
Act-Circ Trav FP (1)	ATTGGCAATGAGCGGTTC	60.1	112
Act-Circ Trav RP (1)	TAGCACAGCCTGGATAGCAA	60	112
Act-Circ FP SS (2)	AGCCTTCCTTCCTGGATCAT	60	101
Act-Circ RP (2)	CCAGTGGTACGGCCAGAG	60	101
Act-Circ FP (3)	AGAGCTACGAGCTGCCTGAC	60	100
Act-Circ RP SS (3)	CAAACATGATCCAGGAAGGAA	60	100
Act-lin FP (1)	GCCATGTACGTTGCTATCCA	60	102
Act-lin RP (1)	GGGCACAGTGTGGGTGAC	60	102
Act-lin FP (2)	CGGACTTCGAGCAAGAGATG	59	138
Act-lin RP (2)	AGGAAGGAAGGCTGGAAGAG	59	138
Act-lin FP (3)	CGGGACCTGACTGACTACCT	60	109
Act-lin RP (3)	GACCCCACAGCTTCTCCTTA	60	109
cgr-ActB (endo) F	TGTCACCAACTGGGACGATA	60	165
cgr-ActB (endo) R	GGGGTGTTGAAGGTCTCAAA	60	165
cgr-Gapdh F	TGGCTACAGCAACAGAGTGG	60	162
cgr-Gapdh R	GTGAGGGAGATGATCGGTGT	60	162
Usp14 FP	GAGCTTCAGGGGAAATGGCT	60	193
Usp14 RP	ACCCAGCACTCGTTAGCATC	60.1	193
Xbp1 FP	CTCCAGAGACGGAGTCCAAG	60	181
Xbp1 FP	ACAGGGTCCAACTTGTCCAG	60	181
Atf4 FP	TTGCAACCTTTTCCCTGTTC	60.1	167
Atf4 RP	TCTGAGGGGGTGTCTTCATC	60	167
Ern1 FP	CTGGAGAGATTCCCCAACAA	60	161
Ern1 RP	GCATGGACCAGCTTCTCTTC	60	161
Hspa5 FP	GTGCAGAAACTTCGTCGTGA	60	162
Hspa5 RP	GGTAGATCGGAACAGGTCCA	59.9	162
ATF6 FP	GCCACTGAAGGAAGACAAGC	60	173
ATF6 RP	TGATGGTTTTTGCTGGAACA	60.1	173
Canx FP	TCTGGCAGCGACCTTTGATT	60	182
Canx RP	GGACCAGAGCTCCAAACCAA	59.9	182
Calr FP	CTCTGGCAGGTCAAGTCTGG	60	199
Calr RP	CCTCAGCCTCTTCCTCCTCT	60	199
Hsp90b1 FP	ACCGGGAAGCAACAGAGAAG	60	105
Hsp90b1 RP	CCGTGAGACGCTGAGATACC	60	105
HYG FP	ATGCTCCGCATTGGTCTTGA	60	131
HYG RP	ATTTGTGTACGCCCGACAGT	60	131

Chapter 6

Sentencing Ribosomes to Tediously Long Messages

Costello, A., Barron, N. * & Clynes, M. *
* authors contributed equally
Manuscript in draft, submitting to *Trends in Biotechnology*.

For this chapter I was the primary author with guidance from Prof. Martin Clynes and Prof. Niall Barron.

ABSTRACT

Circular RNA (circRNA) is a covalently closed loop RNA species. Formed naturally through non-colinear splicing of pre-mRNA and once designated as non-coding (ncRNA) artefacts of splicing errors. It is now accepted that circRNAs are abundant with diverse functions in gene regulation, transcriptional activation and protein coding in eukaryotes. Numerous reports have evaluated the relevance of circRNA in various diseases, yet the potential these molecules have in the production of recombinant protein therapeutics is only now coming into focus. In this review we will highlight applications of circRNA for engineering transgenes and mammalian cell expression systems. Describing strategies for stable and transient or stable exogenous circRNA expression. Discovery of endogenous circRNA targets for genetic engineering and the potential for recombinant protein production from infinite circular open reading frames.

Key words: Circular RNA, Biotechnology, Synthetic Biology, miRNA, 2A, Rolling circle translation

6.1 INTRODUCTION

Originally discovered in the early 1990's (Nigro et al., 1991; Cocquerelle et al., 1993), eukaryotic circular RNA (circRNA) was believed to be nothing more than artefacts of splicing error as very few transcripts were identified. However, in the last 5 years next-generation RNA-sequencing (RNA-seq) has revealed ~10% of expressed genes produce circRNA splice variants (Jeck et al., 2013; Guo et al., 2014; Kelly et al., 2015) with independent studies demonstrating circRNA isoforms exceeding that of associated linear RNA (Burd et al., 2010; Salzman et al., 2012). These RNA transcripts are distinct from other RNAs such as linear mRNA which typically harbours a 5' cap and 3' poly-adenylation (A) tail. circRNAs can be composed of exons (Hansen et al., 2013), introns (Talhouarne and Gall., 2014) or both (Liang & Wilusz, 2015) and are highly stable compared to their linear counterparts (Jeck et al., 2013; Wang and Wang, 2015). The classification of this RNA species is difficult as functional circRNAs can be non-coding (Hansen et al., 2013; Memzcak et al., 2013) or coding circular mRNA (Abe et al., 2015; Wang and Wang, 2015; Pamudurti et al., 2017; Yang et al., 2017). The breadth of function could make these molecules desirable novel targets for engineering mammalian cell expression systems.

Heterologous production of recombinant protein therapeutics by mammalian or bacterial host cells have a common underlining issue. The rare cells that produce desirable high titres are at a competitive disadvantage in terms of growth resulting from their obedience. Dramatic titre improvements in mammalian expression hosts has come largely from process intensification and optimization of chemically defined media. Through this scientists and engineers aim to help cells by facilitating them with optimal environmental conditions. With publication of the Chinese hamster and CHO-K1 genomes (Xu et al., 2011; Lewis et al, 2013; Brinkrolf et al., 2013) our understanding of the systems biology of the dominant mammalian expression host has improved. As our understanding of cell systems evolves, we can stop "helping" the cells and force them perform better. For a comprehensive review of CHO cell engineering efforts, see (Fischer et al, 2015).

Synthetic biology is the most prudent means of engineering cells to control their behaviour (Valdés-Bango Curell and Barron, 2018). Eukaryotic circRNA is still largely understudied and ill-defined. However, desirable attributes of these molecules can be applied to form synthetic circRNA molecules and potentially enhance cellular productivity. In this review we will discuss what is currently know of endogenous circRNA biogenesis, functions and identification. Outline exogenous circRNA expression by synthetically mimicking a natural biogenesis pathway. In addition, we discuss the emerging strategies for host cell and transgene engineering applications of these molecules. Covering the regulatory roles of circRNAs both transcriptionally and post-transcriptionally. Finally, we discuss the successes in exogenous expression of recombinant proteins from circular mRNAs.

6.2 CIRCULAR RNA EXPRESSION

Eukaryotic circRNAs are formed during pre-mRNA splicing and are dependent on spliceosomal machinery (Ashwal-Fluss et al., 2014; Starke et al., 2015). circRNA biogenesis competes with canonical co-linear pre-mRNA splicing (Figure 1A). The exact mechanism of splicing is not known but it has been reported that the "back-splicing" (Figure 1A) required to form circRNA is much less efficient than canonical splicing (Zhang et al., 2014; Ashwal-Fluss et al., 2014; Strake et al., 2015). circRNA formation may be facilitated by *cis* (Ashwal-Fluss et al., 2014; Starke et al., 2015; Liang et al., 2014) or *trans*-elements (Ashwal-Fluss et al., 2014; Conn et al., 2015; Kramer et al., 2015; Ivanov et al., 2015). Exon circularization is associated with flanking complementary intron sequence (Figure 1B) bringing non-colinear splice sites close in proximity. Roughly 38 and 9% of circRNAs identified in C. elengans and human respectfully are formed by the presences of complementary flanking intronic regions (Ivanov et al, 2015). An enrichment of Alu repeats in the flanking introns of exons known to circularize has been reported (Jeck et al, 2013; Zhang et al, 2014). However, their relevance in the circularization of exons is not deemed essential given the identification of circRNA lacking Alu elements (Westholm et al, 2014). Furthermore, circularization albeit minimal, with no flanking intron complement sequence has been observed (Li et al., 2014). The splicing factors Quaking and Musclebind can induce circRNA formation form exons with intronic flanks containing binding motifs to the respective splicing factors (Conn et al, 2015; Ashwal-Fluss et al, 2014). In this case the RNA binding proteins (RBP) interaction bring the exon terminal splice signals close together (Figure 1C). Again, this pathway suggests the necessity of proximity between the splice motifs. However here the complementary repeat sequences are made redundant by the presence of specific RBP motifs.

Thousands of circRNAs can be produced during pre-mRNA splicing of protein coding genes but why and how this occurs is still largely unknown. The likelihood of circRNA expression being an artefact of splicing error however has been rebutted with the findings of Salzman et al., (2013). No correlation between circRNA abundance and expression of associated linear transcripts was found, strongly indicating regulation of these splice variants. circRNA from hundreds of human genes exist as the predominant transcript isoform (Salzman et al., 2012) backs this theory. The increased lifespan of circRNA transcript variants over associated linear RNAs makes it difficult to draw conclusions from this. circRNA expression has been reported as cell type (Salzman et al., 2013; Conn et al., 2015; Strake et al., 2015) and stage (Okholm et al., 2017; Ouyang et al., 2017) specific. Why this occurs is not understood but a report from (Liang et al., 2017) found limiting the pre-mRNA splicing machinery shifts transcripts of protein coding genes to circRNA variants. Stating that variable readthrough of poly(A) signals by RNA polymerase II, as found in cancer progression (Grosso et al., 2015; Kannan et al., 2011; Maher et al., 2009), corelated to circRNA formation. The debate on whether circRNA expression is regulated or results from error

in transcriptional machinery remains a grey area. While differential circRNA formation in disease states and stages would support the theory of splicing error. The tissue specific (Salzman et al., 2012) and dynamism of expression during embryonic muscle development (Ouyang et al., 2017) suggest the contrary. What is known is that these molecules are ubiquitous and that their expression is volatile. Pertaining to mammalian cell engineering, this is a positive as differential expression of circRNA may provide novel targets for cell engineering.

The abundance of eukaryotic circRNAs has remained obscured until recently as probe-based sequencing platforms and enrichment of poly(A) transcripts has blinkered the identification of non-canonical splice variants. It is now accepted that circRNA is abundant in model organisms such as human, mouse, insects and plants (Jeck et al., 2013; Hansen et al., 2013; Lu et al., 2015; Ye et al., 2015; Westholm et al., 2014). While making them interesting genetic engineering targets, the diversity of circRNA structure can make them difficult to identify. circRNAs can arise from both sense and antisense of strands or intergenic regions (Salzman et al., 2013; Hansen et al., 2013; Gao et al., 2015). Discerning circRNA from mis-spliced concatemeric linear RNA is a challenge. Gao & Zhao, (2018) reviewed the current crop of circRNA prediction pipelines for RNA-seq datasets. By re-analysing existing RNA-seq data it could be possible to find novel abundant endogenous circRNAs from differential expression studies. For enrichment of circRNA from total RNA, RNase R has been used (Memzcak et al., 2013). RNase R degrades all linear RNA but not circRNA (Suzuki et al., 2006), this strategy enables identification of even very low abundant circRNAs (Jeck et al., 2013). RNase R has also been widely used to validate circularity of exogenous circRNA expression.

There is no consensus mechanism for endogenous expression of circRNA biogenesis. It is known that circularization of exon sequence requires canonical splice signals (Starke et al., 2015) and that flanking the sequence of interest with complementary introns dramatically improves back-splicing efficiency (Liang & Wilusz, 2014). Through pure serendipity, studies evaluating the functional effects of natural circRNAs (Hansen et al, 2013; Liang & Wilusz, 2014) have established criteria for the artificial circularization of any sequence in mammalian cells. Transgene sequence can be circularized by bracketing the sequence of interest with a splice acceptor (SA) and splice donor (SD) (Figure 2A). This cassette is the further fringed with intronic sequence both 5’ and 3’. Downstream of the 3’ intron, is a sequence of reverse complementarity to the 5’ intron. Upon transcription, the pre-mRNA generated sees hybridization of intron complements forming a large hairpin-like structure. This is an exogenous mechanism to mimic the natural biogenesis of circRNA by intron pairing. The splice signals made proximal by the intron pairing facilitates a BSJ producing nascent circRNA. Where introduction of plasmid DNA is not desirable the circularization of synthesized ssRNA made be done with T4 RNA ligase (Abe et al., 2015) (Figure 2B).

6.3 CIRCRNAS AND GENE REGULATION; RNA LASSOES?

The first reports of circRNA function in human and mouse (Hansen et al., 2011; Hansen et al., 2013; Memzcak et al., 2013) described miRNA sponge activity. Direct Argonaut II (AGO II) mediated cleavage of ae circRNA from the human CDR1 locus via binding of the miR-671 was found to regulate the levels of the CDR1 protein (Hansen et al., 2011). This was the first report of post-transcriptional regulation of a gene by an antisense transcript. It was later reported that this circRNA referred to as "ciRS-7" harboured >70 miRNA binding sites for a second miRNA, miR-7 (Hansen et al., 2013) (Figure 3A). A second circRNA sponge was found to come from the mouse Sry gene with regulation of miR-138 levels (Memzcak et al., 2013). Since these early discoveries many more naturally occurring circRNAs have been found to have miRNA sponge activity (Huang et al., 2015; Zheng et al., 2016; Hsiao et al., 2017; Zheng et al., 2017; Zhang et al., 2017; Chen et al., 2017; Han et al., 2017; Zhu et al., 2017; Jin et al., 2017). With so many of these circRNA sponges existing naturally it begs the question, is there an advantage of expressing a competing endogenous (ceRNA) as a circRNA over a linear ncRNA molecule?

