microRNAs as potential tools for ‘miR’aculous CHO cell phenotypes in Bioprocessing systems

A thesis submitted for the degree of Master of Science

Dublin City University

By

Ankur Solanki, B.Sc. (Hons)

The research work presented in this thesis was performed under the supervision of

Dr. Niall Barron and Prof. Martin Clynes

School of Biotechnology

Dublin City University

National Institute for Cellular Biotechnology (NICB)

January 2016

I hereby certify that this material, which I now submit for assessment on the programme of study leading to the award of Master of Science 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: _____ (Candidate)

Student ID No.: 58110801

Date: _____

This thesis work is dedicated to my loving and caring parents

Sheela Devi & Jasbir Singh Solanki

Acknowledgements

First and most importantly, I would like express my sincere gratitude to one of the most 'humble souls' and my mentor Prof. Martin Clynes for giving me an opportunity to pursue a research masters degree at NICB under his guidance. Secondly, I would like to thank Dr. Niall Barron for supervising and encouraging me, particularly during the times things did not work. It's ineffably hard to describe how great supervisor he is to me and to many I am sure. Moreover, a great human being in every sense of the word. I would like to thank him for sharing his knowledge, which trust me always encouraged me to learn more about science. Not only work related issues but also I would like to thank him for all those pesky visa letters I needed, which he had to write and sign. Trust me he will miss writing those (Nah! Never haha!).

Prof. Martin and Dr. Niall supported me during this whole journey at NICB; the list is endless, both professionally and personally. I would also like to thank them for guiding and suggesting me towards the next step in my life after this Masters and always offering to help anytime required by me.

Special Thanks to Dr. Nga Lao and Dr. Srinivas Rao Suda, who helped me during the troubleshooting of all those molecular biology techniques and made sure I get them to work.

I would like thank Dr. Claire Gallagher for her help during the FACS sort and Dr. Finbarr O'Sullivan with confocal microscope work.

Thanks to Shane McSweeney, my english-oirish buddy for constantly encouraging me that I will finish up in time. I would also like to thank Paul Kelly- Jack of all Trades ☺ for all the help he provided me during this Masters journey.

Thanks to the molecular biology mates from the Chemistry Group: Creina Slator, for helping me the first time with the 'Precious' Guava, Tadhg McGivern (took me time to pronounce your name) for sharing the same boat, same story regarding his and mine thesis progress during the later stages.

I would like to Thank everyone in the CHO group and second floor gang in general: Alan Griffith, Gemma Moore, Shane Kelly, Maria O'Sullivan, Alan Costello, Andrew McCann, Kevin Kellner, Mark Gallagher, Orla Coleman, Edel, Justine Miller, Michael O'Donohoe, Gillian Smith and of course the new Spanish and Indian gang in the centre: Berta, Ricardo, Jesus, Antonio and Krishna & Prashant (Pandatji - Jai Ho) for creating a friendly and fun-filled environment in the centre.

Also, I would like to thank the administrative staff of NICB (Real Heroes): Carol McNamara and Mairead Callan for all their support, take care of the money part and printing those visa letters on the NICB letter headed paper. Fun Task!

Very special thanks to Miss Anita 'Sweet' White (Nitaa) and Miss Rashika Yadav (Miss Energy) for supporting me during the tough time of writing my thesis and encouraging me for the same. Moreover, I would also like to thank Gaurav Verma, Kompal Joshi, Nicole Casiraghi (Weirdo) and Jennifer Casavantes for their lovely support everytime I explained where I am in terms of the finishing up.

Lastly, very very special thanks to my loving and caring family, especially my parents, Sheela Devi and Jasbir Singh Solanki for supporting me throughout my academics, without their constant blessing shower I would never have reached this far. Also, my sweet but mostly annoying sisters and brother: Anshika Rathee (Solanki suits much better btw!), Priyanka Solanki (Nerdie Panda) and Ajay (Johnnie, Nuts) Solanki, specially coming here to support me whilst I was writing.

Table of Contents

List of Abbreviations

2D-DIGE- 2 Dimensional-Difference gel electrophoresis

ACE- Angiotensin-converting enzyme

ADCC- Antibody-dependent cellular cytotoxicity

Ago- Argonaute

ARE- AU-rich elements

AT-III- Antithrombin III

ATCC- American Tissue Culture Collection

ATF4- Activating transcription factor

ATP- Adenosine triphosphate

Bak- B-cell lymphoma antagonist-killer protein

Bax- B-cell lymphoma associated protein

BCC- Business Communications Company

Bcl-2- B-cell lymphoma-2

Bcl-xL- B-cell lymphoma- extra large protein

BFP- Blue Fluorescent Protein

BGH- Bovine Growth Hormone

BMS- Bristol Myers Squibb

Bok- B-cell lymphoma ovarian killer protein

BSA- Bovine serum albumin

C&E- Carbon and Energy

Cas- CRISPR-associated genes

CDK- Cyclin-dependent kinase

cDNA- Complemtary Deoxyribonucleic acid

CHEF-1 CHO-derived elongation factor-1

CHO- Chinese Hamster Ovary

CIP- Calf-intestinal phosphatase

CMV- Cytomegalovirus

COSMC- C1GALT1 specific chaperone 1

CPP- Critical Process Paramter

CRISPR- Clustered regularly-interspaced short palindromic repeats

crRNA- Clustered regularly-interspaced short palindromic repeats ribonucleic acid

CTLA-4- Cytotoxic T-lymphocyte-associated protein 4

deGFP- Destabilized enhanced green fluorescent protein

DGCR8- DiGeorge syndrome critical region gene 8

DHFR- Dihyrofolate reductase

DMEM- Dulbecco's Modified Eagle Medium

DMSO- Dimethyl Sulfoxide

DNA- Deoxyribonucleic acid

DO- Dissolved Oxygen

ds-RNA- Double stranded ribonucleic acid

DSB- Double Stranded Breaks

EDTA- Ethylene Diamine Tetra Acetic acid

EF1α- Elongation Factor 1 aplha

EGF- Epidermal Growth Factor

eGFP- Enhanced Green Fluorescent Protein

EGFR- Epidermal growth factor receptor

ELISA- Enzyme-linked Immunosorbent Assay

EpCAM- Epithelial Cell Adhesion Molecule

EPO- Erythropoietin

ERK- Extracellular signal-regulated kinases

ERp27- Endoplasmic reticulum protein 27

EU- European Union

Exp5- Exportin 5

FACS- Fluorescence Activated Cell Sorting

Fc- Fragment crystallisable region

FCS- Flow Cytometry Standard

FDA- Food and Drug Administration

FITC- Fluorescein isothiocyanate

FSC- Forward Scatter Cytometry

FU- Fluorescence Units

FUT8- Fucosyltransferase 8

G418- Geneticin

GADD34- Growth arrest and DNA damage-inducible protein 34

GAPDH- Glyceraldehyde-3-Phosphate Dehydrogenase

GFP- Green Fluorescent Protein

GlcNAc- N-acetylglucosamine

GLUT5- Glucose transporter member 5

GOI- Gene of Interest

GS- Glutamine Synthetase

GTP- Guanosine triphosphate

HDR- Homology directed repair

HER2- Human epidermal growth factor receptor-2

HIV- Human Immunodeficiency Virus

hnRNP K- Heterologous nuclear ribonucleoprotein K

HoxD10- Homeodomain transcription factor homeobox D10

HR- Homologous Recombination

HRP- Horseradish peroxidase

hTPO- Human thyroid peroxidase antibody

HygB- Hygromycin B

IDT- Integrated DNA Technologies

IgG-1- Immunoglobulin G-1

IL- Interleukin

IMS- Industrial-methylated spirits

Indels- Insertions and Deletions

IRES- Internal ribosome entry site

Kb- Kilobases

KEGG- Kyoto Encyclopaedia of Gene and Genomes

LB- Lysogeny Broth

LDC- Limted Dilution Cloning

LDH- Lactate Dehydrogenase

LF- Laminar Flow

LNA-Locked Nucleic Acid

mAbs- Monoclonal Antibodies

Mcl-1- Myeloid cell leukemia 1

MCS- Multiple Cloning Site

$MgCl_2$- Magnesium chloride

miRNA- micro ribonucleic acid

miRNA*- micro ribonucleic acid star

MODC- Mouse ornithine decarboxylase confers

mRNA- Messenger ribonucleic acid

MS- Mass Spectrometry

MSX- Methionine Sulfoximine

MTX- Methotrexate

NaBu- Sodium Butyrate

NaCl- Sodium chloride

NaOH- Sodium hydroxide

NC- Negative Control

NCBI- National Center for Biotechnology Information

NEB- New England Biolabs

Neu5Ac- N-acetylneuraminic acid

NHEJ- Non-homologous end-joining

NLS- Nuclear localization signal

NMR- Nuclear Magnetic Resonance

Nt- Nucleotide

O/N- Overnight

OD- Optical Density

ORF- Open Reading Frame

PAGE- Polyacrylamide Gel Electrophoresis

PAM- Protospacer adjacent motif

PBS- Phosphate Buffer Saline

PDI- Protein disulphide-isomerase

PGK- Phosphogylcerate kinase

PI- Propidium Iodide

PI3K- Phosphatidyl Inositol 3-kinase

piRNA- piwi ribonucleic acid

PMT- Photo-multiplier tube

PNK- Polynucleotide kinase

PolyA-Polyadenylation

pre-miRNA

pri-miRNA

PS- Penicillin/Streptomycin

Psme3- Proteosome (prosome, macropain) activator subunit 3

r-tPA- Recombinant tissue plasminogen activator

Rad54l- Recombination protein 54-like

RAN- RAs-related Nuclear protein

RANK- Receptor activator of nuclear factor kappa-B ligand

rCHO- Recombinant chinese hamster ovary

rHIgG-1- Recombinant human Immunoglobulin G-1

RISC- RNA-induced silencing complex

RNA- Ribonucleic acid

RNAi- Ribonucleic acid interference

Rpm- Rotation per minute

RT- Reverse Transcription

RT-qPCR- Real Time Quantitative Polymerase Chain Reaction

RVD- Repeat-variable di-residues

S/MAR- Scaffold/matrix attachment regions

SAP- Shrimp alkaline phosphatase

SCC- Single Cell Clone

SDS- Sodium doceyl sulphate

SEAP- Secreted Alkaline Phosphatase

SFM- Serum-free medium

sgRNA- Single guide ribonucleic acid

shRNA- Short hairpin ribonucleic acid

siRNA- Small interfering ribonucleic acid

Skp2- S-phase kinase-associated protein-2

SM- Syl-1 and Mucn18 proteins

SNAP- Synaptosome-associated protein-8

SNARE- Soluble N-ethylmaleimide-sensitive factor attachment protein receptor

snRNA- Small nuclear ribonucleic acid

SOP- Standard Operating Procedure

SV40- Simian Virus 40

TAE- Tris base, acetic acid and EDTA

TALE- Transcription activator-like effector

TALEN- Transcription activator-like effector nuclease

TMB- Tetramethylbenzidine

TMMP- Targeted mediated miRNA protection

tPA- Tissue plasminogen activator

TPO- Thyroid Peroxidase Antibody

tracrRNA- Trans-activating ribonucleic acid

TRBP- TAR-RNA-binding protein

UCOE- Ubiquitous Chromatin Opening Element

UPR- Unfolded protein response

US- United States

UV- Ultraviolet

VAMP8- Vesicle-associated membrane protein 8

VCD- Viable Cell Density

VEGF- Vascular Endothelial Growth Factor

Wt- Wildtype

XBP-1- X-box binding protein 1

XIAP- X-linked inhibitor of apoptosis protein

XRN- Exoribonuclease

ZFN- Zinc Finger Nuclease

ZFP- Zinc Finger Protein

Abstract: miRNAs as potential tools for 'miR'-aculuous CHO cell phenotypes in Bioprocessing systems

Chinese Hamster Ovary (CHO) cells are the biopharmaceutical industry's "mini biofactories" for the production of complex, post-translationally modified therapeutic proteins. In order to address the ever-growing market need for these recombinant proteins, various genetic engineering tools have been employed. Here, we describe genetic approaches to improve CHO cell culture longevity, with a view to increasing overall process yield via the manipulation of two miRNAs: Let-7a and miR-7. Previous miRNA profiling studies in our laboratory and in the published literature helped in the identification of these miRNAs, which have shown to be disregulated in various tumor types and are key regulators of the cell cycle. Therefore, this stimulated the interest of our research group to manipulate these miRNAs in CHO cells with a view to positively impact bioprocess-relevant CHO cell phenotypes.

In the first approach, we used a Let-7 sponge decoy vector to deplete endogenous Let-7 levels with a view to increasing culture longevity and productivity of CHO-K1 SEAP expressing cells. Despite let-7 having a recognised role in deregulated cell growth no improvement was observed in stable, sponge-transfected clones. Out of a panel of 40 clones, we observed only two with improved cellular viability in 24 well plate format, however, the results were not reproduced in a 5 mL scale-up batch study. In the second approach, we targeted a previously verified miRNA for improved CHO cell growth and productivity i.e. miR-7, using a bacterial genome-editing tool, CRISPR-Cas9. A considerable amount of optimisation work was performed to establish the CRISPR system for use in the lab, initially using eGFP as model target gene in CHO cells. Finally we designed single guide RNAs to target Cas9 to the miR-7a-5p genomic locus to disrupt miR-7 in order to enhance growth of a CHO-K1 cell line producing an IgG-1. We estimated ~ 40% targeting efficiency of miRNAs using this approach. After an extensive screen, one stable clone was identified with what appeared to be a heterozygous deletion of one miR-7a copy. We demonstrate that CRISPR-Cas9 can be successfully used to target miRNA loci in the CHO genome but that functional knockout may be more difficult compared to protein coding genes.

Section 1.0

Introduction

1.1 Chinese Hamster Ovary cells – major "biofactories" for the production of biopharmaceuticals

Chinese hamster ovary (CHO) cells are the 'primary workhorse' of the biopharmaceutical industry with more than two-third (~ 70%) of current recombinant therapeutic products on the market produced in CHO cells. Owing to the inherent ability of CHO cells to produce human-like post-translational modifications and glycosylation patterns in the final protein product, they have earned the status of being the primary host choice for the production of recombinant therapeutic products (Jayapal et al. 2007a). CHO cells are favoured host for producing many recombinant therapeutic proteins over microbial counterparts, which lack the ability of to produce recombinant proteins with desired glycosylation patterns that are both compatible and bioactive in humans. The FDA approval of the first recombinant protein produced using CHO i.e. recombinant tissue plasminogen activator (r-tPA), (Activase commercially) in 1987 by Genentech, was the 'launchpad' for the manufacture of other CHO-based therapeutics. By 2011 there were 19 newly FDA approved biopharmaceuticals in the United States (US) and European Union (EU) combined (**Table 1.1** outlines list of some selected FDA/EU approved biologics in CHO) (Walsh 2011 and Walsh 2012). Of these 19 new recombinant products, 14 were produced using CHO expression systems (Walsh 2012). Globally, the current annual sales of recombinant therapeutics reported is ~ US $120 billion, of which US $70 billion sales per annum constituted by biologics produced using CHO cells. This emphasises how greatly the biotechnology industry depends upon CHO cells. Moreover, CHO cells will provide a large contribution to the development of biotechnology industry in future to meet the predicted US $150 billion world annual sales of biologics in 2015 (Jayapal et al. 2007a).

The growth of biopharmaceuticals has overtaken the growth of small molecule therapeutics. Monoclonal antibodies (mAbs) dominate the biologics market currently (**Figure 1.1**) in comparison to other recombinant therapeutic products. According to Business Communications Company's (BCC) research predictions the global sales of mAbs in 2016 will increase to US $58 billion by 2016 from US $44.6 billion in 2011 (BCC research 2012).

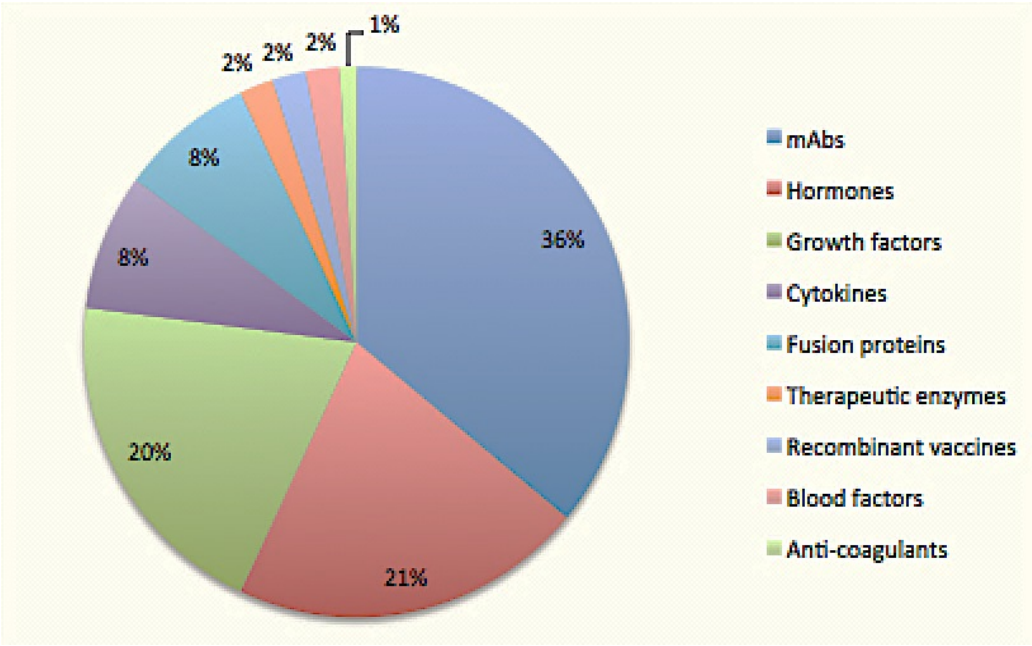


Figure 1.1: Biologics market distribution 2011 data. Data adapted from Aggarwal, S Nature Biotechnol. 29(12) 1083-1089, 2011.

Table 1.1: Recombinant therapeutics products produced using CHO cells as expression systems (Jayapal et al. 2007b, Animal Cell Technology Industrial Platform 2012 & 2013)**.**

Biologics	Type	Therapeutic Application	Company	Approved
Zaltrap (aflibercept)	Fusion Protein consisting of VEGF receptors 1 and 2 binding domains fused to an IgG Fc.	Metastatic colorectal cancer	Regeneron /Sanofi-aventis	2013 (EU) 2012 (USA)
Rixubis	Recombinant coagulation factor IX	Hemophilia B	Baxter	2013 (USA)
Perjeta (pertuzumab)	Humanized mAb specific for HER2	Breast Cancer	Roche (Genentech)	2013 (EU) 2012 (USA)
Ovaleap (follitropin alfa)	Biosimilar recombinant human follicle stimulating hormone.	Infertility /subfertility	Teva	2013 (EU)
NovoEight (turoctocog alfa)	Modified factor VIII	Haemo-philia A	Novo	2013 (EU & USA)
Gazyva (obinutuzumab)	Humanized, glycol-engineered mAb specific for CD20 expressed on B lymphocytes.	Chronic lymphocytic leukemia	Roche (Genentech)	2013 (USA)
Yervoy	Human mAb binds to	Melanoma	Bristol-Myers	2011 (EU & USA)

(Ipilimumab)	cytotoxic T-lymphocyte-associated protein 4 (CTLA-4, a negative regulator of T-cell activation)		Squibb (BMS)	
Nulojix (belatacept)	Fusion protein consisting of extracellular domain of human CTLA-4 fused to IgG Fc.	Prophylaxis of organ rejection following kidney transplant	BMS	2011 (EU & USA)
Eylea (aflibercept)	Fusion protein consisting of extracellular ligand binding domains of VEGF receptor fused to IgG Fc	Neovascular (wet) age-related macular de-generation	Regeneron	2011 (EU & USA)
Adcetris (brentuximabveodtin)	Chimeric mAb conjugate specific for human CD 30 (expressed on the surface of lymphoma cells)	Lymphoma	Seattle genetics	2011 (USA)
Prolia (denosumab)	Human mAb specific for Receptor activator of nuclear factor kappa-B ligand (RANK) ligand	Osteoporo-sis in post-menopausal women	Amgen	2010 (EU & USA)
Lumizyme (alglucosidas-e alfa)	Recombinant human acid-alfa-glucosidase	Pompe disease (glycogen storage disease type III)	Genzyme	2010 (US)
Elonva (corifollitrop-in alfa)	Modified recombinant human follicle stimulating hormone	Controlled ovarian stimulation	N.V Organon	2010 (EU)
Stelara (usetekinumab)	Humanized mAb for interleukin (IL)-12/23	Psoriasis	Centocor/Johnson & Johnson (J&J	2009 (EU & USA)
Abseamed (epoeietin-alfa)	Hormone Red blood cell progenitor	Anaemia	Medice Arzneimittel Putter	2007 (EU)
Vectibix (panitumuma-b)	Humanized mAb Epidermal growth factor (EGF) receptor	Colorectal cancer	Amgen	2006 (USA) 2007 (EU)
Proxinium (catumaxom-ab)	Humanized mAb Epithelial cell adhesion molecule (EpCAM) target	Head and Neck cancer	Viventia Biotech	2005 (EU & USA)
Aldurazyme	Enzyme alfa-L-iduronidase	Mucopolysacch	Genzyme	2003 (EU & USA)

(iduronidase)		aridosis type I (MPS-I)	/Biomarin	
Opgenra (eptotermin alfa	Osteogenic protein 1	Alternative to autograft	Stryker Biotech	2002 (EU) 2001 (USA)
Herceptin (trastuzumab)	Chimeric mAb for human epidermal growth factor 2 (HER2) receptor	Breast cancer	Genentech	2000 (EU) 1998 (USA)
Epogen/ Procrit	Cytokine/receptor antagonist/growth factor	Anaemia	Genentech	1989 (USA)
Activase (alteplase)	Enzyme for tPA as target	Acute myocardial infection	Genentech	2002 (EU) 1987 (USA)

Therefore, in order to meet the ever growing global biologics market demand, the production process for biologics using the well-accepted CHO cell line still needs to be optimized. Over the past two decades, product titres have been improved considerably. This is mainly attributed to medium development as well as optimized feed strategies and approximately 20-fold due to improved vector expression systems and producer clone isolation process. However, the advent of complex therapeutics such as antibody fragments, fusion proteins and others, has presented new challenges to the current process of bio-manufacturing recombinant proteins using CHO cell systems. Therefore, there is still an issue to be addressed in the area of optimization of biologics production process (Hacker, De Jesus and Wurm 2009, De Jesus and Wurm 2011, Butler and Meneses-Acosta 2012)

In the past, there have been various approaches to augment CHO cell performance both at the bioprocess engineering level such as using biphasic mode of CHO cell cultivation and at the molecular biology level i.e. single and/or multi-gene, vector engineering and other approaches to establish a robust and stable CHO cell line. In terms of manipulating the bioprocess and maintaining a stable 'happy' microenvironment, critical process parameters (CPPs) such as pH, temperature, osmolality, dissolved oxygen (D.O), feed supplement are monitored and regulated accordingly (Kaufmann et al. 1999, Yoon, Song and Lee 2003, Fox et al. 2004, Kumar et al. 2008). For instance, specific productivity in the CHO cell culture production processes has been enhanced by using serum-free media (Butler 2005a). The combined use of hyper-osmolality and/or a low temperature strategy with the chemical supplementation of histone deacetylase inhibitors such as sodium butyrate or valproic acid have also been employed to boost specific productivity (Matasci et al. 2009, Kantardjieff et al. 2010a, Sunley and Butler 2010).

At the genetic engineering level, single and/or multi-gene targeting strategies to target various genes in CHO cells with a view to improve specific productivity include cell engineering through anti-apoptosis (Kim and Lee 2000a, Sauerwald, Oyler and Betenbaugh 2003, Sung et al. 2007), autophagy (Hwang and Lee 2009), cell-cycle engineering (Mazur et al. 1998a, Astley and Al-Rubeai 2008). Various protein-based engineering approaches such as chaperone (Hwang, Chung and Lee 2003, Chung et al. 2004a, Mohan et al. 2007), unfolded protein response-based (UPR) (Tigges and Fussenegger 2006), secretion (Florin et al. 2009), metabolic engineering (Zhang et al. 2006, Wlaschin and Hu 2007) and glycosylation engineering (Ferrara et al. 2006)

have also been explored. One of the other important genetic engineering approaches is vector engineering (Wurm 2004, Davies et al. 2013). Most of these approaches have been discussed in separate sections later.

Glycosylation plays a central role in controlling the composition of glycans, which enhance the pharmacological properties of therapeutic recombinant protein products. For example, the effector function of an antibody i.e. antibody-dependent cell cytotoxicity (ADCC), in the case of the major blockbuster drug Herceptin™, has been elevated by reducing the core fucosylation of the biantennary-glycan on the Fc region of antibody. The maintenance and manipulation of the CPPs to cater for a 'happy' culture environment for CHO cells is required since CPPs could impact on the recombinant protein quality, which in turn could seriously impact the bioactivity profile of these biological molecules. Therefore, engineering CHO cells as well as the bioprocess can result in increased volumetric or specific productivity of the therapeutic protein products with desired pharmacological properties (Zanghi et al. 1998, Yang and Butler 2000, Okazaki et al. 2004).

1.2 CHO-mics'

1.2.1 Genomics

The success story of CHO cells, so far, has also been further enhanced by the publication of 'omics data for mouse (*Mus musculus*- as a reference sequence), Chinese hamster (*Cricetulus griseus*) and CHO cells (Xu et al. 2011, Lewis et al. 2013). Owing to the importance of CHO cell lines in the biotechnology industry, there has been a lot of focus on generating CHO genome data. For instance, in an effort of extracting maximal genomic data information, Xu et al. (2011) published the genomic sequence of CHO-K1 parental cell line and assembled 2.45 Gb of genomic sequence predicting 24,383 genes. They connected most of the assembled scaffolds with 21 chromosomes in order to unravel their chromosomal genes locations. Moreover, they reported that the homologs of most human glycosylation-associated genes are present in the CHO-K1 genome, 141 (~ 53 %) of which are not expressed during exponential growth phase. Such information of the glycosylation-associated genes could augment the process of glyco-engineering in CHO cells. With more than 350 therapeutic products, mainly monoclonal antibodies to be produced with CHO cells as hosts,

currently in the pipeline for pre-clinical trial (Reichert 2012 and 2013), the need for the availability of CHO-K1 genomic sequence remains an objective of prime importance. The potential of CHO-mics' is yet to be fully realized.

1.2.2 Other Omics tools

1.2.2.1 CHO-based Transcriptomics

In last decade, a considreable amount of CHO transcriptomic data has been explored. For instance, Birzele et al. (2010) performed CHO expression profiling and identified 13,375 CHO genes and more than 6000 transcripts were predicted to be full-length and covered 95 % of the corresponding mouse orthologs. However, only a part of this data (non-annotated) has been deposited in the National Center for Biotechnology Information's (NCBI) database. The limited annotation and availability of the assembled CHO genomic sequences data deposited and released in public databases continues to be a challenge (Birzele et al. 2010, Becker et al. 2011). In order to address the issue of limited publicly available CHO genome data, Becker et al. (2011) made use of a diverse set of cDNA sequences i.e. ~ 29000 transcripts from CHO cell lines. The researchers assembled and compared these transcripts to *M.musculus* to study the distinction between different isoforms in these two species. The researchers generated 13,187 CHO cell line transcripts with functional annotations, classified them into isogroups (possible genes) and isotigs (splice variants) (Becker et al. 2011).
The utilization of Omics tools assisting in the development of rCHO cell line process have been reviewed in great detail by (Gupta and Lee 2007, Kim, Kim and Lee 2012). Detailed detection of the global changes in DNA/RNA, metabolites, protein and the desired glycosylation patterns is now possible with recent advances in Omics based tools. The impact of these global changes could provide a better understanding of the molecular mechanisms associated with particular CHO cell phenotypes, observed under certain environmental conditions. For instance, the perturbations on specific productivity (q) and specific growth rate (μ) with the addition of chemicals such as NaBu, DMSO, zinc sulphate, cytochalasin, tunicamycin and growth factors have been investigated in many studies (Van Dyk et al. 2003, Hayduk and Lee 2005, Li et al. 2006, Joon et al. 2008, Kantardjieff et al. 2010b). However, due to the lack of sufficient CHO genome sequence information, as mentioned above, non-CHO

specific DNA arrays for mouse and rat have been used for comparative transcriptomic analysis, which are useful but not ideal for CHO transcriptome analyses. But there are reports wherein DNA microarrays have been successfully utilised to study rCHO cells under q enhancing conditions such as low culture temperature (Kaufmann et al. 1999), NaBu treatment (Joon et al. 2008) and high osmotic condition (Shen et al. 2010). Similar studies focused on the enhancement of μ and/or high q by studying different transcripts expressed across other rCHO cell lines have also been performed (Nissom et al. 2006, Wong et al. 2006, Doolan et al. 2008, Doolan et al. 2010). CHO-specific analysis tools could yield more accurate and reliable differential expression information, which could also augment the reliable identification of CHO-specific proteins associated with certain genes (Meleady et al. 2012). In another effort, Courtes et al., (2013) performed a translatome analysis by probing for 13 K annotated CHO-specific genes using high resolution and streamlined profiling technology. The researchers successfully identified key genes, which are potential key players for cellular growth, namely: heterogenous nuclear ribonucleoproteins, protein regulator of cytokinesis 1, glucose-6-phosphate dehydrogenase, UTP6 small subunit processome and RuvB-like protein 1 (Courtes et al. 2013). In a recent study, CHO RNA-Seq data was utilised from sequencing efforts by Vishwanathan et al. (2015) to study the transcriptomic data to provide physiological insights related CHO cell performance during production processes. The researchers assembled and annotated set of RNA-Seq data to identify genes involved in major functional pathways such as metabolism, glycolysis and glycosylation (Vishwanathan et al. 2015).

1.2.2.2 CHO-based ProteOmics

Transcriptomic data is often integrated with proteomic data, thus allowing for better understanding of the molecular mechanisms driving certain types of rCHO cells phenotypes. For example, an approach such as two-dimensional gel electrophoresis (2D-DIGE) combined with mass spectroscopy (MS) has been widely used to identify proteins modified under specific conditions in rCHO cells (Lee, Harrington and Bailey 1996). Proteomic approaches to monitor global protein changes caused by overexpressing and/or down regulation of an effector protein, for example, overexpressing B-cell lymphoma-extra large (Bcl-xL) (Carlage et al. 2009, Baik and Lee 2010) or E2F Transcription factor-1 (E2F-1) (Lee et al. 1996) in recombinant

CHO cells have been performed with a view to impact cell growth positively. A number of proteomic studies in rCHO cells have been carried out in order to identify differentially expressed proteins during hyperosmotic pressure (Lee et al. 2003), apoptosis induction (Baik and Lee 2010), culture in serum-free media with optimized hydrolysates mixtures (Kim et al. 2011a).

1.2.2.3 CHO-based Metabolomics

Another crucial Omics based tool attracting attention is metabolite profiling in rCHO cell cultures. Metabolomics can aid in media development and process feeding strategies. 100-fold product titre enhancement, in the last few decades, is largely attributed to the type of media used and feeding strategies. Metabolomics studies are performed with a combination of tools such as liquid chromatography-MS or nuclear magnetic resonance (NMR to monitor extracellular or intracellular metabolite flux for media development with view to, for example, reduce the likelihood of early onset of apoptosis in rCHO cell culture (Bradley et al. 2010, Goudar et al. 2010). For instance Chong et al. (2010) reported that malate accumulation in the medium could prevent rCHO attaining high cell densities. This revelation and subsequent overexpression of malate dehydrogenase II led to significant growth enhancement in integral viable cell counts (Chong et al. 2010).

1.2.2.4 CHO-based Glycomics

Post-translational modifications regulating recombinant therapeutic protein function also play a key role in affecting cell phenotype. Engineering approaches have also diversified into the field of glycosylational engineering for the improvement. For example, most mAbs possess complex glycoforms with fucosylated bi-antennary glycans with trimannosyl core capped by an N-acteylgalactosamine (GlcNAc), galactose and N-acteylneuraminic (Neu5Ac) residue on each branch. Many other glycoforms found to be afucosylated, have terminal or bisecting GlcNAc, with 5-9 mannose residues for enhanced ADCC (Mimura et al. 2007), while enhanced sialylation of recombinant EPO through addition of two extra glycan sites resulted in enhanced serum half-life of the recombinant protein (Kiss et al. 2010).

These Omics tools, when combined, could also allow for the identification of differentially expressed genes and enzymatic reactions when a parental cell-line

transitions from being a non-producer to producer cell-line (Dietmair et al. 2012). In a nutshell, the integration of various 'Omics data could provide for system biology levels insights into the underlying complexity inherent in CHO cell physiology and shed light on targets for metabolic engineering intervention (**Figure 1.2**).