Stable constitutive suppression of specific endogenous miRNA function in Chinese hamster ovary (CHO) cells has been reported previously as a means of improving industrially relevant cell behaviours (Sanchez et al., 2013; Kelly et al., 2015; Emmerling et al., 2015). In all cases this was achieved by introduction of a synthetic un-translated region (UTR) 3' of a reporter gene encoding miRNA binding sites for specific endogenous targets. As cells appear to express sponge decoys as circRNA, it could be interesting to elucidate any advantages of expressing synthetic miRNA sponge decoys as circRNAs. This could be achieved as schematically depicted in (Figure 2) with the "gene of interest" encoding a miRNA sponge decoy.

A class of nuclear localized exonic, intron retaining circRNA named (EIciRNA) was reported to function in *cis* where EIciRNA splice variants enhance the transcription of their parental genes (Li et al., 2015) (Figure 3B). Candidates were found by RNA-seq for circRNAs enriched with cross-linked immunoprecipitation of RNA polymerase II. Knockdown of EIciRNAs, one from the EIF3J gene and the other from PAIP2 by both siRNA and antisense oligonucleotides (ASO) depleted both the circRNAs and associated linear mRNA transcripts. Forced expression of circEIF3J and circPAIP2 had no effect on the transcription of the endogenous associated genes, suggesting that EIciRNA works in *cis* but not in *trans*. Transcriptional regulation by ncRNA is a fundamental aspect of gene expression (Goodrich & Kugel, 2006; Geisler & Coller, 2013) but cis acting ncRNAs that promote transcription of associated linear mRNA either post or co-transcriptionally is novel. Control of these ncRNA transcriptional activators by siRNA is very interesting as it pertains to the control of gene expression. The nature of miRNA expression in mammalian cell systems is dynamic (Hernandez Bort et al., 2012). Could synthetic EIciRNAs encoded with binding sites of miRNAs known to fluctuate in abundance with response to

bioprocess-controlled variables be employed as new ncRNA mediated inducible gene expression systems? It is perhaps too early to suggest such a thing but the thought of controlled manipulation of gene expression without introduction of recombinant protein trans-activators is tantalizing.

6.4 TRANSLATION OF CIRCULAR MRNA; FINDING THE START OF A CIRCLE

In 1979, Marilyn Kozak, of consensus sequence fame, proposed a theory that eukaryotic translation initiates by ribosomes threading an mRNA like beads on a string (Kozak, 1979). Circularization of a message would abolish its ability to interact with a ribosome should this threading mechanism hold through and would explain the mono-cistronic character of eukaryotic mRNAs. It has since been shown that eukaryotic translation initiation occurs by two distinct mechanisms either cap-dependent or cap-independent via an internal ribosome entry site (IRES) (Stoneley & Willis, 2004). The canonical model of eukaryotic translation initiation describes a process of assembly of the 48S ribosome complex in the 5' UTR of an mRNA (Figure 4A). The process is highly regulated, involving at least nine eukaryotic initiation factors (Jackson et al., 2010) prior to 48S recruitment and scanning. Culminating in the formation of the elongation component 80S ribosome complex when a Met-tRNA anticodon loop is paired to the P-site of a ribosome 40S subunit. The umbrella term of "IRES-mediated translation initiation" is less well defined due to the lack of consensus RNA-protein binding motifs between viral and eukaryotic IRES variants. Viral IRES sequences are large ranging from ~200-450nt in length (Jackson et al., 2010). Picornavirus IRES sites from poliovirus and encephalomyocarditis virus (Pestrova et al., 1996; de Breyne et al., 2009) facilitate translation initiation by direct interaction with eIF4G in the absence of the cap-dependent factor eIF4E (Figure 4B).

The first evidence of eukaryotic translation of circular RNA containing a viral IRES site was observed in the mid-1990s but proved less efficient than a linear RNA control in rabbit leucocyte lysate and the absence of an IRES in the circular RNA resulted in undetectable translated products (Chen & Sarnow, 1995). A recent study by Yang et al., (2017), demonstrated the ability of circular RNA translation in the absence of a viral or eukaryotic IRES. The un-expected translation of "negative control" circular RNAs with no deliberate IRES site lead them to identify a common six nucleotide motif. The hexamer closely resembled the most abundant internal RNA modification (Li & Mason, 2014), N^6-methyladenosine (m^6A) (Wei et al., 1975). Replacement of the IRES structures and in absence of a consensus Kozak sequence, a single m^6A motif permitted translation of a reporter gene from a circular mRNA with un-detectable protein in the adenosine-free control. Furthermore, the discovery that m^6A motifs are enriched in endogenous circular RNAs suggests a natural IRES-like function of this motif in eukaryotes (Yang et al., 2017). The presence of a single m^6A motif in the 5'UTR of an mRNA can directly bind the

eukaryotic initiation factor 3 (eIF3), (Figure 4C), which is sufficient to recruit the 43S complex and initiate translation in the absence of the cap-binding factor eIF4E (Meyer et al., 2015). While the presence of an m^6A motif in the 5' UTR of an mRNA significantly improved translation over a non-methylated control in all cases, the position of the m^6A relative to the start codon has an effect also (Meyer et al., 2015). Though cap-independent translation initiation is said to be less efficient in eukaryotes than cap-dependent, this could be the result of a low intracellular concentration of relevant initiation factors (Jackson et al., 2010). Highly elevated levels of eIF4G is associated with many advanced breast cancers and leads to the efficient IRES-mediated translation of p120-catenin and VEGF mRNAs (Silvera et al., 2009). The translation of exogenous circular mRNA could perhaps be greatly enhanced by finding a host with highly abundant relevant cap-independent initiation factors or exogenous up-regulation of such factors in an otherwise suitable host.

6.5 IMPROVING TRANSLATION EFFICIENCY BY MRNA CIRCULARIZATION

Although the initiation of circular mRNA translation, while still largely undefined, differs from that of cap-dependent linear mRNA. The mechanisms by which elongation and translational termination occur should in theory be unchanged. For a circular mRNA molecule comprised of the following; a gene open reading frame with 5' start codon and 3' stop codon, a UTR and some motif in the UTR to recruit translational machinery. In simple terms translation occurs by, recruitment and assembly of the 80S ribosome complex in the UTR upstream of the start codon. The 80S then reads the transcript in a 5' to 3' direction generating a polypeptide chain until terminating at a stop codon (Figure 5A). In 2015, Wang and Wang demonstrated the ability of human cells to produce active GFP from a circular mRNA of such structure (Wang & Wang, 2015). The translation of circular RNA only was comprehensively controlled using a split open reading frame, wherein the transcript read *5'Splice acceptor – C-terminus – IRES – N-terminus -3' Splice Donor*. In this instance, only a circular RNA molecule would permit the translation of an intact open reading frame. The aim of the study was to investigate if translation of such an RNA construct would occur and no investigation into the relative translational efficiencies between linear and circular mRNAs was discussed.

Factors that contribute to an mRNAs translation efficiency include its initiation, codon usage, stability, rate of turnover and finally termination. Linear mRNAs suffer from very short half-lives ranging from minutes to hours (Kaczmerek et al., 2016). A lack free ends, protects circular RNAs from several mechanisms of RNA turnover resulting in extended lifespans over associated linear mRNA (Jeck et al., 2013; Chen et al., 2015; Wang & Wang, 2016; Enuka et al., 2016; Wesselhoeft et al., 2018). Independent studies have shown circular RNAs to survive up to 48hrs

(Jeck et al., 2013, Wang & Wang, 2015). This favourable attribute of increased RNA stability was evaluated for implications in enhancing recombinant protein production, therapeutic and otherwise (Wesselhoeft et al., 2018). Indeed, prolonged lifespan of a circular mRNA over a linear control produced greater quantities of protein for a longer duration. The study demonstrated circularization of up to 5kb with their engineered group 1 catalytic intron-based circularization. Wesselhoeft et al., (2015) also provided a method of purifying circular RNAs by HPLC. This makes it feasible to segregate the circular RNA generation and protein production phases and could prove useful in the bulk production of circular mRNA for prokaryotic cell-free expression platforms.

The closed loop structure of circular RNA offers a unique characteristic over linear mRNA. By encoding a gene open reading frame lacking a stop codon on a circular RNA molecule, the message becomes infinite. Should the nucleotides of the molecule be perfectly divisible by three, the circular mRNA acts like a polycistron repeating the same message indefinitely, yielding large polyprotein chains (Figure 5B). The first investigating of this phenomenon however, was conducted in the late 1990s (Perriman & Ares, 1998). The removal of the terminator sequence from a GFP open reading frame provided a model for producing long repeating polyprotein sequences in *E. coli*. The objectives of this study were two-fold. Firstly, to overcome the fractious nature of expressing multiple repeating amino acid sequences in recombinant systems (Prince et al., 1995) and secondly, a model for studying IRES-mediated cap-independent translation initiation in bacteria (Chen & Sarnow, 1995). Although this system was able to generate multimeric forms of GFP exceeding 300kDa in size, the expression from the circRNAs was poor and generated very heterogenous products with only the monomeric forms being fluorescent. Postulating that the low-level protein expression from circular mRNAs seen by (Perriman & Ares, 1998) was the result of poor translation initiation alone (Abe et al., 2013) looked to solve this issue. By introducing chemically synthesized RNA circularized by T4Rnl1 (Figure 2B) to a cell free *E. coli* lysate "rolling circle translation" (RCT) a principle not unlike rolling circle amplification of DNA (Fire & Xu., 1995), of infinite small repeating circular mRNAs improved translation by >100 – fold (Abe et al., 2013). Again however, the protein was very heterogenous. As a means of proving the expression of protein from circular mRNAs, Abe et al., (2015) expressed repeating FLAG-tag sequences from a synthetically circularized infinite mRNA in cell-free and live human cells. Like the work of (Perriman & Ares, 1998) infinite ORFs in mammalian cells can produce long repeating polyproteins >250kDa. These synthetic circular mRNAs lacked IRES, ploy(A) and cap structures suggesting cap-independent translation initiation does not require any of these (Abe et a., 2015). RTC was used in this case to prove the translation of circular mRNAs in human cells. However, as with the work by (Perriman & Ares,

1998) the spooling polyprotein products were heterogenous and un-predictable. Though the improved translational yield is promising, the products need to be homogenous and functional.

So, is this RCT action producing polyproteins a nice trick or is there any means of perpet-CHO-ating (*I know it's a bad joke*) translation of an infinite message, yielding homogenous quality protein? For a mammalian expression system there are two strategies. If we think of our infinite circular mRNA not as an endless loop but rather a concatemeric series of genes or a polycistronic mRNA there are existing solutions to this issue that nature has kindly provided. Unlike polycistrons however, in the case of rolling circle translation we want to avoid the rate limiting step of translational initiation at each start codon and instead have the ribosome traverse messages unabated indefinitely. The first strategy for producing homologous proteins from this mechanism is the post-translational cleavage of the polyprotein chains by host cell proteins (Figure 5C&D). A limiting step in the production of bioactive therapeutic proteins in mammalian cells was the proteolytic cleavage of precursor molecules. This was overcome by bridging the precursor and gene sequences with a Furin cleavage motif (Ayoubi et al., 1996). Inclusion of a cleavage motif in the between the start and end of a transgene open reading frame on a circular RNA would allow for post-translational processing of the polyprotein strand. In the case of secreted proteins, the presence of a signal peptide immediately-proceeding the product sequence is will remain at the C-terminus of a nascent protein from the growing polyprotein chain unless a cleavage peptide is included (Figure 5D).

The second solution is co-translational cleavage in the form of ribosome arrest peptides (RAPs). RAPs are present in a diverse range or organisms (Ito & Chiba, 2013). The best known comes from foot and mouth disease virus. The ssRNA+ve genome of this virus encodes a single polyprotein which is co-translationally processed to produce multiple proteins. Two proteins "2A" and "2B" are processed by a process of ribosome "skipping" (Donnelly et al., 2001) or "Stop – carry on" (Shamra et al., 2012) (Figure 5E). At the C-terminus of the 2A protein is a highly conserved proline – glycine – proline residue. Steric hindrance of the final proline sees it egress from the ribosome. The growing polypeptide chain is dissociated by release factors and translation continuous from the proline to translate the downstream chain (Doronina et al., 2008). 2As work broadly in eukaryotes (Ryan et al., 1991; Donnelly et al., 2001; de Felipe et al., 2003) but not in prokaryotes (Donnelly et al., 1997). Uniting both co and post-translational polyprotein processing elements 2A and Furin has been executed to produce 1:1 molar ratios of heavy and light chains of a monoclonal antibody from a polycistronic linear mRNA in mammals (Fang et al., 2005). The "pause and carry-on" manner of 2A-self cleavage peptides may introduce a rate limiting step themselves. The homogeneity of proteins produced by post-translational cleavage may be affected by the rate at which the growing chain is processed resulting in multimeric product variants. (Possibly what we have seen in our work as EPO has a natural signal peptide?).

How does a mammalian cell combat the notion of perpetual ribosome motion? There has to be an underlying quality control mechanism regulating the ribosome performance. Ribosomes after all are composed of proteins, when proteins miss fold or malfunction, they are tagged for proteasomal degradation. A considerable cellular resource is designated to produce ribosomes and t-RNAs with a typical mammalian cell having 10^6-10^7 ribosomes accounting for ~10% of its protein mass (An & Harper, 2018). An & Harper (2018) reported that under basal conditions ribophagic flux in mammalian cells is low, at <1% over a 24hr period. Work by Pulk et al., (2010) reported that prokaryotic ribosomes can substitute defective 70S ribosomal subunits in situ. Thus, it could be postulated that while poor quality product could trigger degradation of the nascent polypeptide. In situ repair of a defective ribosome complex could allow it to remain locked onto an infinite circular mRNA and continue translation upon repair. Think of it as the infinite message as a journey across the universe and the ribosome as your spacecraft. You may need to repair the ship every now and again to keep you going. Now all we have to figure out is lightspeed or better yet, warp speed.

6.6 CONCLUDING REMARKS AND PERSPECTIVE

CircRNA and circular mRNAs pose novel means of enhancing the biotherapeutic productive potential of mammalian cell and cell-free expression platforms. The abundance of natural RNA circles has only come to light of late. The complexity and diversity of endogenous circRNA expression and function could perhaps make them favourable targets for host cell engineering. Application of the desirable attributes of these molecules; stability, translational potential and form synthetic circRNA or circular mRNAs has shown promise. Process intensification through platform or chemical engineering has done wonders for improving protein titres. Synthetic circRNA and circular mRNA could see process intensification on the molecular level by rounding out the cell engineering toolbox. To close the loop, “indefinite” translation of a highly stable (Jeck et al., 2013; Wang & Wang et al., 2015) infinite ORF (Perriman & Ares, 1998; Abe et al., 2014) producing functional protein by the incorporation of RAPs, the potential to improve recombinant protein yields are endless.

6.7 REFERENCES

1. Abe, N., Hiroshima, M., Maruyama, H., Nakashima, Y., Nakano, Y., Matsuda, A., Sako, Y., Ito, Y. & Abe, H. 2013, "Rolling circle amplification in a prokaryotic translation system using small circular RNA", *Angewandte Chemie International Edition*, vol. 52, no. 27, pp. 7004-7008.

2. Abe, N., Matsumoto, K., Nishihara, M., Nakano, Y., Shibata, A., Maruyama, H., Shuto, S., Matsuda, A., Yoshida, M. & Ito, Y. 2015, "Rolling circle translation of circular RNA in living human cells", *Scientific Reports*, vol. 5, pp. 16435.

3. An, H. and Harper, J.W., 2018. Systematic analysis of ribophagy in human cells reveals bystander flux during selective autophagy. *Nature Cell Biology*, 20(2), p.135.

4. Ashwal-Fluss, R., Meyer, M., Pamudurti, N.R., Ivanov, A., Bartok, O., Hanan, M., Evantal, N., Memczak, S., Rajewsky, N. & Kadener, S. 2014, "circRNA biogenesis competes with pre-mRNA splicing", *Molecular Cell*, vol. 56, no. 1, pp. 55-66.

5. Ayoubi, T.A., Meulemans, S.M., Roebroek, A.J. and Van de Ven, W.J., 1996. Production of recombinant proteins in Chinese hamster ovary cells overexpressing the subtilisin-like proprotein converting enzyme furin. *Molecular Biology Reports*, 23(2), pp.87-95.

6. Bort, J.A.H., Hackl, M., Höflmayer, H., Jadhav, V., Harreither, E., Kumar, N., Ernst, W., Grillari, J. & Borth, N. 2012, "Dynamic mRNA and miRNA profiling of CHO-K1 suspension cell cultures", *Biotechnology Journal*, vol. 7, no. 4, pp. 500-515.

7. Brinkrolf, K., Rupp, O., Laux, H., Kollin, F., Ernst, W., Linke, B., Kofler, R., Romand, S., Hesse, F. & Budach, W.E. 2013, "Chinese hamster genome sequenced from sorted chromosomes", *Nature Biotechnology*, vol. 31, no. 8, pp. 694-695.

8. Burd, C.E., Jeck, W.R., Liu, Y., Sanoff, H.K., Wang, Z. & Sharpless, N.E. 2010, "Expression of linear and novel circular forms of an INK4/ARF-associated non-coding RNA correlates with atherosclerosis risk", *PLoS Genetics*, vol. 6, no. 12, pp. e1001233.

9. Chen, C.Y. & Sarnow, P. 1995, "Initiation of protein synthesis by the eukaryotic translational apparatus on circular RNAs", *Science (New York, N.Y.)*, vol. 268, no. 5209, pp. 415-417.

10. Chen, J., Li, Y., Zheng, Q., Bao, C., He, J., Chen, B., Lyu, D., Zheng, B., Xu, Y., Long, Z., Zhou, Y., Zhu, H., Wang, Y., He, X., Shi, Y. & Huang, S. 2017, "Circular RNA profile identifies circPVT1 as a proliferative factor and prognostic marker in gastric cancer", *Cancer Letters*, vol. 388, pp. 208-219.

11. Cocquerelle, C., Mascrez, B., Hetuin, D. & Bailleul, B. 1993, "Mis-splicing yields circular RNA molecules", *FASEB Journal : official publication of the Federation of American Societies for Experimental Biology*, vol. 7, no. 1, pp. 155-160.

12. Conn, S.J., Pillman, K.A., Toubia, J., Conn, V.M., Salmanidis, M., Phillips, C.A., Roslan, S., Schreiber, A.W., Gregory, P.A. & Goodall, G.J. 2015, "The RNA binding protein quaking regulates formation of circRNAs", *Cell*, vol. 160, no. 6, pp. 1125-1134.

13. de Breyne, S., Yu, Y., Unbehaun, A., Pestova, T.V. and Hellen, C.U., 2009. Direct functional interaction of initiation factor eIF4G with type 1 internal ribosomal entry sites. *Proceedings of the National Academy of Sciences*, 106(23), pp.9197-9202.

14. de Felipe, P., Hughes, L.E., Ryan, M.D. & Brown, J.D. 2003, "Co-translational, intraribosomal cleavage of polypeptides by the foot-and-mouth disease virus 2A peptide", *The Journal of Biological Chemistry*, vol. 278, no. 13, pp. 11441-11448.

15. Donnelly, M.L., Hughes, L.E., Luke, G., Mendoza, H., ten Dam, E., Gani, D. & Ryan, M.D. 2001, "The 'cleavage'activities of foot-and-mouth disease virus 2A site-directed mutants and naturally occurring '2A-like'sequences", *Journal of General Virology*, vol. 82, no. 5, pp. 1027-1041.

16. Donnelly, M., Gani, D., Flint, M., Monaghan, S. & Ryan, M.D. 1997, "The cleavage activities of aphthovirus and cardiovirus 2A proteins.", *Journal of General Virology*, vol. 78, no. 1, pp. 13-21.

17. Doronina, V.A., Wu, C., de Felipe, P., Sachs, M.S., Ryan, M.D. & Brown, J.D. 2008, "Site-specific release of nascent chains from ribosomes at a sense codon", *Molecular and Cellular Biology*, vol. 28, no. 13, pp. 4227-4239.

18. Emmerling, V.V., Fischer, S., Stiefel, F., Holzmann, K., Handrick, R., Hesse, F., Hörer, M., Kochanek, S. & Otte, K. 2016, "Temperature-sensitive miR-483 is a conserved regulator of recombinant protein and viral vector production in mammalian cells", *Biotechnology and Bioengineering*, vol. 113, no. 4, pp. 830-841.

19. Enuka, Y., Lauriola, M., Feldman, M.E., Sas-Chen, A., Ulitsky, I. and Yarden, Y., 2015. Circular RNAs are long-lived and display only minimal early alterations in response to a growth factor. *Nucleic Acids Research*, 44(3), pp.1370-1383.

20. Fang, J., Qian, J., Yi, S., Harding, T.C., Tu, G.H., VanRoey, M. & Jooss, K. 2005, "Stable antibody expression at therapeutic levels using the 2A peptide", *Nature Biotechnology*, vol. 23, no. 5, pp. 584.

21. Fire, A. & Xu, S.Q. 1995, "Rolling replication of short DNA circles", *Proceedings of the National Academy of Sciences of the United States of America*, vol. 92, no. 10, pp. 4641-4645.

22. Fischer, S., Handrick, R. & Otte, K. 2015, "The art of CHO cell engineering: A comprehensive retrospect and future perspectives", *Biotechnology Advances*, vol. 33, no. 8, pp. 1878-1896.

23. Gao, Y., Wang, J. & Zhao, F. 2015, "CIRI: an efficient and unbiased algorithm for de novo circular RNA identification", *Genome Biology*, vol. 16, no. 1, pp. 4.

24. Gao, Y. & Zhao, F. 2018, "Computational Strategies for Exploring Circular RNAs", *Trends in Genetics*.

25. Geisler, S. & Coller, J. 2013, "RNA in unexpected places: long non-coding RNA functions in diverse cellular contexts", *Nature Reviews Molecular Cell Biology*, vol. 14, no. 11, pp. 699.

26. Goodrich, J.A. & Kugel, J.F. 2006, "Non-coding-RNA regulators of RNA polymerase II transcription", *Nature Reviews Molecular Cell Biology*, vol. 7, no. 8, pp. 612.

27. Grosso, A.R., Leite, A.P., Carvalho, S., Matos, M.R., Martins, F.B., Vítor, A.C., Desterro, J.M., Carmo-Fonseca, M. and de Almeida, S.F., 2015. Pervasive transcription read-through promotes aberrant expression of oncogenes and RNA chimeras in renal carcinoma. *eLife*, 4, p.e09214.

28. Guo, J.U., Agarwal, V., Guo, H. & Bartel, D.P. 2014, "Expanded identification and characterization of mammalian circular RNAs", *Genome Biology*, vol. 15, no. 7, pp. 409.

29. Han, D., Li, J., Wang, H., Su, X., Hou, J., Gu, Y., Qian, C., Lin, Y., Liu, X. & Huang, M. 2017, "Circular RNA MTO1 acts as the sponge of miR-9 to suppress hepatocellular carcinoma progression", *Hepatology*, 66(4), pp.1151-1164.

30. Hansen, T.B., Jensen, T.I., Clausen, B.H., Bramsen, J.B., Finsen, B., Damgaard, C.K. & Kjems, J. 2013, "Natural RNA circles function as efficient microRNA sponges", *Nature*, vol. 495, no. 7441, pp. 384-388.

31. Hansen, T.B., Wiklund, E.D., Bramsen, J.B., Villadsen, S.B., Statham, A.L., Clark, S.J. & Kjems, J. 2011, "miRNA-dependent gene silencing involving Ago2-mediated cleavage of a circular antisense RNA", *The EMBO Journal*, vol. 30, no. 21, pp. 4414-4422.

32. Hsiao, K.Y., Lin, Y.C., Gupta, S.K., Chang, N., Yen, L., Sun, H.S. & Tsai, S.J. 2017, "Noncoding Effects of Circular RNA CCDC66 Promote Colon Cancer Growth and Metastasis", *Cancer Research*, vol. 77, no. 9, pp. 2339-2350.

33. Huang, G., Zhu, H., Shi, Y., Wu, W., Cai, H. & Chen, X. 2015, "cir-ITCH plays an inhibitory role in colorectal cancer by regulating the Wnt/β-catenin pathway", PloS One, vol. 10, no. 6, pp. e0131225.

34. Ito, K. & Chiba, S. 2013, "Arrest peptides: cis-acting modulators of translation", *Annual Review of Biochemistry*, vol. 82, pp. 171-202.

35. Ivanov, A., Memczak, S., Wyler, E., Torti, F., Porath, H.T., Orejuela, M.R., Piechotta, M., Levanon, E.Y., Landthaler, M. & Dieterich, C. 2015, "Analysis of intron sequences reveals hallmarks of circular RNA biogenesis in animals", *Cell Reports*, vol. 10, no. 2, pp. 170-177.

36. Jackson, R.J., Hellen, C.U. and Pestova, T.V., 2010. The mechanism of eukaryotic translation initiation and principles of its regulation. *Nature Reviews Molecular Cell Biology*, 11(2), p.113.

37. Jeck, W.R., Sorrentino, J.A., Wang, K., Slevin, M.K., Burd, C.E., Liu, J., Marzluff, W.F. & Sharpless, N.E. 2013, "Circular RNAs are abundant, conserved, and associated with ALU repeats", *RNA (New York, N.Y.)*, vol. 19, no. 2, pp. 141-157.

38. Jin, H., Jin, X., Zhang, H. & Wang, W. 2017, "Circular RNA hsa-circ-0016347 promotes proliferation, invasion and metastasis of osteosarcoma cells", *Oncotarget*, vol. 8, no. 15, pp. 25571-25581.

39. Kaczmarek, J.C., Kowalski, P.S. and Anderson, D.G., 2017. Advances in the delivery of RNA therapeutics: from concept to clinical reality. *Genome Medicine*, 9(1), p.60.

40. Kannan, K., Wang, L., Wang, J., Ittmann, M.M., Li, W. & Yen, L. 2011, "Recurrent chimeric RNAs enriched in human prostate cancer identified by deep sequencing", *Proceedings of the National Academy of Sciences of the United States of America*, vol. 108, no. 22, pp. 9172-9177.

41. Kelly, P.S., Breen, L., Gallagher, C., Kelly, S., Henry, M., Lao, N.T., Meleady, P., O'Gorman, D., Clynes, M. & Barron, N. 2015, "Re-programming CHO cell metabolism using miR-23 tips the balance towards a highly productive phenotype", *Biotechnology Journal*, vol. 10, no. 7, pp. 1029-1040.

42. Kelly, S., Greenman, C., Cook, P.R. & Papantonis, A. 2015, "Exon skipping is correlated with exon circularization", *Journal of Molecular Biology*, vol. 427, no. 15, pp. 2414-2417.

43. Kozak, M., 1979. Inability of circular mRNA to attach to eukaryotic ribosomes. *Nature*, 280(5717), p.82.

44. Kramer, M.C., Liang, D., Tatomer, D.C., Gold, B., March, Z.M., Cherry, S. & Wilusz, J.E. 2015, "Combinatorial control of Drosophila circular RNA expression by intronic repeats, hnRNPs, and SR proteins", *Genes & Development*, vol. 29, no. 20, pp. 2168-2182.

45. Lewis, N.E., Liu, X., Li, Y., Nagarajan, H., Yerganian, G., O'Brien, E., Bordbar, A., Roth, A.M., Rosenbloom, J. & Bian, C. 2013, "Genomic landscapes of Chinese hamster ovary cell lines as revealed by the Cricetulus griseus draft genome", *Nature Biotechnology*, vol. 31, no. 8, pp. 759-765.

46. Li, S. and Mason, C.E., 2014. The pivotal regulatory landscape of RNA modifications. *Annual Review of Genomics and Human Genetics*, 15, pp.127-150.