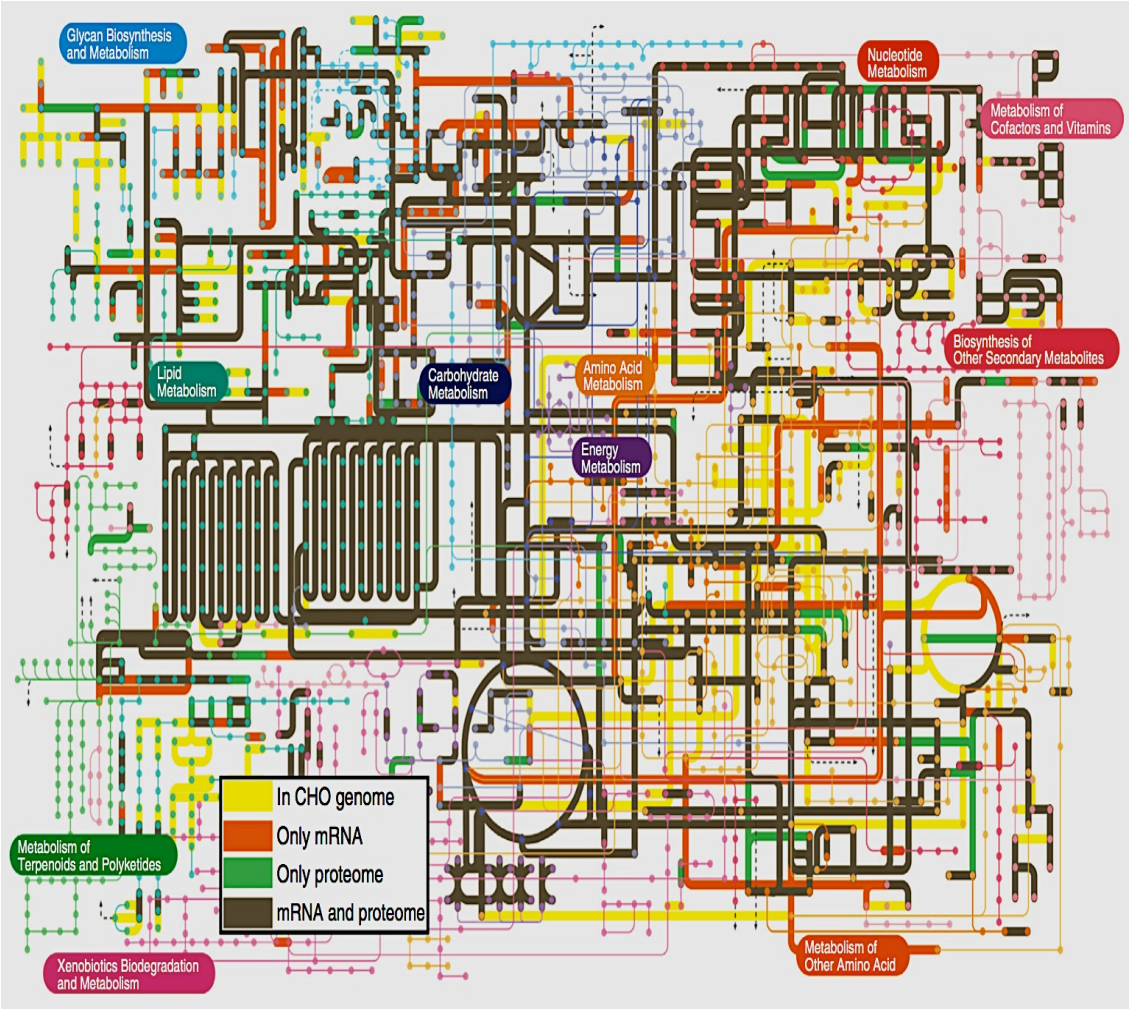


Figure 1.2 represents a system biology mapping with the genome, transcriptome and proteome of CHO cell-lines mapped to all metabolic pathways in the Kyoto Encyclopedia of Genes and Genomes (KEGG). Image was sourced from a review by (Kildegaard et al. 2013).

1.3 Improving CHO cell productivity

1.3.1 Stable GOI integration – An uphill task

Generating a therapeutic protein producing cell line starts with the process of stably transfecting the cell line with the gene encoding the protein of interest. Most of the

cell lines exhibit variations in expression of proteins and more often these variations are associated with the genetically engineered cell-lines in contrast to parental cell-lines (Browne and Al-Rubeai 2007a). In genetically engineered cells, random integration of the recombinant GOI gives rise to variable expression patterns due to 'positional' effects and present a serious bottleneck both in terms of time (i.e. by forcing extensive screening of large number of clones) and resources during recombinant CHO cell line generation. This increased heterogeneity is particularly due to the fact that only 0.1 % of the genome is considered 'hot-spot' or 'transcriptionally active' (Little 1993), therefore the decreasing the frequency of GOI insertion into such a rare transcriptionally active zone of the genome.

1.3.2 Selection and Amplification Systems for CHO cell line development

In the biopharmaceutical sector, mostly mutant CHO cell lines are utilised for biologics production processes following the GOI transfection into the cells. **Table 1.3.2a** provides a list of selected commercially available CHO cell lines. These CHO cell lines are either dihyrdofolate reductase negative (DHFR- mutated or deleted versions of the DHFR gene) or glutamine synthase negative (GS-) (Liu et al. 2010b). The DHFR and GS gene-based selection and amplification systems are the two most commonly employed during the CHO cell line development. The methodologies of these two systems have been described in detail in a review from Camire (2000). The selection of positive transfectants is attained through the introduction of GOI along with a copy of DHFR or GS gene that complements the auxotrophy. Medium used for GS and DHFR-based are medium without glutamine and hypoxanthine supplements, respectively. Cells are further isolated with stringent selection through addition of methotrexate (MTX) or methionine sulfoximine (MSX) selective agents in increasing concentrations. The advantage of using GS-based system for gene amplification is that the high-producer clones can be produced within 3 months in comparison to DHFR-based system, which takes 5-6 months. Moreover with the GS-based system, another advantage is that little optimisation is needed when supplementing the culture media with glutamine (Butler 2005b).

There are alternative approaches to these selection methods including geneticin (G418), hygromycin B (HygB), zeocin, blasticidin or puromycin genes conferring resistance to these selective antibiotic reagents in culture (Butler 2005a). Li et al.

(2010) reviewed and presented a list of selection markers used in the **Table 1.3.2b**.

Table 1.3.2a: List of selected commercial CHO cell-lines used for research and industrial purposes.

Commercial CHO cell lines	Source
CHO-DXB11 or CHO-DUK-XB11 or CHO-DUKX	CHO-K1 strain mutagenized to generate this DHFR deficient cell-line
CHO-K1	Parental CHO cell-line
CHO DP12	Co-expresses light and heavy chains of murine 6G4.2.5 mAb (American Type Culture Collection (ATCC)-HB-11722).
CHO DG44	Parental CHO cell-line (ATCC) mutagenized to cause deletions in both the DHFR alleles.
CHO-S	CHO cell-line adapted to serum-free culture conditions and produces high levels of recombinant protein.

Table 1.3.2b: Selected list of Selection Markers used for CHO cell-based Gene Selection and Amplification process.

Selectable Marker	Selective Reagent
Metabolic selectable marker (most commonly utilized)	
Dihydrofolate reductase (DHFR)	Methotrexate (MTX)
Glutamine synthase (GS)	Methionine sulfoximine (MSX)
Antibiotic selectable marker	

Puromycin acetyltransferase	Puromycin
Blasticidin deaminase	Blasticidin S
Histidinol dehydrogenase	Histidinol
Hygromycin phosphotransferase	Hygromycin B
Zeocin resistance gene	Zeocin
Bleomycin resistance gene	Bleomycin
Aminoglycoside phosphotransferase	Geneticin (G418)

1.3.3 Automated systems for the isolation of CHO clones

The process of selecting clones for large-scale production is done more efficiently via high-throughput cell screening techniques, reviewed in great details in (Browne and Al-Rubeai 2007b, Kim, Kim and Lee 2012), and includes FACS, the ClonePix™ system (Genetix), the Cell Xpress™, the CellCelector™ system (Aviso) and the Laser-Enabled Analysis and Processing (LEAP™) system (Cyntellect). All these advanced technologies allow for the selection of clones with high GOI expression levels, whose performances are further tested in serum-free suspension fed-batch cultures.

1.3.4 Vector Engineering

An area of optimization for increased product yield include vector engineering, which facilitates rCHO cell line development by modulating transcriptional activity via insertion of various DNA elements into a vector. These techniques are discussed further in **Section 1.3.4 and 1.4** in more detail. The surroundings around the chromosome exert strong influences on the promoter (e.g. CpG islands are CG rich DNA regions wherein gene promoters are often located) that could potentially affect GOI expression. CpG islands are epigenetic regulatory hotspots for methylation (West and Fraser 2005). The availability of the CHO genomic sequence should facilitate the identification and pre-tagging of transcriptionally ‘active’ zones using site-specific integration techniques. Recombinase-mediated cassette exchange utilizes Cre & Flp

recombinases recognizing loxP and FRT sequences and artificial-chromosome engineering (ACE) system are examples of this (Huang et al. 2007a, Kennard et al. 2009, Kameyama et al. 2010).

1.3.4.1 Remodeling chromatin in an ‘Active-state’

In the quest to generate a stable hyper-producer cell-line, other techniques to remodel chromatin in an active state for increased stable transgene expression and increasing the gene copy number, have also been entertained (**Figure 1.3)**. These techniques include transposable elements (Ley et al. 2013), *trans*- and *cis*- genetic elements such as scaffold/matrix attachment regions (S/MARs), and ubiquitous chromatin opening elements (UCOEs). Other techniques and strategies that allow for selection marker attenuation offer greater probability of isolating high-producer clones from low-producers. For example, AU-rich elements (ARE) and murine ornithine decarboxylase (MODC) PEST regions combo-use for mRNA and protein destabilization, respectively have yielded positive results by reducing the selection marker during MTX amplification. Moreover, when both ARE and MODC PEST used with MTX selection system, the specific productivity increased from 14-fold to 27-fold (Ng, Wang and Yap 2007). In a similar attempt to reduce the selectable marker gene expression downstream, an attenuated internal-ribosome entry site (IRES) element as a substitute for ARE, has also been employed for high recombinant protein product titers in CHO cells (Ng et al. 2012).

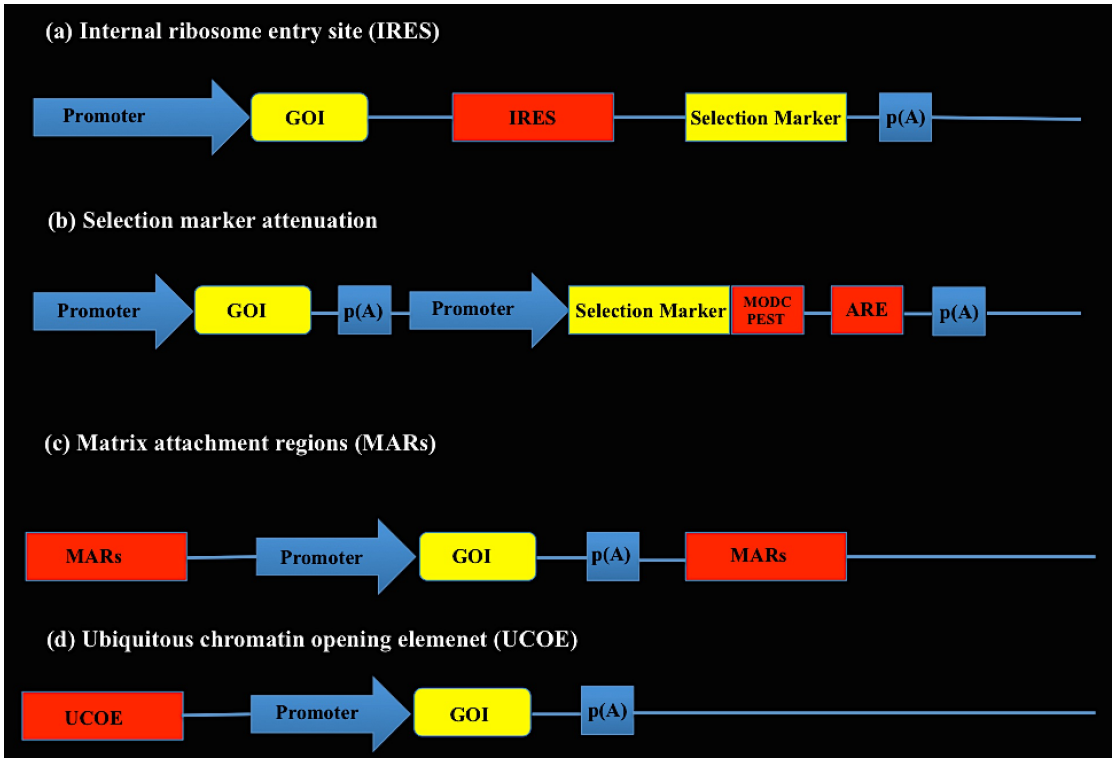


Figure 1.3 outlines vector-engineering approaches to address the issue of low-production capacity of CHO cells by stable and sustained expression of transgenes using IRES, MARs, UCOEs and other *cis-trans* chromatin remodeling techniques. Image reproduced from a review by (Lai, Yang and Ng 2013).

1.3.4.2 CHO Promoters for high transcriptional activity

The choice of promoter used is crucial in terms of achieving consistently high expression of heterologous genes within CHO cells. The most commonly employed promoters in vector expression systems in mammalian cells include cytomegalovirus (CMV) and simian viral 40 (SV40) promoters for ensuring high transcriptional rates (Spenger et al. 2004). Moreover, human elongation factor 1α (EF1α) is also a strong promoter for gene expression in mammalian cells (Deer and Allison 2004, Rita Costa et al. 2010). However, there are issues associated with the use of some the virus-based constitutive promoters since the over-expression of the GOI could lead to permanent cell-stress and could even lead to gene silencing as a common host response, which can lead to heterogeneity in expression patterns within a cell population (Kim et al. 2011b). Such unwanted effects can be prevented with the use of endogenous promoters derived from house keeping genes such as CHO-derived elongation factor-1 (CHEF-1) promoter. CHEF-1 has been reported to maintain high expression levels

for a multitude of genes in different CHO cell-lines (Deer and Allison 2004). Additionally, native CHO promoters such as *s100a6* (calcyclin), which is a cold-shock responsive promoter reported to be an excellent alternative choice to inducible promoter systems (that require addition of chemicals to the media). Calcylin promoter has been reported to induce 3-4 fold increase basal productivity in CHO cells (upon temperature shift to 33 °C) in comparison to SV40 promoter (Thaisuchat et al. 2011).

1.4 Bioprocess-relevant CHO cell engineering

1.4.1 Cell Cycle pathway engineering

The use of small chemical molecules as selective cell cycle inhibitors can mediate a complete and sustained G0/G1 growth phase arrest without impacting the G2/M phase by directly targeting cyclin dependent kinases (CDK 4/6) (Du et al. 2014). This approach works even better than the commonly employed physical parameter manipulation approaches such as low temperature and hyper-osmolality. Chemical perturbations (NaBu) to induce cell cycle arrest and eventually enhancing specific productivity (Sunley and Butler 2010) have also been explored by researchers. Using a selective inhibitor to induce growth arrest has been utilized to optimize growth, q (i.e. increased q to 110 pg/cell/day (Du et al. 2014)) and glycosylation profile (i.e. decreased mannose content (Du et al. 2014)) across different rCHO cell lines. Similar studies with cell-cycle mediators such as $p27^{KIP1}$ and $p21^{CIP1}$ have been performed in the past to induce rCHO-K1 SEAP cell growth arrest with the aim to achieve high q (Fussenegger, Mazur and Bailey 1997, Mazur et al. 1998b, Meents et al. 2002b).

1.4.2 Apoptosis pathway engineering

Apoptotic signals within cells are mediated by caspase-cascade signalling system. Therefore, the suppression or by-pass of caspase activation can be a fruitful strategy to positively influence and/or enhance rCHO cell growth. Overexpression of anti-apoptotic proteins such as those from the B-cell lymphoma-2 (Bcl-2) family proteins-Bcl2, Bcl-xL and Myeloid cell luekemia-1 (Mcl-1) and the suppression of pro-apoptotic Bcl-2 associated x protein (Bax-like) proteins (Bax, Bcl-2 antagonist killer protein (Bak) and Bcl-2 ovarian killer protein (Bok)) have been reported to result in increased recombinant protein production (Kim and Lee 2000b, Meents et al. 2002a, Chiang and Sisk 2005) For instance, Majors et al. (2009) observed that Mcl-1

overexpression in CHO cells led to a significant 90 % improved cell viability for 14 days. This improved viability was coupled with 20-30 % increase in antibody production (Majors et al. 2009). Moreover, Cost et al. (2010) used CHO cell lines with two important Bcl-2 family proteins deleted, resulting in significant higher cell densities in large-scale culture (Cost et al. 2010a). Combinatorial strategies of anti-apoptosis and secretion engineering have been reported to yield a CHO cell line that expresses X-box binding protein-1 (XBP-1), which is a potent stimulator of promoters of secretory pathway genes leading to increased protein synthesis (as a result enhancing q, however, reported to reduce cell viability too) and x-linked inhibitor of apoptosis (XIAP) protein (whose overexpression can alleviate the negative impact of XBP-1) to inhibit apoptosis. This resulted in 60 % increased production titres (Shaffer et al. 2004, Becker et al. 2010). RNA-interference technology has also been employed in providing resistance to apoptosis in rCHO cells. Lim et al. (2006) has reported an enhancement in cellular viability and productivity of CHO cells using small hairpin RNA (shRNA) technology to knockdown Bax and Bak of the Bcl-2 protein family.

1.4.3 Secretion engineering

Cell line engineering approaches have also been employed for the enhancement of protein secretion since protein productivity by rCHO cells is often hindered by a problem in the saturated secretory pathway. Soluble N-ethylmaleimide-sensitive factor attachment protein receptors (SNAREs) play a central role in vesicle trafficking (Jahn and Scheller 2006, Toonen and Verhage 2003). Syl1/Munc18 (SM) proteins are essential key regulators of SNARE-mediated fusion event. In secretion engineering in rCHO cells components associated with the secretary pathway i.e. SNAREs synaptosome-associated protein of 23kDa (SNAP-23) and vesicle-associated membrane protein 8 (VAMP8) and SM proteins (Syl 1 and Munc18c) were introduced in an effort to successfully elevate q (Peng and Fussenegger 2009, Peng, Abellan and Fussenegger 2011). In another study, researchers used phosphorylation-resistant ceramide transfer protein (CERT) along with XBP-1 (overexpressed) in rCHO cells. CERT is involved in the trafficking of ceramide from endoplasmic reticulum (ER) to Golgi. A 35 % increase in specific productivity of t-PA ($q_{t\text{-}PA}$) was observed in rCHO cells (Rahimpour et al. 2013).

1.4.4 Chaperone engineering

Enhancing q as well as efficiently producing an authentic therapeutic protein is compromised by several processes including transcription, translation and proper protein folding. The repercussions of any over-expression strategy could potentially compromise the protein folding machinery resulting in increased production of misfolded proteins. Misfolded proteins in turn could cause potentially cytotoxic protein aggregation. However, mammalian cells are equipped with molecular chaperones in the ER to ensure newly synthesized proteins are correctly folded (Schröder, Schäfer and Friedl 2002, McClellan et al. 2005, Tartaglia et al. 2007). rCHO cells overexpressing protein disulphide isomerase (PDI) enhanced q in a product-specific manner. For example, rCHO cells producing antibody have benefitted from chaperone engineering, however, in the case of rCHO cells producing thrombopoietin (TPO), interleukin-15 (IL-15), or a tumour necrosis factor receptor: F_c fusion protein, similar improvements could not be achieved (Davis et al. 2000, Borth et al. 2005, Mohan et al. 2007). On the other hand, CHO cell engineered to overexpress PDI isoforms e.g. endoplasmic reticulum protein 27 (ERp27) and its co-overexpression with calnexin/calreticulin showed improved TPO production (Hwang, Chung and Lee 2003, Chung et al. 2004b). Numerous other reports have suggested mixed views on the benefits of PDI overexpression as an engineering target with different cell lines and various expression systems.

1.4.5 Unfolded protein response machinery engineering

The primary signalling network induced in response to the build-up of misfolded and/or unfolded proteins in ER is the UPR. UPR activation is due to elevated transcriptional rates of the genes encoding ER-resident chaperones, with the likes of endoplasmin and PDI associated with improved recombinant productivity in CHO (Smales et al. 2004). However, ER-stress (due to overexpression of therapeutic proteins and/or bioprocess related insults) in CHO could potentially compromise the overall productivity of therapeutic proteins since ~ 30 % of newly synthesized proteins can be degraded due to improper protein folding (Schubert et al. 2000, Schröder and Kaufman 2005, Yewdell and Nicchitta 2006, Chakrabarti, Chen and Varner 2011). Studies have reported that UPR activation is dynamically regulated during the production phase of cell culture. In this regard, rCHO engineering has been focused on the engineering of ER-localized quality control proteins. For example, the

activating transcription factor (ATF4) improved the specific IgG production rate by 2.4-fold and DNA-damage inducible protein 34 (GADD34) have been reported to enhance the q to 28 pg/cell/day of anti-thrombin III (AT-III) in rCHO cells (Omasa et al. 2008, Ohya et al. 2008). In another study, the widely explored XBP-1s (spliced form of XBP-1) overexpression resulted in 2.5-fold increase in q for various therapeutic proteins such as mAbs, EPO, interferon-gamma (Ku et al. 2010).

1.4.6 Metabolic engineering

The two major carbon and energy (C & E) source for CHO cell that are absolutely vital for their growth are glucose and glutamine. However, the utilisation of a mole of glucose and glutamine concomitantly results in the accumulation of a mole of lactate and 2 moles of ammonia, respectively, as by-products. The accumulation of such by-products has been reported to have negative impact on CHO cell growth profile, and consequently to the quality and quantity of the recombinant proteins produced (Ozturk, Riley and Palsson 1992, Lao and Toth 1997, Zeng, Deckwer and Hu 1998). In order to address these issues, rCHO have been engineered to efficiently utilize metabolites at the same time reducing waste by-product accumulation in the culture. For example, the catalytic enzyme lactate dehydrogenase (LDH) (responsible for glucose to pyruvate to lactate conversion) has been genetically manipulated in hybridoma and CHO cells to reduce the specific production of lactate in culture (Chen et al. 2001, Jeong et al. 2001). The small interfering RNA (siRNA)-mediated suppression of LDH-A (subunit of LDH) resulted in 45-79 % reduced lactate production without impairing the cellular growth rate and human TPO (hTPO) production (Kim and Lee 2007). In an alternative method, CHO cells engineered with overexpression of fructose-specific transporter (GLUT-5) have been observed to utilize fructose as the alternate C & E source, hence, causing reduced lactate production (Wlaschin and Hu 2007).

1.5 Genome-editing technology

Genome editing tools such as the use of endonucleases to cause functional knockouts of unwanted genes in the CHO genome is a topical area of exploration. Tools such as chemical nucleases, zinc-finger nucleases (ZFNs), mega-nucleases and transcription activators like effector nucleases (TALENs) and bacterial CRISPR-Cas9 (clustered

regularly interspaced short palindromic repeats-Cas9) systems offer exciting new opportunities in terms of genome engineering for CHO cell line developmental purposes. The best known of these nucleases-based technologies to cause a functional knockout in virtually any genome of plants, animals and bacteria are ZFNs, TALENs and CRISPR-Cas9 system. In the following sections mainly these nucleases will be discussed.

1.5.1 ZFNs and TALENs

Complex genome modifications at single nucleotide resolution have been the biggest goal of scientists working in applied genetics, including gene therapy, synthetic biology, drug development and biotechnology. In the past, genome editing was achieved by traditional chemical mutangenesis (Eeken and Sobels 1983, Solnica-Krezel, Schier and Driever 1994) and transposon-mediated mutagenesis methods (Marx 1982, Spradling and Rubin 1982). These methods were intrinsically limited since they lacked specificity. In the past two decade methods such as homologous recombination (HR) based specific gene targeting (Gaj, Gersbach and Barbas 2013), ϕC31-mediated integration systems (Groth et al. 2004) and ZFNs (Bibikova et al. 2002, Bibikova et al. 2003) began to surface. However, the use of these techniques were limited by several factors, including the low efficiency at which the genetically engineered constructs were introduced to target a specific chromosomal locus, time-consuming engineering of these constructs and offsite-targeting issues (Gaj, Gersbach and Barbas 2013). An alternative solution for these techniques emerged in the form RNAi technology, which provided researchers with a rapid, inexpensive and high-throughput solution to gene knockdown. However, it was acknowledged that RNAi knockdown of genes is incomplete, varies between experiments and laboratories performing research using this technology. Moreover, there were reports of unpredictable off-target effects and only temporary inhibition of gene function, hence, limiting the practical application of RNAi technology (Gaj, Gersbach and Barbas 2013). Better genetic tools to address the aforementioned issues, came to light in the form of chimeric nucleases i.e. ZFNs and TALENs which are composed of a programmable, sequence-specific DNA-binding motif linked to a non-specific DNA cleavage domain. Both these chimeric nucleases have been used to introduce broad range of genetic modifications. Both these chimeric nuclease-based techniques work

by inducing double-strand DNA breaks (DSBs) that stimulate error-prone non-homologous end joining (NHEJ), as well as HR repair mechanisms, hence resulting in a mutated DNA locus. NHEJ is the frequent repair mechanism employed by mammalian cells, in contrast to HR, which occurs at a several orders of magnitude lower frequency than NHEJ (Barnes 2001, van den Bosch, Lohman and Pastink 2002, Lieber 2010). ZFNs and TALENs discovery and application made a revolutionary contribution to the genome-editing field and this rightly earned them the status of "Method of the Year" in 2011 by *Nature Methods* (Kim and Kim 2014a).

A ZFN has a modular structure comprising two main domains: a Cys_2-His_2 tandem array of zinc finger protein (ZFP) domain (most common in eukaryotes) (Tupler, Perini and Green 2001) and a *FokI* type II restriction enzyme-derived nuclease domain (**Figure 1.5.1 (A) and (B)**). ZFPs can be used to replace the DNA-binding domain of *FokI* in order to create a ZFN. There is an absolute requirement for two ZFN monomers in order to form an active nuclease to cleave the target DNA. Therefore, the two monomer domains (separated by 5-7 bp spacers) of *FokI* must dimerize for this active configuration enhancing ZFNs specificity (Kim, Cha and Chandrasegaran 1996, Bitinaite et al. 1998). The sequence specificity of ZFNs is dictated by the ZFPs and each 'finger' recognizes a 3-bp sequence, therefore, 3-6 zinc fingers are used to generate a single synthetic ZFN that targets a 9-18 bp DNA target sequence (Wolfe, Nekludova and Pabo 2000, Liu et al. 1997). They have been widely employed in the CHO cell engineering, for example, Cost et al. performed Bak and Bax knockouts using ZFN technology to increase CHO cells resistance to apoptosis (Cost et al. 2010a).

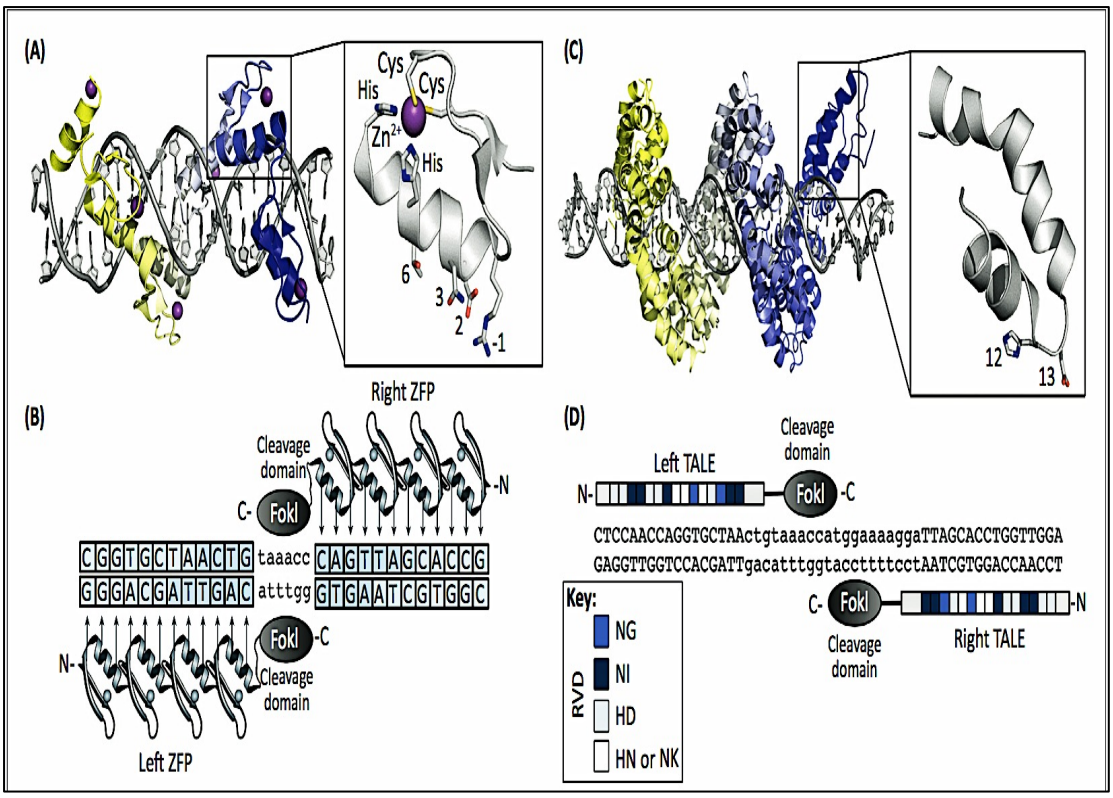


Figure 1.5.1 Structure of zinc-finger and transcription activator-like effectors nucleases. (A) (Top Left) Customized zinc-finger protein in complex with target DNA (grey). (B) (Bottom Left) ZFN dimer bound to target DNA. ZFN sites consist of two zinc-finger binding sites separated by 5-7 bp spacer sequence recognized by FokI cleavage domain. Custom designed ZF proteins recognize unique 'left' and 'right' half-sites. (C) (Top Right) TALE protein in complex with the target DNA. (Bottom Right) TALEN dimer bound to target DNA. TALEN target sites consist of two TALE binding sites separated by a spacer sequence of variable length of 12-20 bp. Custom designed TALEs recognize unique 'left' and 'right' half-sites. Image reproduced from a review by (Gaj, Gersbach and Barbas 2013).

Moreover, their utilization by Fan et al. (2012) to generate GS knockout CHO cells in order to study the impact of GS selection efficiency in cell line generation has been one of the greatest applications of this genome-editing tool (Fan et al. 2012). In the biomedical field ZFNs have been designed as drugs against human-immunodeficiency virus (HIV) and currently being tested for phase 2 clinical trials. However, ZFNs have limitations. This is due to the fact that the construction of ZFN protein with high selectivity is a costly, laborious and time-consuming task (Bibikova et al. 2002, Bibikova et al. 2003, Urnov et al. 2010, Cradick et al. 2011), requiring very specific expertise.

Similar to ZFNs, another alternative platform for genetic engineering was discovered in the form of second-generation artificial TALENs. TALENs differ from ZFNs only in that in TALENs DNA-binding domain is not zinc finger domain, instead multiple 33-35 amino-acid-repeat domains that recognize a single base pair. The construction and production of these nucleases is more easier and efficient than the ZFNs. TALEs were first discovered in bacteria *Xanthomonas oryzae pv. oryzae (Xoo)* and *Xanthomonas oryzae pv. oryzicola (Xooc)* (Zhang et al. 2014). These pathovars are bacteria invasion strategies to infect plant (Bonas, Stall and Staskawicz 1989). Owing to their similarity with eukaryotic transcription factors, they are termed transcription activator-like effectors. These are comprised of a group of key effector proteins, which contain N- and C- termini and a central domain for specific DNA binding that varies with 5-30 (average 17.5) tandem monomer repeats. TALENs specificity is determined by two hypervariable amino acids known as the repeat-variable di-residues (RVDs) (Boch and Bonas 2010, Miller et al. 2011a) (**Figure 1.6.1 (C) and (D)**). The length of DNA sequence recognized by TALENs is usually 17-18 bp, however, for artificial TALENs is 14-20 bp. Similar to ZFNs, TALENs also have a *FokI* component to them, which works in a dimeric fashion, and thus makes it necessary to design a pair of TALENs to effect a DSB.

Artificial TALEN-mediated genome editing was optimized to be a rapid, efficient and specific method for genome modifications (Cermak et al. 2011, Bedell et al. 2012, Garg et al. 2012, Sung et al. 2013, Zu et al. 2013). TALEN-based genetic manipulation could be used for the insertion of exogenous sequences (e.g. FRT/loxP sequences) in CHO cells genome for instance with a view to enhance cell growth and protein productivity. Moreover, TALENs have been utilized in the construction of transgenic animals, for *in-vivo* gene therapy and treatment of diseases such as sickle cell anaemia and neurodegenerative disorders (Hsu and Zhang 2012, Schiffer et al. 2012, Sun et al. 2012). However, before considering TALEN application in other areas of clinical research, factors such as specificity and efficiency of RVDs, off-target effects and structural stability of its repeat modules must be heeded. Moreover, the delivery method chosen and cytotoxicity of TALEs require detailed examination, as well as the fact that these are derived from plant pathogens, therefore, immune responses in clinical treatment are among other concerns (Schiffer et al. 2012). TALENs are a great tool to have in the genome-editing toolbox, however, due to the

labour and time invested to construct these, limits their practical applicability (Wei et al. 2013).

1.5.2 CRISPR-Cas system for genome engineering

The latest tool that provides greater specificity and efficiency, with reduced off-target effects and decreased cytotoxicity when applied to mammalian cells, is the bacterial CRISPR-Cas9 system (Cong et al. 2013, Mali et al. 2013b).

1.5.2.1 Types of CRISPR-Cas systems

CRISPR-associated (*cas*) genes constitute a natural adaptive-immunity response in bacteria upon invasion of viruses and other parasites in order to protect the bacteria from infection (Gasiunas et al. 2012). In CRISPR-Cas system, CRISPR repeat-spacer arrays are transcribed into long primary transcripts, which are further processed into a set of short CRISPR RNAs (crRNAs) that are composed of a conserved repeat fragment and a variable spacer sequence (referred to as the guide RNA) complementary to the sequence of the invading nucleic acid (Carte et al. 2008, Hale et al. 2009). The *cas* genes are also present in the vicinity of CRISPR arrays (Jansen et al. 2002, Makarova et al. 2006). The combination of crRNAs and the Cas proteins results in the formation of an effector complex that recognizes the complementary sequence of the invasive species using the guide RNA functional specificity, inducing sequence-specific cleavage and hence, rendering the foreign species harmless by preventing the proliferation and propagation of foreign nucleic acids (Jore et al. 2011) (**Figure 1.5.2.1**).

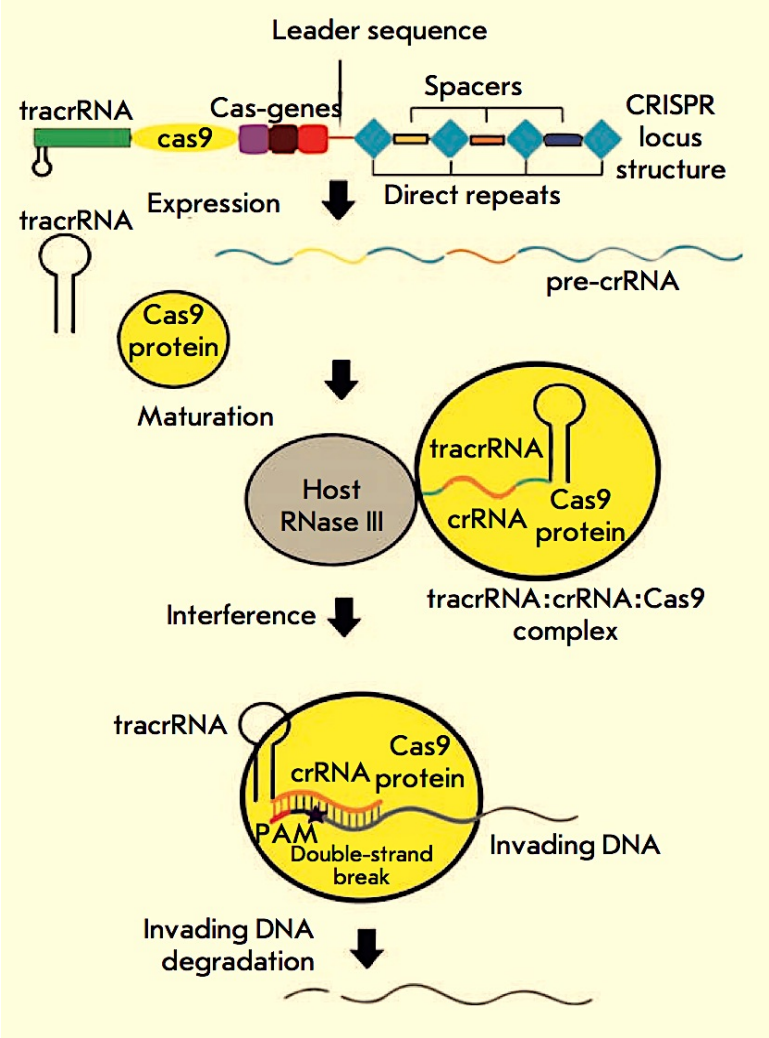


Figure 1.5.2.1 Bacterial CRISPR-Cas9 mechanism of action to degrade foreign parasite's and/or virus's nucleic acids, and consequently rendering them harmless.
Image reproduced from (Nemudryi et al. 2014).

There are various distinct forms of CRISPR-Cas systems. They only differ in the Cas nucleoprotein type required to induce specific DNA cleavage (Wiedenheft, Sternberg and Doudna 2012). The type I-E system as noticed in *E.coli* has a crRNA located in a multi-subunit, which binds to complementary DNA sequence and degrades the target DNA using the Cas3 nuclease (Beloglazova et al. 2011, Sinkunas et al. 2011). In certain types of bacteria such as *Sulfobulus solfataricus* and *Pyrococcus furiosus* i.e. in type III CRISPR-Cas system, the crRNA recognizes and cleave synthetic RNA *in-vitro*, however, the type III CRISPR-Cas system of *Staphylococcus epidermis* targets DNA *in-vivo* (Marraffini and Sontheimer 2008, Hale et al. 2012). The type II CRISPR-Cas system of *S.thermophilus* has four *cas* (i.e. *cas9, cas1, cas2 and csn2*) genes. The DNA disruption is accomplished by Cas9 protein (Makarova et al. 2011). Very recently, a novel class II CRISPR-Cas system has been described in literature.

This CRISPR-Cas system uses a class II effector protein, Cpf1- a single RNA-guided endonuclease lacking trans-activating RNA (tracrRNA) (Shmakov et al. 2015).

1.5.2.2. CRISPR-Cas9 System and its application

The type II CRISPR-Cas9 system of *S.pyogenes* is one of the better-characterized genome editing CRISPR-Cas systems (Magadán et al. 2012). In this CRISPR-Cas type, the crRNA array encodes the guide RNAs and a required auxiliary tracrRNA. The tracrRNA facilitates processing of the crRNA array into discrete units (Garneau et al. 2010, Gasiunas et al. 2012). Each crRNA unit contains a 20 nt guide sequence that guides Cas9 to cleave the DNA target in a site-specific manner. In addition to these key components of the CRISPR-Cas9 genome editing machinery, there is an absolute requirement that the target site must immediately precede a 5' - NGG protospacer-adjacent motif (PAM) sequence (N can be any nucleotide of the four A, C, G, T) (Jinek et al. 2012). The Cas9 nuclease uses the conserved HNH and RuvC nuclease (Sapranauskas et al. 2011, Cho et al. 2013, Ran et al. 2013) domains in order to induce double strand breaks (DSBs) in the DNA. The RuvC catalytic domain can be mutated to yield a mutated version of Cas9 i.e. Cas9 nickase that performs single-strand breaks in DNA that can be employed to reduce off-target effects (Sapranauskas et al. 2011, Cho et al. 2013, Ran et al. 2013). Lately, studies have been conducted to design chimeric RNAs that contain both crRNA and tracrRNA fused together as a unit i.e. 'single guide RNA' (sgRNA) (**Figure 1.5.2.2.a**), which can be customized to target 17 nt, 18 nt, 19 nt or 20 nt sequences in a gene (Chang et al. 2013, Cong et al. 2013, Dicarlo et al. 2013, Mali et al. 2013b). These customized small chimeric sgRNA are combined into a plasmid vector that also contains codon-optimized Cas9 nuclease for the heterologous expression of all these elements from a single expression construct (Cho et al. 2013, Cong et al. 2013, Jinek et al. 2013, Mali et al. 2013b).

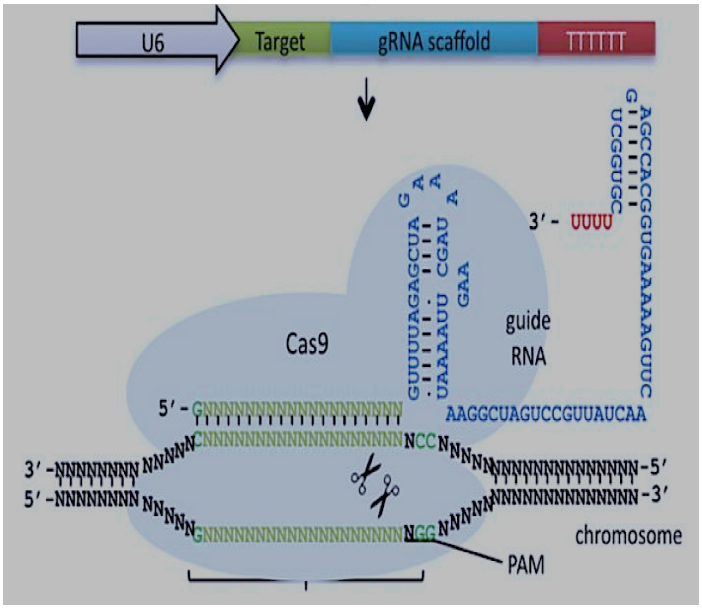


Figure 1.5.2.2.a shows a single chimeric sgRNA to induce DSBs in target loci (Green sequence). Single sgRNA construct consists of a guide RNA scafflod to guide the Cas9 nuclease to the target DNA site adjacent to the short PAM (NGG in Sea Green) site at 3' of the protospacer.
Image source: *Addgene.org*

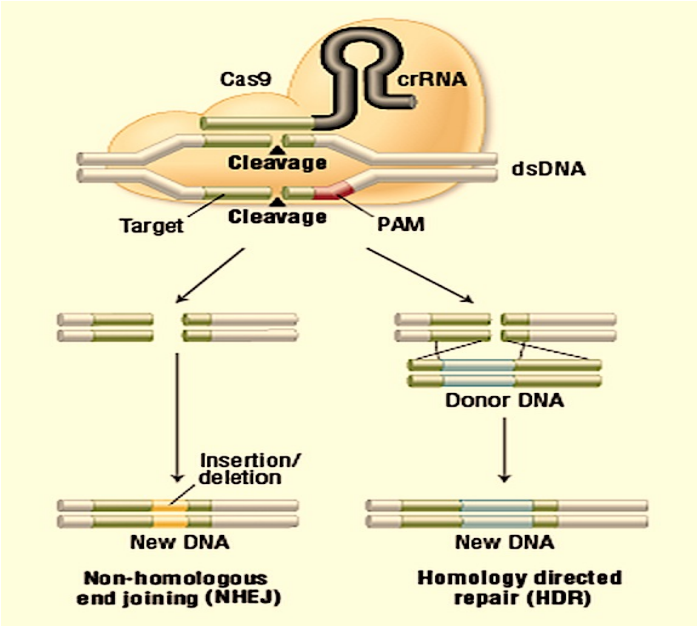


Figure 1.5.2.2.b shows the employment of the cell's natural DNA repair system upon DSBs break in the target DNA by nucleases. The Left side of the image shows error-prone NHEJ leading to frame shift mutations, while the Right side of the image shows the less-frequent HR-mediated repair and, in the presence of donor DNA
Image source: www.neb.com

CRISPR-Cas9 system works through the guide RNA sequence that guides the Cas9 nuclease to the complementary target sequence. Once bound, the Cas9 performs site-specific cleavage to cause DSBs. This is followed by the cell natural DNA repair system deployment to repair the DSBs in the DNA via NHEJ or HR. Since NHEJ occurs at frequencies several orders of magnitude more than HR, it is the most likely method of repair (Sedivy and Sharp 1989). NHEJ is an imperfect repair process that gives rise to mutations (e.g. indels, mismatch repair) at the site of DSBs during repair, potentially disturbing the reading-frame of a protein-coding gene and leading to pre-mature STOP codon formation, hence, rendering it functionless (Perez et al. 2008, Santiago et al. 2008, Galetto, Duchateau and Pâques 2009, Miller et al. 2011b)

(**Figure 1.5.2.2.b**). Moreover, the use of CRISPR-Cas9 system opens up a vast array of opportunities since the introduction of a DSB at a given locus increases the frequency of HR by several orders of magnitude. Therefore, if CRISPR-Cas9 targeting is combined with donor DNA with homology to the nucleotide sequence at the break/cleavage site, then homology directed repair (HDR) takes place as a repair mechanism with the insertion of the donor DNA (Yang et al. 2013). With such great opportunities to genetically modify the sequence of virtually any gene, the CRISPR-Cas9 system has wide applications in gene engineering. For example, development of isogenic human stem cells (Horii et al. 2013), exploiting HDR for precise genetic manipulation to correct mutant cell phenotype (Schwank et al. 2013), the study of the functional relationships between large groups of genes (Shalem et al. 2014, Wang and Quake 2014) and screening for the active regions of a genome in living cells (Chen et al. 2013) can be conducted via this technology.

In regard to the biopharmaceutical industry, CRISPR-Cas9 systems have been employed to engineer CHO cells and have been reported lately by several groups (Ronda et al. 2014, Bachu, Bergareche and Chasin 2015, Grav et al. 2015, Lee et al. 2015). Ronda et al. (2014) genetically manipulated CHO cells using the CRISPR-Cas9 system to enhance protein quality by designing sgRNA to target COSMC (a C1GALT1C1 encoding C1GALT1-specific chaperone 1-essential for correct O-glycosylation (Wang et al. 2010) and FUT8 (the α1,6-fucosyltransferase-catalyzing the addition of fucose on IgG antibodies that could reduce ADCC). The researchers achieved 47.3 % indel frequency in the COSMC gene, while a combination of CRISPR-Cas9 system with a lectin selection application attained 99.7 % indel frequency in the FUT8 gene. These results demonstrate how efficient and convenient the CRISPR-Cas9 technology is for targeting genes in CHO cells to improve their characteristics. Generating a stable CHO cell line that expresses the therapeutic protein gene invariably and/or generating a stable CHO cell line that has multiple knockouts of unwanted genes is challenging but desirable. Previously, ZFNs were used to achieve knock out of multiple undesirable genes, however, only sequentially (Cost et al. 2010b, Grav et al. 2015). Moreover, the time, labor and cost of generating these knockouts were exorbitant. For this purpose, the CRISPR-Cas9 systems can offer advantages in terms of designing highly efficient and effective multiplex sgRNA units, which can target several parts of the genome simultaneously with greater

efficiency. Such multiplexed CRISPR-Cas9 systems have been reported in the literature and have proven their efficacy. These all-in-one CRISPR-Cas9 vector systems could potentially provide an effective targeting strategy for multiplex epigenome editing and simultaneous activation/repression of multiple complex mammalian genes, activators and promoters (Gilbert et al. 2013, Sakuma et al. 2014, Hilton et al. 2015). In CHO cells, very recently, Grav et al. (2015) reported one-step generation of a CHO cell line with a triple knockout using a CRISPR-Cas9 system followed by FACS sorting to enrich (i.e. by linking GFP to the Cas9 nuclease via 2A peptide) for multiple knockout CHO clones with Bax and Bak gene disruptions, in addition to FUT8 knockout. The functional analysis of these triple knockout clones showed resistance to apoptosis and decreased fucosylated content in IgG-1 antibodies, in comparison to a wild-type cell line (Grav et al. 2015).

CRISPR-Cas9 offers several advantages over traditional overexpression or insertion of GOI, or RNAi-mediated knockdown (Krämer, Klausing and Noll 2010, Jadhav et al. 2013). Transient expression of sgRNA and Cas9 nuclease can result in permanent gene modifications. Furthermore enrichment methods such as GFP linked via 2A peptide can be used for the selection of positive clones, therefore, saving labour and time-intensive screening for positive clones. This also provides an alternative to the use of selection markers, hence, reducing the transcriptional burden of expressing the selection marker gene. Multiplex genome engineering through the design of an all-in-one CRISPR vector (with customizable sgRNAs targeting multiple genes) could also solve the problem of co-transfection of single sgRNA CRISPR-Cas9 systems, whose transfection may not be as high as a multi-sgRNA CRISPR plasmid. Multiplexing CRISPR systems has completely revolutionized genome modification in mammalian cells and could play a huge role in engineering CHO cells to improve recombinant therapeutic protein production.

With the advent of global CHO gene expression profiling, the CRISPR-Cas9 system would not just be confined to knocking out undesirable genes but also for integration of transgenes into transcriptionally ‘active zones’ of the CHO genome (Duda et al. 2014). As discussed in **Sections 1.3 and 1.4**, the process of recombinant CHO cell-line development has been seriously limited due to variable/or inconsistent expression of transgenes. This is particularly due to the ‘positioning-effect’ (Kim and Kim

2014b), which results from the random integration of GOI in the CHO genome. Traditional methods such as Cre/loxP system, Flp/FRT system and ϕ C31/34 integrases have been applied in order to mitigate the issue of uncontrollable transgene insertion. However, these systems are limited in that they require a prior establishment of a platform recombinant cell line with the insertion of recombination site into transcriptional 'hot-spot' regions of the genome in order to retarget and generate a cell-line for the expression of GOI (Kito et al. 2003, Huang et al. 2007b, Lieu et al. 2009). For this reason, CRISPR-Cas9 technology can be preferred choice for the insertion of GOI in one-go over the traditional methods. For instance, Lee et al. (2015) recently reported a successful 3.7 kb gene expression cassette at a specific loci in CHO cells using the CRISPR-Cas9 system. The researchers successfully manipulated the endogenous loci of the CHO genome and precisely replaced the knocked out sites with homologous pieces of endogenous/exogenous donor DNA using the leveraged HDR-mediated targeted transgene integration systems (Lee et al. 2015). Combining the CRISPR-Cas9 system to induce site-specific gene cleavage and the subsequent integration of GOI using leveraged HDR system (less frequent DNA repair system than NHEJ) by suppressing key molecules of NHEJ system (i.e. KU70, KU80 or DNA ligase by gene silencing in order to increase HDR events frequency) (Chu et al. 2015, Maruyama et al. 2015) could assist researchers to attain the elusive task of generating isogenic CHO cell lines. Moreover, successful NHEJ-mediated plasmid integration studies have also surfaced lately (Maresca et al. 2013, Pan et al. 2013, Auer et al. 2014). For example, targeted recombinant plasmid integration via NHEJ was performed in CHO cells with insertion of donor DNA (non-homologous arms) with 0.45 % efficiency (Bachu, Bergareche and Chasin 2015).

In summary, CRISPR-Cas9 genome editing tool offers great promise and flexibility for genome modification. In comparison to other genome modification tools such as ZFNs and TALENs, CRISPR-Cas9 system completely outshines them in particular, regarding its simple inexpensive design and highly efficient site-specific targeting in virtually any gene and/or even multiple gene (s).

1.6 microRNAs

1.6.1 miRNAs biological relevance and potential rols in CHO cell engineering

Efforts in the field of RNAi could provide for an alternative in order to knockdown the expression of target gene(s) in a sequence-dependent manner (Wu 2009). However, knockdown using RNAi approaches can be variable and only induce temporary inhibition of the target gene. A potentially investing option is the exploitation of endogenous miRNAs and the manipulation of their expression levels to increase CHO cell growth and productivity. Unlike single or multi- gene engineering, the use of miRNAs for gene manipulation does not put stress on cell translational machinery (Müller, Katinger and Grillari 2008, Barron et al. 2011c, Jadhav et al. 2013).

Mature miRNAs are endogenous non-coding, 19-22 nt single-stranded RNA molecules that regulate gene expression globally in cells at the post-transcriptional level. miRNAs were first discovered in *Caenorhabditis elegans*. Similar to certain transcription factors and in contrast to siRNAs, single miRNAs post-transcriptionally regulate more than 10-100 of mRNAs potentially affecting cellular behavior by globally modifying gene expression patterns (Fire et al. 1998). This global gene regulation is exerted wherein mature miRNAs targets their mRNAs in the 3’-untranslated region (3’-UTR) using Watson-Crick base-pairing, hence inhibiting translation of the targeted mRNAs and/or catalyzing mRNAs cleavage depending upon the complementarity between miRNA and the 3’ UTR of the gene. These small RNA molecules have been reported to regulate many biological functions and/or pathways in the published reports. For instance, Pasquinelli and Ruvkun (2002) reported miRNAs role in the regulation of developmental timing in *C.elegans* (Pasquinelli et al. 2000), and miRNA roles in other cellular functions such as differentiation, cell growth and apoptosis and cell cycle (Cheng et al. 2005, Chivukula and Mendell 2008). Some studies have reported their effect on metabolism (Xu et al. 2003, Gao et al. 2009, Lin et al. 2009) and deregulation in multiple diseases, including cancer of many types (Xi 2013). miRNAs also play a role in fine-tuning cellular response due to environmental insults via a well-orchestrated regulation of genes in a coordinated manner. They act to buffer cells against any stochastic

fluctuations due to any internal/external cellular perturbations and providing ‘robustness’ to cell. Hence, they act as rheostats to maintain cellular homeostasis (Li et al. 2009, Ebert and Sharp 2012, Vidigal and Ventura 2015).

1.6.2 miRNA discovery

The first miRNA, lin-4 was discovered in a nematode *C.elegans* whilst studying the post-embryonic developmental timing in 1993. Transcription of lin-4 negatively regulated LIN-14 protein levels at the L1 larval stage of development (Lee, Feinbaum and Ambros 1993). Following the discovery of the first miRNA, miRNAs have been reported in array of different organisms including worm (Lau et al. 2001) fly, rodents, human (Lagos-Quintana et al. 2001, Wang et al. 2008, Sanchez et al. 2014) even plants, green algae and viruses (Griffiths-Jones et al. 2006).

1.6.3 miRNA biogenesis and processing

Small non-coding RNAs ranging from 20-30 nt in length have been classified into three categories: siRNA, piwi-interacting RNAs (piRNAs) and miRNAs. These RNA molecules share common links to certain pathways, however, their biogenesis and mode of gene regulation is different from one-another (Bartel 2004a). Mature miRNAs are single stranded short 20-23 nt RNA molecules derived from long primary miRNA (pri-miRNA) units (several thousand kilobases transcripts transcribed by RNA pol II). Initially, the pri-miRNA is processed by the Microprocessing complex comprising DROSHA-DGCR8 (DiGeorge Syndrome Critical Region 8) to yield a 70 nt hairpin stem loop structure or preliminary miRNA (pre-miRNA) (**Figure 1.6.3**). This processing to yield the pre-miRNA is mediated by RNAse type III DROSHA enzyme and its double stranded RNA-binding adaptor protein DGCR8 and other nuclear components (Gregory et al. 2004, Han et al. 2004, Landthaler, Yalcin and Tuschl 2004). These processed pre-miRNAs are then exported into the cytoplasm by the nuclear protein Exportin-5 (Exp5) in a RAN-GTP dependent manner (Yi et al. 2003, Bohnsack, Czaplinski and Görlich 2004, Lund et al. 2004). The Exp5 recognition of pre-miRNA is independent of pre-miRNA sequence or the loop structure. Moreover, only defined lengths of ds-stem and 3’ overhangs molecules are successfully bound to Exp5 in order to ensure that only correctly processed pre-miRNAs are shuttled from the nucleus to the cytoplasm (Zeng

and Cullen 2004, Lund and Dahlberg 2006). Following this, the pre-miRNA is further processed by Dicer (RNAse type III enzyme), which in conjunction with ds-RNA binding protein (TRBP- TAR RNA binding protein) removes the loop and generates a ~ 22 nt RNA duplex. This RNA duplex interacts with Ago family proteins to form a pre-RISC (pre- RNA induced silencing complex), with guide strand of the duplex bound to the Ago2 protein (endo-nucleolytic activity) (Peters and Meister 2007).

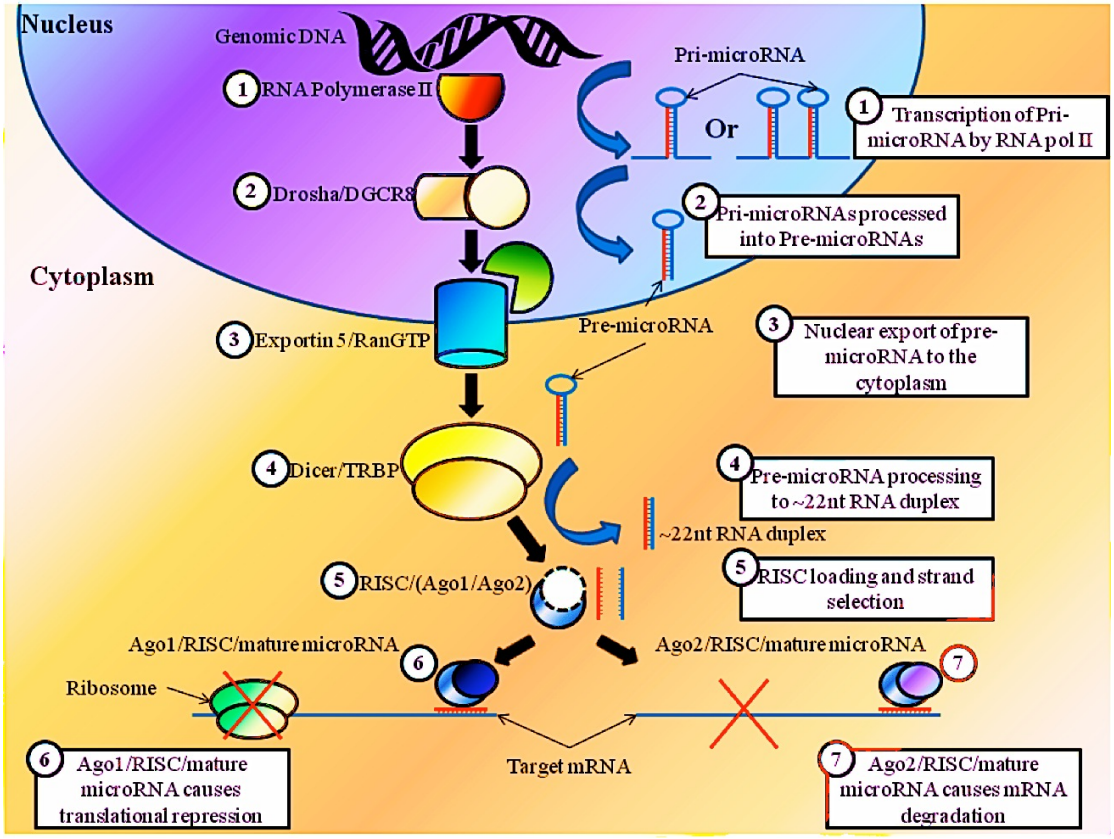


Figure 1.6.3: The biogenesis and processing of mature miRNAs. (1) Initially, the process begins with the translation of a nascent transcript, pri-miRNA (2) Microprocessor complex with Drosha/DGCR8 cleaves the overhanging ssRNA liberating pre-miRNA processed transcript (~70 nt in length) (3) Shuttling of the pre-miRNA transcript is carried out by Exportin-5 in a Ran-GTP manner from the nucleus to the cytoplasm (4) The pre-miRNA transcript is further processed by Dicer machinery to generate a hairpin loop structured ~ 22 nt RNA duplex (5) Following this, a preferential strand selection occurs upon RISC loading, wherein the guide strand loaded onto the RISC complex targets the respective mRNA (6) Two possibilities i.e. either (6) translational repression occurs, or (7) mRNA degradation occurs. These two events are dependent on the sequence complementarity of the miRNA: mRNA interaction and the type Ago protein fused to RISC complex. This image was reproduced from (Barron et al. 2011c).

The other 'star' or passenger strand (miRNA*) is thought to be unwound from the

duplex and is degraded leading to the formation of mature RISC complex (Chendrimada et al. 2005). The preferential association of the guide miRNA strand in the RISC complex is based on the thermodynamic stability of its 5'-end and is dependent on the abundance of target transcripts, which drives miRNA arm selection (Schwarz et al. 2003).

1.6.4 miRNA mode of action: Translational repression and/or mRNA degradation.

The mature RISC complex targets the mRNA by binding to the 3'-UTR of the mRNA. The Ago2 protein is the main driver of the miRNA-mediated mRNA cleavage. There exists a number of Ago proteins in mammals, namely Ago1-4. miRNA-mediated RNA interference has two outcomes i.e. either translational repression or target mRNA degradation, with both leading to translational inhibition (Baek et al. 2008). Post-transcriptional regulation by miRNA is mediated through the base pairing of the 'seed' region (2-8 nt at the 5'-end). This region is conserved in most vertebrates. The translational repression efficiency is dependent on factors such as AU-rich composition near the seed region and miRNA binding site position (Grimson et al. 2007). The extent of the miRNA: mRNA 3'-UTR complementarity also influences which of the two aforementioned outcomes. For instance, low complementarity (i.e. sequences with bp mismatches) of miRNA: mRNA leads to translational repression via deadenylation and subsequent decapping of target mRNAs (Bagga et al. 2005, Eulalio et al. 2009) while high-complementarity results in target mRNA degradation.

1.6.5 Nomenclature of miRNAs

With the rapid identification of the new miRNAs, it was imperative that these non-coding small RNA molecules to have a database with an easy to understand nomenclature platform followed universally by researchers across the globe (Ambros et al. 2003). One such database is the "miRBase", wherein miRNAs are catalogued with a classification system in order to cater for the information regarding the miRNA identity along with sequence information, annotation, species, affiliation, genome site origin, target prediction resources and quick-hyperlink to literatures.

1.6.6 Genomic Organization of miRNAs

Based on their genomic locations, miRNA genes are categorized as (i) *Intergenic*, (ii) *Intronic*, (iii) *Mirtronic*, and (iv) *Exonic* (Rodriguez et al. 2004), **Figure 1.6.6**)

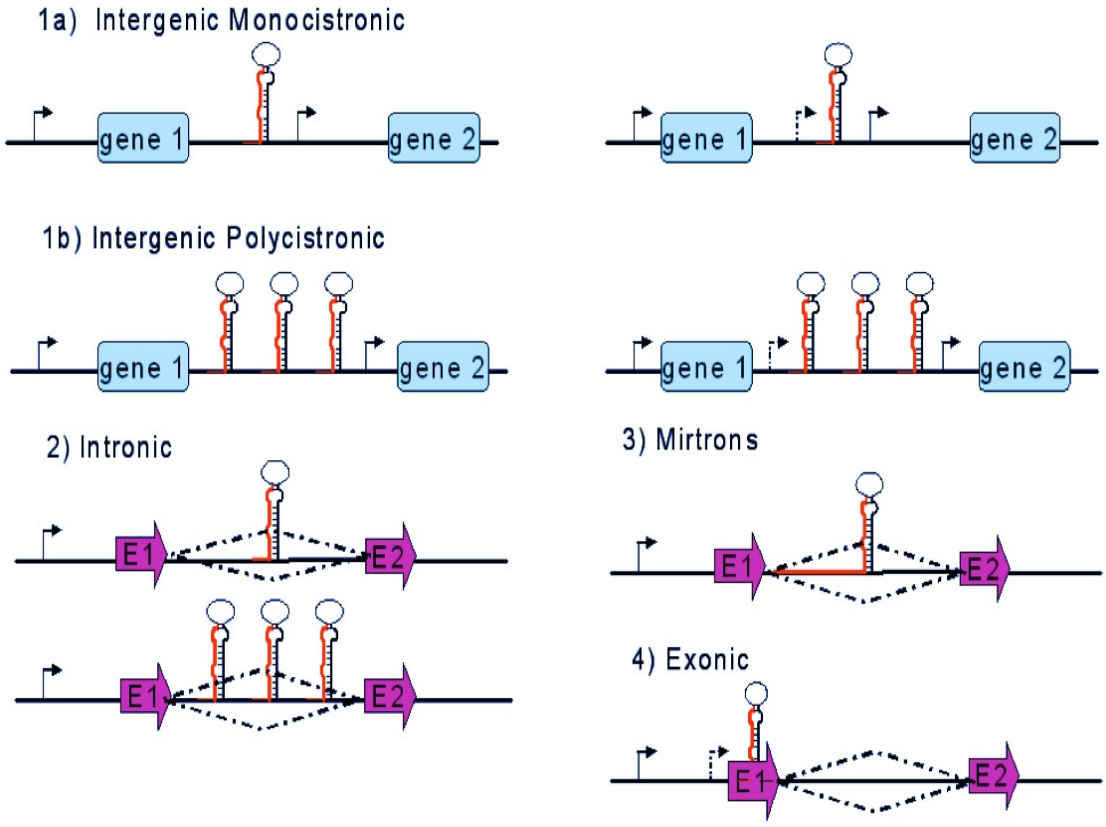


Figure 1.6.6 Four classes of genomic organization of miRNAs. (1) Intergenic miRNAs located in isolation from other genes and occur either as (1a) monocistronic or (1b) polycistronic units (2) Intronic miRNAs within the intron of protein-coding genes, again both either as mono- or poly-cistronic units (3) Mirtrons are intronic miRNAs that exist in short introns, however, only differ from intronic miRNAs in that they can by-pass Drosha processing (4) Exonic miRNAs are located in an exon and are independently transcribed from their own promoter. Legends list: Black arrow- Host gene promoter start site, Black arrow with lines- miRNA promoter start site, Rhombus dotted lines- Intron region, Blue rectangle- Gene, Purple big arrows- Exon. This image was sourced from a review on miRNA structural organization by (Barron et al. 2011c)

1.6.7 Functional Validation of miRNAs

Discovering the regulatory significance of the specific miRNA-target association involves various lose-and gain-of function approaches to confirm and verify the functionality of miRNAs.

1.6.7.1 Loss of function approaches

Techniques that use molecules for targeted miRNA inhibition: Antagomirs are antisense miRNA oligonucleotides commonly employed for miRNA inhibition (Krützfeldt et al. 2005, Kaur, Babu and Maiti 2007, Henke et al. 2008). There are chemically modified versions of antagomirs i.e. 2’-O-methyl- or 2’-O-methoxyethyl, Locked Nucleic acid (LNA) and multiple-targeting (MTg) technology (for miRNA cluster’s functional validation) oligonucleotides. These modified antagomirs have been developed to antagonize miRNA *in vivo* (Elmén et al. 2008, Lu et al. 2009). These molecules are transient tools for verifying miRNA functions. The use of MTg-antagomirs is particularly suited for overcoming issues of miRNA inhibition compensation from other members of the cluster (Lu et al. 2009). However, these antisense oligonucleotide molecules are not suitable for long-term inhibition. Alternative approaches such as the use of simple expression vectors and/or the “sponges” or “decoy-vectors” can be particularly fruitful for long-term stable knockdown of miRNAs (discussed in more detail in **Section 3.1** and **3.2**) (Yang et al. 2012a, Druz et al. 2013).

1.6.7.2 Gain of function approaches

Techniques that work by overexpression of miRNA mimics and/or vector based expression systems to modulate miRNA function: Overexpression of miRNAs is a prominent alternative for studying the miRNA function. In this regard, two commonly employed techniques for the ectopic expression are: miRNA mimic strategy (Wang 2011) and/or vector based transcription of miRNA precursors, which are then further processed to yield mature miRNA of interest (Jadhav et al. 2013). miRNA mimics can be in high throughput screens for miRNA function via large synthetic libraries covering the entire miRNome for the specific organism. However, due to the nature of being transient with short-term effects, limits their application for functional validation of miRNAs. To overcome this, shRNA vector approaches for ectopic expression of mature miRNA transcripts utilizing the endogenous miRNA maturation pathway can be used since such constructs works well both in transient and stable miRNA transfection studies (Jadhav et al. 2013).