47. Li, Z., Huang, C., Bao, C., Chen, L., Lin, M., Wang, X., Zhong, G., Yu, B., Hu, W. & Dai, L. 2015, "Exon-intron circular RNAs regulate transcription in the nucleus", *Nature Structural and Molecular Biology*, vol. 22, no. 3, pp. 256.

48. Liang, D., Tatomer, D.C., Luo, Z., Wu, H., Yang, L., Chen, L., Cherry, S. & Wilusz, J.E. 2017, "The output of protein-coding genes shifts to circular RNAs when the pre-mRNA processing machinery is limiting", *Molecular Cell*, vol. 68, no. 5, pp. 940-954. e3.

49. Liang, D. & Wilusz, J.E. 2014, "Short intronic repeat sequences facilitate circular RNA production", *Genes & Development*, vol. 28, no. 20, pp. 2233-2247.

50. Lu, T., Cui, L., Zhou, Y., Zhu, C., Fan, D., Gong, H., Zhao, Q., Zhou, C., Zhao, Y., Lu, D., Luo, J., Wang, Y., Tian, Q., Feng, Q., Huang, T. & Han, B. 2015, "Transcriptome-wide investigation of circular RNAs in rice", *RNA (New York, N.Y.)*, vol. 21, no. 12, pp. 2076-2087.

51. Maher, C.A., Kumar-Sinha, C., Cao, X., Kalyana-Sundaram, S., Han, B., Jing, X., Sam, L., Barrette, T., Palanisamy, N. & Chinnaiyan, A.M. 2009, "Transcriptome sequencing to detect gene fusions in cancer", *Nature*, vol. 458, no. 7234, pp. 97.

52. Memczak, S., Jens, M., Elefsinioti, A., Torti, F., Krueger, J., Rybak, A., Maier, L., Mackowiak, S.D., Gregersen, L.H., Munschauer, M., Loewer, A., Ziebold, U.,

Landthaler, M., Kocks, C., le Noble, F. & Rajewsky, N. 2013, "Circular RNAs are a large class of animal RNAs with regulatory potency", *Nature*, vol. 495, no. 7441, pp. 333-338.

53. Meyer, K.D., Patil, D.P., Zhou, J., Zinoviev, A., Skabkin, M.A., Elemento, O., Pestova, T.V., Qian, S.B. and Jaffrey, S.R., 2015. 5′ UTR m 6 A promotes cap-independent translation. *Cell*, 163(4), pp.999-1010.

54. Nigro, J.M., Cho, K.R., Fearon, E.R., Kern, S.E., Ruppert, J.M., Oliner, J.D., Kinzler, K.W. & Vogelstein, B. 1991, "Scrambled exons", *Cell*, vol. 64, no. 3, pp. 607-613.

55. Okholm, T.L.H., Nielsen, M.M., Hamilton, M.P., Christensen, L., Vang, S., Hedegaard, J., Hansen, T.B., Kjems, J., Dyrskjøt, L. & Pedersen, J.S. 2017, "Circular RNA expression is abundant and correlated to aggressiveness in early-stage bladder cancer", *NPJ Genomic Medicine*, vol. 2, no. 1, pp. 36.

56. Ouyang, H., Chen, X., Wang, Z., Yu, J., Jia, X., Li, Z., Luo, W., Abdalla, B.A., Jebessa, E. & Nie, Q. 2017, "Circular RNAs are abundant and dynamically expressed during embryonic muscle development in chickens", *DNA Research*, 25(1), pp.71-86.

57. Pamudurti, N.R., Bartok, O., Jens, M., Ashwal-Fluss, R., Stottmeister, C., Ruhe, L., Hanan, M., Wyler, E., Perez-Hernandez, D. & Ramberger, E. 2017, "Translation of circRNAs", *Molecular Cell*, vol. 66, no. 1, pp. 9-21. e7.

58. Perriman, R. & Ares, M. 1998, "Circular mRNA can direct translation of extremely long repeating-sequence proteins in vivo", *RNA*, vol. 4, no. 9, pp. 1047-1054.

59. Pestova, T.V., Hellen, C.U. and Shatsky, I.N., 1996. Canonical eukaryotic initiation factors determine initiation of translation by internal ribosomal entry. *Molecular and Cellular Biology*, 16(12), pp.6859-6869.

60. Prince, J.T., McGrath, K.P., DiGirolamo, C.M. & Kaplan, D.L. 1995, "Construction, cloning, and expression of synthetic genes encoding spider dragline silk", *Biochemistry*, vol. 34, no. 34, pp. 10879-10885.

61. Pulk, A., Liiv, A., Peil, L., Maiväli, Ü., Nierhaus, K. and Remme, J., 2010. Ribosome reactivation by replacement of damaged proteins. *Molecular Microbiology*, 75(4), pp.801-814

62. Ryan, M.D., King, A.M. & Thomas, G.P. 1991, "Cleavage of foot-and-mouth disease virus polyprotein is mediated by residues located within a 19 amino acid sequence", *Journal of General Virology*, vol. 72, no. 11, pp. 2727-2732.

63. Salzman, J., Chen, R.E., Olsen, M.N., Wang, P.L. & Brown, P.O. 2013, "Cell-type specific features of circular RNA expression", *PLoS Genetics*, vol. 9, no. 9, pp. e1003777.

64. Salzman, J., Gawad, C., Wang, P.L., Lacayo, N. & Brown, P.O. 2012, "Circular RNAs are the predominant transcript isoform from hundreds of human genes in diverse cell types", *PloS One*, vol. 7, no. 2, pp. e30733.

65. Sanchez, N., Kelly, P., Gallagher, C., Lao, N.T., Clarke, C., Clynes, M. & Barron, N. 2014, "CHO cell culture longevity and recombinant protein yield are enhanced by depletion of miR-7 activity via sponge decoy vectors", *Biotechnology Journal*, vol. 9, no. 3, pp. 396-404.

66. Sharma, P., Yan, F., Doronina, V.A., Escuin-Ordinas, H., Ryan, M.D. & Brown, J.D. 2011, "2A peptides provide distinct solutions to driving stop-carry on translational recoding", *Nucleic Acids Research*, vol. 40, no. 7, pp. 3143-3151.

67. Silvera, D., Arju, R., Darvishian, F., Levine, P.H., Zolfaghari, L., Goldberg, J., Hochman, T., Formenti, S.C. and Schneider, R.J., 2009. Essential role for eIF4GI overexpression in the pathogenesis of inflammatory breast cancer. *Nature Cell Biology*, 11(7), p.903.

68. Starke, S., Jost, I., Rossbach, O., Schneider, T., Schreiner, S., Hung, L. & Bindereif, A. 2015, "Exon circularization requires canonical splice signals", *Cell Reports*, vol. 10, no. 1, pp. 103-111.

69. Stoneley, M. and Willis, A.E., 2004. Cellular internal ribosome entry segments: structures, trans-acting factors and regulation of gene expression. *Oncogene*, 23(18), p.3200.

70. Suzuki, H., Zuo, Y., Wang, J., Zhang, M.Q., Malhotra, A. and Mayeda, A., 2006. Characterization of RNase R-digested cellular RNA source that consists of lariat and circular RNAs from pre-mRNA splicing. *Nucleic Acids Research*, 34(8), pp.e63-e63.

71. Talhouarne, G.J. & Gall, J.G. 2014, "Lariat intronic RNAs in the cytoplasm of *Xenopus tropicalis* oocytes", *RNA (New York, N.Y.)*, vol. 20, no. 9, pp. 1476-1487.

72. 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*, , pp. 1700220.

73. Wang, Y. & Wang, Z. 2015, "Efficient backsplicing produces translatable circular mRNAs", *RNA (New York, N.Y.)*, vol. 21, no. 2, pp. 172-179.

74. Wei, C.M., Gershowitz, A. and Moss, B., 1975. Methylated nucleotides block 5′ terminus of HeLa cell messenger RNA. *Cell*, 4(4), pp.379-386.

75. Wesselhoeft, R.A., Kowalski, P.S. and Anderson, D.G., 2018. Engineering circular RNA for potent and stable translation in eukaryotic cells. *Nature Communications*, 9(1), p.2629.

76. Westholm, J.O., Miura, P., Olson, S., Shenker, S., Joseph, B., Sanfilippo, P., Celniker, S.E., Graveley, B.R. & Lai, E.C. 2014, "Genome-wide analysis of drosophila circular RNAs reveals their structural and sequence properties and age-dependent neural accumulation", *Cell Reports*, vol. 9, no. 5, pp. 1966-1980.

77. Xu, X., Nagarajan, H., Lewis, N.E., Pan, S., Cai, Z., Liu, X., Chen, W., Xie, M., Wang, W. & Hammond, S. 2011, "The genomic sequence of the Chinese hamster ovary (CHO)-K1 cell line", *Nature Biotechnology*, vol. 29, no. 8, pp. 735-741.

78. Yang, Y., Fan, X., Mao, M., Song, X., Wu, P., Zhang, Y., Jin, Y., Yang, Y., Chen, L. & Wang, Y. 2017, "Extensive translation of circular RNAs driven by N6-methyladenosine", *Cell Research*, vol. 27, no. 5, pp. 626-641.

79. Ye, C., Chen, L., Liu, C., Zhu, Q. & Fan, L. 2015, "Widespread noncoding circular RNAs in plants", *New Phytologist*, vol. 208, no. 1, pp. 88-95.

80. Zhang, H., Wang, G., Ding, C., Liu, P., Wang, R., Ding, W., Tong, D., Wu, D., Li, C. & Wei, Q. 2017, "Increased circular RNA UBAP2 acts as a sponge of miR-143 to promote osteosarcoma progression", *Oncotarget*, vol. 8, no. 37, pp. 61687.

81. Zhang, X., Wang, H., Zhang, Y., Lu, X., Chen, L. & Yang, L. 2014, "Complementary sequence-mediated exon circularization", *Cell*, vol. 159, no. 1, pp. 134-147.

82. Zheng, J., Liu, X., Xue, Y., Gong, W., Ma, J., Xi, Z., Que, Z. & Liu, Y. 2017, "TTBK2 circular RNA promotes glioma malignancy by regulating miR-217/HNF1β/Derlin-1 pathway", *Journal of Hematology & Oncology*, vol. 10, no. 1, pp. 52.

83. Zheng, Q., Bao, C., Guo, W., Li, S., Chen, J., Chen, B., Luo, Y., Lyu, D., Li, Y. & Shi, G. 2016, "Circular RNA profiling reveals an abundant circHIPK3 that regulates cell growth by sponging multiple miRNAs", *Nature Communications*, vol. 7, pp. 11215.

84. Zhu, X., Wang, X., Wei, S., Chen, Y., Chen, Y., Fan, X., Han, S. & Wu, G. 2017, "hsa_circ_0013958: a circular RNA and potential novel biomarker for lung adenocarcinoma", *The FEBS Journal*, vol. 284, no. 14, pp. 2170-2182.

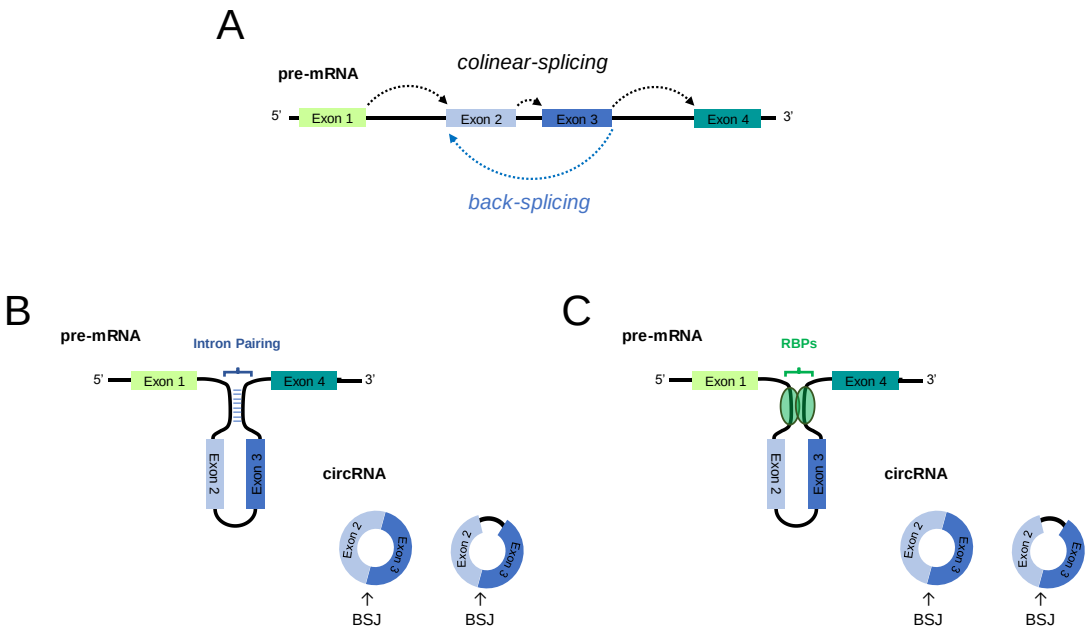


Figure 1. Circular RNA biogenesis

(**A**) Canonical splicing of pre-mRNA to remove intronic sequence occurs in a co-linear fashion, joining exons in sequential order. During this process "back-splicing" can occur joining the splice signals of downstream exons to upstream sequence. Circular RNA (circRNA) arises from back-splicing pre-mRNA sequence. (**B**) Intron complementarity brings non-colinear splice junctions close together creating a back-splice junction (BSJ). (**C**) This can also result from RNA binding proteins (RBP) interacting with the intronic sequence. circRNA products can be exonic or intron retaining.