1.6.8 Online tools and experimental prediction/validation methods for miRNA

targets

The ability of a single miRNA to regulate a multitude of genes poses challenges in predicting target gene in the absence of experimental data due to the small seed region of miRNAs. However, Bioinformaticians have developed algorithms and/software tools that help in the prediction of targets, although they tend to suffer from large numbr of false positive predictions. **Table 1.6.8** provides a synopsis of these online tools and target predicting software. These tools are mainly categorized into two: (i) One method searching targets on the basis of the complementarity between the miRNA seed region and 3’-UTR of the mRNA and their thermodynamic stability of duplex formation, (ii) while other method provides list of predicted targets based on in-silico machine-learning approaches using aforementioned rules as well as considering extra algorithms to expand the criteria such as miRNA’s multiple binding events, G: U wobble pairs near 6-8 nt seed region, miRNA with functional redundancy , target site accessibility, evolutionary conservation of binding sites and seed regions etc. (Li et al. 2010). Although these tools help in predicting the potential ‘real’ biological targets, however, are not foolproof. Therefore, experimental approaches such as transcriptomic and proteomic profiling following the up-or down-regulation of miRNAs needed to be performed in order to validate these predicted targets (Baek et al. 2008, Selbach et al. 2008, Muniyappa et al. 2009). Experimental approaches such as hybridization techniques: northern blotting, RNAse protection assays, signal-amplifying ribozyme (Lee et al. 2002, Hartig et al. 2004, Baskerville and Bartel 2005a) Real Time quantitative polymerase chain reaction (RT-qPCR) (Tang et al. 2006) microarray (Kantardjieff et al. 2009) and next generation sequencing (Hackl et al. 2011a) can be utilized to verify predicted direct/indirect miRNA targets.

Table 1.6.8 provides a list of prediction algorithms available and the binding criteria used for miRNA target prediction. The table below is sourced from review by (Barron et al. 2011c).

Program	Species specificity	Algorithm description	Webserver
miRNA target EMBL	Drosophila	Complementarity with 3′UTR	http://www.russell.embl-heidelberg.de/miRNAs/
miRNAanda	Flies, vertebrates	Complementarity with 3′UTR, thermodynamic stability, duplex species conservation	http://www.microrna.org/microrna/home.do
RNAhybrid	Any	Complementarity with 3′UTR, thermodynamic stability, binding conservation	http://bibiserv.techfak.uni-bielefeld.de/rnahybrid/
TargetBoost	Worm and fruit fly	miRNA-mRNA binding site characteristics	https://demo1.interagon.com/targetboost/
miTarget	Any	Thermodynamic stability and sequence complementarity	http://cbit.snu.ac.kr/~miTarget/
Pictar	Flies Vertebrates Worm	Perfect and partial complementary sequence with 3′UTR, Thermodynamic stability	http://pictar.mdc-berlin.de/
RNA22	Any	miRNA-mRNA binding sites characteristics, Complementarity with 3′UTR, no cross-species conservation	http://cbcsrv.watson.ibm.com/rna22.html
MicroTar	Any	Complementarity with 3′UTR and thermodynamic stability	http://tiger.dbs.nus.edu.sg/microtar/
EIMMo	Humans, mice, fish, Flies, worms	miRNA binding sites conservation	http://www.mirz.unibas.ch/ElMMo3/
GenMiR++	Any	Sequence complementarity, based on expression data sets	http://www.psi.toronto.edu/genmir/
PITA	Any	Target site accessibility thermodynamic	http://genie.weizmann.ac.il/pubs/mir07/mir07_data.html
NBmiRNATar	Any	Sequence and duplex characterisitics, no sequence conservation	http://wotan.wistar.upenn.edu/NBmiRTar/login.php
Sylamer	Any	Based on microarray data to identify 3′UTR sites	http://www.ebi.ac.uk/enright/sylamer/
MiRTarget2	Vertebrates	Based on microarray data to identify 3′UTR sites	http://mirdb.org/miRDB/
TargetScan TargetScan S	Vertebrates	Complementarity with 3′UTR, thermodynamic stability, duplex species conservation	http://www.targetscan.org/
DIANA-microT	Any	Complementarity with 3′UTR, thermodynamic stability, duplex species conservation, combined experimental data sets	http://diana.cslab.ece.ntua.gr/microT/

1.6.9 miRNAs as potential tools for enhancing CHO cell phenotypes

For the overexpression of a multitude of CHO cell growth-favoring and protein-producing genes, co-transfection could be one alternative, however, such a method imposes stress on the cell's translational machinery and can also compete with the product. The burden on the translational machinery is doubled when other transgenes, such as selection marker transgenes, also needed to be expressed. An alternative route to the coding-gene engineering is via the use of small non-coding miRNAs (Pasquinelli and Ruvkun 2002). Their abilities to regulate hundreds of genes and to orchestrate complex cellular pathways (e.g. cell-cycle, cell metabolism, secretion and apoptosis) in a coordinated fashion without over-charging the cell's translational machinery make them an attractive tool to be utilized for CHO cell engineering (Bratkovic et al. 2012). miRNA profiling studies of cancers of different types have unveiled the important role of miRNAs in cell cycle regulation. These studies have identified potential miRNA candidates, which when up- or down-regulated affect cell growth behaviour, hence, stimulating an interest in using these non-coding RNA molecules to positively enhance CHO cell phenotypes. **Table 1.7.8** provides a list (not exhaustive) of potential miRNA candidates for CHO cell engineering.

Table 1.6.9 provides a selected list of potential miRNA candidates for CHO cell engineering

Cellular process	microRNA	Cluster	Effect	Verified/Predicted targets	References
Cell cycle	Let-7	miR-200/let-7a-2	Tumour suppressor	E2F2, CCND2	(Dong et al. 2010)
	cgr-miR-7	-	Tumour suppressor	EGFR, AKT pathway, Skp2, Psme3	(Kefas et al. 2008, Sanchez et al. 2013)

	cgr-miR-17	miR-17~92	Oncogenic	c-myc, E2F1, PTEN, CDKN1a/p21	(O'Donnell et al. 2005, Scherr et al. 2007, Ivanovska et al. 2008)
	miR-23 miR-24	miR-23a/27a~24-2 miR-23b-24-1	Tumour suppressive	Myc, E2F2	(Cheng et al. 2005, Gammell et al. 2007, Lal et al. 2009)
	miR-31	-	Oncogenic	p53, LATS, PPP2R2A	(Creighton et al. 2010, Liu et al. 2010a)
Apoptosis	miR-15a/16	miR-15a~16-1	Pro-apoptotic	Bcl2, Bcl-xL	(Weber and Fussenegger 2007, Calin et al. 2008, Yang and Paschen 2008)
	miR-21	-	Oncogenic	PDCD4, Caspases 3 and 7	(Chan, Krichevsky and Kosik 2005, Gammell et al. 2007)
	miR-34a	miR-34b/miR-34c	Pro-apoptotic	p53, Bcl2, SIRT1 deacetylase	(Raver-Shapira et al. 2007, Hermeking 2010)
	miR-133	miR-1/133	Pro-apoptotic	Caspase 9, HSP	(Xu et al. 2007)
	miR-144/155	-	Pro-apoptotic	Caspase 3	(Druz et al. 2011)
	miR-218	-	Pro-apoptotic	ECOP	

	miR-297	miR-297-669	Pro-apoptotic	Bcl2L2, DAD1, BIRC6, STAT5a, SMO	
Protein productivity and Quality	miR-34a	-	Post-translation modification-core fucosylation	FUT8	(Bernardi et al. 2013)
	miR-30b/d	miR-30b/d	O-glycosylation	GALNT7	(Gaziel-Sovran and Hernando 2012)
	miR-148b	-	N-glycosylation	C1GALT1	(Coppo and Amore 2004)
	miR-30c	miR-30c-2*	UPR	XBP-1	(Byrd, Aragon and Brewer 2012)
	miR-410	miR-323b/154/496/377/541/409 miR-412/369/656	Protein secretion	Not known yet	(Hennessy et al. 2010)
Energy metabolism and other Cellular stresses	Let-7 family	Let-7 a/d/f	Glucose metabolism	INSR, IGF1R, IRS2, HMGA2	(Zhu et al. 2011)
	miR-23	miR-23a~24-1	Glutamine metabolism	GLS	(Rathore et al. 2012)
	miR-31	-	Hypoxia	HIF	(Liu et al. 2010a)

	miR-23	miR-23a/27a~24-2	Hypoxia	PU1	(Dang 2010)
	miR-126	-	Shear stress	Bcl2, FOXO3, IRS1	(Zhou et al. 2013)
	miR-375	-	Lactate production	LDHB	(Kinoshita et al. 2012)

Abbreviations: Akt protein kinase B protein, Bcl2 B-cell lymphoma 2, Bcl2L2 Bcl 2-like 2, Bcl-xL B-cell lymphoma-extra large, BIRC6 Baculoviral containing repeat 6, C1GALT1 Core-1 synthase glycoprotein N-acetylgalactosimine 3-beta-galactosyltransferase 1, CCND2 Cyclin D2 protein, CDKN1a/p21 Cyclin-dependent kinase inhibitor 1a, DAD1 Duo and Dam 1-interacting protein 1, ECOP EGFR co-amplified and overexpressed protein, EGFR Epidermal growth factor receptor, E2F1 and E2F2 Transcription factors, FUT8 Fucosyltransferase 8, FOXO3 Forkhead box O3, GALNT7 N-acetyl-galactosaminyltransferase 7, HMGA2 High mobility group AT-hook 2 protein, HIF Hypoxia inducible factor, HSP Heat-shock protein, IGF1R Insulin-like growth factor 1 receptor, IRS1 and IRS2 Insulin receptor substrate 1 and 2, INSR Insulin Receptor, LATS Large tumour suppressor, LDHB Lactate dehydrogenase B, PDCD4 Programmed cell death 4, PPP2R2A PPP2A regulatory subunit B alpha isoform, Psme3 Proteosome activator subunit 3, PTEN Phosphatase and tensin homolog, PU.1 Transcription factor for B-cell activation or differentiation, SIRT1 Sirtuin 1, Skp2 S-phase kinase-associated protein 2, SMO Smoothened gene, STAT5a Signal transducer and activator of transcription 5A.

Right from their discovery, miRNAs were thought to be critical players in cell-cycle control. Deregulation of their expression was associated with both tumour-suppressive or oncogenic phenotypes and their differential expression associated with cell cycle, cell-growth and cell death (Lee and Dutta 2006, Blenkiron et al. 2007, Porkka et al. 2007, Solomides et al. 2012). Bort et al. (2012) performed a detailed transcriptomic analysis in CHO cells, which identified 10 miRNAs as down regulated and their corresponding mRNA targets up regulated in a batch culture. These miRNAs played critical roles in cell cycle and apoptosis, making them targets for regulating cell growth and proliferation during late stage of the culture period. miR-7 was identified as a candidate for manipulation in CHO cell culture with a view to enhance longevity and protein productivity (Barron et al. 2011a, Sanchez et al. 2014). In addition to the positive impact of depletion of miR-7 on CHO cell culture growth and productivity, miR-7 overexpression impact on the global protein levels in CHO cells was also analyzed by Meleady et al (2012a). They reported up-regulation of the protein folding machinery associated proteins as well as protein secretion; however, targets regulating translation of proteins and nucleic acid were down regulated. In another transcriptomic study that utilized CHO-specific microarrays, Doolan et al. (2013) performed an extensive screening protocol on CHO-K1-mAb secreting cell-lines spanning a range of growth rates derived from single CHO cell line development. With the use of *in-silico* functional analysis, they identified a 416 transcripts associated with high growth rates (196 of which were up-regulated, while 226 down-regulated) (Doolan et al. 2013). These above mentioned in-depth differential miRNA expression profiling studies were performed to get a better understanding of the role of miRNAs in CHO cell-cycle regulation in serum-containing and serum-free CHO cell culture (Hackl et al. 2011a), at different culture time points (Bort et al. 2012) and with individual clones with different growth rates (Clarke et al. 2012, Doolan et al. 2013).

Additionally, many other groups have highlighted the importance of these global gene expression regulating RNA molecules as potential tools to manipulate desired phenotypes, with improved protein titers and protein with desired attributes in CHO cells. With the flood of genome sequence data for CHO in the last few years and the advantage of using miRNAs could prove very promising in optimizing CHO cell growth and productivity, in comparison to single-gene CHO cell engineering

approaches. With this very aim, this thesis is also an attempt to screen for the two potential miRNAs i.e. Let-7a of the big Let-7 family, and miR-7a-5p knockout role in impacting CHO cell growth behaviours.

1.6.9.1 Let-7 family

Lethal-7 (let-7) was the first miRNA discovered in *C.elegans* and was reported to be involved cell fates at different larval transitions (Reinhart et al. 2000). Following the study of this miRNA in *C.elegans*, the discoveries of mature let-7 miRNAs in simple organisms such as fly, zebrafish as well as complex organisms such as mouse and human began. Interestingly, not only are the mature let-7 sequences and the timing of expression of many of let-7 family members in these organisms conserved but also it was observed that the genomic organization and clustering of many let-7 members is conserved as well (Sempere et al. 2002, Bashirullah et al. 2003, Sempere et al. 2003, Weber 2005). The let-7 family often present in multiple copies in a genome and the nomenclature of these miRNAs is a special one, in contrast to other miRNAs. For instance, multiple isoforms of let-7 miRNAs are indicated by a letter at the end of let-7 to indicate slight difference in the sequence and a number after the letter usually denotes that the same sequence is present in multiple genomic locations. For example, ten mature let-7 sequences are present in humans, transcribed and processed from 13 precursors. Three separate precursors produce mature let-7a sequence (let-7a-1, let-7a-2, let-7a-3) and precursors from two different genomic loci produce the let-7f (let-7f-1 and let-7f-2) sequence. Although the sequence and function of let-7 family is conserved across species, with few exceptions, the size of the let-7 family also differs between organisms. For example, *D.melanogaster* (fly) has only one let-7, *D.rerio* (zebrafish) has 11 mature let-7 sequences in 19 genomic locations, 9 in *C.elegans*, 12 members of let-7 in 8 loci in *M.musculus* (Roush and Slack 2008).

Let-7 miRNAs have been implicated in the regulation of stem-cell differentiation in *C.elegans* (Reinhart et al. 2000), neuromuscular development and adult behaviour in flies (Caygill and Johnston 2008, Sokol et al. 2008), limb development in chicken and mouse (Mansfield et al. 2004, Schulman, Esquela-Kerscher and Slack 2005) and to regulate cell-cycle i.e. role in cell proliferation and differentiation (Hatfield and Ruohola-Baker 2008). In addition to this, most let-7 family members have been linked to development of cancers of many types (Takamizawa et al. 2004, Johnson et

al. 2005, Thomson et al. 2006) and have been reviewed extensively in literature by (Thomson et al. 2006) and (Boyerinas et al. 2010). Most let-7 family members are reported to act as tumour suppressors (Esquela-Kerscher and Slack 2006, Slack and Weidhass 2006). For instance, let7a, let-7d and let-7f cluster is directly down regulated by *myc* overexpression (*myc* is an oncogenic transcription factor often deregulated in cancers), hence, promoting tumorigenesis. A study on the negative feedback loop of let-7a reports that let-7a represses *myc* expression in a Burkitt lymphoma cell line indicating the possibility of a negative-feedback loop (Rodriguez et al. 2004). However, it was demonstrated that the DNA methylation state, particularly, in the unusual oncogenic let-7a-3 miRNA is also associated with cancerous phenotypes. let-7 expression is reported to be regulated by DNA methylation. The methylation state of CpG islands is an epigenetic mechanism for the regulation of gene expression. The hypomethylated let-7a-3 miRNA on chromosome 22q13.31 is associated with CpG islands and resulted in the up-regulation of let-7a-3 levels, thereby inhibiting the growth of tumour cells in lung adenocarcinomas. In contrast to this observation, hypermethylation of let-7a-3 resulted in down-regulation of its expression in epithelial ovarian cancer (Brueckner et al. 2007).

cgr-let-7a-3p	CTGTACAGCCTCCTAGCTTTCC
cgr-let-7a-5p	TGAGGTAGTAGGTTGTATAGTT
cgr-let-7b-3p	CTATACAACCTACTGCCTTCCT
cgr-let-7b-5p	TGAGGTAGTAGGTTGTGTGGTT
cgr-let-7c-1-3p	CTGTACAACCTTCTAGCTTTCC
cgr-let-7c-1-5p	TGAGGTAGTAGGTTGTGTGGTTAA
cgr-let-7c-2-3p	CTATACAATCTACTGTCTTTCC
cgr-let-7c-2-5p	TGAGGTAGTAGGTTGTATGGTT
cgr-let-7d-3p	CTATACGACCTGCTGCCTTTCT
cgr-let-7d-5p	AGAGGTAGTAGGTTGCATAGTT
cgr-let-7f-3p	CTATACAATCTATTGCCTTCCT
cgr-let-7f-5p	TGAGGTAGTAGATTGTATAGTT
cgr-let-7g-3p	CTGTACAGGCCACTGCCTTGC
cgr-let-7g-5p	TGAGGTAGTAGTTTGTACAGTT
cgr-let-7i-3p	CTGCGCAAGCTACTGCCTTGC
cgr-let-7i-5p	TGAGGTAGTAGTTTGTGCTGTTA

Figure 1.6.9.1: List of different members of Let-7 and their mature sequences in Let-7 family (CHO cell). Image sourced from data by Hackl et al. (2011).

Additionally, a report by Johnson et al. (2007) assembled a list of direct targets of let-7 involved in controlling the cell cycle, cell division and cell proliferation genes in humans. These targets included 8 transcription factors, consistent with *C.elegans* let-7 direct targets (Großhans et al. 2005). The assessment of these targets via RT-qPCR when performing let-7 overexpression studies can be useful for validation purposes. More importantly in regard to our project, recently a study using anti-let-7a oligos to increase CHO cell specific productivity (by > 60 %) and improved cell viability (late stages of culture period) in two mAb producing CHO cell-lines surfaced. The researchers also reported an increased expression of let-7a targets involved in cell cycle related pathways and apoptosis upon endogenous let-7a inhibition by anti-let-7a

miRNAs (Greenlees et al. 2014).

1.6.9.2 miR-7

miR-7 (specifically, miR-7-1) is a mirtonic non-coding RNA molecule found in the intron of the hnRNP K (heterologous nuclear ribonucleoprotein K) gene on chromosome 9. miR-7 is a highly conserved miRNA, in terms of both genomic structure and function, among diverse species (e.g. annelids to humans), suggesting its significance in coordinating various physiological processes. The expression profile of most intronic miRNAs are highly correlated with the expression profile of their host gene, however, miR-7 has been reported to have a differential expression to that of its host gene (Aboobaker et al. 2005, Baskerville and Bartel 2005b). This evolutionary conserved miRNA is derived from 3 miRNA precursors- miR-7-1, miR-7-2, miR-7-3 in the human genome. The other two isoforms miR-7-2 and miR-7-3 are located in the intergenic region on chromosome 15 and in another host gene (non-protein coding), respectively (NCBI Database, hsa-miR-7). The mouse miR-7 homologs originate from mmu-mir-7a and mmu-miR-7b precursors. The mature miR-7 processed by cytoplasmic RNAase are miR-7-5p, miR-7-1-3p and mir-7-2-3p.

cgr-miR-7a-3p	CAACAAATCACAGTCTGCCATA
cgr-miR-7a-5p	TGGAAGACTAGTGATTTTGTTGTT
cgr-miR-7b-5p	TGGAAGACTTGTGATTTTGTTGTT

Figure 1.6.9.2: List of different members of miR-7 and their mature sequences (CHO cell). Image sourced from data by Hackl et al. (2011).

1.6.9.2.1 miR-7 role in cell proliferation and apoptosis

Lately, several studies have been conducted to disclose the regulation activities of miR-7 in influencing cell-cycle progression and apoptotic mechanisms. A larger context of these studies was mainly focused on understanding various cancer etiologies and therapeutic potential. Multiple studies have reported miR-7 as a potential player in mediating cell proliferation and apoptosis and its deregulation

(acting as a tumour suppressor) is associated with different types of cancers, including lung cancer (Rai et al. 2011, Xiong et al. 2014), breast cancer (Webster et al. 2009), gastric cancer (Zhao et al. 2013), glioblastoma (Karsy, Arslan and Moy 2012), colorectal cancer (Zhang et al. 2013), melanoma (Giles et al. 2013), cervical cancer (Liu et al. 2013), tongue squamous cell cancer (Jiang et al. 2010) and hepatocellular carcinoma (Fang et al. 2012). Many other contrasting studies have reported miR-7 function as an oncogene. For instance, Gu et al. (2015) reported that miR-7 was up regulated in renal cell carcinoma where cell migration and proliferation were suppressed using a synthetic inhibitor that caused miR-7 down regulation (Gu, Huang and Tian 2015a)

The transcriptional regulation of endogenous miR-7 expression is mediated by the homeodomain transcription factor homeoboxD10 (HoxD10). HoxD10 positively regulates the endogenous miR-7 expression by directly interacting with the putative miR-7 gene promoter region and the loss of HoxD10 is associated with increased invasiveness (Carrio et al. 2005, Reddy et al. 2008). Another transcription factor, *c-myc* has been reported to control the transcription of miR-7 by binding to the miR-7 promoter and enhancing its activity, thereby inducing lung cancer. Other targets of miR-7 such as EGFR, PI3K could also stimulate miR-7 expression by utilizing Ras/extracellular signal-regulated kinase (ERK/myc) and phosphatidylinositol-3 kinase (PI3K/v)-Akt murine thymoma viral oncogene homolog (Akt) signaling pathways (Chou et al. 2010). Therefore, it can be concluded that miR-7 coordinately mediate several signaling pathways at multiple levels, including a number of cellular pathways, which are cell cycle and apoptosis relevant both in normal and tumour cells.

1.6.9.2.2 Potential of miR-7 manipulation in CHO cells

Over 28,645 miRNAs have now been registered and catalogued in miRBase database, with over 1800 miRNAs being identified in the human genome. The use of next-generation sequencing techniques such as Illumina has led to the discovery of more than 400 conserved mature miRNA sequences, with a single study yielding 387 miRNA identifications, from 6 biotechnologically relevant CHO cell-lines (Hackl et al. 2011a, Johnson et al. 2011). The potential of mir-7 as a tool to enhance CHO cell productivity was demonstrated when transient overexpression of miR-7 induced

growth arrest, thus maintaining high and sustained viability, mimicking the temperature shift effect of CHO cell culture to increase production of recombinant therapeutic proteins (Barron et al. 2011b). The importance of miR-7 as a potential cell cycle regulator has also been reported by Sanchez et al. (2013). Sanchez et al. reported miR-7 is a key regulator of the transition from G1 to S phase of the cell cycle, and can cause cell-cycle arrest by targeting an array of cell-cycle regulators components such as Skp2, *myc* and $p27^{KIP1}$. In a follow up study, recently, Sanchez et al. reported over 3-fold improved CHO proliferation and ~ 30 % cell viability increase and 2-fold increase in protein productivity by using miR-7 sponge decoy vector to sequester endogenous miR-7 levels in CHO cells (Sanchez et al. 2014). Hence, demonstrating miR-7 as a very interesting candidate to be further explored in enhancing bioprocess relevant CHO cell phenotypes. With this objective in mind, the second half of the project is focused on completely knocking out mature miR-7a-5p locus sequence in industrially relevant CHO cell-line in order to enhance CHO cell growth and productivity. While the first half aims to manipulate endogenous Let-7 levels using miRNA sponges approach in CHO-K1 SEAP cells to achieve the same objective.

Objectives of Thesis

1. Investigation of Let-7 miRNA as a potential candidate for CHO cell engineering

To:

- Stable depletion of endogenous Let-7 miRNAs in CHO-K1 SEAP cells using Let-7 sponge 'decoy' vector.
- Evaluate the impact of stable Let-7 depletion on CHO-K1 SEAP cells growth and productivity characteristics.
- Validate the Let-7 sponge vector technology using:
    - Gain-of-Function approaches through the use of Let-7 specific mimics.
    - To measure endogenous Let-7a levels using RT-qPCR.
    - Quantitating the expression levels of downstream targets of Let-7 using RT-qPCR

2. Use the novel CRISPR-Cas9 system to knock out miR-7a-5p locus in an industrially relevant CHO cell-lines

To:

- As a Proof-of-Concept Experiment: Design CRISPR-Cas9 vectors containing sgRNA(s) targeting a stably integrated eGFP in CHO-eGFP cell line.
- Evaluate and assess the indels frequencies induced by the eGFP targeting CRISPR-Cas9 systems using Surveyor Assay and Sanger sequencing.
- To target the miR-7a-5p locus using CRISPR-Cas9 vector systems in an industrially relevant CHO cell line.
- Evaluate and assess the indels frequencies generated by the CRISPR-Cas9 systems employed.
- Evaluate and assess mature miR-7a expression levels in single cell clones.
- Isolate a clone containing homozygous deletion of miR-7a.

Section 2.0

Materials and Methods

2.1 General Cell Culture Techniques

2.1.1 Ultrapure Water

All the media, solution and reagents' dilutions were performed using Ultrapure water from the Elgastat Maxima Water Purification System, 12-18 MΩ/cm. This water purification unit combines four processes – reverse osmosis, adsorption (using activated carbon), deionization and photo oxidation (UV irradiation to destroy micro-organisms of any sorts), yielding water quality of highest purity.

2.1.2 Sterilisation of Glassware and other consumables

All the lab glasswares and lids were initially soaked in a 2% (v/v) RBS-25 (VWR International) de-proteinising solution for ~ 1 hour (hr). Following this, the glasswares were washed and rinsed using tap water, before treating them with detergent. All the glasswares were then washed and rinsed with distilled water followed by rinse with ultrapure water. Finally, all the glasswares were sterilised by autoclaving at 121 °C for 20 minutes under one bar pressure.

The spinner vessels were cleaned in the same manner as mentioned above, however, included an extra step of 1M NaOH treatments overnight in order to ensure the elimination of cell debris from the inner surface of the vessels. These vessels were subsequently washed twice using ultrapure water. The thermo-labile solutions (i.e. 10% DMSO, serum) were filtered through 0.22 μm sterile filter (Millipore, millex-gv, SLGV-025BS).

2.1.3 Cell culture cabinets

Class II laminar airflow (LF) cabinets (Holten) were used during the cell culture work. The cleaning of LF cabinets was carried out by using 70% industrial methylated spirits (IMS) before and after use. Before using the LF cabinet, the laminar was turned on for 15 minutes to allow for the airflow to acclimatize. After 15 minutes, the laminar was cleaned with 70% IMS solution before performing any cell

culture task. Items brought into the LF cabinets were sprayed with 70% IMS solution and regular spraying of the gloved-hands (while in and out of the cabinet) with IMS was performed to maintain sterile conditions. The cabinets were always used with single cell line type work to avoid any cross-contamination. When working with a different cell line subsequently, the cabinets were cleared-out with all the previous cell line type and waste-tips/pipettes and cleaned with 70% IMS. The empty cabinets were allowed to run for 15 minutes before to be used again for a different cell line. This standard operating procedure (SOP) was adhered in order to prevent any cross-contamination between cell lines/clones. These LF cabinets were weekly cleaned using Virkon detergent solution (Virkon, Antec International; TEGO, TH. Goldschmidt Ltd.), followed by ultrapure water and 70% IMS.

2.1.4 Incubators

Kuhner (ISF1-X, Climo-Shaker) incubators with speed ranging 130-170 rpm were used to maintain suspension cells at 37 °C in an atmosphere of 5 % CO_2 and 80 % humidity. Adherent cells were also maintained at the conditions mentioned above, however, not in shaker incubators. The incubators were cleaned weekly adhering to the same protocol as described for cell culture cabinets.

2.2 Subculture of cell lines

2.2.1 Anchorage-dependent cells (Monolayer)

The spent media in the flask was removed using pipette and discarded into a sterile waste bottle. Following this, the flask was rinsed using 2-4 mL of PBS (autoclaved/sterile) to remove any residual media. Depending upon the size of the flask, 2-5 mL of trypsin was added to the flask forming thin film enough to cover the monolayer cells surface and incubated at 37 °C for ~ 2-5 minutes (Note: Did not over-trypsinize) until all the cells have detached. A volume of serum equivalent to the amount of trypsin added was added in order to deactivate the trypsin once all the cells were detached. The cell suspension was then pipetted into a sterile 30 mL universal and centrifuged at 1000 rpm for 5 minutes in order to form a cell pellet. The

supernatant from the universal was discarded and the cell pellet was resuspended in fresh basal media. A small aliquot was then taken and cell count was performed. Finally, a required aliquot from this cell suspension was used to seed a fresh flask at a desired cell density. Routinely, cells were passaged once every 3-4 days at ~ 90-100 % confluency.

2.2.2 Suspension cells

30-50 mL shake flasks (working volume 5 mL) were used for suspension culture cells at 37 °C in a shaker incubator (as described in section **2.1.4**) at 170 rpm. The cells were passaged routinely every 3-4 days. A sample aliquot was taken from the flask and used for counting viable cell number using the Trypan blue exclusion method. Fresh flasks were then seeded at 2 x 10^5 cells/mL density.

2.3 Cell counting and viability determination

2.3.1 Trypan Blue Exclusion method

The Trypan blue exclusion method allows one to distinguish between live and dead cells. The rational behind the technique is that the trypan dye penetrates and stains the dead cells blue through the damaged/compromised cell membrane. Viable healthy cells with an intact cell membrane appear non-blue making it easier to distinguish the two when observed under microscope. A cell sample was taken from the flask and equal amount of Trypan blue was added (Note: The sample was diluted accordingly as required i.e. if too concentrated) to make a final mix of 100 µl (with diluent)). 10 or 20 µl of this mix was pipetted onto the hemocytometer coverslip-on carefully to avoid any air-bubbles formation or release. The number of cells was then counted by the cell count average of four grids (corners) in the hemocytometer. This number was then multiplied by 10^4 (volume of the grid) and by the dilution factor to get the number of cells/mL in the original sample.

2.3.2 Flow cytometry

2.3.2.1 Guava Viacount® Assay

A greater number of samples and their viability can be assessed using the Guava ViaCount assay in comparison to the Trypan Blue Exclusion method. The assay also discriminates between viable and non-viable cells based on differential permeability of two DNA-binding dyes in the Guava ViaCount® reagent (Merck-Millipore). The membrane permeant-dye (LDS 751) stains all the nucleated cells producing a fluorescent signal detected by the photomultiplier tube 2 (PMT2). The other dye penetrates the damaged/compromised membrane of non-viable cells producing a different fluorescent signal detected by PMT1. After detection by PMT, viable cells are recorded as single positive events when the fluorescent signal is accompanied by a signal of forward-scattered light (FSC) and is of equivalent intensity to that produced by a cell of particular size. A low FSC event is counted as cellular debris and is unaccounted within the nucleated population count. The Guava Flow cytometer instrument allows setting a threshold prior to the analysis in order to exclude the low FSC events. The membrane impermeant-dye (Propidium Iodide, PI) stains only the damaged/compromised membrane non-viable/dead cells. This combination of the dyes in the Guava ViaCount® reagent differentiates between viable, non-viable and apoptotic cells.

Before performing any analysis on the Guava Flow Cytometer instrument, a calibration run (easyCheck) was performed in order to perform assay correctly. The samples for anaylsis were pipetted into round bottom 96-well plate and the minimum sample volume was 100 µl/well. It was important that the cell count per well has no less than 10 cells/µl and no more than 500 cells/µl since too dilute or concentrated samples could lead to error-prone read-outs. The Guava ViaCount® reagent was 1:1 ratio with sample volume, even if the sample was diluted with sterile PBS. For example, for 50 µl of cell sample + 50 µl of sterile PBS = 100 µl diluted cell sample requires 100 µl of reagent. The Guava ViaCount® was allowed to incubate at room temperature 30 minutes prior use.

2.3.2.2 EasyFit Analysis

EasyFit analysis is a method in-built in the Guava software, which groups all the events into three individual categories i.e. viable cells, dead/compromised/apoptotic cells and cellular debris. This analysis using its own grouping method allows for the discrimination between these three populations. In order to set a threshold, manually a method can be executed, for example for 'viable cells' by changing the FSC threshold and hence, completely discriminating between cellular debris and viable cells.

2.3.2.3 GFP expression analysis

GFP positive cells were assessed for mean GFP and % GFP expression using the bench-top flow cytometer Guava's Guava ExpressPlus programme. Firstly, a GFP negative cell line was used to gate the instrument. Forward scatter and side scatter were set to observe the cells on the histogram. The fluorescein isothicyanate (FITC) channel was manipulated depending upon the cells to be analysed for GFP expression.

2.4 Cryopreservation, Thawing and Storage of cells

2.4.1 Cryopreservation

Cells to be preserved in liquid nitrogen dewar were supplemented with suitable medium with 5 % fetal calf serum (FCS) and 5 % DMSO (Sigma-Aldrich, D5879). Firstly, a separate solution of basal medium supplemented with 10 % DMSO/FCS mix was prepared. For sterility purposes this solution was filtered through 0.22 μm bell filter (Gelman, 121-58) and kept on ice until used. DMSO has cytotoxic effects at room temperature. It was imperative that cells to be cryopreserved were harvested during the log phase of their growth cycle and ~ 1×10^6 cells/mL per cryovial (Greiner, 201151) were frozen down. After counting and centrifuging, the cells were resuspended in half the 1 mL volume of suitable medium making the density 2×10^6 cells/mL in a cryovial. On top of this cell suspension, 500 μl of the sterile filtered DMSO/FCS mixture was pipetted. Finally, the cryovial was slowly frozen and was

kept in -20 °C for ~ 1 hour, then moved to -80 °C overnight (O/N) and the next day into liquid nitrogen dewar at-196 °C.

2.4.2 Thawing

The simple methodology of 'slow cooling (while cryopreserving cells) and rapid thawing' was followed. In order to thaw the cryopreserved cells, they were removed from the liquid nitrogen, immediately placed on ice and then slightly thawed at 37 °C. 5 mL of pre-warmed fresh medium was kept in a sterile universal. Using a sterile Pasteur pipette, 1 mL by 1 mL of the pre-warmed medium was used to resuspend the cell (in the form of ice) and pipetted into a separate sterile universal until all of the cell-ice was resuspended. This resuspended cells solution from the cryovial was then centrifuged at 1000 rpm for 5 minutes in order to remove the toxic DMSO supernatant. The cell pellet was then resuspended in fresh pre-warmed desired medium in desired cell culture format. The freshly revived cells were not passaged with the selective media for the first passage.

2.5 Other Cell-culture related techniques

2.5.1 Limited Dilution Cloning (LDC)

In order to attain single-cell clone, cells were diluted using a serial dilution method in 96-well plates. Cells were harvested at ~ 90 % confluency during the log growth phase and were trypsinized as per **section 2.1.5.1**. These cells were then fed with fresh medium prior to the dilution process. When the cells were pelleted, the pellet was resuspended and mixed by pipetting up and down several times in 10 mL of media to ensure precise dilution. An aliquot from this 10 mL cell suspension was then counted 3-4 times for accuracy using the Trypan blue exclusion method (**section 2.1.6.1**). Serial dilution was performed with pre-warmed medium with sterile filtered conditioned medium (extracted 24-48 hours after culture). At the end of dilution, the final concentration in 50 mL of fresh pre-warmed medium and condition medium mix was 5 cells/mL. Using an automatic pipette pump, aliquots of 100 µl were transferred to a 96-well plate already containing the desired pre-warmed medium to get one clone

per two wells. The cells were then incubated at 37 °C for a week and were intermittently observed under microscope for scoring clones in each well. The medium used during this procedure contained 50 μg sterile-filtered Penicillin/Streptomycin solution in order to avoid Gram-negative and Gram-positive bacterial contaminations) and appropriate Selective Agents.

2.5.2 Fluorescence-Activated Cell Sorting (FACS)

An automated machine such as FACS is a more advanced, timesaving alternative to LDC. It accurately and positively dispenses one clone per well into a 96-well plate. FACS is a flow-cytometer technology used for sorting out cells based on their size, shape and fluorescence properties. The process starts when cells are directed into a stream that forms droplets after which a laser is directed at the droplets to measure these properties. The droplets may contain fluorescent cells that, when detected are sorted based on the level of fluorescence i.e. low, medium and/or high by the FITC channel into 96-well plates containing conditioned medium along with desired basal medium supplemented with serum. The instrument is initially gated using GFP negative cells to allow the software to distinguish between GFP positive and GFP negative cells.

2.6 Transfection

2.6.1 Transfection of plasmid DNA

Several reagents were utilized for transfecting DNA. Certain transfection reagents work with greater efficiency depending upon the cell lines used.

2.6.1.1 Lipofectamine 2000

Prior to transfection cells were seeded 5×10^5 cells/mL in adherent mode 24 hours before transfection in suspension culture at a density of $0.5 - 1 \times 10^6$ cells/mL. A separate solution containing DNA and DMEM F-12 or CHO-SFM (50 μl) media was prepared to dilute DNA. Lipofectamine™ (Invitrogen, Cat. No. 11688-019) was diluted in a separate 50 μl DMEM F-12 or CHO-SFM media. Both these mixtures

were then gently mixed (making the volume 100 µl) and incubated at room temperature for 25-30 minutes, allowing for enough time to form complexes of DNA and Lipofectamine. The ratio of DNA to Lipofectamine used was 1:2. After 30 minutes of incubation, the 100 µl of complex mixture was pipetted drop-wise in circular manner and gently mixed by shaking the plate up and down and left-right. Since this transfection reagent could be toxic to cells, therefore, the media after 4- 6 hours post-transfection was changed.

2.6.1.2 TransIT-2020 reagent

Similarly, the complexes were formed for the Trans-IT®-2020 transfection reagent in CHO-SFM media in 250 µl volume. 1 µl of Trans-IT®-2020 (Mirus, Cat. No. MIR-5400) was added to pre-warmed 250 µl CHO-SFM media containing 1 µg plasmid DNA to be transfected. The contents were mixed and incubated at room temperature for 20-30 minutes. Following the incubation period, the mixture containing DNA/transfection reagent complex was added drop-wise to the cells. Trans-IT®-2020 is a non-toxic transfection reagent; therefore, it was not mandatory to change the media post-transfection.

2.6.1.3 Electroporation

Prior to the transfection, the required materials such as plasmid DNA, cells, buffer, cuvettes were incubated on ice for 30 minutes. In a 1.7 mL eppendorf tube, 800 µl of chilled filtered sterilised electroporation buffer (2 mM HEPES – 15 mM K^+PO_4, 1 mM $MgCl_2$ and 250 mM Mannitol in 50 mL final volume) 8×10^6 cells were seeded. The desired volume of pre-warmed media, depending upon the culture volume format (e.g. 5 mL spin tube or 10 mL T75 flask with 1 % serum if original culture was in serum-free media), was used to re-suspend cells post electroporation. The pulsing machine was set at standard 500 µF and 200 volts. Mammalian cells require high capacitance (> 50 µF). The electroporated cells were then finally revived using pre-warmed desired media of choice by gently pipetting up and down and incubated at 37 °C.

2.6.2 Transfection of miRNAs

The following recipe for transfection of miRNA is valid for a single transfection reaction. The PolyPlus-transfection™ reagent INTERFERin™ (PolyPlus-transfection™, PPLU-409-01) was utilized for CHO cells transfection as per manufacturer's guidelines. 1×10^5 cells/mL per well of a 6-well plate were transfected with a range of 100-200 nM per 2 mL volume of CHO-S-SFM II (Gibco®, 12052-098) media in 6-well plate. The reagents such as INTERFERin™ and miRNA (mimics and/or inhibitors) were kept on ice until used. INTERFERin™ was brought to room temperature 5-10 minutes prior to use. Initially 110 µl of pre-warmed SFM media was aliquoted in sterile eppendorf tube and 2.2 µl of mimic/inhibitor miRNA was pipetted into the eppendorf with 110 µl SFM media and mixed gently by pipetting up and down. Finally, 1.1 µl of INTERFERin™ reagent was added to the aforementioned mixture of miRNA and SFM media. This final mixture was incubated at room temperature for 15-20 minutes. Post incubation. 100 µl of the complex mix was pipetted drop-wise into the well of 6-well plate and swirled briefly for uniform mixing. The transfected cells were then allowed to grow at 37 °C post-transfection in the incubator. INTERFERin™ reagent is a non-toxic transfection reagent and does not lead to cytotoxicity; therefore, the media change was not mandatory.

2.7 Molecular Biology Techniques

2.7.1 DNase Treatment and RNase-free/Sterile tips

For the below mentioned techniques, only RNase-free/Sterile pipette tips were used in order to avoid false-positive results and/or contamination. Moreover, the DNase (RNase-free kit, M0303S, New England Biolabs®) treatment of RNA samples is useful for the removal of contaminated genomic and plasmid DNA from transfection performed during the early stages of experiment. The contaminated DNA sources are highly abundant post-transfection and during RNA isolation protocols as carry-over sources.

2.7.2 RNA extraction using Tri Reagent®

RNA was isolated using the Tri Reagent®, which is a mixture of guanidine thiocyanate and phenol in a monophase solution. This reagent dissolves the DNA, RNA and protein on homogenization or lysis of the cell sample. 1 mL of Tri Reagent® was added to sufficiently lyse 5/6 × 10^6 cells/mL and then vortexed for 10-15 s. After vortex, 200 µl of chloroform was added and vortexed for 10 s. The sample was then centrifuged at 12000 rpm at 4 °C for 10-15 minutes to separate the DNA, RNA and protein into three different phases: a upper aqueous phase (RNA), the interphase (DNA) and a lower organic (proteins).

The upper aqueous RNA phase was then decanted carefully (i.e. without disturbing the interphase) into a fresh eppendorf. RNA was then precipitated by adding 500 µl of iso-propanol and centrifuging at 12000 rpm at 4 °C for 10 minutes. The supernatant was discarded and the RNA pellet was then washed with 70 % ethanol and again spun down at 12000 rpm at 4 °C for 10 minutes. The supernatant was carefully decanted-off and the RNA pellet was air-dried under vacuum for 10 minutes and finally resuspended in 200-300 µl of Nuclease-free water or appropriate elution buffer. The RNA sample can be stored without compromising the sample integrity at -80 °C.

2.7.3 High Capacity cDNA reverse transcription

RNA was reverse transcribed into cDNA using High Capacity cDNA reverse transcription kit (Applied Biosystems, 4368814) as per manufacturer's protocols. 2 µg of RNA sample was made upto 10 µl in a 0.5 mL PCR tube using Nuclease-free water. A master-mix as suggested in the protocol was prepared in 1.7 mL sterile eppendorf (depending upon the reactions to be performed). Table below shows a single reverse transcription reaction master mix recipe in accordance to manufacturer's guidelines:

Component	Volume per reaction (µl)	
	Kit with RNase inhibitor	**Kit without RNase inhibitor**
10X Reverse Transcription (RT) Buffer	2	2
25X dNTP Mix (100 mM)	0.8	0.8
10X RT Random Primers	2	2
Multiscribe™ Reverse Transcriptase	1	1
RNase inhibitor	1	-
Nuclease-free H_2O	3.2	4.2
Total per reaction	**10.0**	**10.0**

10 µl of the RT master mix was added to 10 µl of the above prepared RNA sample and mixed using a pipette. The bench top thermocycler was set as per the below tabulated temperature profile:

	Step 1	**Step 2**	**Step 3**	**Step 4**
Temperature (°C)	25	37	85	4
Time (mins)	10	120	5	∞

The reverse transcribed cDNA is ready to use and can be stored at -20 °C for future use.

2.7.4 Polymerase Chain Reaction (PCR)

Phusion™ High-Fidelity PCR Master Mix with High-Fidelity buffer kit (Life Technologies, F-531S) was used to perform most of the PCR amplification reactions as per manufacturer's specification. The Master Mix was thawed on ice prior to setting-up reaction until used. A single PCR reaction was made up to a volume of 50 µl in 0.5 mL PCR tube. Below mentioned is the recipe for single PCR amplification

reaction, with the Phusion™ Master Mix added at the very last after all other components were added.

Component	Volume/Concentration
Phusion™ PCR Master Mix	25 µl
Forward Primer	0.1 – 1.0 µM
Reverse Primer	0.1 – 1.0 µM
Template DNA	1 – 100 ng
Nuclease-free H_2O	To 50 µl
Total Volume	**50 µl**

Once all the components were added, the PCR tube was spun and vortexed briefly. After this, the thermocycler was used to amplify the product of interest. The cycler was manipulated according to the annealing temperature (T_m) for each set of primer used and length of amplicon to be amplified. Below is the programme used to perform the polymerisation reaction cycle.

Step	Temperature (°C)	Time	No. of cycles
Initial Denaturation	98	1 – 3 minutes	1
Denaturation	98	30 s	25 - 30
Annealing	T_m - 5	30 s	25 - 30
Extension	72	1 minutes/kb	25 - 30
Final Extension	72	5 – 15 minutes	1
HOLD	4	∞	1

Following PCR completion, samples are ready to be used and can be stored at -20 °C until further use.

2.7.5 Fast SYBR® Green Real-Time Quantitative PCR

RT-qPCR from Applied Biosystems™ Fast SYBR® Green RT-qPCR Master Mix (Applied Biosystems™, Cat. No. 4385612) was used to perform qPCR reactions for gene expression quantification, in accordance with manufacturer's specification. The quantification of the gene of interest is based on the binding affinity of SYBR® Green for double-stranded (ds) DNA. As the PCR amplification progresses by the AmpliTaq® DNA polymerase, the target sequence is amplified, resulting in increased number of ds-amplicon to bind to SYBR® Green. As the PCR amplification proceeds, the fluorescent intensity is proportional to the increase in ds-amplicon.

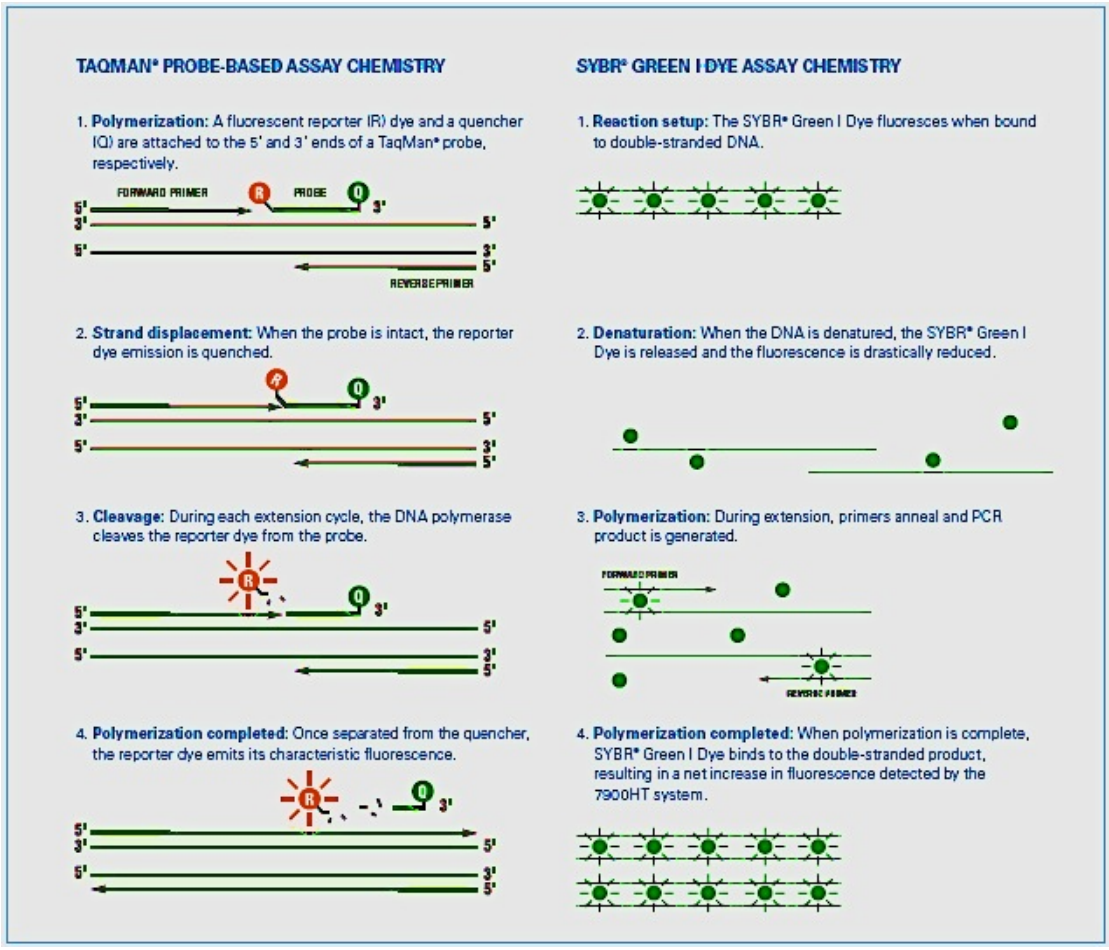


Figure 2.1 shows RT-qPCR technique for two types of assay. One based on Fast SYBR® Green RT-qPCR and other based on Taqman Probe based RT-qPCR.

Prior to plate set-up, master-mixes for all samples (i.e. test samples and endogenous controls such as beta-actin and glyceraldehyde 3-phosphate dehydrogenase (GAPDH), in triplicate) were prepared. A single PCR reaction volume was kept to 20

μl as per manufacturer's guidelines and the reaction was performed in MicroAmp™ Fast Optical 96-well plate. The following recipe is for a single 20 μl PCR volume: 10 μl of Fast SYBR® Green Master Mix (2X) was added to an eppendorf with 1μl of primers- forward and reverse (200 mM each). 18 μl each was added to each well of the MicroAmp™ 96-well reaction plate from the master mix for triplicate reactions for both control and test samples. Using a fresh pipette tip for each time, 2 μl of cDNA template (~ 20 ng to achieve optimal detection of signal) was aliquoted to each of the wells of the MicroAmp™ 96-well reaction plate already containing 18 μl of Master Mix (aforementioned). Finally, the reaction plate was sealed using a MicroAmp™ Optical adhesive film to avoid any reaction volume loss through evaporation during the PCR run. The PCR run was performed on the Applied Biosystems RT-qPCR 7500. The data analysis was based on the Relative Quantitation ($2^{-\Delta\Delta Ct}$) and the relative expression of the target gene was compared to the expression of endogenous gene as control.

2.7.6 miRNA based Molecular Biology techniques

2.7.6.1 Reverse Transcription for miRNA(s)

Applied Biosytems Taqman microRNA Reverse Transcription kit (ABI, 4366596) was used to prepare separate RT master mixes for each sample according to manufacturer's specification as tabulated below:

Component	Master Mic (15 μl/reaction)
100 mM dNTPs (incl. dTTP)	0.15
MultiScribe™ Reverse Transcriptase (50 U/μl)	1.0
10X Reverse Transcriptase Buffer	1.5
5X RT Primer	3.0
RNase Inhibitor (20 U/μl)	0.19
Nuclease-free water	4.16
Total	**10**

RNA samples were diluted to a working concentration of 2 ng/µl, made in nuclease-free water. 5 µl (20 ng) of RNA sample from this working concentration was added to each of the above prepared 10 µl RT-master mix, which was then gently mixed and centrifuged. In order to validate for any genomic contamination of the primers and reaction components, a separate master mix reaction without the MultiScribe™ Reverse Transcriptase for each primer set was prepared within the assay.
Finally, the cycler programme was used as per manufacturer's guidelines:

Step	Time (minutes)	Temperature (°C)
HOLD	30	16
HOLD	30	42
HOLD	5	85
HOLD	∞	4

cDNA samples were in ready-to-be-used condition for qPCR quantitification of miRNA using specific primers and/or could be stored at -20 °C until further needed.

2.7.6.2 RT-qPCR for miRNA expression analysis

The Taqman 2X Universal PCR master mix (Applied Biosystems, ABI, 4324018) was utilized for individual miRNA expression analysis. According to manufacturer's recipe specifications (tabulated below), a single reaction was prepared:

Component	**20 µl/reaction**
Taqman miRNA Assay (20X)	1.00
Product from RT reaction (Min. dilution 1/15)	1.33
Taqman 2X Universal PCR Master Mix, No AmpErase® UNG)	10.00
Nuclease-free water	7.67
Total	**20**

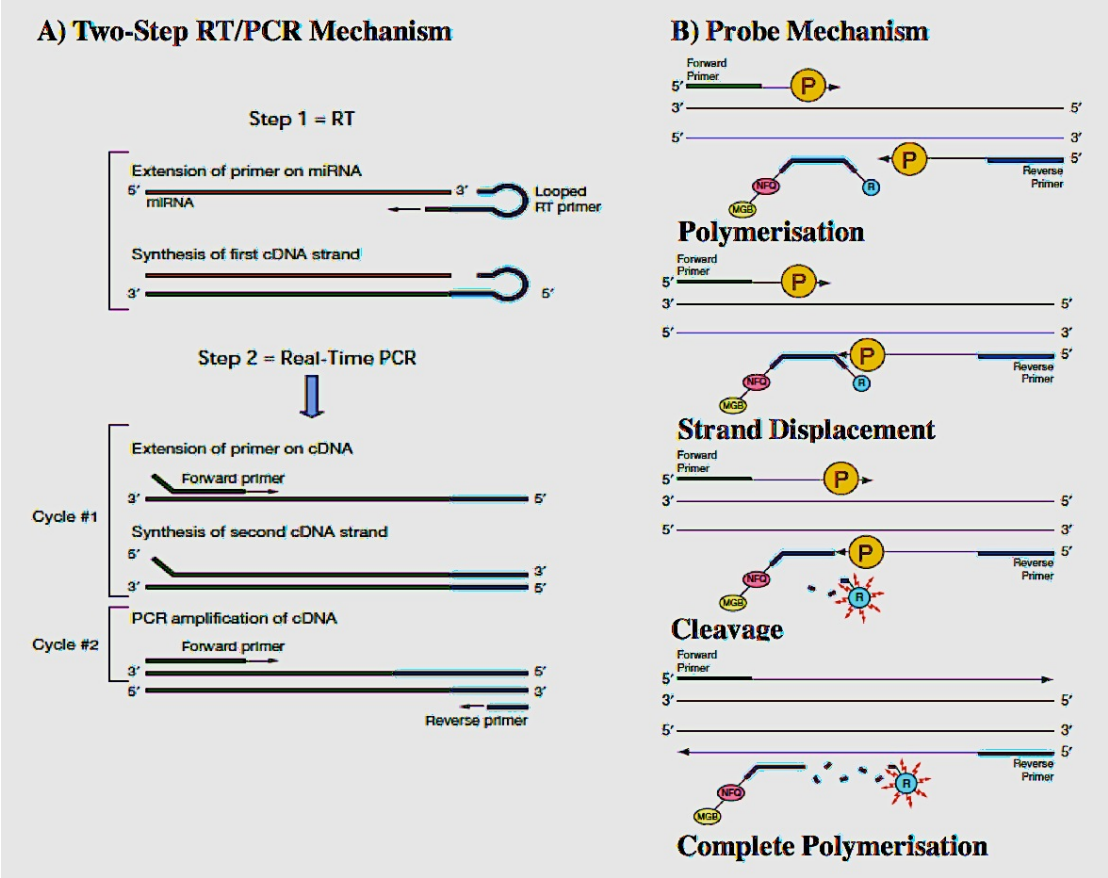


Figure 2.2 showing the two-step RT/PCR mechanism (i.e. reverse transcription followed by PCR for cDNA conversion and subsequent amplification of the cDNA) (A), wherein in Step 1 – stem loop RT primers anneal to miRNA targets and extended further by the reverse transcriptase enzyme and Step 2 – PCR reactions progress with miRNA-specific forward and reverse primer and detection of signal via the Taqman® probe. Figure 2.2 (B) shows the mechanism by which the Probe functions. The quantification of miRNAs is based on the C_t values. MGB: minor groove binder; NFQ: non-fluorescent quencher; R: reporter dye (FAM™ dye).

The MicroAmp™ optical reaction plate was used for the qPCR. From the above-mentioned master-mix, 16 µl, containing the appropriate primer (or here the Taqman miRNA Assay), was aliquoted into the plate. The plate was then sealed with standard optical adhesive film and incubated on ice for 10 minutes prior to the qPCR-run. Finally, using the 9600 emulsion mode of thermocycler, the qPCR was performed in the HT7600 RT-qPCR machine with conditions tabulated below:

Step	AmpliTaq Gold® Enzyme Activation	PCR	
	HOLD	Cycles (40)	
		Denatured	Anneal/Extend
Time (minutes)	10	15 s	60 s
Temperature (°C)	95	95	95

Below mentioned are the assay kits used during the course of this Master's Degree.

Applied Biosystems™ miRNA assay kits		
miRNA	**Cat. No.**	**Assay ID**
U6 snRNA (Endogenous control)	4427975	001973
hsa – Let7a – 5p	4427975	000377
hsa – miR7a-5p	4427975	000386

2.8 Cloning Techniques and other basic reagents preparations

2.8.1 Gel and Lysogeny Broth (LB) media preparations

2.8.1.1 Agarose gel preparation

Agarose (Sigma-Aldrich, A9539) was aliquoted into 1X TAE (Tris base, acetic acid and EDTA) (Thermo, #B49) buffer as desired volume for desired gel resolution. For example, 2 % (w/v) being considered as high percentage for resolution for ~ 100 bp DNA bands. This solution of agarose and 1X TAE buffer was then heated until all of the agarose was completely dissolved and then cooled to hand-hot condition. Following this, 1 µl of ethidium bromide (0.5 µg.mL) was added and mixed as a stain for DNA in the gel when observed under UV trans-illuminator. Finally, the hand-hot

agarose, TAE and dissolved ethidium bromide was poured into a desired gel tray for further use.

2.8.1.2 LB media preparation

For 500 mL of LB media, 5 g of Tryptone (Sigma-Aldrich, #T7293), 2.5 g of Yeast Extract (Sigma-Aldrich, #92144) and 5 g of NaCl (Sigma-Aldrich, #S9888) was dissolved in 500 mL of UHP and subsequently autoclaved.

2.8.1.3 LB agar plate preparation

For the LB agar plate media, the recipe aforementioned was the same except addition of 7.5 g of Select Agar (Sigma-Aldrich, #A5306) in 500 mL of UHP and then autoclaved. When the LB agar was hand-hot, desired antibiotics were pipetted into it and then, finally, poured into 100 mm petri dishes (plastic material).

2.8.2 Cloning techniques

2.8.2.1 Restriction enzyme digestion

Digestion reactions were performed in 0.5 mL PCR tube with the recipe outlined below:

Component	**Volume (µl)**
Plasmid DNA	2 µg
NEB Buffer, 1-4 (10X)	2
BSA (100X)	0.2
Restriction enzyme	1 U (0.5 – 1 µl)
Nuclease-free water	To 20 µl reaction volume

The plasmid DNA was digested at 37 °C in the hot-hole plate for the specified temperature according to enzymes' manufacturers specifications. Both single and double digestions can be carried out with various enzymes under specified conditions.

2.8.2.2 Alkaline phosphatase treatment

Alkaline phosphatase, New England Biolabs (NEB), enzymes such as Calf-intestinal phosphatase (CIP) and/or Shrimp alkaline phosphatase (SAP) catalyse the removal of 5'-phosphate group from DNA/RNA and prevents the CIP or SAP treated DNA to self-re-ligate after the circularised plasmid has been linearized. 2 μg of digested (and/or linearized) plasmid DNA were transferred to 0.5 mL PCR tube containing 5 μl of 10X CIP enzyme (20 U/μl) in a final volume reaction volume of 50 μl. This PCR tube containing the reaction mix was incubated at 37 °C for one hour.

2.8.2.3 Large *Klenow* fragment treatment

A 30 μl of reaction was set-up using 25 μl of digested plasmid DNA in 10X NEB buffer 2, 33 μM of dNTPs and 3X *Klenow* enzyme (5 U/μl). This reaction tube was incubated at 25 °C for 15 minutes, followed by a further incubation at 75 °C for 20 minutes. Finally, the reaction was stopped by adding EDTA at 10 mM concentration.
DNA polymerase I, large (*Klenow*) fragment is a proteolytic product of *E.coli* DNA polymerase I that retains the enzyme's 3' → 5' exonuclease acitivity and polymerisation ability, however, has lost the 5' → 3' exonuclease acitivity. This treatment is necessary in order to trim the overhangs in order to make them compatible with other restriction enzymes. Moreover, *Klenow* treatment also inhibits re-ligation of linearized plasmid DNA.

2.8.2.4 DNA ligation

T4 DNA ligase enzyme kit (ROCHE®, Cat. No. 10 481 220 001) was used for all the ligation reactions that were carried out. In order to ensure efficient ligation reaction, the recommended ratios of 1:3 and/or 1:5 (with insert to vector saturation) were used. The vector backbone used per reaction was in the range of ~ 50-150 ng, depending upon yields from purification processes. Each reaction was carried out with a reaction volume of 20 μl but also scaled up to 50 μl if required.
The following recipe was set-up for each ligation reaction: 2 μl of 10X T4 DNA ligase reaction buffer was added to the reaction mix (containing the backbone and

insert) and made 19 µl with nuclease-free water. Finally, 1 µl of T4 DNA ligase enzyme (1 U) was added to the reaction mix. This final reaction mix of 20 µl was incubated O/N on temperature gradient in an ice bucket (with floater eppendorf-holder). The controls such as 'No-ligase', 'NO insert' and others were also set-up with all the other mentioned components in order to account for DNA backbone's self-ligation frequency as well as incomplete digestion of the vector backbone resulting in false-positive results.

2.8.2.5 Kinase treatment

T4 Polynucleotide Kinase (T4 PNK) is used for the phosphorylation of the DNA at 5'-ends. 2 µl of 10X T4 kinase buffer was added to 0.5 µl (already added) of both the sense and anti-sense strands (annealed previously) in the 0.5 mL PCR tubes. 17 µl of nuclease-free water was added to make 20 µl reaction volume and at last 0.4 µl of Kinase enzyme (ROCHE®) was added to the 20 µl reaction mix. The reaction mix with kinase enzyme was incubated at 37 °C for 30 minutes. The kinase-treated DNA can be stored at – 20 °C for up to one year later.

2.8.2.6 Bacterial Transformation for cloning

Transformation experiments were done in One Shot® Max Efficiency® DH5α™-T1® competent cells. 50 µl of DH5α competent cells were thawed on ice prior to use for each transformation reaction. After thawing, 1 µl of ligation reaction/plasmid vector and/or DNA was added to 50 µl of competent cells. The mixture was then incubated on ice for about 20-25 minutes, followed by 42 °C heat-shock for ~ 30 seconds in a hot-water bath. The heat-shocked cells were then revived on ice for 2 minutes. Following this 300 µl of pre-warmed S.O.C rich medium was added to the vial and incubated at 37 °C at 225 rpm for one hour in shaker incubator. After incubation, the vial was centrifuged at 2000 rpm to make cell pellet and 85-90 % of the supernatant was discarded off. The rest of the supernatant was used to re-suspend the cell pellet and plated onto LB agar plate (containing the desired antibiotic) using a sterile spreader and incubated at 37 °C O/N.

2.8.2.7 TOPO-TA vector cloning

The TOPO® TA Cloning® kits provide a highly efficient 5-minute, one step cloning strategy for the direct insertion of *Taq*-polymerase amplified PCR products into a plasmid vector for sequencing. For this kit, no ligase, post-PCR procedures, or PCR primers containing specific sequences are required.

In this kit, a plasmid vector (pCR™4-TOPO®) is supplied linearized with (i) single 3' thymidine (T) overhangs for TA Cloning®, and (ii) Topoisomerase covalently bound to the vector (referred to as "activated" vector) (**Figure 2.3 and 2.4**). *Taq* polymerase has a non-template-dependent terminal transferase activity that adds a single deoxyadenosine (A) to the 3' ends of the PCR products. The "activated" linearized vector in the kit has single, overhanging 3' deoxythymidine (T) residues bound to it (Shuman 1991, Shuman 1994). This allows for the PCR inserts to ligate efficiently with the vector.

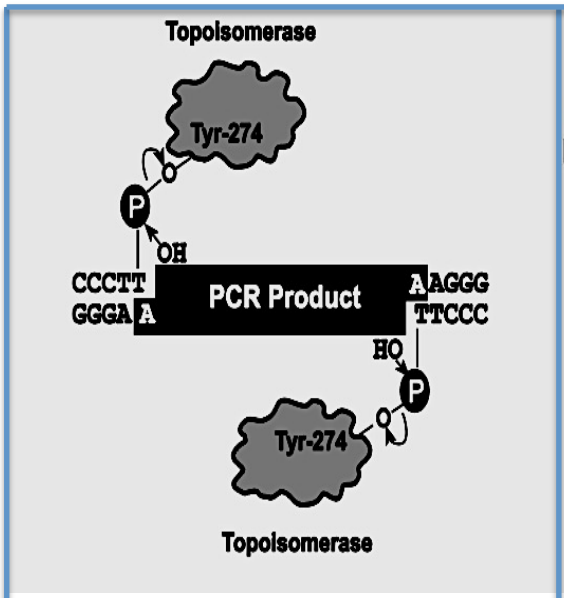


Figure 2.3 showing the "activated" vector with covalently bound Topoisomerase enzyme, which allows to efficiently clone PCR products into the linearized vector.

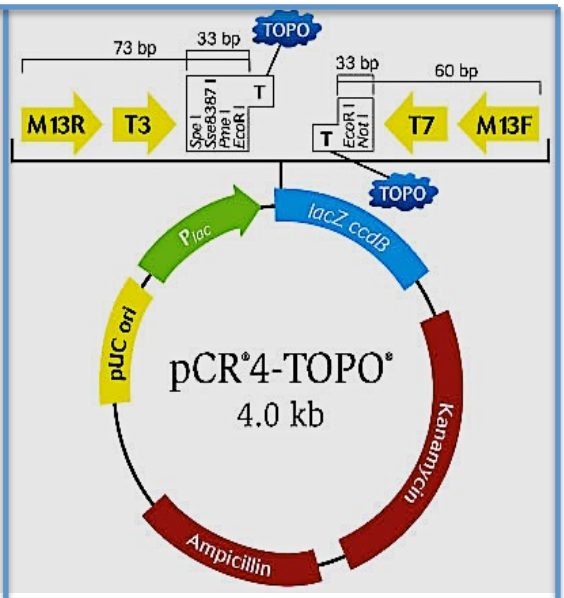


Figure 2.4 shows the complete vector map of the linearized pCR™4-TOPO®

After amplifying the PCR products using the *Taq* polymerase enzyme, they were cloned into the linearized pCR™4-TOPO® TA vector as manufacturer's specification. The recipe is tabulated below:

Reagents	Volume
Fresh PCR product (10-100 ng)	0.5 – 4 µl
Salt Solution	1 µl
Nuclease-free water	Add to 5 µl
pCR™4-TOPO® TA vector (last)	1 µl
Final Volume	**6 µl**

This reaction mixture was then incubated at room temperature (preferably at 20 °C) for 5-10 minutes. After incubation, the reaction was ready for transformation (**according to section 2.8.2.6**) into One Shot® TOP10 and/or DH5α™-T1^R competent cells and plated onto LB agar plates (with either Kanamycin or Ampicillin antibiotic). Post-transformation, colonies were pipetted into 5 mL LB agar (Ampicillin or Kanamycin) containing media for DNA Miniprep purposes. Post DNA Miniprep (as per **section 2.8.3.1**), appropriate volume of DNA sample was sent for sequencing.

2.8.2.8 SURVEYOR Assay

Surveyor® Mutation Detection Kit for Standard Gel Electrophoresis™ (Integrated DNA Technologies, IDT, Cat. No. 706020) was used for detecting mutations in DNA sequences with great sensitivity. This kit uses a mismatch-specific DNA endonuclease that scans mutations and polymorphisms in heteroduplex DNA. The Surveyor Nuclease is an endonuclease that cleaves DNA with high specificity at sites where insertions, deletions (indels) or any frame-shift mutations are present in the DNA sequence. The Surveyor DNA endonuclease cut both strands of a DNA heteroduplex on the 3'-side of the mismatch site. Indels and other base-substitution mismatches are recognized, however, the efficiency of cleavage varies with the sequence of mismatch (Qiu et al. 2004). Surveyor assay has been reported to be robust and reproducible tool

for the detection of mutations accurately in human, other mammalian, bacterial and plant genomes (Qiu et al. 2004, Tsuji and Niida 2008).