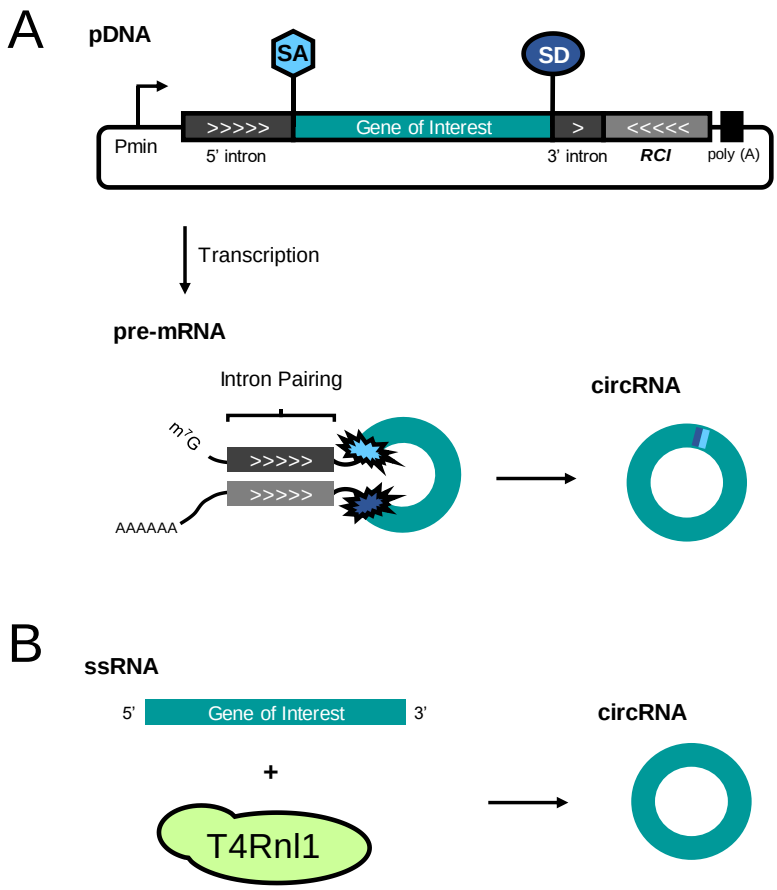


Figure 2. Exogenous expression of circRNA

(**A**) Expression of circRNA from plasmid DNA (pDNA). The gene or sequence of interest is directly flanked 5' with a splice acceptor (*SA*) and a splice donor (*SD*) to the 3'. These splice motifs are further flanked by intronic sequence. Arrows (>>> /<<<) indicate the direction of the intron sequence. Upon transcription the complementary intronic sequence will hybridize mirroring the natural intron pairing pathway. This intron pairing makes the SA and SD proximal, facilitating a back-splice junction to produce a circRNA. (**B**) The formation of circRNA can also be achieved without cell machinery. Synthesized RNA can be circularized with T4 RNA ligase 1 (T4Rnl1).

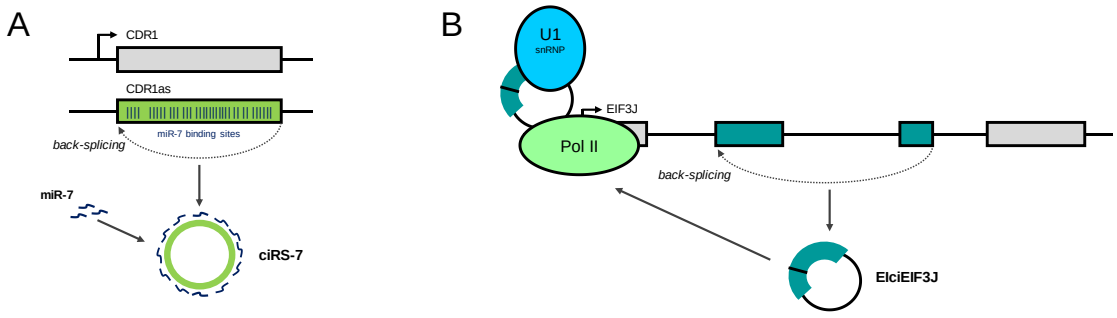


Figure 3. Gene regulatory functions of natural RNA circles

(**A**) The antisense to the human CDR1 gene is expressed as a circRNA. Named ciRS-7, this circRNA harbours greater than 70 miR-7 binding sites and has been demonstrated to function as an efficient miRNA sponge decoy. (**B**) EIciEIF3J is an intron retaining circRNA produced by back-splicing of the eukaryotic translation initiation factor 3J (EIF3J). It functions to enhance transcription of the EIF3J promoter by interacting with U1 small nuclear ribonucleoprotein (U1 snRNP).

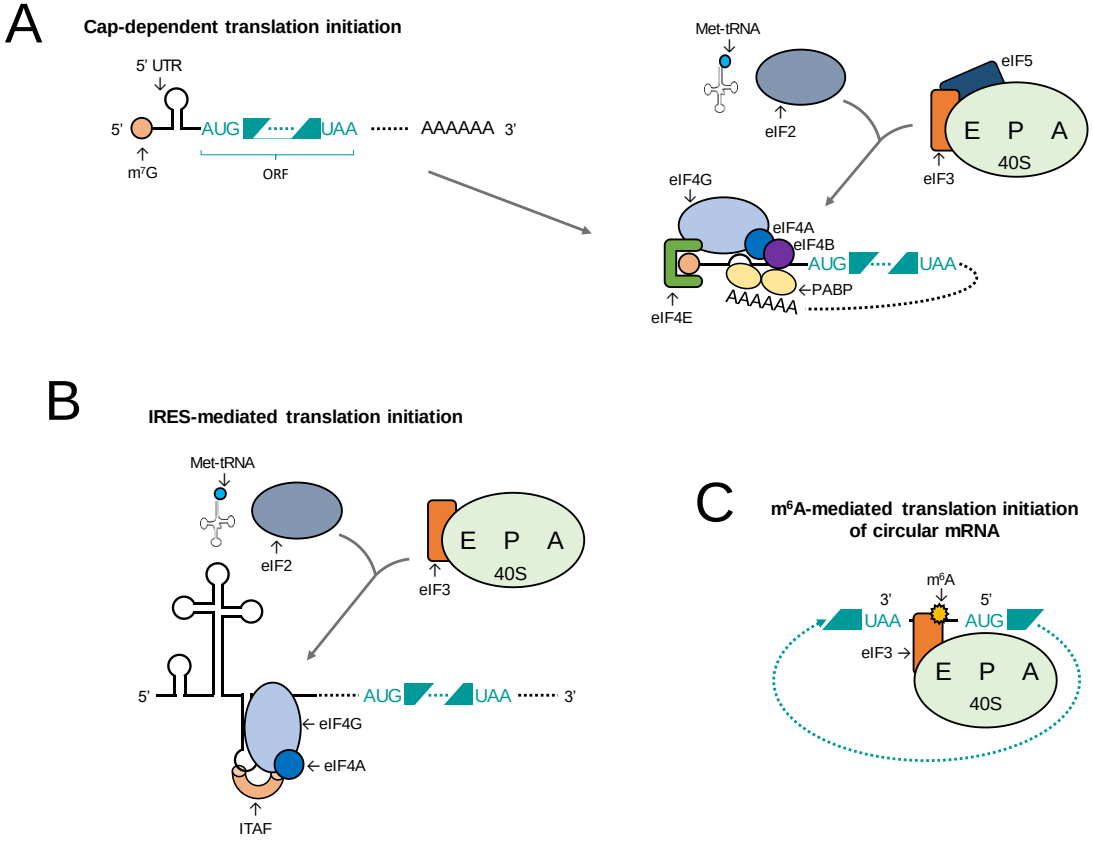


Figure 4. Cap-dependent and cap-independent translation initiation

(**A**) The canonical pathway of eukaryotic translation initiation is a highly regulated process of ribosome assembly in the 5' un-translated region of an mRNA. The mRNA substrate is bound by poly-A binding proteins (PABP) and the eukaryotic initiation factor 4F (eIF4F) complex comprised of eIF4G, eIF4A and the m^7GpppG (m^7G) cap-dependent factor eIF4E. Together with eIF4B this complex activates the mRNA and recruits the components of the ribosome 48S complex. (**B**) Internal ribosome entry sites (IRES) are RNA elements that recruit eukaryotic initiation factors in a cap-independent manner. The secondary structure of IRESs can vary greatly depending on source, be it viral or eukaryotic. The localization of initiation factors can be stabilized by IRES trans-acting factors (ITAFs). (**C**) N^6-methyladenosine (m^6A) can directly bind to eIF3, a component of the 48S complex, to direct translation at internal locations in mRNA. The mechanism by which this works is not yet well defined.

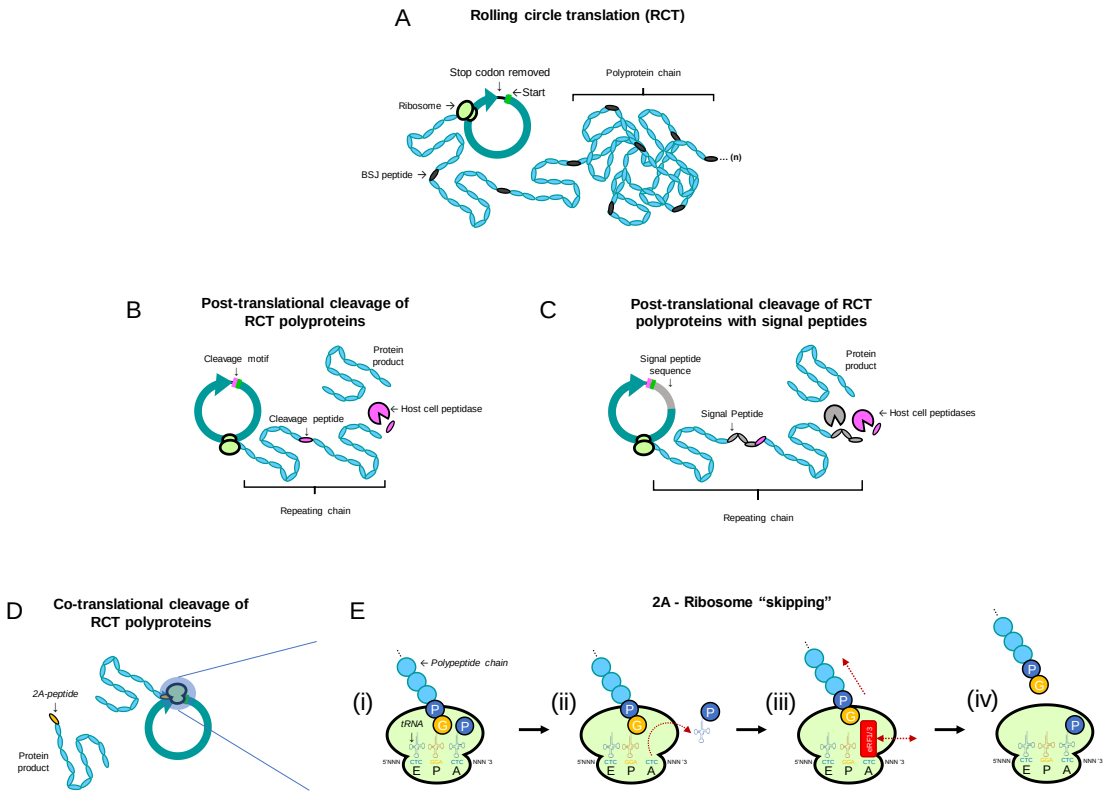


Figure 5. Co and post-translational cleavage of rolling circle translation products

The principle of circular mRNA translation is the same as that of linear mRNA. The ribosome complex is formed in the un-translated region (UTR) followed by elongation through the message until termination upon recognition of a stop codon. (**A**) By removing the stop codon from the gene open reading frame in a circular mRNA, the message becomes infinite. This facilitates indefinite circling of the mRNA by the ribosome, known as rolling circle translation (RCT). RCT produces long concatemeric spools of repeating polyprotein. These polyproteins contain unique peptides encoded by the translation of the back-splice junction (BSJ). (**B**) The repeating polyprotein can be post-translationally processed by host cell peptidases to yield monomeric proteins by the incorporation of an endoproteolytic cleavage motif at the 3' of the message. (**C**) For secreted proteins, the presence of a signal peptide offers a challenge. In the case of ribosome circling the signal peptide of one product will remain on the C-terminus of another. This could be again overcome by the incorporation of an endoproteolytic cleavage motif. (**D**) The cleavage of polyproteins can also occur co-translationally by encoding a viral 2A – "self-cleavage" motif. (**E**) (**i**) The C-terminus of 2A motifs have a conserved Proline – Glycine – Proline sequence. (**ii**) Steric hindrance of the final Proline forces the propyl-tRNA to egress. (**iii**) Translational release factors (eRF1/3) enter the A-site of the ribosome and release the nascent polypeptide chain. (**iv**) Ingression of the propyl-tRNA follows release of the nascent polypeptide enabling translation to continue. This mechanism of "self-cleavage" is also referred to as ribosome – "skipping".

Chapter 7

Improved Yield of rhEPO in CHO Cells with Synthetic 5' UTR

Costello, A., Lao, T.N., Barron, N. * & Clynes, M. *
* authors contributed equally
Biotechnology Letters, DOI: 10.1007/s10529-018-2632-2, PMID: 30506229
Impact Factor (2018): 1.846

I was the primary contributor to all aspects of experimental work and manuscript preparation described in Chapter 7. Dr Nga T. Lao contributed with discussion of appropriate controls. Prof. Martin Clynes and Prof. Niall Barron contributed to experimental design and redaction of the manuscript.

ABSTRACT

The impact of local structure on mRNA translation is not well-defined pertaining to the 5' UTR. Reports suggest structural remodelling of the 5' UTR can significantly influence mRNA translation both in *cis* and *trans* however a new layer of complexity has been applied to this model with the now known reversible post-transcriptional chemical modification of RNA. N^6-methyladenosine (m^6A) is the most abundant internal base modification in mammalian mRNA. It has been reported that mRNAs harbouring m^6A motifs in their 5' UTR have improved translation efficiency. The present study evaluated the addition of putative m^6A motifs to the 5' UTR of a model recombinant human therapeutic glycoprotein, Erythropoietin (EPO), in a direct comparison with an A to T mutant and a no adenosine control. The m^6A construct yielded significantly improved EPO titer in transient batch culture over no adenosine and m^6T controls by 2.84 and 2.61-fold respectively. This study highlights that refinement of transgene RNA elements can yield significant improvements to protein titer.

Key words: 5' UTR, N^6-methyladenosine, m^6A, CHO, Bioprocessing, Synthetic biology

7.1 INTRODUCTION

RNA forms complex secondary and tertiary structures with known examples of structure-directed function (Leppek et al., 2017). Predicting the impact of cis-regulatory sequences of non-coding scripts, such as un-translated regions (UTR), on mRNA translation in higher eukaryotes remains a challenge. Recent growing appreciation for functional heterogeneity between eukaryotic ribosomal complexes and its associated proteins (Simsek et al., 2017; Shi et al., 2017) coupled with evidence of mRNA-structure-influenced translation raises the question: how much regulatory potential is owed to mRNA structure and structure-directed recruitment of RNA-binding protein (RBP) translation-driving factors? (Topisirovic et al., 2011; Pichon et al., 2012; Manning & Cooper, 2016). For this reason, the structure of UTR sequence in transgene expression cassettes may offer "playgrounds" for synthetic engineering of transgene expression. Early studies to synthetically remodel the 5' un-translated region (UTR) of transgene mRNA (Kozak, 1986; Grens & Scheffler, 1990) indicated highly-stable GC-rich secondary duplexes (ΔG = -50 kcal/ mol) upstream of the initiation AUG triplet prevented scanning of the 40S ribosome subunit leading to reduced protein yield. Complete removal 5' UTR sequence comes with the caveat of omitting known translational enhancer elements, impacting negatively on mRNA translation (Stein et al., 1998).

Reversible epigenetic modifications to both genomic DNA (Suzuki & Bird, 2008) and histone proteins (StrahI & Allis, 2000) have been known to influence gene expression and regulate cellular behaviour. Recent discovery of the intermediary member of the central dogma, RNA, having similar reversible chemical tuning has redefined how RNA metabolism is viewed (He, 2010; Fu et al., 2014). The most abundant internal mRNA modification in mammalian cells is *N*^{6}-methyladenosine (m^{6}A) (Wei et al., 1975). Transcriptome analysis of human cells and mouse tissue have revealed m^{6}A consensus motifs to be in the form of DRACH, [D=G/A/U] [R=G>A] m^{6}AC [H=U>A>C]. m^{6}A motifs are present in coding regions, 3' or 5' UTRs of mRNAs (Dominissini et al., 2012; Meyer et al, 2012) and in non-coding RNA (ncRNA) scripts (Wei et al., 1975) with enriched abundance at stop codons (Huang et al., 2018). This discovery of RNA chemical tuning adds a new layer of complexity to our understanding of how UTR sequence and structures can regulate mRNA translation.

In mammalian cells, m^{6}A is generated by a multicomponent methyltransferase "writer"-complex comprised of; METTL3, METTL14 and WTAP (Liu et al., 2014; Ping et al., 2014). This reaction is reversible with the "eraser" proteins FTO (Jia et al., 2011) and ALKBH5 (Zheng et al., 2013). Both writer and eraser modifications occur in the cell nucleus. The molecular mechanisms underpinning how m^{6}A motifs are recognised by different known m^{6}A-RNA binding proteins (RBPs) and regulate m^{6}A-RNA fate is still poorly understood. It has been reported that the stability of m^{6}A-modified mRNA is under dynamic regulation by three cytoplasmic "reader"

proteins of the YTH domain family, YTHDF1-3 (Wang et al., 2015). The crystal structure of YTHDF2s YTH domain revealed the m^6A recognition site (Zhu et al., 2014) as an aromatic cage (Li et al., 2014). The m^6A reader YTHDF2 recognises m^6A marked mRNA and selectively reduces the stability of such messages (Wang et al., 2014). The mechanism by which YTHDF2 destabilises mRNA harbouring m^6A was revealed to be mediated by accelerated transcript deadenylation via the CCR4-NOT complex (Du et al., 2017). While YTHDF2 controls the turnover of m^6A containing mRNA, YTHDF1-mediates enhanced translational efficiency of m^6A containing scripts (Wang et al., 2015). This mechanism of epigenetic post-transcriptional control permits the cell to prioritise the protein synthesis of m^6A marked mRNAs. Contrary to the mRNA-decay-promoting action of YTHDF2, IGF2BPs 1-3 have been identified as a family of m^6A readers that promote target mRNA stability and translation (Huang et al., 2018). The consensus m^6A recognition motif differs slightly between YTHDF and IGF2BPs with the nucleotide at position -3 to the A being A/U for YTHDF and U/C for IGF2BP. This selective recognition of different m^6A-RBP families adds to the complexity of post-transcriptional epigenetic regulation.

Independent studies have demonstrated that the presence of consensus putative m^6A motifs in the 5' UTR of transgene mRNA significantly improves the transcripts' translation efficiency in both cap-independent (Meyer et al., 2015; Yang et al., 2017) and cap-dependent (Wang et al., 2015) manners in human cells. Moreover, the location of a single m^6A motif in the 5' UTR significantly impacts the mRNAs' translation with close-proximity to the gene start codon giving the highest protein production (Meyer et al., 2015). An understudied area in industrial biotechnology has been improving the protein coding potential of transgene mRNA in mammalian production hosts. While codon optimisation has improved the rate of elongation, selectively biasing the host to initiate transgene protein synthesis and or increase mRNA stability could significantly enhance recombinant protein yields.

7.2 MATERIALS AND METHODS

7.2.1 Plasmid construction and transcript analysis

The human EPO open reading frame was synthesised (GenScript) with one of; a putative m^6A - GGACTAAAGCGGACTTGT, m^6T – GGTCTAAAGCGGTCTTGT, or no adenosine control sequence - CGGTGCCGGTGC, taken from (Yang et al., 2017). These constructs were cloned into in a pcDNA-3.1 (+) Hygromycin (Invitrogen) backbone using HindIII and EcoRV sites. RNA secondary structure was predicted with RNAFold (http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi). Tertiary RNA structure was predicted using RNAComposer (http://rnacomposer.cs.put.poznan.pl/) and sequence features highlighted using PyMOL V2.2.

7.2.2 Cell culture and transfection

CHO-K1 (ATCC) were grown in Balan CD Growth A (Irvine Scientific) with (2.52g L^{-1}) polyvinyl alcohol (PVA), and sub-cultured every 72-96hrs. Cultures were maintained in a Climo-Shaker (Kühner), 37°C, 80% humidity and 5% CO_2 with routine seeding at $2x10^5$ cells mL^{-1} in 50mL spin tube (TPP). Cell growth and viability was monitored using the Guava EasyCyte ViaCount programme (Merck Millipore). Prior to transfection, cells were washed twice with CHO S SFM II (Gibco) and seeded at $1x10^6$ cells mL^{-1} in 2 mL of CHO S SFM II. Transfection was performed in accordance with the MIRUS Transit 2X protocol. In brief, a complex of a [1:2] ratio µg pDNA to µL transfection reagent was formed in 100µL of pre-warmed, 37°C, CHO S SFM II, for 30mins. Transfection complexes were then added to the cell suspension and incubated for 24hrs. All transfections were done using a ratio of 1µg of plasmid DNA per $1x10^6$ cells. 24hrs post transfection the cultures were washed twice in Balan CD Growth A with PVA and seeded at $3x10^5$ cells mL^{-1} in triplicate 5mL cultures.

7.2.3 RNA isolation

Transfected cells were harvested by centrifugation at 91 x g for 5 minutes. Cell pellets were washed in phosphate buffered saline (PBS) and centrifuged again. Total RNA was isolated from $1\text{-}5x10^6$ cells using 1mL of Tri-reagent (Ambion) with no divergence from the manufacturer's protocol. RNA quantification and quality were evaluated by NanoDrop (Thermo Scientific). To remove contaminating plasmid DNA, RNA samples were treated with DNase I (Sigma Aldrich) as per manufacturer's protocol.

7.2.4 qPCR

Reverse transcription of total RNA was performed with the High Capacity cDNA Reverse Transcription Kit (Applied Biosystems) in accordance with the manufacture's protocol. RT-qPCR was performed on an AB7500 (Applied Biosystems) using Fast SYBR Green Master Mix (Applied Biosystems). The 2X SYBR master mix was combined with 20 ng of cDNA, 200 nM forward and reverse primers, and water made up to 20 µL final reaction volume. Each biological

replicate sample was run in technical triplicate wells. Primer sequences used are as follows; Gapdh FW - TGGCTACAGCAACAGAGTGG, Gapdh RV - GTGAGGGAGATGATCGGTGT, hEPO FW - ACTTGTCTCGAGATGGGGGT, hEPO RV - AGGTACCTCTCCAGGACTCG, HYG RW - ATGCTCCGCATTGGTCTTGA, HYG RV - ATTTGTGTACGCCCGACAGT. Relative quantification of PCR products was calculated using the ddCt method with Gapdh as an endogenous control and HYG as a control of transfection efficiency. Error was calculated as the standard deviation of three biological replicates.

7.2.5 ELISA

Secreted EPO was quantified by enzyme linked immunosorbent assay (ELISA) of cell culture supernatant. Culture supernatant was harvested by centrifugation of cell suspension at 91 x g for 5 minutes. Nunc-Immuno™ MicroWell™ 96 well solid plates (M9410 Sigma-Aldrich) were coated with 100 µL of capture antibody MAB-287 (R&D Systems), diluted 1:500 in coating buffer (C3041-50CAP) (Sigma-Aldrich) for 2 hrs at 4°C. Plate washing consisted of 3 x 100 µL per well washes of ELISA wash buffer (T9039) (Sigma-Aldrich). Blocking was done for 1 hr at room temperature with (T6789) Blocking Buffer (Sigma-Aldrich). Standards of recombinant human EPO (Cat no. 329871) (Merck Millipore) were diluted to range from 8000-0 pg mL^{-1} in sample dilution buffer (T6789 + 0.05% Tween20) (Sigma-Aldrich). Samples were diluted to fall within the range of standard curve and 100 µL incubated on the plate at 37°C for 1 hr. Rabbit Immunoglobulin-HRP (P0448) (Dako, Agilent) diluted 1:2000 was used for detection, 100 µL per well was incubated for 1 hr at room temperature. 100 µL of development solution; 1.2 mL of 3,3',5,5'-Tetramethylbenzidin at 1 mg mL^{-1} in DMSO, 2.4 µL 30 % (v/v) hydrogen peroxide and 10.8 mL Phosphate-Citrate buffer (P4809) (Merck), was added to each well and incubated for 30 minutes. The reaction was stopped by adding 100 µL 0.18 M H_2SO_4 to each well. Protein concentration was determined by reading on a Multi-scan Go (Thermo Fischer Scientific) and samples concentrations quantified using linear regression of the standard curve. Analysis consisted of six technical replicates over three biological replicates.

7.2.6 Statistical Analysis

The values reported represent the mean of three independent experiments with error displayed as standard deviation. Statistical analysis was carried out by a two-tailed homoscedastic student t-test, ($p \leq 0.05$ *, $p \leq 0.01$**, $p \leq 0.001$ ***) of three independent experimental replicates.

7.3 RESULTS

It has been previously reported that m^6A motifs are significantly enriched in known circRNAs (Yang et al., 2017). In their study, Yang et al., (2017) demonstrated dramatic increases in the translation efficiency of circular mRNAs containing m^6A motifs upstream of a reporter-gene start codon in direct comparison to an A to T mutant and to a no adenosine control (No A CTRL). To investigate the potential of m^6A motifs improving the production efficiency of recombinant protein therapeutics in CHO cells, we cloned three variants up-stream of a recombinant human Erythropoietin (rhEPO) open reading frame (Fig. 1).

The three transcript variants; No A – CTRL, m^6T and m^6A were assessed by transient batch performance in a CHO-K1 host. There was no significant difference in peak cell density between the transcript variants and un-transfected (CHO-K1) and transfection reagent only (vehicle) controls (Fig. 2a). Nor were there any deleterious effects on cell viability (Fig. 2b). Secreted EPO was monitored by ELISA of culture supernatant harvested at 48 hr intervals. Cell specific productivity (Qp) was highest in the m^6A variant (Fig. 2c). The mean Qp of CHO-K1 cells was improved to (9.31 pg cell^{-1} day^{-1}) with m^6A from (3.84 pg cell^{-1} day^{-1}) in the No A control ($p = 0.00048$) and (3.56 pg cell^{-1} day^{-1}) m^6T ($p = 0.00019$). Accordingly, mean EPO titer was improved significantly in the m^6A variant at all time points of analysis over both m^6T and No A controls (Fig. 2d). The m^6A construct increased the mean secreted EPO titer to (13.6 µg mL^{-1}) from that of m^6T (5.5 µg mL^{-1}) ($p = 0.000487$) and No A (5.7 µg mL^{-1}) ($p = 0.000185$) at 48 hrs into culture (Fig. 2d). Expression of EPO mRNA with the m^6A variant culminated in a final increase at 144 hrs of 2.61-fold over m^6T ($p = 0.0042$) and 2.84-fold over the No A control ($p = 0.0029$) (Fig. 2d).