Surveyor Mutation detection kit allows easy detection of mutational DNA sequences after nuclease cleavage and its subsequent visual analysis by agarose gel electrophoresis or PAGE. DNA size range between 200-4000 bp can be easily analysed using agarose gel and/or PAGE. A detailed step-by-step protocol is explained below:

Step 1 Preparing PCR amplicons from mutant (test) and wild-type (reference) DNA.

The most critical requirements for the successful nuclease digestion in this step are: PCR yield is sufficiently high (i.e. > 25 ng/μl), has low background (i.e. specific single PCR bands of correct sizes, with no non-specific bands) and the amplicons are free from any PCR primer-dimer artefacts. Other important requirements are outlined in the protocol of the Surveyor Nuclease Assay kit and were strictly adhered to while performing the experiments.

The amount of purified PCR amplicons (using PCR purification kit **as per section 2.8.3.3)** used in the reaction mix was 400 ng in total, which is optimal substrate required for the range of Surveyor Nuclease S enzyme used (i.e. 0.5-2 μl).

Step 2 Formation of homo/hetero-DNA duplexes

The appropriate volumes for 200 ng and 400 ng DNA for both the test and reference DNA were worked out. In order to detect homogeneous mutation in a test sample, the PCR product must be hybridized with a wild-type reference PCR product to generate mismatches for Nuclease S cleavage. Both the test sample and wild-type reference DNA were amplified separately with same PCR primers and then mixed in 1:1 ratio (i.e. 200 ng: 200 ng, 400 ng in total) to maximize the heteroduplexes formation during hybrization protocol (tabulated below).

Similarly, the heterogeneous DNA (containing pools of fragments derived from genetically different homozygous or heterozygous sources) was also PCR amplified according to ***Step 1***. Such heterogeneous PCR amplicon samples were also hybridized without mixing them with a wild-type reference DNA. The proportion of mutant to reference DNA in the DNA populations should be more than 5-10 % for efficient gel electrophoresis analysis.

Equal amounts of test sample and reference PCR products in a 0.2 mL tube were mixed. A separate 0.2 mL tube was used for reference DNA as control. For efficient annealing the reaction volume was kept at least 10 µl. (Note: Each tube should contain total DNA ~ 400 ng.

Table below shows a summarised view of homo and hetero mixes of DNA or PCR amplicons for one reaction, including the respective Controls. It also shows the possible results if experiments performed adhering to all aspect of sample preparations according to manufacturer's specifications.

Experiment	**DNA amount**	**Purpose and Expected result**
Control 1: Wild-type reference DNA (Nuclease positive)	400 ng	Nuclease +ve control: Nuclease cleaves no part of the Homo-duplexes, which in theory have no mutations. Therefore, Single Band on agarose gel.
Control 2: Wild-type reference DNA (Nuclease negative)	400 ng	Nuclease -ve control: No Nuclease S → no digestion of the homo and hetero duplexes. Therefore, Single Band on agarose gel
Test Control 1: Test + Reference wild-type DNA (Nuclease negative)	200 ng + 200 ng = 400 ng	No Nuclease test control: No Nuclease S→ no digestion of the homo and hetero duplexes. Therefore, Single Band on agarose gel.
Test: Test + Reference wild-type DNA (Nuclease positive)	200 ng + 200 ng = 400 ng	Test: Nuclease positive test sample → Nuclease S digests the hetero duplexes of reference DNA and test DNA. Therefore, resulting in Two Bands on agarose gel

The thermocycler set-up for hybridization for the formation of homo/heteroduplexes is tabulated below:

Temperature (°C)	Time (minutes)	Temperature ramp (°C/s)
95	10	
95 to 85		-2.0
85	1	
85 to 75		-0.3
75	1	
75 to 65		-0.3
65	1	
65 to 55		-0.3
55	1	
55 to 45		-0.3
45	1	
45 to 35		-0.3
35	1	
35 to 25		-0.3
25	1	
4	∞	

Following the cycler hybridization programme, the heteroduplex samples (as mentioned in the table were treated with Nuclease S enzyme on ice in ***Step 3***.

Step 3 Treatment with Surveyor Nuclease S

In this step, the heteroduplex test sample DNA from ***Step 2*** was cleaved with appropriate amount Nuclease S enzyme and Enhancer S (amount depending on PCR-type used, as per manufacturer's specifications), along with necessary reference DNA and controls

The hetero/homoduplex DNA experimental samples and any reference DNA samples were kept in separate tubes. The internal G and C controls supplied with the kit were also set-up in separate tubes according to manufacturer's specifications.

All the reactions were carried out using Phusion High-Fidelity PCR Master Mix with HF buffer (ThermoFisher Scientific, F531S). Therefore, 2 µl of Nuclease S and 1 µl of Enhancer S was used each time, without the need for the STOP solution (supplied in the kit) after Nuclease S digestion. The nuclease digestion was performed at 42 °C for exactly 60 minutes.

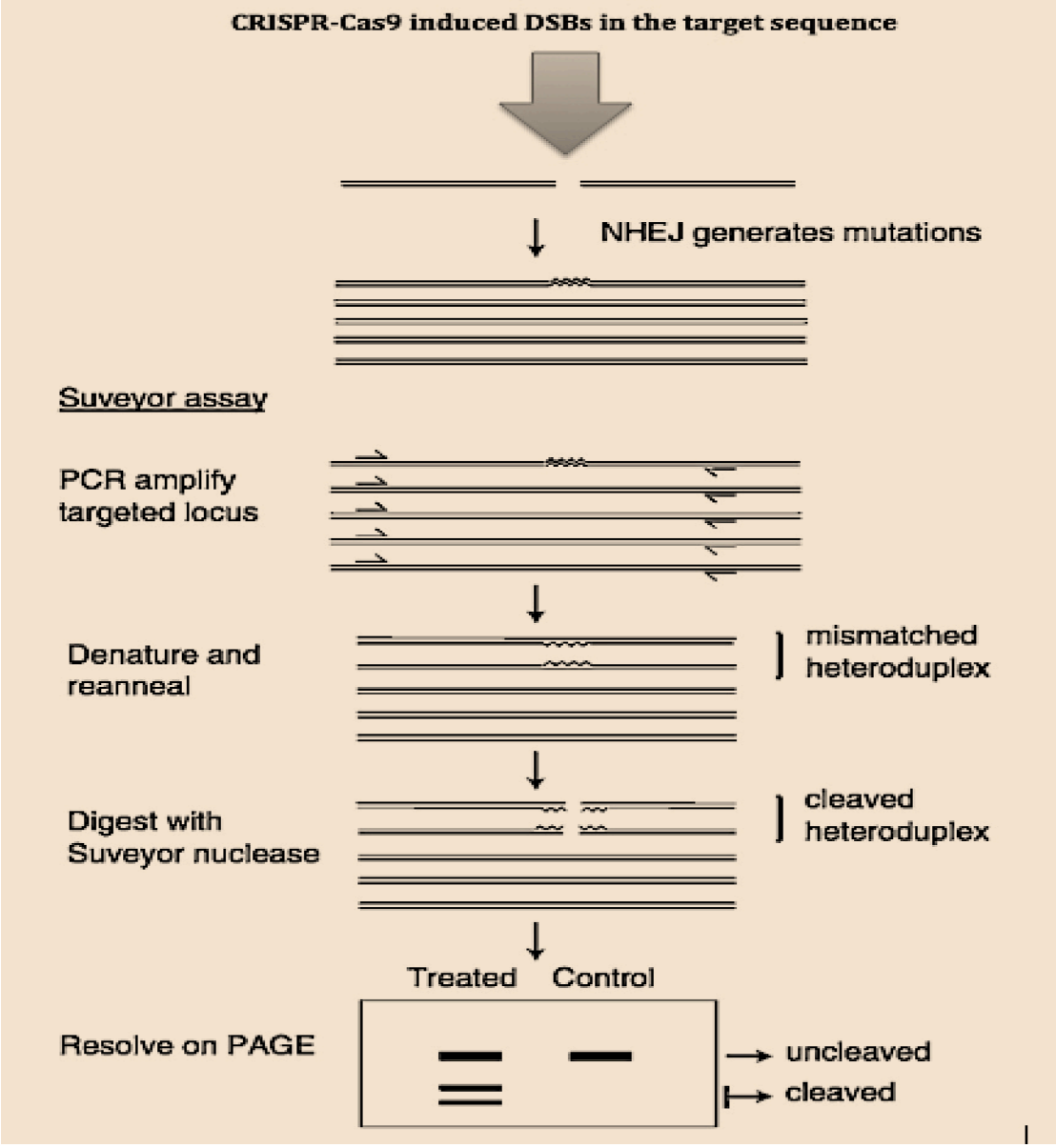


Figure 2.5 Overview of the Surveyor Experiment. CRISPR-Cas9 induces double stranded breaks in the target DNA, which are repaired by cells natural machinery via the error prone NHEJ mechanism. NHEJ is an imperfect DSBs repair method of the cell and often yield mutations in the sequence. These mutations are detected with the highly mutation-sensitive Surveyor Assay. The Surveyor Assay begins with PCR amplification of the targeted locus and formation of the homo-/hetero-duplexes of the equimolar amounts of wild-type (reference DNA) and the test PCR amplicons. The cross-annealing of the mutated (test) amplicons with the reference or wild-type amplicons, converts mutations into mismatched duplexes. Following this, the re-annealed

amplicons are digested with Surveyor Nuclease S enzyme, which is mis-match specific and cleaves the distorted duplexes at the site of mismatch, generating digested products. Finally, these digested products can then be visualized on agarose or poly-acrylamide gels at appropriate resolutions. Legends: *Horizontal half arrows* → PCR primers; *Straight double-lines* → wild-type PCR amplicons and *Jagged lines* → Mutated amplicons.

Step 4 Analysis of DNA fragments

DNA fragments in 1000-4000 bp range were resolved on 1.2 % Agarose Gel using 1X TAE buffer. The heteroduplex test sample digested with Nuclease S will yield two products of different/same sizes.

2.8.3 DNA purification

2.8.3.1 Plasmid DNA purification MiniPrep

Post-transformation reaction, 20 μg of plasmid DNA was purified using the QIAprep Spin Miniprep kit (QIAGEN Cat. No. 27104). After overnight (O/N) incubation of the LB agar plate, a positive transformant colony was picked to inoculate a 5 mL LB medium culture supplemented with appropriate antibiotic concentration. This culture was then inoculated at 37 °C O/N at 225 rpm in a shaker incubator. The following day, the bacterial culture was centrifuged at 4000 rpm for 10 minutes at 4 °C. The supernatant discarded-off and the pellet was resuspended using 250 μl of resuspension P1 buffer (supplied in the kit). The resuspended cell pellet mixture was then transferred to a 1.7 mL eppendorf into which equal amount of cell-lysis buffer P2 was aliquoted. Cells were lysed using the lysis buffer and then mixed vigorously by inverting the eppendorf up down several times, immediately after the addition of lysis buffer (Note: The lysis was not allowed to go more than 5 minutes). Following the lysis of cells, 350 μl of neutralization buffer was added to enhance precipitation of DNA and the tube was inverted several times immediately after the addition of neutralization buffer and spun at 13000 rpm for 10 minutes. The supernatant was then applied to QIAprep spin column and spun at 13000 rpm to bind the DNA to the column for 3 minutes. Following this, the column was washed with 500 μl of PB buffer to remove any traces of nuclease activity. This was followed by addition of 750

μl PE buffer wash, then spinning at 13000 rpm for 1 minute. The column after first spin was spun again at 13000 rpm in order to remove any traces of PE buffer.
DNA was then eluted from the column by applying 50 μl of luke-warm EB buffer (pH 8) in the center of column filter. The column was incubated at room temperature after the EB buffer addition for 2-3 minutes, and then centrifuged for 1 minute at 13000 rpm to elute DNA into a fresh eppendorf.

2.8.3.2 Plasmid DNA purification: Maxiprep

Post-transformation reaction, 500 μg of plasmid DNA was purified using the QIAprep Spin Miniprep kit (QIAGEN Cat. No. 27104). After O/N incubation of the LB agar plate, a positive transformant colony was picked to inoculate a 25 mL LB medium culture supplemented with appropriate antibiotic concentration. This culture was then inoculated at 37 °C O/N at 225 rpm in a shaker incubator. The following day, the bacterial culture was centrifuged at 6800xg for 15 minutes at 4 °C. The supernatant was discarded-off and the pellet was resuspended using 10 mL of pre-chilled resuspension P1 buffer (supplied in the kit). After cell-pellet resuspension, 10 mL of lysis buffer P2 was added to lyse the cells and the tube was mixed vigorously by inverting up and down and incubated at room temperature for not more than 5 minutes. Subsequently, 10 mL of neutralization buffer N3 was added to neutralize the lysis buffer and the tube was mixed vigorously, immediately. The cell lysate was transferred to a QIAfilter cartridge and incubated at room temperature for 10 minutes. The lysate was then filtered through cartridge in a separate centrifuging tube by inserting a plunger to the cartridge. During the incubation time when the lysate was applied to the QIAfilter cartridge, a QIAgen-tip 500 was equilibrated with 10 mL of QBT equilibration buffer in order to reduce the surface tension to allow easy entry of fluid through the filter. The filtrate (containing the DNA) was applied to the equilibrated QIAgen-tip 500, which was then washed with QC buffer twice to remove any possible contaminants. Finally the DNA was eluted using 15 mL of QN buffer. The eluted DNA was then precipitated with 10.5 mL of room temperature isopropanol in pyrogen-free centrifuge glass tube. The glass-tube was spun at 15000 rpm for 30 minutes at 4 °C. After centrifuging, the supernatant was decanted-off and the DNA pellet was washed with 5 mL of 70 % ethanol to remove precipitated salts. This 5 mL

volume containing the washed DNA pellet was then again centrifuged at 15000 rpm for 10 minutes at 4 °C. Finally, the DNA pellet was air-dried for 5-10 minutes and the purified DNA redissolved in lukewarm EB buffer.

2.8.3.3 PCR Purification

Amplified PCR products were purified using QIAGEN® QIAquick PCR purification kit (QIAGEN®, 28104) in accordance with manufacturer's specifications. Other purifications of single/double-stranded DNA from PCR and enzymatic reactions such as Kinase/CIP/*Klenow* were also carried out using this kit.

Firstly, 5 volumes of PB buffer (with pH indicator) were added to the 1 volume of PCR reaction mix. The mixture will turn yellow, however, if the mixture appears orange or violet then 10 µl of 3M sodium acetate was added and mixed for 15 s using vortex. This reaction mix was then applied to the QIAquick spin column and incubated at room temperature for 2-3 minutes. After incubation, the spin column was spun at 13000 rpm for 1 minute to bind the DNA to the column. A wash step with 750 µl of buffer PE was performed after the DNA binding step, and the column was spun at 13000 rpm for 1 minute. The flow-through after this spin was discarded and QIAquick column was placed into the collection tube and spun again at 13000 rpm for 1 minute in order to get rid-off any residual wash buffer PE. Following this, 30-40 µl of lukewarm nuclease-free water or EB buffer was pipetted into the center of the column filter and incubated at room temperature for 2-3 minutes. After incubation, the column was spun down with a fresh eppendorf as the collection tube at 13000 rpm for 1 minute. The purified DNA can be stored at – 20 °C until further used.

2.8.3.4 NanoDrop® Spectrophotometer

The purified DNA/RNA samples were quantified using the NanoDrop® ND-1000 Spectrophotometer (NanoDrop Technologies). The pedestal of the NanoDrop was wiped using a lint-free tissue with wet UHP, before analysing any DNA/RNA sample. 2 µl of UHP was loaded to 'Initialize' the system and wiped after initializing the system, then another 2 µl UHP was then used to 'Blank' the instrument. The programmes stored to measure the DNA/RNA samples were DNA-50 and RNA-50,

respectively at 260 nm. The concentration of RNA/DNA was calculated by software using the following formula:

$$OD_{260nm} \text{ x Dilution factor x } 40 = \text{ng/µl RNA}$$

Abs_{260}/Abs_{280} and Abs_{260}/Abs_{230} ratio for each sample was recorded. Ratios ~ 2 of Abs_{260}/Abs_{280} and ~ 1.8-2.2 of Abs_{260}/Abs_{230} are indicative of pure DNA/RNA samples. Samples with absorbance values outside these ratio ranges were indicative of potential contaminants due to protein or phenol or ethanol and of bad quality DNA/RNA.

2.9 Proteomic-based techniques and assays

2.9.1 Bradford assay

A Bovine serum albumin (BSA) standard (Bio-Rad, Cat. No. 500-0206) starting with 1 mg/mL from the stock concentration of 2 mg/mL was serially diluted using UHP at 0.5, 0.25, 0.125 and 0 mg/mL. The Standards and samples were measured in triplicates and ~ 5 µl of each sample were pipetted into the corner of the base of a well in a 96-well plate. 250 µl Quick start™ Bradford 1X Dye Reagent (Bio-Rad, Cat. No. 500-0205) was added to each well and incubated in the dark for 10-15 minutes. The absorbance readings were measured on a spectrophotometer at a wavelength of 595 nm.

2.9.2 Western Blot

Sodium dodecyl sulphate - Polyacrylamide gel electrophoresis (SDS-PAGE) was used to resolve proteins analysed from Western Blotting. The cell pellet was lysed on ice for 20 minutes using the lysis buffer containing urea and vortexed. After 20 minutes, the lysate was centrifuged at high speed for 15 minutes at 4 °C to remove any insoluble debris. The 2X Laemlli buffer (Loading dye from Sigma-Aldrich, S#3401) and appropriate volumes of lysis buffer were used to dilute protein samples. The protein samples were quantified according to aforementioned, Bradford protocol.
The samples were boiled for 5 minutes and cooled on ice, before loading onto the gel. 4-12 % of NuPAGE Bis Tris precast gradient gel (Life Technologies, NP0322BOX)

was used for loading 5-20 μg of protein and the molecular weight marker (NEB). Electrophoretic transfer, blocking and development of western blots was carried out. The membranes were probed with the appropriate antibody of choice diluted in Tris-buffer saline containing 0.1 % - 20 % Tween (TBS-T). An anti-mouse GAPDH mAb (Abcam, #ab8245) was used as an internal control.

2.9.3 SEAP assay

Lipscomb et al., (2005) reported an enzymatic assay method for the quantification of SEAP (secreted alkaline phosphatase). According to the assay, 50 μl of supernatant was transferred to a flat-bottom well of a 96-well plate. On top of this 50 μl of 2X SEAP reaction buffer (10.5 diethanolamine (100 %), 50 μl of 1M $MgCl_2$ and 226 mg of L-homoarginine in a total of 50 mL volume) was pipetted. Following the 96-well plate was incubated at 37 °C in order to increase enzymatic activity. A phosphate substrate (158 mg of p-nitro-phenol-phosphate) in 5 mL of 1X SEAP reaction buffer was made fresh for each use (adjusted to volume to suit the number of samples to be tested). 10 μl of this substrate was added to each well (Note: Avoid direct contact with light). A kinetic absorbance assay was performed in the Micro-titer plate reader in the KC4 instrument using the Gen5^{TM} programme (Biotek® Instruments, Inc.). This measures absorbance readings and the changes in the readings per minute, which are indicative of the amount of SEAP present in the sample as the reaction progresses (i.e. the substrate depletion by SEAP per minute). Mean V values were taken as the change in the average of the kinetic slope.

2.9.4 Enzyme-linked immunosorbent assay (ELISA)

Bethyl Laboratories Human IgG ELISA Quantitation set (Bethyl Laboratories Inc., E80-104) was used to quantify CHO-secreted IgG antibody as per manufacturer's specifications. CHO cells were centrifuged at 1000 rpm for 5 minutes and the supernatant was harvested in a microcentrifuge tube. In order to remove any trace amount of residual cells, the supernatant was further spun again at 13000 rpm for 5 minutes before the assay.

The coating antibody (goat anti-human IgG-affinity purified) was first diluted (1 μl in 100 μl Coating Buffer) and used at 100 μl/well to coat the immunoassay plate (Sigma) wells. Following this, in order to avoid any non-specific binding, the plate was blocked with 200 μl/well of 1 % BSA blocking solution for at least one hour at room temperature. The test sample supernatant was then applied to each well of the plate in order to allow interaction with the coating antibody. CHO-IgG antibodies were detected by addition of 100 μl/well of the secondary antibody (goat anti-human IgG-linked to horse radish peroxidase (HRP)). 3-5 washes were performed using a wash solution between each step of the assay. 100 μl of 3,3',5,5'-tetramethylbenzidine (TMB) solution was used to prepare the enzyme substrate reaction, which develops a deep blue product after 15 minutes upon interaction with HRP conjugates in ELISA. The reaction is stopped by a H_3PO_4 solution at 2M (100 μl/well), which gives a yellow colour as a signal. A standard curve with purified recombinant human IgG1 antibody (rHIgG1) was used to quantify the amount of rHIgG1 present in the samples. The absorbance readings of the reader plate were measured at 450 nm.

Section 3.1 & 3.2

Stable Depletion of endogenous Let-7 microRNA levels in CHO-K1 SEAP cells using “sponge” decoy vectors

3.1.1 miRNA "sponge" decoy vectors

For loss-of-function studies, miRNA sponge decoy vectors were employed for stable long-term gene knockdown *in vivo*. These decoy vectors were constructed by cloning miRNA anti-sense sequences into an appropriate backbone vector. The key advantage of using miRNA sponges is that they potentially inhibit a whole miRNA family (sharing the common seed region sequence) by using the complementary heptameric seed sequence. miRNA sponge technology was first explored by Ebert and colleagues (Ebert, Neilson and Sharp 2007).

3.1.2 Let-7 sponge design and description

As mentioned in **Section 1.6.9.1**, Let-7a manipulation resulted in enhanced cellular viability and specific productivity in two different CHO cell lines. Moreover, previously our lab reported 4-fold improvement in CHO cell normalised productivity following Let-7e manipulation (unpublished data). Previously our laboratory also reported Let-7 family members being differentially expressed in CHO-K1 cells cultivated at two different temperatures, suggesting their importance in regulating growth (Gammell et al. 2007). These positive observations stimulated our interest to observe the impact of stable depletion of endogenous Let-7 in CHO-K1 SEAP cells using a Let-7 miRNA sponge decoy vector (henceforth referred to as Let-7 sponge). In particular, we were interested in improving CHO cell growth rate and as a result, the overall protein yield.

Briefly, the endogenous mature Let-7 miRNA recognizes its mRNA targets by sequence complementarity between the seed region of miRNA (2-8 nt of 5'-end) and binding sites in the 3'-UTR of the mRNA targets and inducing post-transcriptional repression. Let-7 sponge was designed to cause efficient endogenous Let-7 inhibition (by 'soaking-up' the endogenous Let-7 levels in the cells like a 'sponge') by competing with mRNA targets for Let-7 binding and diverting the endogenous Let-7 function of post-transcriptional gene regulation. As outlined in **Figure 3.1.1**, four Let-7 binding sites were inserted in tandem in the 3'-UTR of a destabilized enhanced green fluorescent protein (deGFP) i.e. at *XhoI* and *EcoRI* sites (**Table 3.1 for Sequence details**) (**Fig. 3.1.2**). The deGFP reporter has a short half-life of 2 hours and any change in the deGFP fluorescence induced by the binding of the miRNA

(Let-7 in this case) to the sponge is correlated to the change in miRNA expression. In simple terms, post binding of endogenous Let-7 to the sponge, the miRNAs prevent deGFP from being expressed (or translated), hence, the deGFP acts as a reporter of miRNA depletion. The Let-7 sponge sequences incorporated in the sponge vector were designed to be almost fully complementary to Let-7. However, a few base mismatches (wobble/bulge) at nucleotides 9-11 were introduced in order to prevent RNAi type cleavage. Moreover, the miRNA binding sites were kept to four since including more than ten miRNA binding sites have been reported to promote sponge degradation (Ebert, Neilson and Sharp 2007). The modification to introduce a "wobble" increases the functional half-life of the sponge construct (Ebert, Neilson and Sharp 2007, Ebert and Sharp 2010, Kluiver et al. 2012). A negative control sponge vector (henceforth referred to as NC sponge) was also designed in a similar manner, however, instead of Let-7 binding sites, the NC sponge contained non-specific sequence that should not bind any known miRNAs.

Table 3.1: miRNA sponge sequence details for Let-7 sponge design

5' to 3' Let-7 miRNA sponge sequence					
Sponge	*XbaI* site	*XhoI* site	Let-7 sponge sequence	*EcoRI* site	*XbaI* site
Let-7 sponge	TCTGA	CTCGAG	acgcgAACTATACAATGATCTACCTCAacgcgAACTATACAATGATCTACCTCAacgagAACTATACAATGATCTACCTCAtgacgtAACTATACAATGATCTACCTCAtcatc	GAATTC	TCTAGA
NC-sponge	TCTGA	TCTGA	tcgagCCGGAAGTTTTCAGAAAGCTAACAccggAAGTTTTCAGAAAGCTAACAccggAAGTTTTCAGAAAGCTAACAccggAAGTTTTCAGAAAGCTAA CAccgg	GAATTC	TCTAGA

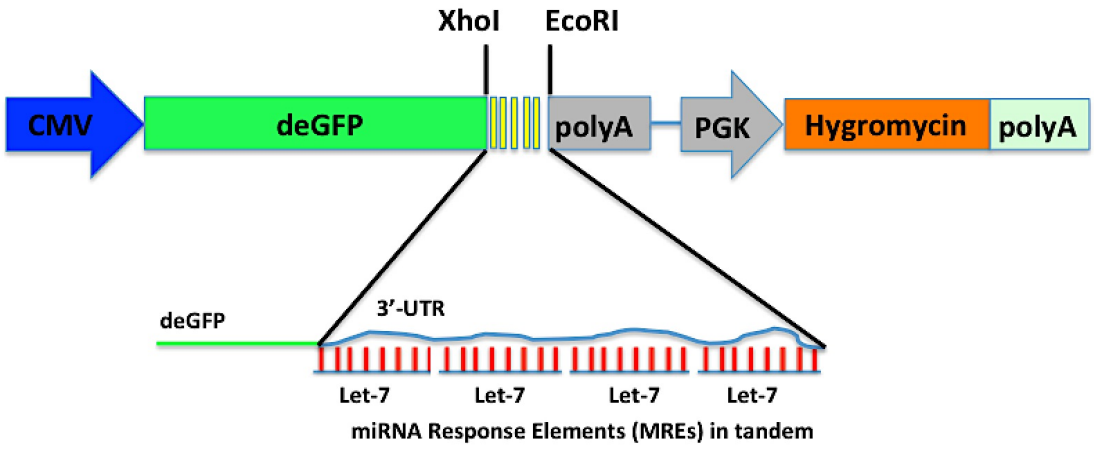


Figure 3.1.1: Let-7 sponge cassette. CMV → Cytomegalovirus promoter; deGFP → destabilized enhanced green fluorescent protein; polyA → polyadenylation signal for transcription termination; PGK → human phosphoglycerate kinase promoter; Hygromycin → Selective Marker gene. *XhoI/EcoRI* → restriction enzyme sites used for cloning. Yellow vertical rectangular boxes represent miRNA response elements in tandem.

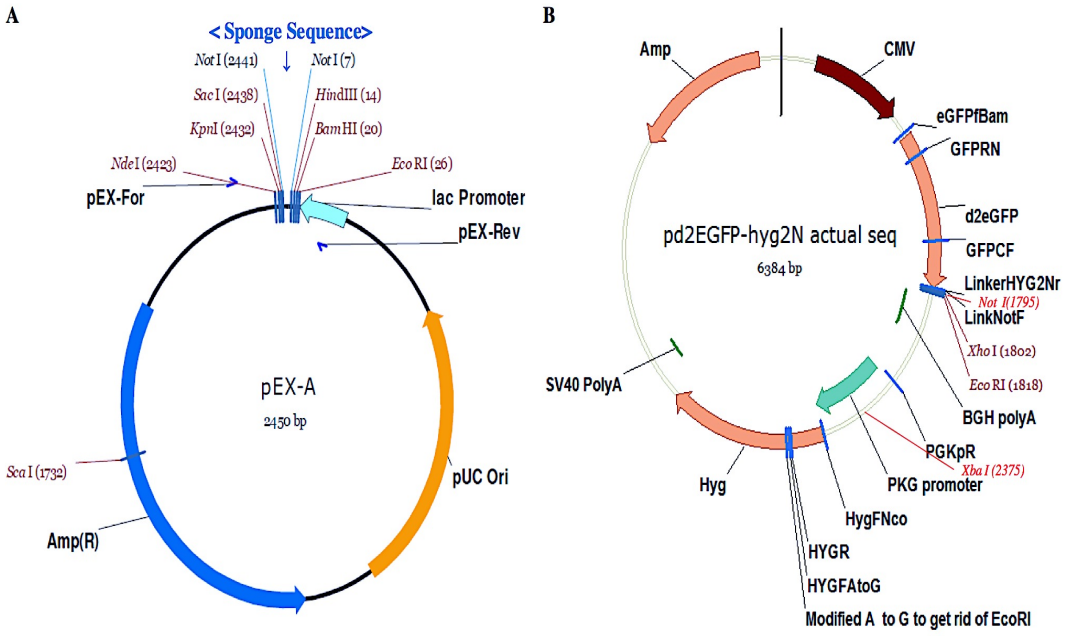


Figure 3.1.2: Plasmid constructs' maps for pEX-A (with miRNA sponge sequences) and miRNA sponge backbone vector. (A) The miRNA sponge concatamers were cloned into pEX-A bacterial expression vector with Ampicillin as Selective Marker gene (sourced from MWG Eurofins) and various other transgene elements, as well as restriction enzyme's recognition sites marked. (B) The final cloned vector with miRNA backbone containing multiple cloning site (MCS), marked with restriction enzymes sites, deGFP reporter sequence and Hygromycin B as Selective Marker gene and other transgene elements included along with restriction enzyme's recognition sites marked.

3.1.3 Stable transfection of NC and Let-7 sponge in CHO-K1 SEAP cells

CHO-K1 SEAP cells (henceforth referred to as SEAP cells) were transfected with NC and Let-7 sponge vectors separately in two wells of a 6-well plate (refer to **Section 2.6.1.2**). 24 hours post-transfection, cells were selected using hygB antibiotic for the next 2-3 passages. At this point we had mixed pools for both NC and Let-7 sponge CHO-K1 SEAP cells.

The NC and Let-7 sponge mixed pools exhibited varying degrees of GFP intensities as assessed by Guava. In addition to copy number, this variation in GFP expression was due to the random integration of the sponge cassette constructs within the genome and not a reflection on the transfection efficiency. The control vector transfected into SEAP cells during the same experiment demonstrated 65 % transfection efficiency (data not shown). Let-7 sponge mixed cells exhibited ~ 4-fold lower GFP expression due to the interaction of endogenous Let-7 with the Let-7 sponge repressing deGFP translation, whilst in the control sponge mixed cells, the NC sponge did not bind to any miRNA and showed higher mean GFP expression (**Fig.3.1.3**). This observation was further validated in **Section 3.3.1** in the transient transfection study.

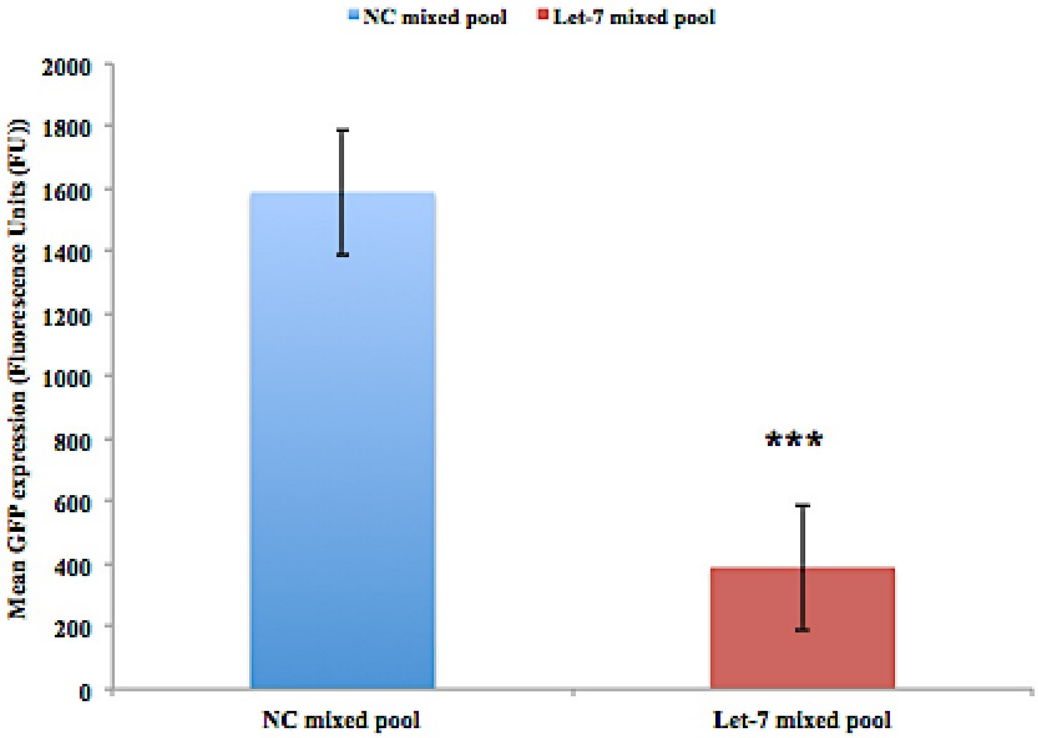


Figure 3.1.3: Stable transfection of SEAP cells with NC and Let-7 sponge vector. Mean GFP expression analysis of NC and Let-7 sponge mixed SEAP pools. Experiments were performed in biological triplicate and GFP expression was analysed using Guava. A standard student t-test was performed to calculate the statistical significance of the GFP data in both the mixed pools of SEAP cells (p-value < 0.0001, n = 3).