The exact function of m^6A in eukaryotic translation is not fully understood with independent reports describing its role in an mRNA's translation efficiency (Wang et al., 2015) and mRNA turnover or stability (Wang et al., 2014; Du et al., 2016; Huang et al., 2018). The vectors used in this study drive the expression of a resistance gene Hygromycin-B-phosphotransferase (HYG) under a PKG promoter. Relative quantification (RQ) of HYG mRNA was assessed at 48, 96 and 144hrs and normalized to that of the No A control at each time point with respect to the endogenous control Gapdh in all cases (Fig. 3a). There was no significant difference between any of the three constructs at any time point indicating similar transfection efficiencies in all cases (Fig. 3a). The relative EPO mRNA abundance was 4.14-fold ($p = 1.39x10^{-5}$) and 4.99-fold ($p = 1.09x10^{-5}$) higher with m^6T and m^6A, respectively over the No A control construct at 48hrs (Fig. 3b). Relative to the No A control at each time point the m^6T transcript variant remained higher with a 4.16-fold ($p = 7.1x10^{-5}$) and 3.45-fold ($p = 1.76x10^{-6}$) increase in mRNA abundance at 96 and 144hr time points respectively (Fig. 3b). Similarly, EPO mRNA abundance in the m^6A variant remained higher than the No A control with a 3.43-fold ($p = 0.00027$) and 3.84-fold ($p =$

$1.46x10^{-5}$) increase in mRNA abundance at 96 and 144hr time points respectively (Fig. 3b). With transient transfection transgene expression decreases over time with the dilution of plasmid DNA as cells divide. There was no significant difference in the rate of HYG mRNA abundance decreasing over time between constructs (Fig. 3c). The Log_2 fold decrease in EPO mRNA abundance showed a significantly different profile for the No A control, from (-0.5941) to (-0.015) with m^6T ($p = 5.47x10^{-6}$) and (-0.0487) with m^6A ($p = 4.3x10^{-5}$) between 48 and 96 hrs into culture and from (-1.1084) in the No A control to (-0.5284) for m^6T and (-0.4013) m^6A variants, ($p = 1.29x10^{-5}$), ($p = 7.923x10^{-6}$) respectively between 96 and 144 hrs (Fig. 3d).

The three transcript variants used in this study contain subtle sequence differences, yet the relative mRNA abundance and subsequent translation differed greatly. The secondary structure of all transcript variants was analysed using RNAFold (Fig. 4a). There was similar structure present in the m^6T and m^6A constructs, but these differed greatly from the distinct open structure of the No A control (Fig. 4a). RNAComposer was used to predict the tertiary structure of these differing structures. The open loop structed remained in the tertiary model of the No A control (Fig. 4b). Again, there was a highly similar appearance between the m^6T (Fig. 4c) and m^6A variants (Fig. 4d).

7.4 DISCUSSION

We present herein an observation of significantly improved recombinant titer with modifications of the 5' UTR of transgene mRNA. Employing a putative m^6A motif previously reported to significantly improve the cap-independent translation of circular mRNAs (Yang et al., 2017) in human cell lines. Here we report it to have a similar effect on the translation of linear EPO mRNA in CHO-K1 cells with higher Qp (Fig. 2c) resulting in increased secreted EPO titer (Fig. 2d). In transient batch culture we observed no deleterious effects on peak cell density (Fig. 2a) or cell viability (Fig. 2b) associated with any of the three transcript variants. It would be interesting to see if this positive impact on product titer would hold through in stable mixed pools or clonal studies. Also, the sequence variants lack a consensus Kozak motif, which is known to enhance translation. Future work should incorporate the use of a Kozak sequence in the 5' UTR design. Simple addition of the Kozak motif may not be sufficient as the inclusion of extra nucleotides would disrupt the 5' UTR secondary structure.

Perhaps the most interesting finding was the dramatic increase in EPO mRNA transcript variants harbouring putative m^6T and m^6A motifs (Fig. 3b). Normalizing for transfection efficiency with HYG mRNA, the m^6T and m^6A EPO mRNA variants had relative increases in abundance in the order of 4.14 and 4.99-fold, respectively over the No A control (Fig. 3b). The rate of m^6T and m^6A mRNA loss with transient delivery was significantly reduced compared to that of the No A

control (Fig. 3d). Given that the loss of HYG mRNA was similar in all cases (Fig. 3c) and that RNA methylation is said to occur post-transcriptionally (Fu et al., 2014), this suggests that the increase in mRNA abundance is due to enhanced stability rather than improved transcription. Analysing the 5' UTR secondary folding of the three transcript variants (Fig. 4) revealed a highly similar structure for both m^6T and m^6A but these two differ greatly from the No A control. Variable translation efficiencies between the m^6A and m^6T suggest that the differences can be attributed to the putative methylation of a single nucleotide substitution. Recent evidence by Liu et al., (2015) described the presence of m^6A conferred changes to mRNA secondary structure. Their data suggests that methylation and demethylation resulted in dynamic remodelling of local mRNA structures terming this "the m^6A switch" (Liu et al., 2015). Restructuring of the m^6A transcript variant to include methylation is not possible with current RNA folding software but this could explain the improved translation observed with the m^6A over the m^6T transcript variants.

To date, apart from circularization of transcript mRNA (Wesselhoeft et al., 2018), efforts to improve the stability of linear transgene mRNA have yielded modest returns (Ferizi et al., 2015; Holtkamp et al., 2006). The m^6A motif used in this study contains a U/C at the -3 position which was found to be an enriched motif from the consensus with IGF2BPs (Huang et al., 2018). As described by Huang et al., (2018), recognition of m^6A by IGF2BPs enhanced mRNA stability. This could be an explanation for why the putative m^6A mRNA in this study had such improved relative mRNA abundance. Although we must point out this may also stem from the highly stable hairpin structures formed in the m^6T ($\Delta G = -32.30$) and m^6A ($\Delta G = -30.40$) variants over that of the No A control ($\Delta G = -17.90$). It is probable this increased mRNA abundance is a combination of both for the m^6A variant and can be attributed to the highly stable hairpin in the m^6T script.

Reports of reversible m^6A writing, erasing and reading (Liu et al., 2014; Ping et al., 2014; Jia et al., 2011; Zheng et al., 2013; Wang et al., 2014; Huang et al., 2018) point to this mechanism of post-transcriptional regulation being highly tuneable. As of yet it has not been reported if m^6A writers, erasers and readers are dynamic between different stages of culture in CHO cells. If this is the case, transgene mRNA encoding selective putative m^6A motifs could be conditionally translated under defined culture stages. With a report of targeted epigenetic editing of genomic DNA using a dCas9-GCN4 peptide fusion (Morita et al., 2016) and newly reported ssRNA targeting CRISPR proteins, Cas13b (Cox et al., 2017) the potential for targeted epigenetic regulation of transgene mRNA in mammalian hosts could offer new components to the mammalian cell engineering toolbox.

FUNDING

This work was conducted under the financial support of Scientific Foundation of Ireland (SFI) grant numbers [13/IA/1963] and [13/IA/1841].

7.5 REFERENCES

Cox, D.B.T., Gootenberg, J.S., Abudayyeh, O.O., Franklin, B., Kellner, M.J., Joung, J. & Zhang, F. 2017, "RNA editing with CRISPR-Cas13", *Science (New York, N.Y.)*, vol. 358, no. 6366, pp. 1019-1027.

Dominissini, D., Moshitch-Moshkovitz, S., Schwartz, S., Salmon-Divon, M., Ungar, L., Osenberg, S., Cesarkas, K., Jacob-Hirsch, J., Amariglio, N. & Kupiec, M. 2012, "Topology of the human and mouse m 6 A RNA methylomes revealed by m 6 A-seq", *Nature*, vol. 485, no. 7397, pp. 201.

Du, H., Zhao, Y., He, J., Zhang, Y., Xi, H., Liu, M., Ma, J. & Wu, L. 2016, "YTHDF2 destabilizes m 6 A-containing RNA through direct recruitment of the CCR4–NOT deadenylase complex", *Nature communications*, vol. 7, pp. 12626.

Ferizi, M., Leonhardt, C., Meggle, C., Aneja, M.K., Rudolph, C., Plank, C. & Rädler, J.O. 2015, "Stability analysis of chemically modified mRNA using micropattern-based single-cell arrays", *Lab on a Chip*, vol. 15, no. 17, pp. 3561-3571.

Fu, Y., Dominissini, D., Rechavi, G. & He, C. 2014, "Gene expression regulation mediated through reversible m6A RNA methylation", *Nature Reviews Genetics*, vol. 15, no. 5, pp. 293-306.

Grens, A. & Scheffler, I.E. 1990, "The 5'- and 3'-untranslated regions of ornithine decarboxylase mRNA affect the translational efficiency", *The Journal of biological chemistry*, vol. 265, no. 20, pp. 11810-11816.

He, C. 2010, "Grand challenge commentary: RNA epigenetics?", *Nature chemical biology*, vol. 6, no. 12, pp. 863.

Holtkamp, S., Kreiter, S., Selmi, A., Simon, P., Koslowski, M., Huber, C., Tureci, O. & Sahin, U. 2006, "Modification of antigen-encoding RNA increases stability, translational efficacy, and T-cell stimulatory capacity of dendritic cells", *Blood*, vol. 108, no. 13, pp. 4009-4017.

Huang, H., Weng, H., Sun, W., Qin, X., Shi, H., Wu, H., Zhao, B.S., Mesquita, A., Liu, C. & Yuan, C.L. 2018, "Recognition of RNA N 6-methyladenosine by IGF2BP proteins enhances mRNA stability and translation", *Nature cell biology*, vol. 20, no. 3, pp. 285.

Jia, G., Fu, Y., Zhao, X., Dai, Q., Zheng, G., Yang, Y., Yi, C., Lindahl, T., Pan, T. & Yang, Y. 2011, "N6-methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO", *Nature chemical biology*, vol. 7, no. 12, pp. 885.

Kozak, M. 1986, "Influences of mRNA secondary structure on initiation by eukaryotic ribosomes", *Proceedings of the National Academy of Sciences of the United States of America*, vol. 83, no. 9, pp. 2850-2854.

Leppek, K., Das, R. & Barna, M. 2018, "Functional 5′ UTR mRNA structures in eukaryotic translation regulation and how to find them", *Nature Reviews Molecular Cell Biology*, vol. 19, no. 3, pp. 158.

Li, F., Zhao, D., Wu, J. & Shi, Y. 2014, "Structure of the YTH domain of human YTHDF2 in complex with an m 6 A mononucleotide reveals an aromatic cage for m 6 A recognition", *Cell research*, vol. 24, no. 12, pp. 1490.

Liu, J., Yue, Y., Han, D., Wang, X., Fu, Y., Zhang, L., Jia, G., Yu, M., Lu, Z. & Deng, X. 2014, "A METTL3–METTL14 complex mediates mammalian nuclear RNA N 6-adenosine methylation", *Nature chemical biology,* vol. 10, no. 2, pp. 93.

Liu, N., Dai, Q., Zheng, G., He, C., Parisien, M. & Pan, T. 2015, "N 6-methyladenosine-dependent RNA structural switches regulate RNA–protein interactions", *Nature,* vol. 518, no. 7540, pp. 560.

Liu, Y., Beyer, A. & Aebersold, R. 2016, "On the dependency of cellular protein levels on mRNA abundance", *Cell,* vol. 165, no. 3, pp. 535-550.

Manning, K.S. & Cooper, T.A. 2017, "The roles of RNA processing in translating genotype to phenotype", *Nature Reviews Molecular Cell Biology,* vol. 18, no. 2, pp. 102.

Meyer, K.D., Patil, D.P., Zhou, J., Zinoviev, A., Skabkin, M.A., Elemento, O., Pestova, T.V., Qian, S. & Jaffrey, S.R. 2015, "5′ UTR m 6 A promotes cap-independent translation", *Cell,* vol. 163, no. 4, pp. 999-1010.

Meyer, K.D., Saletore, Y., Zumbo, P., Elemento, O., Mason, C.E. & Jaffrey, S.R. 2012, "Comprehensive analysis of mRNA methylation reveals enrichment in 3′ UTRs and near stop codons", *Cell,* vol. 149, no. 7, pp. 1635-1646.

Morita, S., Noguchi, H., Horii, T., Nakabayashi, K., Kimura, M., Okamura, K., Sakai, A., Nakashima, H., Hata, K. & Nakashima, K. 2016, "Targeted DNA demethylation in vivo using dCas9–peptide repeat and scFv–TET1 catalytic domain fusions", *Nature biotechnology,* vol. 34, no. 10, pp. 1060.

Pichon, X., A Wilson, L., Stoneley, M., Bastide, A., A King, H., Somers, J. & E Willis, A. 2012, "RNA binding protein/RNA element interactions and the control of translation", *Current Protein and Peptide Science,* vol. 13, no. 4, pp. 294-304.

Ping, X., Sun, B., Wang, L., Xiao, W., Yang, X., Wang, W., Adhikari, S., Shi, Y., Lv, Y. & Chen, Y. 2014, "Mammalian WTAP is a regulatory subunit of the RNA N6-methyladenosine methyltransferase", *Cell research,* vol. 24, no. 2, pp. 177.

Shi, Z., Fujii, K., Kovary, K.M., Genuth, N.R., Röst, H.L., Teruel, M.N. & Barna, M. 2017, "Heterogeneous ribosomes preferentially translate distinct subpools of mRNAs genome-wide", *Molecular cell,* vol. 67, no. 1, pp. 71-83. e7.

Simsek, D., Tiu, G.C., Flynn, R.A., Byeon, G.W., Leppek, K., Xu, A.F., Chang, H.Y. & Barna, M. 2017, "The mammalian ribo-interactome reveals ribosome functional diversity and heterogeneity", *Cell,* vol. 169, no. 6, pp. 1051-1065. e18.

Stein, I., Itin, A., Einat, P., Skaliter, R., Grossman, Z. & Keshet, E. 1998, "Translation of vascular endothelial growth factor mRNA by internal ribosome entry: implications for translation under hypoxia", *Molecular and cellular biology,* vol. 18, no. 6, pp. 3112-3119.

Strahl, B.D. & Allis, C.D. 2000, "The language of covalent histone modifications", *Nature,* vol. 403, no. 6765, pp. 41.

Suzuki, M.M. & Bird, A. 2008, "DNA methylation landscapes: provocative insights from epigenomics", *Nature Reviews Genetics,* vol. 9, no. 6, pp. 465.

Topisirovic, I., Svitkin, Y.V., Sonenberg, N. & Shatkin, A.J. 2011, "Cap and cap-binding proteins in the control of gene expression", *Wiley Interdisciplinary Reviews: RNA*, vol. 2, no. 2, pp. 277-298.

Wang, X., Lu, Z., Gomez, A., Hon, G.C., Yue, Y., Han, D., Fu, Y., Parisien, M., Dai, Q. & Jia, G. 2014, "N 6-methyladenosine-dependent regulation of messenger RNA stability", *Nature*, vol. 505, no. 7481, pp. 117.