3.1.4 Impact of Let-7 depeltion on bioprocess relevant phenotypes in CHO-K1 SEAP mixed sponge pools in 5 mL culture spin tubes

3.1.4.1 Impact of Let-7 sponge on growth characteristics

Following the transfection and selection of mixed SEAP cells with hygromycin, the mixed pools of NC and Let-7 sponge cells were assessed for growth in a 5 mL batch suspension culture. Cell growth and viability were monitored over a time-period of six days. It was observed that the Let-7 sponge expressing mixed SEAP pool showed significantly increased viable cell density (VCD) (p-value = 0.01) of 5.26×10^6 cells/mL on Day 4 i.e. 1.23-fold higher in comparison to NC sponge control mixed SEAP pool of 4.27×10^6 cells/mL. The VCD of Let-7 sponge mixed pool seemed to peak at 5.25×10^6 cells/mL on Day 3 and Day 4 (**Fig. 3.1.4A**), however, in the NC sponge mixed SEAP pool the VCD dropped from 5.18×10^6 cells/mL on Day 3 to 3.66×10^6 cells/mL on Day 5 (**Fig. 3.1.4A**). On Day 6, Let-7 sponge mixed pool also exhibited higher VCD than the NC sponge mixed SEAP pool. In addition to the VCD data, we also observed a significant increase in the % cell viability (p-value = 0.01) between the Let-7 sponge mixed pool i.e. ~ 56 % and 38 % on Day 5 and Day 6, respectively, as opposed to NC sponge mix pool of cells, which displayed 46 % and 21 % cell viability on Day 5 and Day 6, respectively (**Fig. 3.1.4B**). Based on these observations, we inferred that the Let-7 sponge expressing mixed SEAP cells seemed to have advantageous features of increased VCD and enhanced % cell viability in comparison to NC sponge mixed SEAP expressing cells. In a follow-up experiment, we also performed SEAP productivity analysis in the two mixed SEAP cell pools.

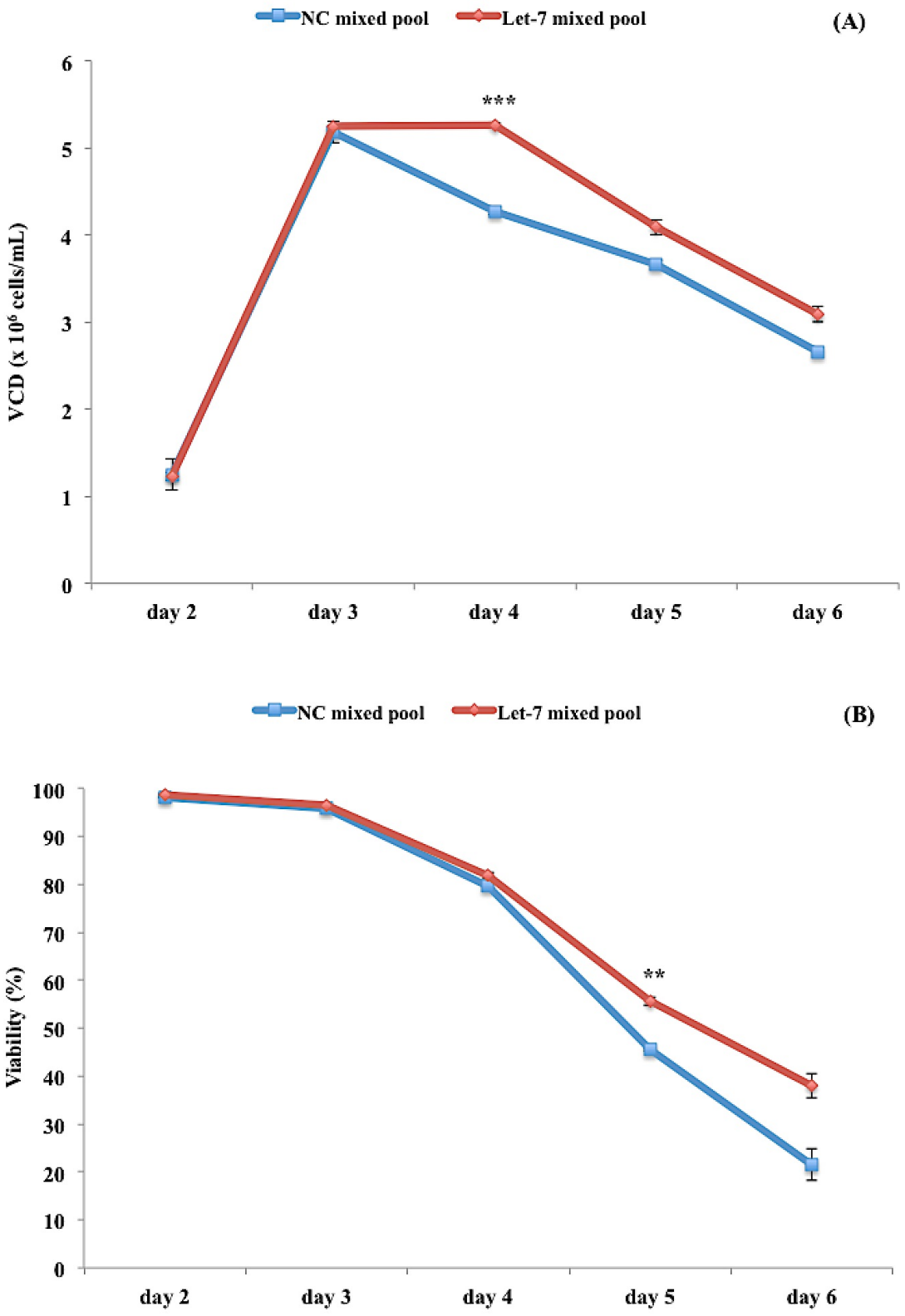


Figure 3.1.4: Viable Cell density (A) and % cell viability (B) comparison between NC and Let-7 sponge mixed pools. Cells were seeded at 2 × 10^5 cells/mL in a 5 mL working volume in a 50 mL spin tube in biological triplicate in serum-free medium. Samples were harvested from Day 2 onwards. Cell counts were performed using Guava Viacount reagent and Guava. A standard student t-test was performed to analyse statistically significant data (p-value ≤ 0.05, n = 3).

3.1.4.2 Impact of Let-7 sponge on productivity characteristics

The volumetric productivity or titer each day of the culture period in the Let-7 sponge cells was significantly higher than the NC sponge mixed SEAP cells (p-value = 0.01). The volumetric productivity of Let-7 sponge cells was ~ 1.5-fold higher than the NC sponge cells on Day 4, Day 5 and Day 6 of the culture period (**Fig. 3.1.5**). We inferred that this higher volumetric SEAP productivity in Let-7 sponge cells was due to enhanced specific SEAP productivity (data not shown) on Day 3 and Day 4, which reflected upon the improvement in volumetric SEAP productivity (**Fig. 3.1.5**). These findings of improved VCD, % cell viability and improved volumetric stimulated our interest in replicating the observed improved phenotypes in stable Let-7 sponge clones.

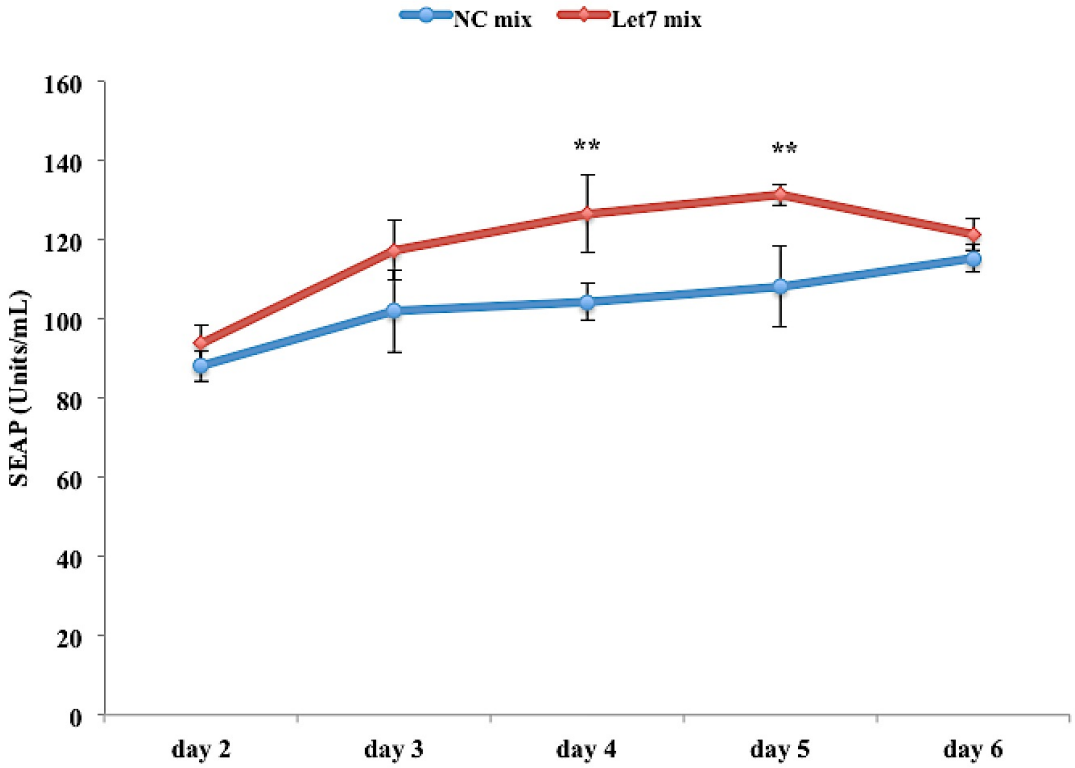
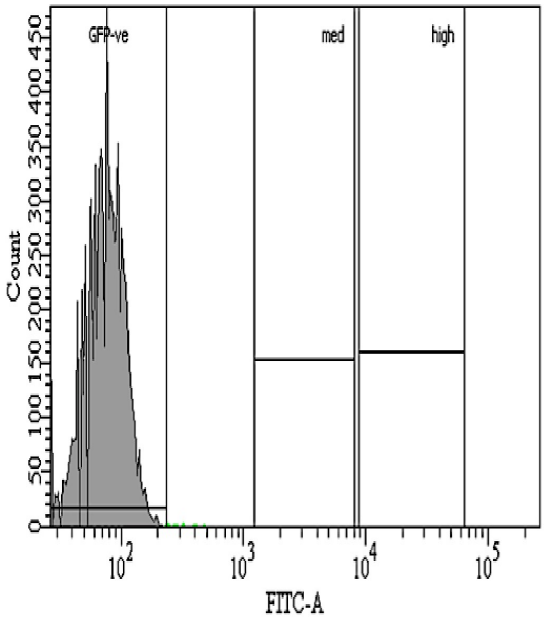


Figure 3.1.5: Volumetric SEAP productivity comparison between NC and Let-7 sponge mixed pools. Cells were seeded at 2 × 10^5 cells/mL in a 5 mL working volume in a 50 mL spin tube in biological triplicate in serum-free medium. Samples were harvested from Day 2 onwards. A standard student t-test was performed to analyse statistically significant data (p-value < 0.05, n = 3).

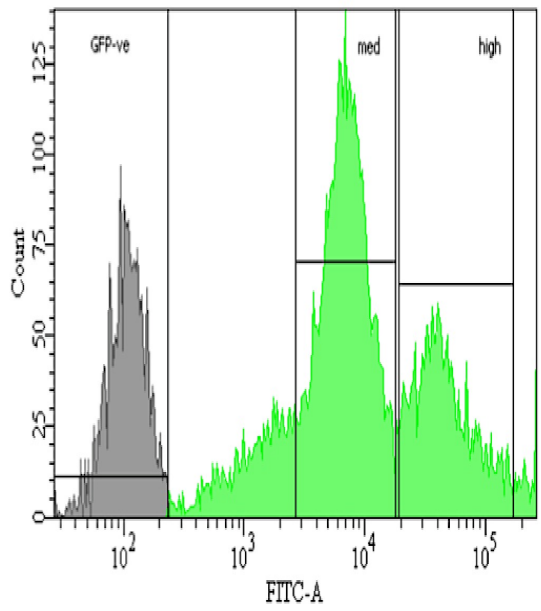
3.1.5 Generation of clones from stable mixed pools of CHO-K1 SEAP cells

3.1.5.1 FACS sorting NC and Let-7 sponge mixed pools

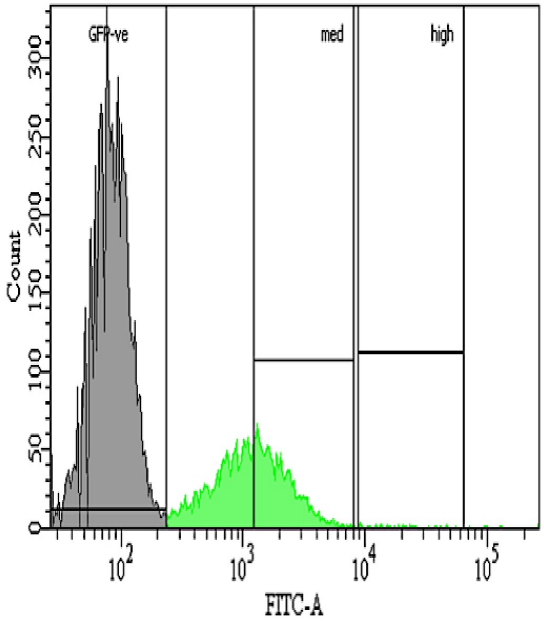
Our preliminary experiments suggested a possible positive impact of Let-7 sponge in SEAP expressing cells, however, these experiments only represented an average across a heterologous population. Therefore, the phenotypic effect of Let-7 sponge expression in stable single cell clones from each NC and Let-7 sponge pool was investigated. Firstly, a panel of stable clones from NC and Let-7 mixed populations were sorted using FACS. The approach was based on sorting medium to high GFP expressing clones from both the NC and Let-7 mixed populations in order to ensure that they all are expressing excess sponge transcript above the level of saturation by endogenous Let-7 (**3.1.6b and 3.1.6c**). A GFP negative parental CHO-K1 SEAP cell line was used to set a strict cut-off gate for GFP negative versus GFP positive cells (**3.1.6a**). Following FACS clone isolation, 24 single cell clones (SCC) (12 high GFP & 12 medium GFP) from each NC and Let-7 stable clonal pool were generated. These were expanded and routinely passaged in serum-containing media containing G418 and HygB. However, the adaptation of these stable clones to suspension serum-free media resulted in the loss of few candidates i.e. reduced the number of candidates to 20 stable clones from each NC and Let-7 group.



(a) CHO-K1 SEAP parental cell-line



(b) NC sponge mixed pool



(c) Let-7 sponge mixed pool

Figure 3.1.6: FACS sort of NC and Let-7 sponge mixed CHO-K1 SEAP pools. (a) A GFP negative cell line (i.e. CHO-K1 SEAP, Grey) was used for gating the FACs instrument. (b) NC and (c) Let-7 sponge mixed pools of cells were sorted based on medium-high range of GFP intensity. The GFP expression was analysed using the FITC channel (Emission 525 nm; Exicitation at 488 nm) using the flow-cytometer (FACS, BD Biosciences).

3.1.5.2 GFP expression in stable clonal panels in 24-well plate format

As mentioned above, the original 24 stable clones for each NC and Let-7 panel was now reduced to 20 stable clones for each. Once all the clones were fully adapted to suspension culture and serum-free medium, they were routinely passaged with selective and non-selective media separately to ensure both the SEAP and sponge transgene expression was maintained. Following this, GFP expression across the 40 stable clones was analysed using the Guava. A range of GFP expression from high to low was observed in both the NC and Let-7 clonal panels (**Fig 3.1.7A and 3.1.7B**). This range could be due to random sponge construct integration within the genome as indicated by our data, in addition to copy number influence. Furthermore, it was observed that mean GFP expression in Let-7 sponge clones (**Fig. 3.1.7B**) was lower than in the NC sponge individual clones (**Fig. 3.1.7A**). The average GFP expression for the Let-7 sponge clonal panel was 88.2 FU in comparison to 281.7 FU for the NC sponge panel. This reduced GFP expression in Let-7 sponge clonal panel was potentially due to endogenous Let-7 binding to the four binding sites in the Let-7 sponge vector leading to deGFP translational repression. The NC sponge clones exhibited higher GFP expression due to the absence of miRNA binding sites, hence, more deGFP expression resulting in greater magnitude of GFP expression and intensity.

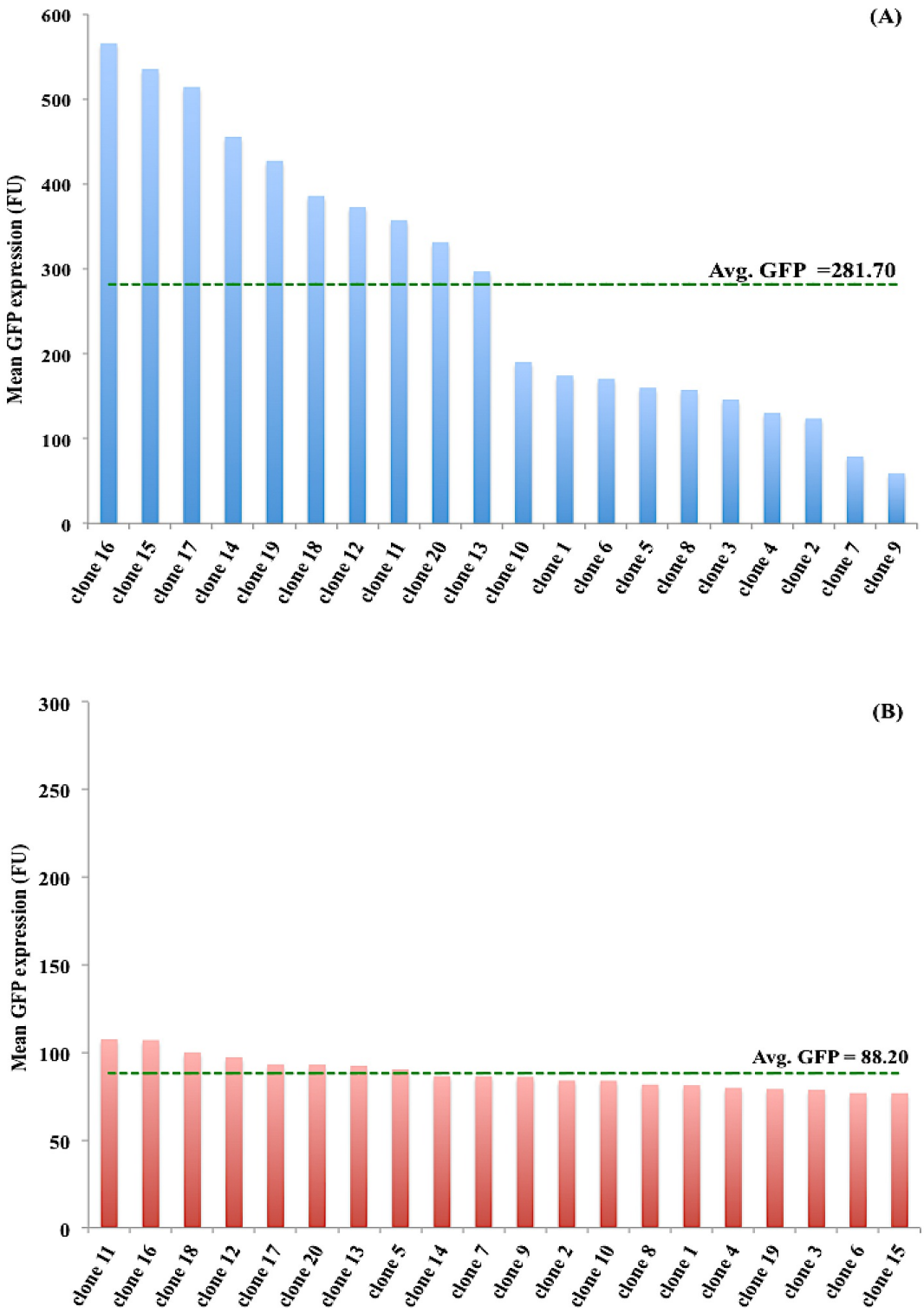


Figure 3.1.7: Mean GFP expression of 20 stable clones isolated from both the (A) NC and (B) Let-7 sponge mixed pools using FACS. Cells were seeded at 2 × 10^5 cells/mL in 24-well plate suspension format in 1 mL/well working volume. Cells were harvested on Day 3 for GFP expression analysis using Guava. A gate for GFP negative cells (using CHO-K1 parental cell line) versus the GFP positive cells was performed prior to mean GFP expression analysis. The dashed line in **Green** shows the average mean GFP expression across each clonal panel.

3.1.6 Impact of Let-7 depletion on bioprocess relevant phenotypes in a panel of CHO-K1 SEAP stable clones

3.1.6.1 Growth characteristics of stable clonal panels in 24-well plate format

The main objective of this experiment was to discover whether we could identify exceptionally good performing stable clones from the Let-7 sponge clonal panel with enhanced cell culture longevity and viability compared to the control panel. Therefore, the test and control clones were subjected to 24-well plate suspension batch culture for their phenotypic characterisation. The NC and Let-7 stable clones were cultured in 24-well plate format with serum-free media. The cell samples were harvested on alternate days for a 7-day period batch culture. It was observed that in contrast to the positive impact of the Let-7 sponge in mixed SEAP cells, the same phenotype was not reflected in the behaviour of the individual clones. The growth profile for almost all of the Let-7 sponge clones (**Fig. 3.1.8B**) was similar to the NC throughout the batch culture period (**Fig. 3.1.8A**). The viable cell densities of all 20 stable Let-7 sponge clones were not significantly different than their control counterparts (p-value > 0.05). Moreover, not even a single clone performed better (or worse either) than any of the NC clone as the average cell densities across both panels were very similar on Day 3 (~ 1.7×10^6 cells/mL) and Day 5 (~ 1.27×10^6 cells/mL). However, similar to the results with mixed pools an increase in cell viability was observed in Let-7 sponge clone 2, 6 and 7. All had at least 50 % cell viability on Day 6, while the highest in the control panel was less than 30 % on Day 6. Furthermore, in the Let-7 sponge panel, clone 6 and 7 both showed ~ 80 % cell viability (**Fig. 3.1.8C**). Overall statistically significant difference in the cell viability was observed between Let-7 and NC clones (p-value ≤ 0.05). The reason for the enhanced % cell viability in the Let-7 sponge clone 6 and clone 7 could be the fact that clone no. 6 maintained VCD of 1.2×10^6 cells/mL on Day 3, which only rose to 1.6×10^6 cells/mL on Day 5 resulting in improved cellular viability on Day 6.

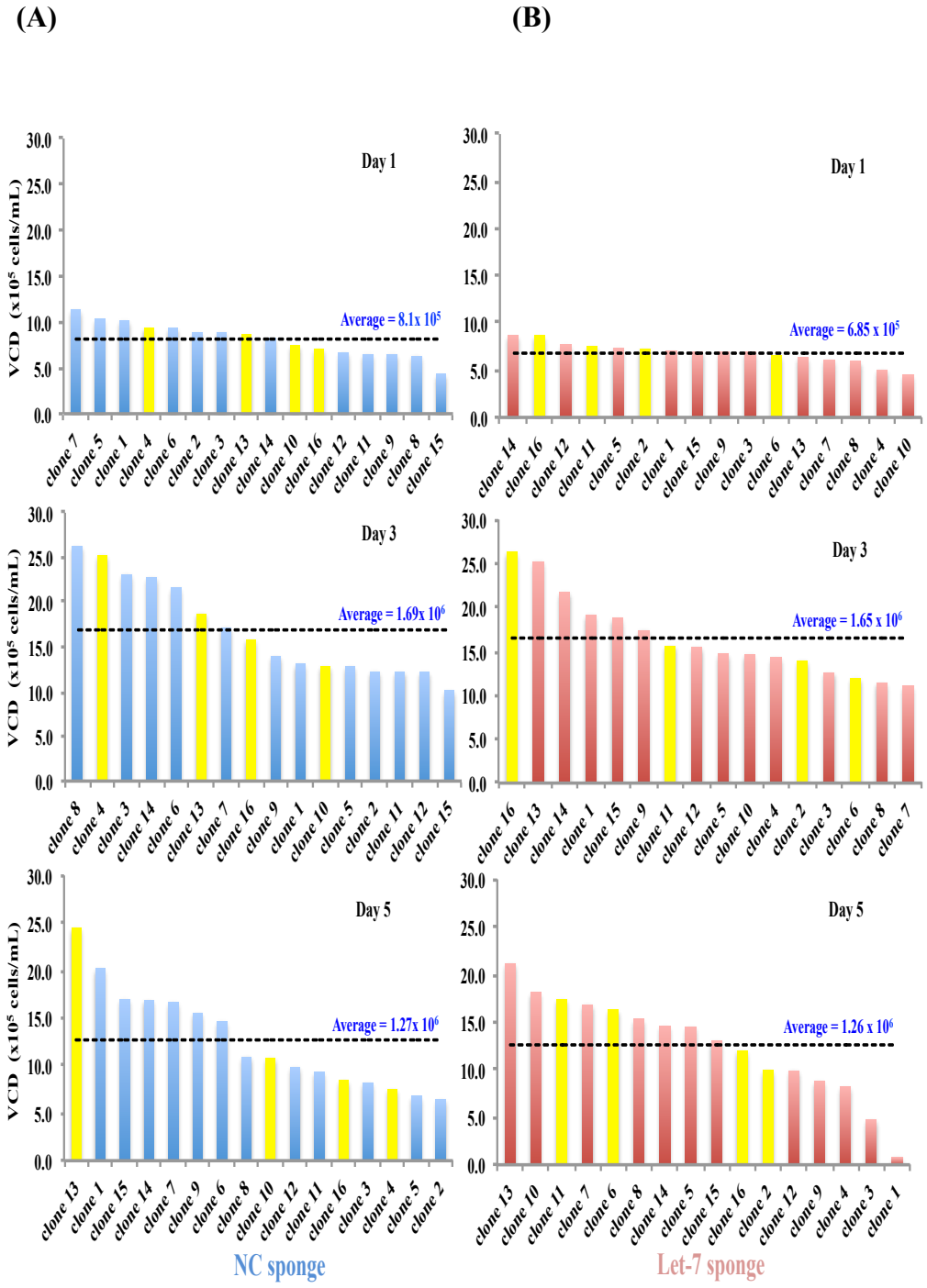


Figure 3.1.8: Batch Growth analysis of NC (A) and Let-7 sponge (B) clonal panels in 24-well plate (sorted from highest to lowest VCD). Each well of two 24-well plates was seeded at a cell density of 2 × 10^5 cells/mL in 1 mL suspension culture volume (biological duplicate, n = 2). VCD on Day 6 was measured, however, not presented as the culture was in decline. Cell counts were performed using Guava Viacount reagent and Guava. Yellow bars highlighted in each clonal panel set represents selected clones chosen for further Scale-up growth and productivity analysis later. The dashed **Black** line represents average cell density across the clonal panels. A standard student t-test was performed to analyse statistically significant data (p-value > 0.05).

(C)

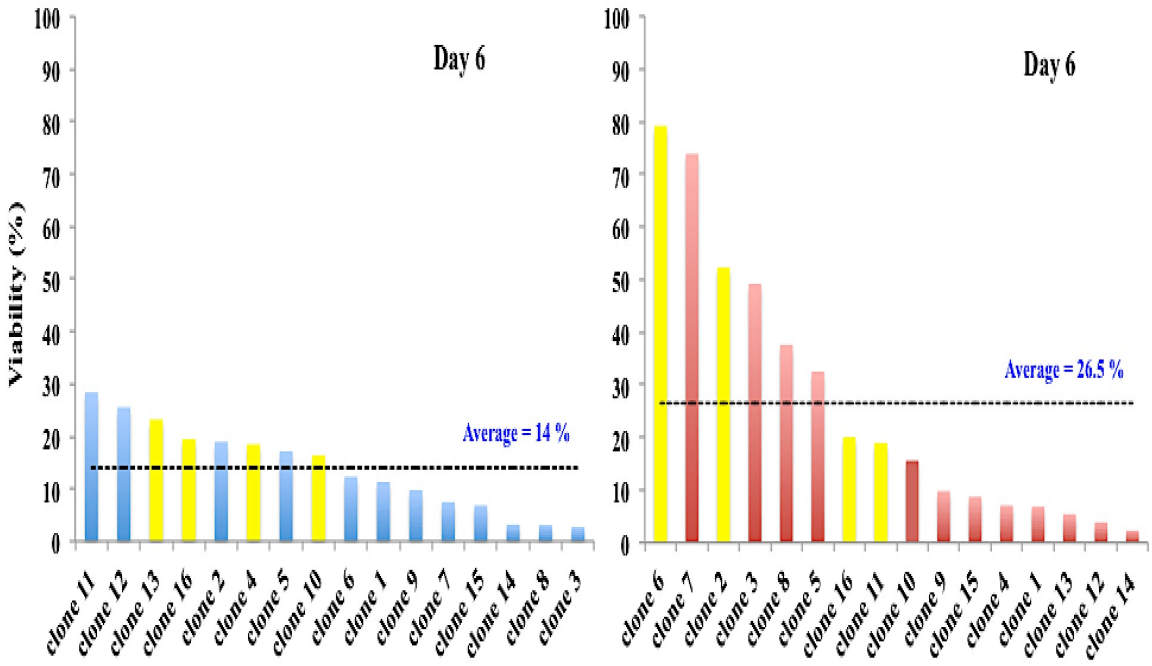


Figure 3.1.8C: Batch Growth analysis- % cell viability comparison between NC (Blue bars) and Let-7 sponge (Red bars) clonal panels in 24-well plate on Day 6 (sorted from highest to lowest % cell viability. Each well of two 24-well plates was seeded at a cell density of 2×10^5 cells/mL in 1 mL suspension culture volume (biological duplicate, n = 2). Cell counts were performed using Guava Viacount reagent and the Guava flow-cytometer. Yellow bars highlighted in each clonal panel set represents selected clones chosen for further Scale-up growth and productivity analysis later. The dashed **Black** line represents average cell density across the clonal panels. A standard student t-test was performed to analyse statistically significant data (p-value ≤ 0.05).

3.1.6.2 Productivity characteristics of stable clonal panels in 24-well plate format

The amount of SEAP protein produced by the NC and Let-7 sponge clones was measured using a kinetic absorbance based SEAP assay (refer to **Section 2.9.3**). However, it was observed that none of the NC and Let-7 sponge clones produced SEAP protein in measurable amounts (**Fig. 3.1.9A and Fig. 3.1.9B**). Typical volumetric SEAP production lies in the range of 100-450 units/mL for CHO-K1 SEAP cell densities in the range of 2-50 × 10^5 cells/mL, however, none of the NC or Let-7 clones surpassed 10 SEAP units/mL.

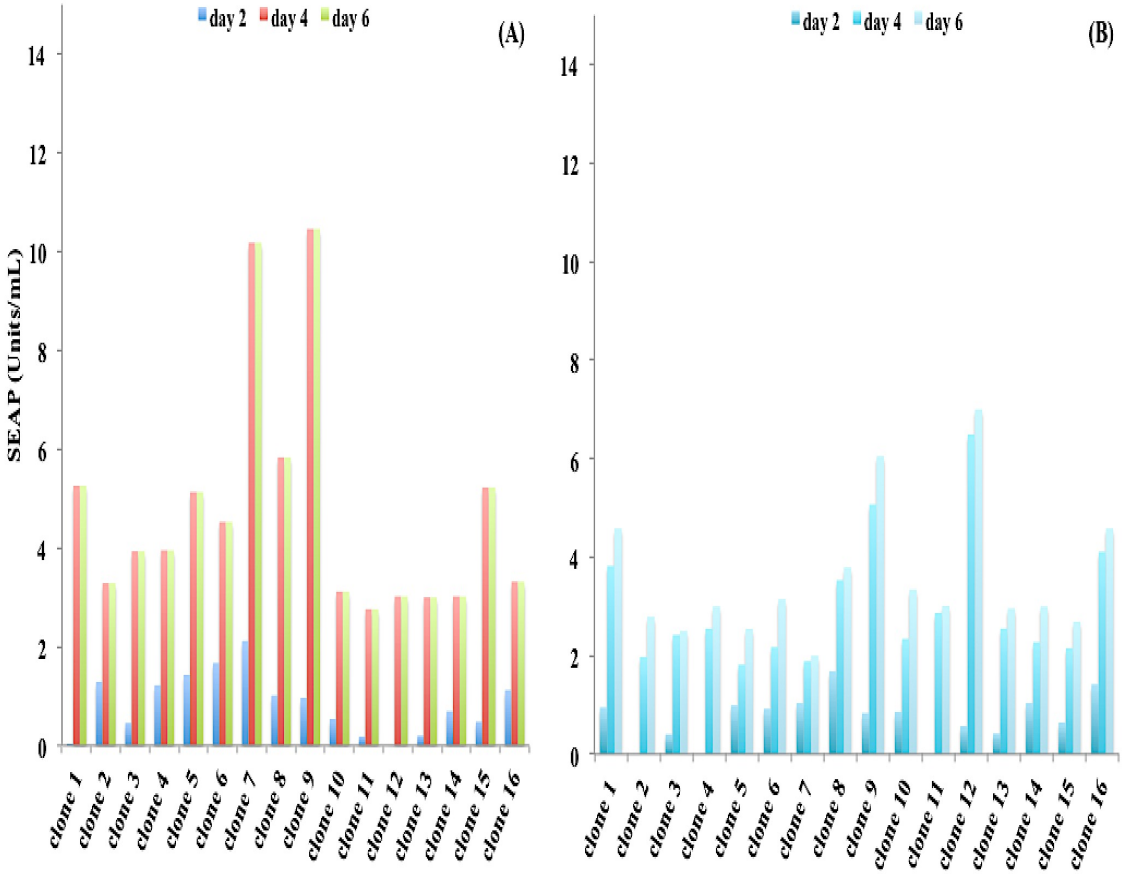


Figure 3.1.9: Volumetric SEAP productivity of NC (A) and Let-7 (B) sponge clonal panels. Cells were seeded at 2 × 10^5 cells/mL in a 1 mL working volume in a 24-well plate well in biological duplicate in serum-free medium. Samples were harvested from Day 2 onwards. A standard student t-test was performed to analyse statistically significant data (p-value > 0.05).