Wang, X., Zhao, B.S., Roundtree, I.A., Lu, Z., Han, D., Ma, H., Weng, X., Chen, K., Shi, H. & He, C. 2015, "N 6-methyladenosine modulates messenger RNA translation efficiency", *Cell*, vol. 161, no. 6, pp. 1388-1399.

Wei, C., Gershowitz, A. & Moss, B. 1975, "Methylated nucleotides block 5′ terminus of HeLa cell messenger RNA", *Cell*, vol. 4, no. 4, pp. 379-386.

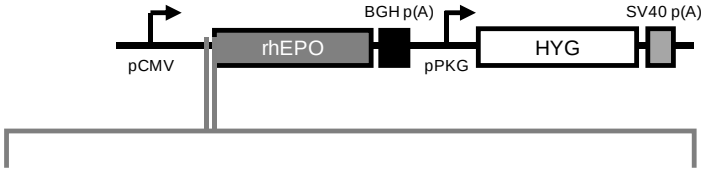
Wesselhoeft, R.A., Kowalski, P.S. & Anderson, D.G. 2018, "Engineering circular RNA for potent and stable translation in eukaryotic cells", *Nature communications*, vol. 9, no. 1, pp. 2629.

Yang, Y., Fan, X., Mao, M., Song, X., Wu, P., Zhang, Y., Jin, Y., Yang, Y., Chen, L. & Wang, Y. 2017, "Extensive translation of circular RNAs driven by N6-methyladenosine", *Cell research*, vol. 27, no. 5, pp. 626-641.

Zheng, G., Dahl, J.A., Niu, Y., Fedorcsak, P., Huang, C., Li, C.J., Vågbø, C.B., Shi, Y., Wang, W. & Song, S. 2013, "ALKBH5 is a mammalian RNA demethylase that impacts RNA metabolism and mouse fertility", *Molecular cell*, vol. 49, no. 1, pp. 18-29.

Zhu, T., Roundtree, I.A., Wang, P., Wang, X., Wang, L., Sun, C., Tian, Y., Li, J., He, C. & Xu, Y. 2014, "Crystal structure of the YTH domain of YTHDF2 reveals mechanism for recognition of N6-methyladenosine", *Cell research*, vol. 24, no. 12, pp. 1493.

a



>No A - CTRL
AAGCTTCGGTGCCGGTGCCTCGAGATG...

>m^{6}T
AAGCTTGGTCTAAAGCGGTCTTGTCTCGAGATG...

>m^{6}A
AAGCTTGGACTAAAGCGGACTTGTCTCGAGATG...
* *

b

ATGGGGGTGCACGAATGTCCTGCCTGGCTGTGGCTTCTCCTGTCC
CTGCTGTCGCTCCCTCTGGGCCTCCCAGTCCTGGGCGCCCCACCA
CGCCTCATCTGTGACAGCCGAGTCCTGGAGAGGTACCTCTTGGAG
GCCAAGGAGGCCGAGAATATCACGACGGGCTGTGCTGAACACTGC
AGCTTGAATGAGAATATCACTGTCCCAGACACCAAAGTTAATTTC
TATGCCTGGAAGAGGATGGAGGTCGGGCAGCAGGCCGTAGAAGTC
TGGCAGGGCCTGGCCCTGCTGTCGGAAGCTGTCCTGCGGGGCCAG
GCCCTGTTGGTCAACTCTTCCCAGCCGTGGGAGCCCCTGCAGCTG
CATGTGGATAAAGCCGTCAGTGGCCTTCGCAGCCTCACCACTCTG
CTTCGGGCTCTGGGAGCCCAGAAGGAAGCCATCTCCCCTCCCGAT
GCGGCCTCAGCTGCTCCACTCCGAACAATCACTGCTGACACTTTC
CGCAAACTCTTCCGAGTCTACTCCAATTTCCTCCGGGGAAAGCTG
AAGCTGTACACAGGGGAGGCCTGCAGGACAGGGGACAGATAA

Fig. 1 To investigate the effect of m6A on protein translation two putative m6A motifs (GGACT) were inserted upstream of the human EPO open reading frame. **a** An A to T mutant was also tested (m6T) and a no adenosine control (No A – CTRL). The sequence variants were cloned between HindIII and XhoI (underlined) in a pcDNA3.1(+) Hygromycin backbone. The sites of methylation are indicated by an asterisk and the EPO start codon highlighted in grey. **b** Nucleotide sequence of EPO used for evaluation of different 5' UTR sequences.

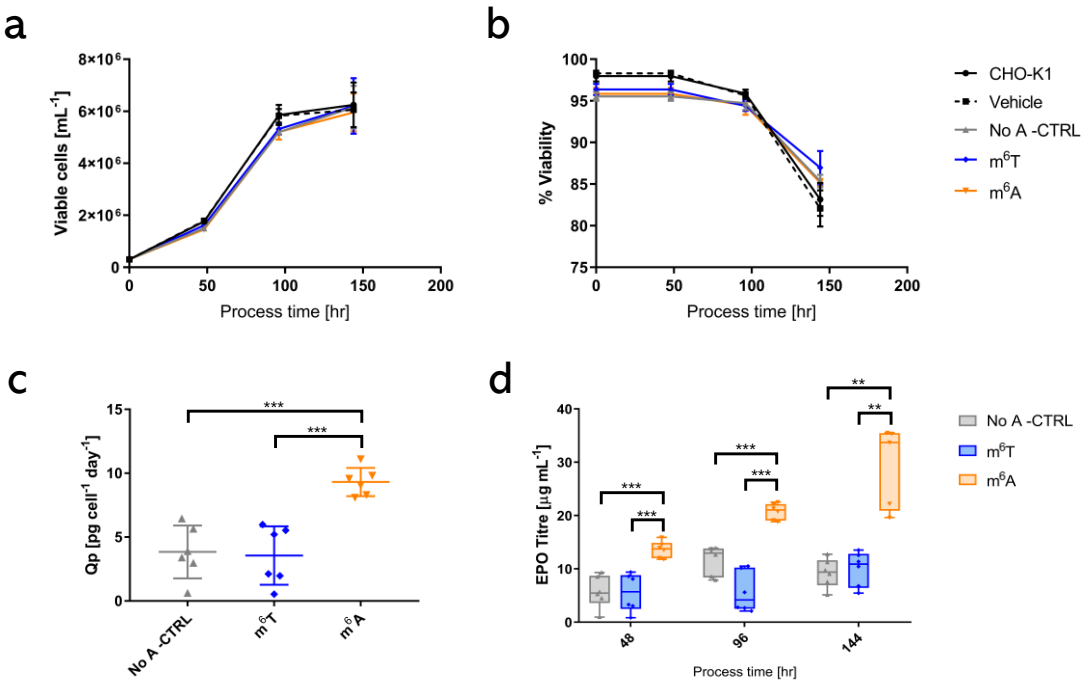


Fig. 2 CHO-K1 were transiently transfected with m6A motif variants and batch performance assessed. **a** Cell growth post transfection is displayed as viable cells per mL of culture. **b** The % viability of transfected cells during batch culture. **c** Cell specific productivity (Qp) of cells from 0 – 48 hr of culture. **d** Secreted rhEPO titer (µg mL^{-1}) of transcript variants was measured using supernatant harvested every 48 hr of culture. The values are based on six technical replicates over three biological replicates represented as the mean with standard deviation. Statistical analysis was done using a homoscedastic student t-test, ($p \leq 0.05$ *, $p \leq 0.01$**, $p \leq 0.001$ ***).

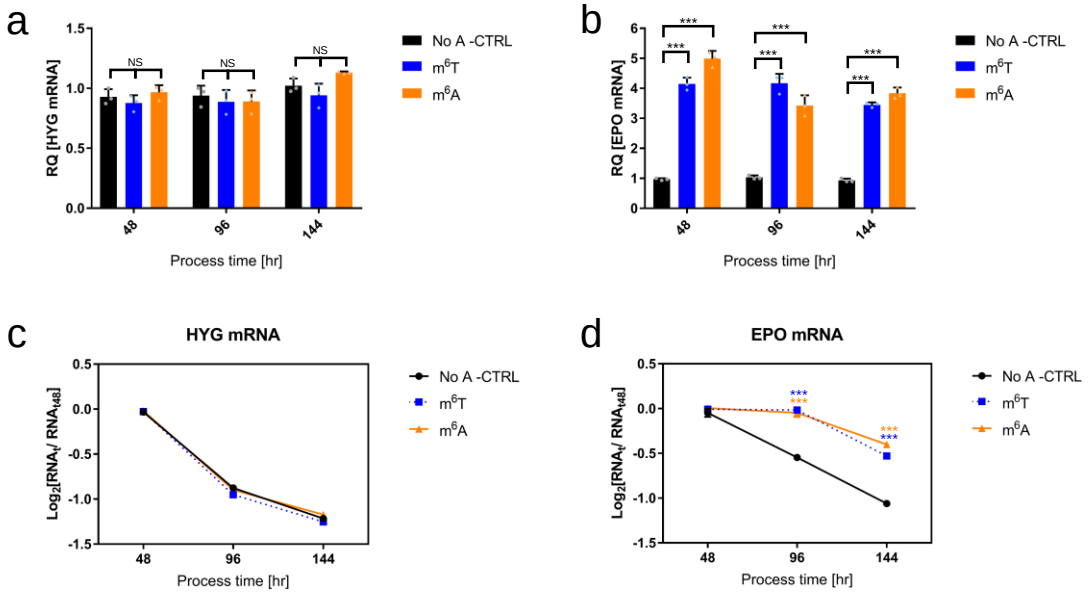


Fig. 3 qRT-PCR analysis of total RNA harvested from CHO-K1 cells transiently expressing one of the No A, m^6T or m^6A EPO mRNA transcript variants was run. **a** Relative quantification (RQ) of Hygromycin (HYG) mRNA was used as a control of transfection efficiency. RQ was done using the No A control as a calibrator at each time point of analysis. **b** RQ of EPO mRNA expression for all three transcript variants was calculated with respect to the No A control at each time point of analysis. **c** The rate of HYG mRNA loss over time during transient batch was calculated by the Log_2 fold reduction in relative mRNA abundance at a time point with respect to the initial reading at t = 48 hrs. **d** The Log_2 fold reduction in EPO mRNA loss over transient batch culture. Values seen are the mean of three biological replicates with error calculated by standard deviation. Statistical analysis was done using a homoscedastic student t-test, ($p \leq 0.05$ *, $p \leq 0.01$**, $p \leq 0.001$ ***).

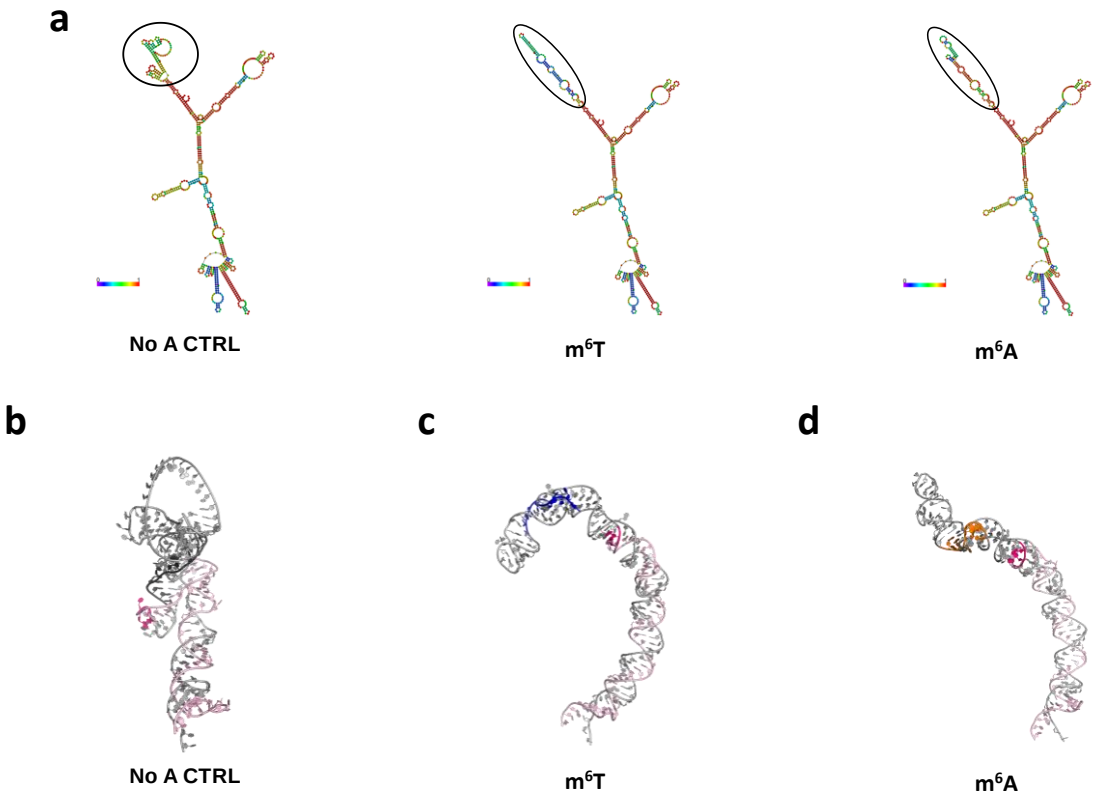


Fig. 4 The 5' UTR secondary and tertiary structures of the three sequence variants was evaluated. **a** RNAFold predictions of the 5' UTR and full EPO open reading frame for each transcript variant. Regions differing in secondary structure are circled. In all cases the difference occurred between the 5' UTR and the first 50bp of the EPO open reading frame. The tertiary structure from the transcriptional start site through the first 50bp of the open reading frame was analysed using RNAComposer and PyMOL. **b** Tertiary structure of the No A CTRL 5' UTR. The EPO start codon is highlighted in hot pink and the first 50bp of the open reading frame light pink. **c** Tertiary structure of the m^6T 5' UTR with m^6T motifs coloured blue. **d** Tertiary structure of the m^6A 5' UTR with putative m^6A sites coloured orange.