We suspected a potential loss of the SEAP transgene in both the NC and Let-7 sponge clonal panels. In order to confirm, we performed RT-qPCR to assess the expression of

SEAP transcript in a subset of clones picked from each NC and Let-7 panel (**Fig. 3.1.9C**). We observed a 25-fold down regulation in SEAP gene expression in both NC and Let-7 sponge clones in contrast to NC and Let-7 mixed pools. This observation was rather surprising since the clones were maintained by routinely passaging and selecting with hygromycin and geneticin containing media to ensure that they were maintaining the expression of both the sponge cassette, as well as the SEAP encoding transgene. Given the tendency for CHO cell lines to be genetically instable and susceptible to gene silencing, we suspected that this loss of SEAP expression could be a case of transcriptional gene silencing by methylation (Chusainow et al. 2009, Yang et al. 2010). This is further discussed in the **Section 4.2**. To investigate the root cause for the loss of SEAP expression in both the NC and Let-7 sponge panels was beyond the scope of this project, therefore, no further investigative study was conducted.

(C)

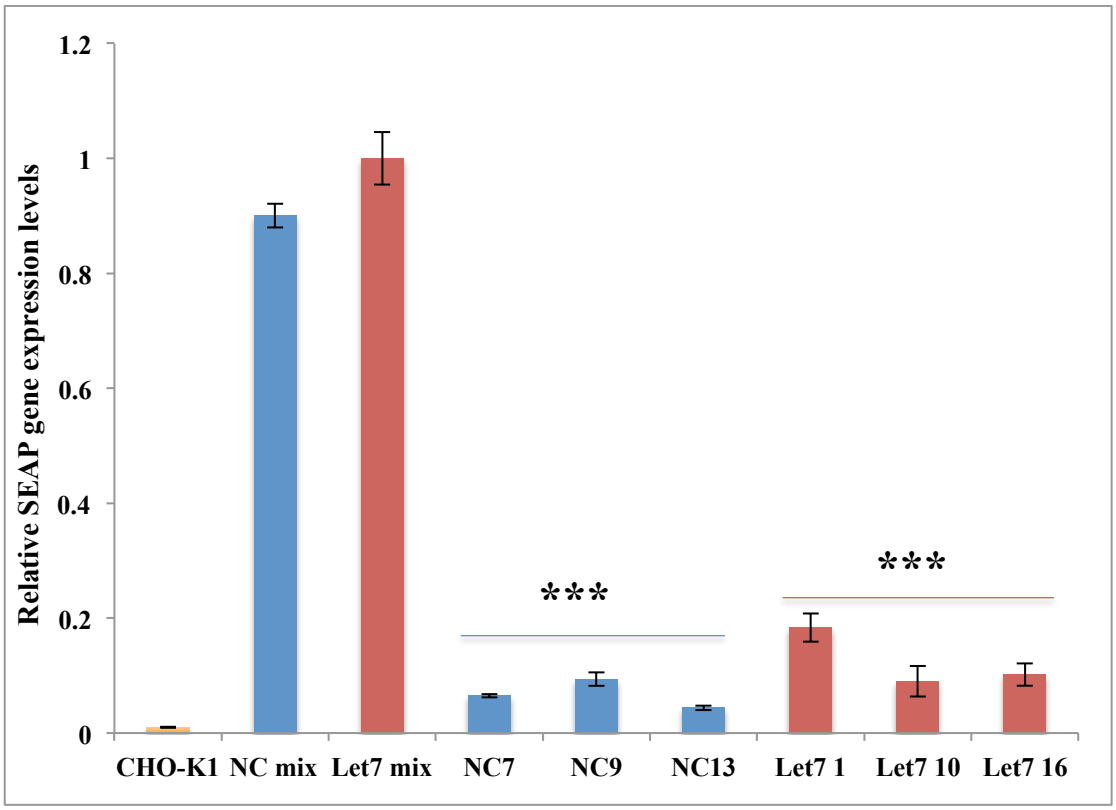


Figure 3.1.9C: Relative expression levels of SEAP transcript in NC and Let 7 sponge clones calculated using RT-qPCR. A parental CHO-K1 cell line was used as negative control for SEAP expression, while mixed pools of NC and Let-7 SEAP cells were chosen as the positive controls for SEAP expression. A standard student t-test was performed to calculate the statistical significant data (p-value < 0.0001). The levels of SEAP encoding gene were normalised to endogenous beta-actin control.

The absence of SEAP productivity data for NC and Let-7 clones stymied the process of identifying better-performing clones for further scale-up analysis. Therefore, we based our choice of clones on the basis of improved VCD on Day 3 and % cell viability on Day 6. We selected four clones from each NC and Let-7 sponge panel (**Highlighted in Yellow bars in Fig 3.1.8A, Fig. 3.1.8B and Fig. 3.1.8C**). From the NC sponge panel, we selected clone number 4, 10 (based on the VCD on Day 3 **Fig. 3.1.8A and Fig. 3.1.8B**), 13 and 16 (based on the highest % cell viability in comparison to other clones within the panel on Day 6 **Fig. 3.1.8C**). From the Let-7 sponge panel, clone number 2, 6 (based on the highest % cell viability in comparison to other clones within the panel on Day 6 **Fig. 3.1.8C**), 11 and 16 (based on the VCD on Day 3 **Fig. 3.1.8B**) were chosen for further scale-up growth and productivity analysis.

3.2 Impact of Let-7 depletion on bioprocess relevant phenotypes in a panel of selected CHO-K1 SEAP clones in 5 mL culture spin tubes

As described in the previous section, in total 8 stable clones were chosen for analysis in 5 mL. The clones from both the panels i.e. NC and Let-7 sponge were seeded at 2 × 10^5 cells/mL in a 5 mL working volume in 50 mL vented spin tubes. The experiment was monitored over a week and samples were harvested on Day 2, Day 3, Day 4 and Day 5. It was observed that the Let-7 clones 6 and 11 failed to demonstrate the same enhanced growth phenotypes in 5 mL batch conditions as observed in the 24-well plate format. The VCD of clone 6 and 11 only rose to 1.24 × 10^6 cells/mL (Day 3) and 1.43 × 10^6 cells/mL (Day 3), respectively followed by declining numbers (**Fig. 3.1.10B**) versus any of the NC clones, which were > 2 × 10^6 cells/mL on Day 3 and Day 4 (**Fig. 3.1.10A**). Let-7 clones 2 and 16 were observed to have higher VCDs i.e. 2.55 × 10^6 cells/mL and 2.79 × 10^5 cells/mL, respectively on Day 3 (**Fig. 3.1.10B**). However, Let-7 clone 2 did not show any sign of improved % cell viability as observed in the 24-well plate format and declined to < 70 % cell viability on Day 4. Let-7 clone 16 was found to have cell viability of 75 % on Day 5 of the culture, which was higher than the rest of the Let-7 sponge clones (**Fig. 3.1.11B**). However, this increased cell viability was not significantly higher than three of the NC sponge clones on Day 5 (p-value > 0.05) (**Fig. 3.1.11A**). Therefore, it can be concluded that Let-7 sponge overexpression had no positive impact on growth of these clones in 5 mL batch culture. Moreover, all the NC clones were observed to have a better growth and viability profiles (**Fig. 3.1.10A and Fig. 3.1.11A)** than the Let-7 clones (**Fig. 3.1.10B and Fig. 3.1.11B**).

Hence, Let-7 depletion using the Let-7 sponge did seem not to confer any positive impact on CHO-K1 SEAP cell growth as inferred in this particular experiment.

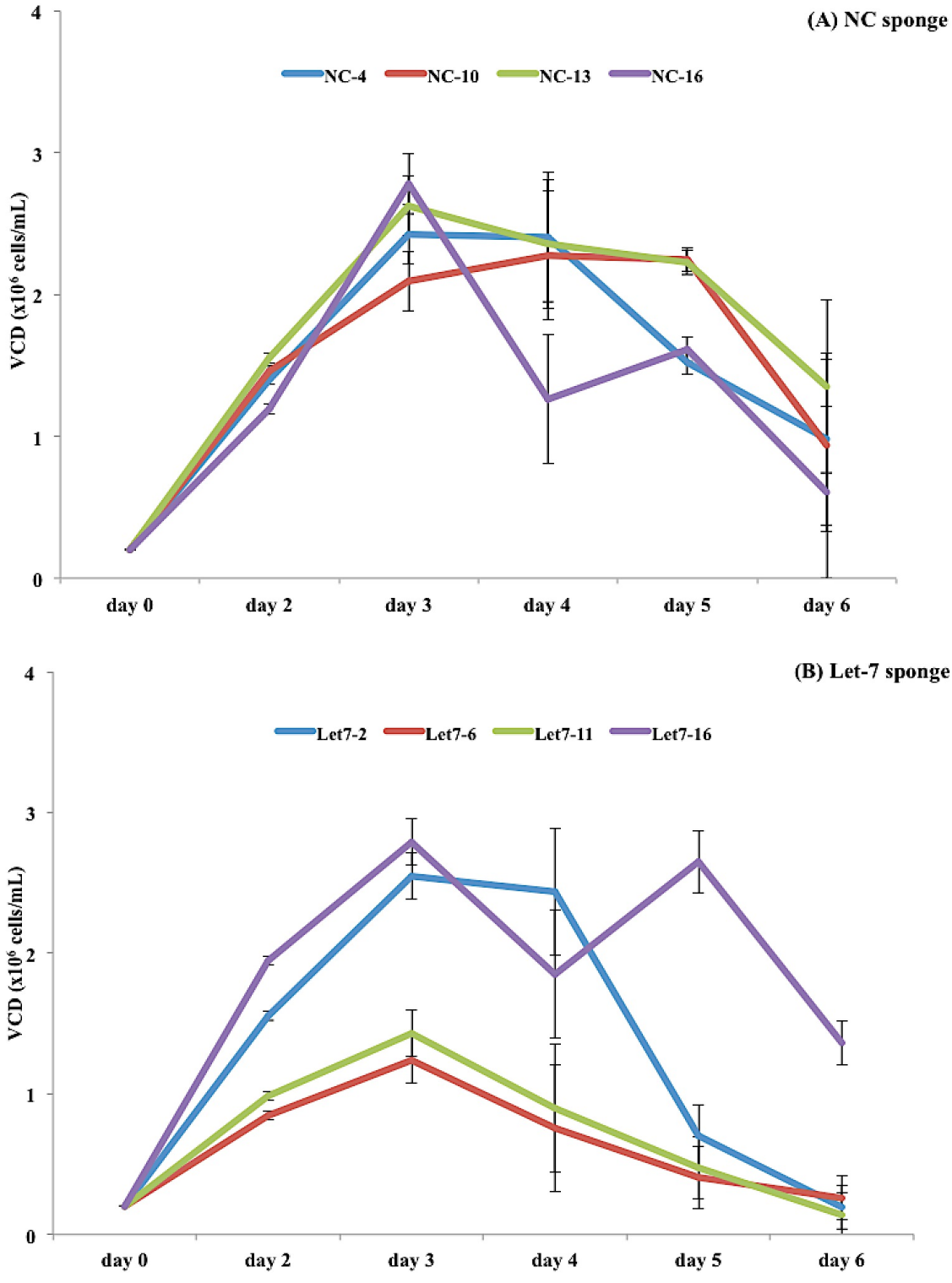
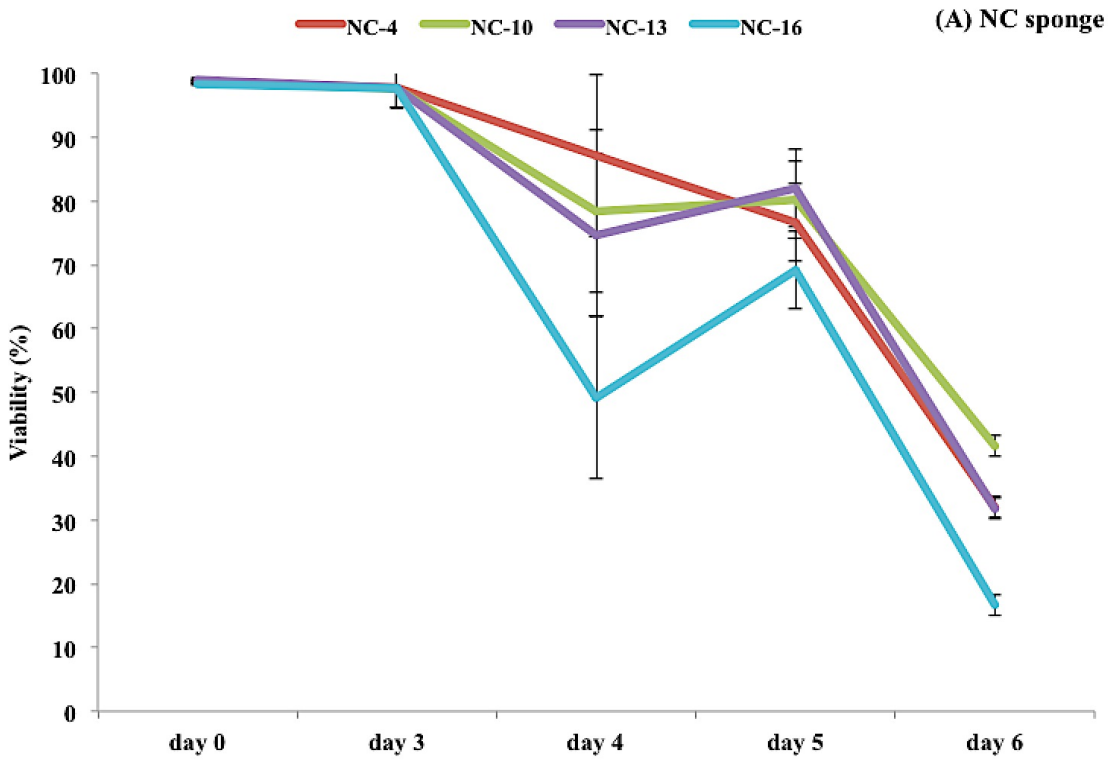


Figure 3.1.10: Batch culture growth analysis of (A) NC and (B) Let-7 selected sponge clones in 5 mL spin tubes. Cells were seeded at 2×10^5 cells/mL in a 5 mL working volume in a 50 mL spin tube in biological triplicate for each NC and Let-7 selected clones in serum-free medium. Cell counts were performed using Guava Viacount reagent and Guava. A standard student t-test was performed to analyse statistically significant data (p-value > 0.05).



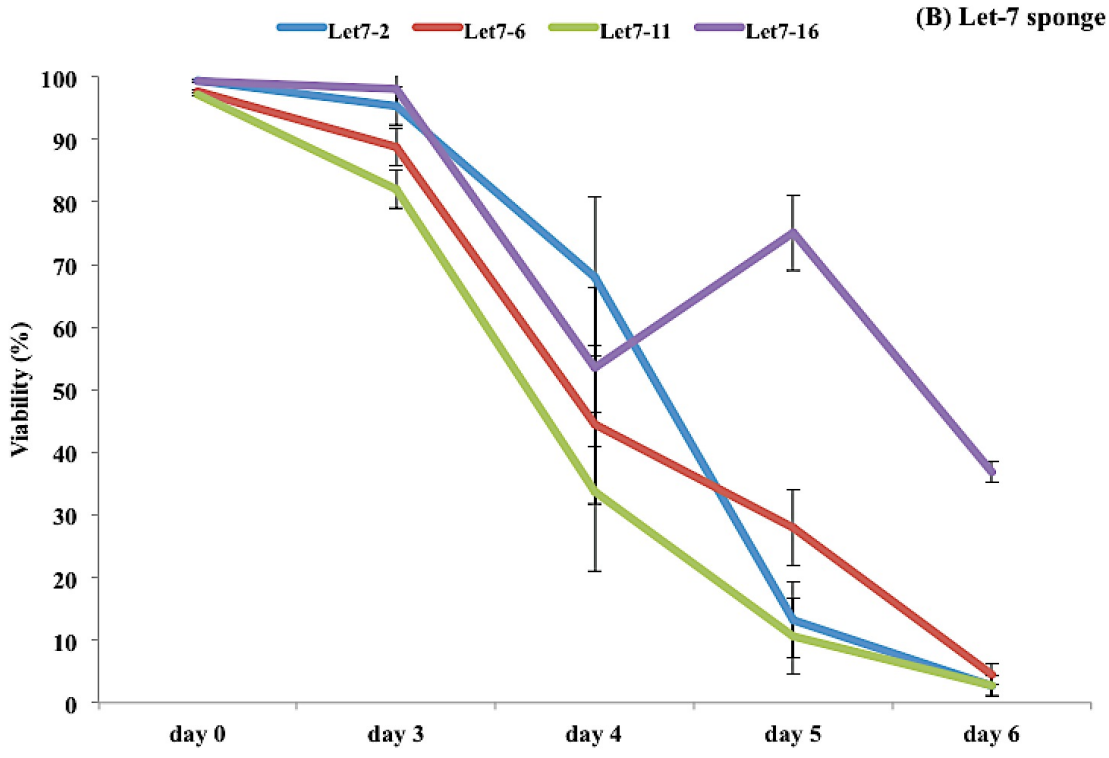


Figure 3.1.11: Batch culture % cell viability analysis of (A) NC and (B) Let-7 selected sponge clones in 5 mL spin tubes. Cells were seeded at 2×10^5 cells/mL in a 5 mL working volume in a 50 mL spin tube in biological triplicate for each NC and Let-7 selected clones in serum-free medium. Cell viability was analysed using Guava viacount reagent and Guava. A standard student t-test was performed to analyse statistically significant data (p-value > 0.05, n = 3)

3.3 Validation of Let-7 sponge technology

3.3.1 Let-7 sponge efficiency and specificity

Before performing any other large-scale batch and/or fed-batch based phenotypic analysis, it was imperative to validate and assess the Let-7 sponge's efficiency and specificity in binding endogenous Let-7 family members. Therefore, for this purpose we transiently transfected Let-7 mimics molecules into two i.e. one high and one low GFP expressing NC (clone number 10 & 17) and Let-7 (clone number 11 & 19) sponge clones.

As mentioned in **Section 3.1.3** Let-7 sponge clones expressing the sponge constructs exhibited lower GFP expression than the NC sponge clones. To verify that this was indeed due to Let-7 binding to the sponge, all four clones picked for this study were transiently transfected with non-specific miRNA mimics (pm-neg) and Let-7 specific mimics (pm-7), separately. Three days post-transfection, this up-regulation of Let-7 induced further significant reduction in GFP fluorescence intensity in the Let-7 sponge clone 11 as the GFP units reduced from 128 FU in pm-neg to 64 FU in pm-7. Clone 19 showed the same effect in the reduction of GFP fluorescence intensity as the GFP units reduced from 84 FU in pm-neg to 50 FU in pm-7. However, the GFP expression in NC sponge clone 10 and 17 remained unchanged, largely (**Fig. 3.2.1**). This further reduction in the mean GFP expression upon pm-7 transient transfection confirmed that the Let-7 sponge was efficient in its job to specifically bind to Let-7 family members and hence, reduce the levels available to bind to its endogenous target mRNA cohort. Based on this observation, wherein both the Let-7 clones displayed significantly reduced GFP levels upon exogenous Let-7 transfection, it was safe to say that theoretically all the stable clones contained enough Let-7 sponge transcript copies to fully soak-up the endogenous Let-7 levels and the lack of phenotypic outcome was not due to ineffective Let-7 depletion in the cells.

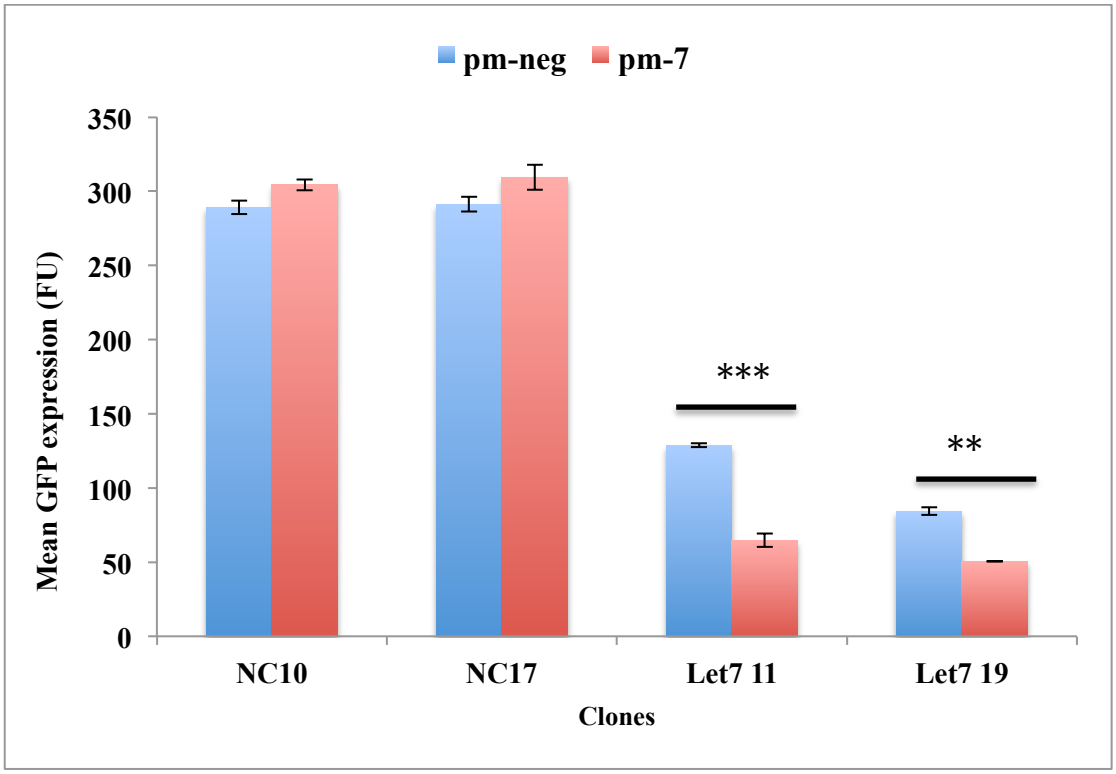


Figure 3.2.1: Mean GFP expression analysis post-transfection of two NC and two Let-7 sponge clones with non-specific miRNA mimics (pm-neg) and Let-7 specific mimics (pm-7). Cells were seeded at 1×10^5 cells/mL per well of three 6-well plates in biological triplicate and transfected with 30 nM of mimic molecules. GFP expression was analysed using Guava post 72 hours transfection. A standard student t-test was performed to analyse the statistically significant data (p-values 0.025 and ≤ 0.05).

3.3.2 Expression levels of downstream targets of Let-7 upon Let-7 depletion

For further validation, another experiment was conducted in order to confirm that Let-7 was deflected from its downstream targets by quantifying the expression levels of known Let-7 targets in NC and Let-7 sponge clones. We chose Bcl2 based on miRTARbase & miRwalk predicted list of targets (Garzon et al. 2007) and cyclin-dependent kinase-6 (CDK6), which has been shown to be a direct target of Let-7 in humans (Johnson et al. 2007). We quantitated their expression levels in NC (Clone 10, 13 and 16) and Let-7 sponge clones (Clone 2, 11, and 16) using RT-qPCR. We observed that the Bcl2 expression levels in all 3 Let-7 sponge clones were not significantly different than in the NC clones and the parent cell line (p-value > 0.05) (**Fig 3.2.2A**). Moreover, no difference in the expression levels of CDK6 in Let-7 sponge clones in comparison with NC clones (p-value > 0.05) was observed (**Fig 3.2.2B**). This experiment and from the findings in **Section 3.3.1**, we suspected that cells were somehow compensating for the lack of Let-7 via certain feedback loop mechanism (Ebert and Sharp 2012). Finally, we were interested in quantitating the endogenous Let-7 expression levels in the Let-7 stable sponge clones by qPCR. We surmised upon Let-7 depletion we could observe increased expression levels of Bcl2 and CDK6 in Let-7 sponge clones but there was no impact on mRNA transcripts of these two genes. However, we could not rule the possibility on any impact at protein level.

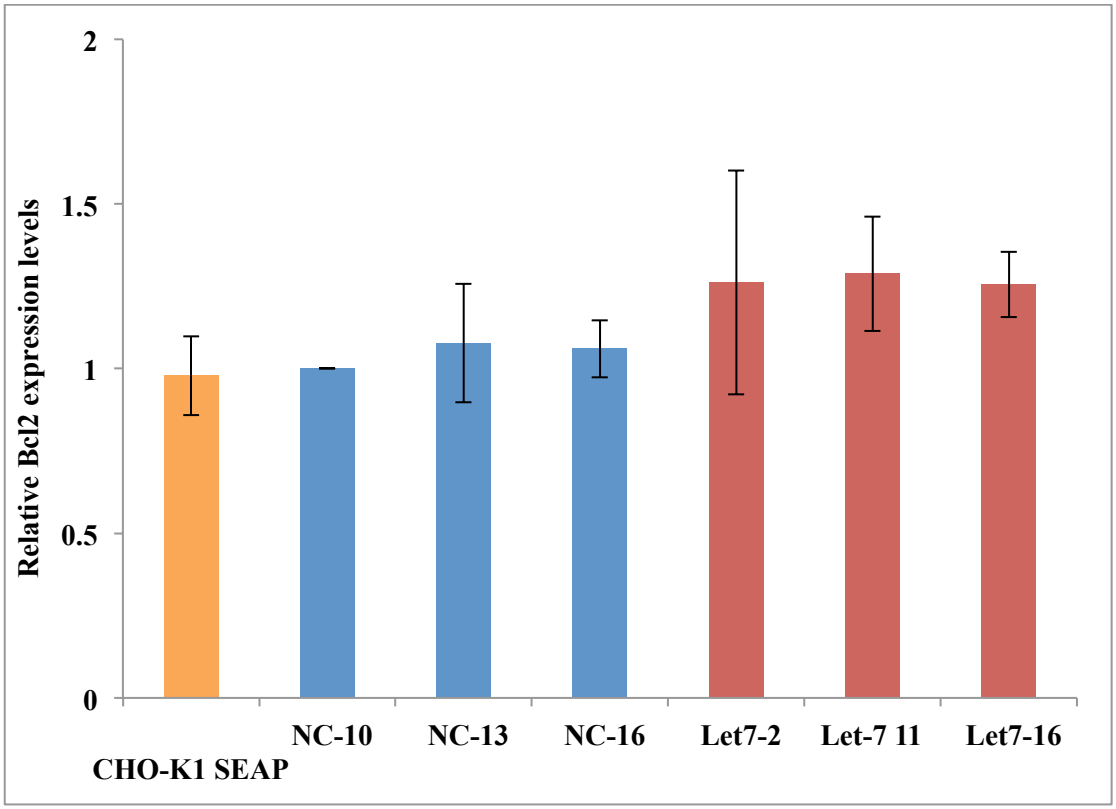


Figure 3.2.2A: Relative expression levels of Bcl2 measured using real-time qPCR in NC and Let-7 sponge clones. . A standard student t-test was performed to analyse statistically significant data (p-value > 0.05). The levels of Bcl2 were normalised to endogenous beta-actin control.

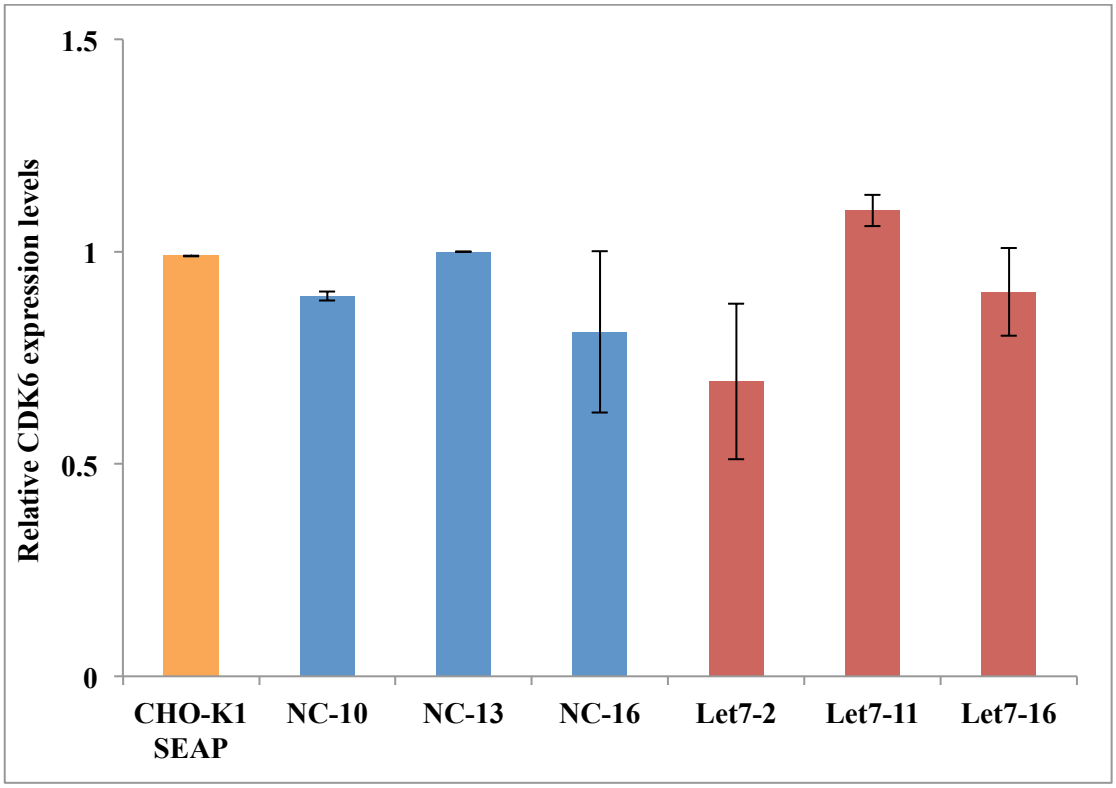


Figure 3.2.2B: Relative expression levels of CDK6 measured using real-time qPCR in NC and Let-7 sponge clones. A standard student t-test was performed to analyse statistically significant data (p-value > 0.05). The levels of expression of CDK6 were normalised to endogenous beta-actin control.

3.3.3 Endogenous Let-7 expression levels in stable sponge clones

This study was important in order to address the hypothesis that post Let-7 depletion in the Let-7 clones, there was a compensation mechanism occurring in the Let-7 sponge clones to maintain the normal levels of endogenous Let-7, hence resulting in no apparent phenotypic outcome. We measured endogenous Let-7a levels across a subset of the stable Let-7 and NC sponge clones using RT-qPCR. We observed that the mature Let-7a levels in all three of the Let-7 sponge clones (Clone 2, 6 and 11) were very similar to the expression levels in NC sponge clones (p-value > 0.05) (**Fig. 3.2.3**). This observation could explain why we observed no phenotypic outcome, however, was not in agreement with the data in **Section 3.3.1**. These observations point towards the possibility of target-mediated miRNA protection (TMMP), whereby increased mRNA target availability could potentially lead to accumulation of miRNAs due to protection from exoribonuclease turnover through target association (Chatterjee et al. 2011) (discussed in more detail in **Section 4.3**).

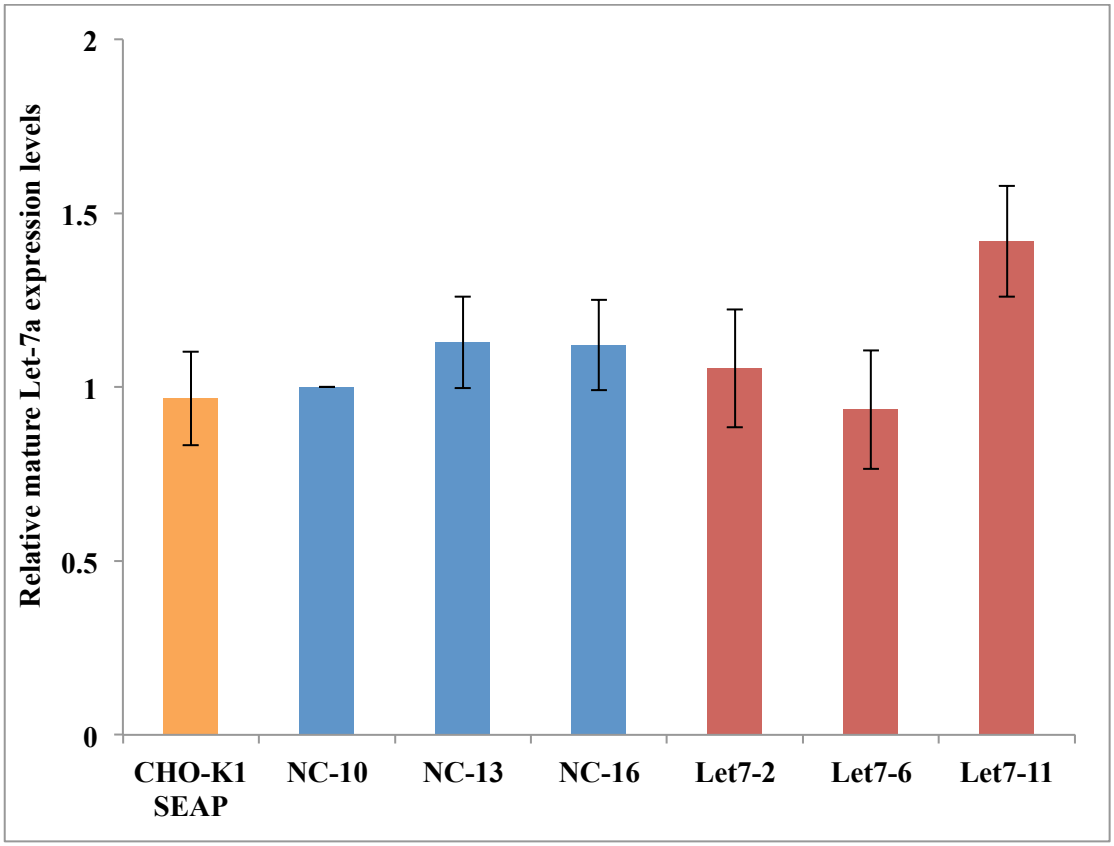


Figure 3.2.3: Relative Expression levels of the mature Let-7 levels in NC and Let-7 sponge clones assessed using real-time qPCR. A standard student t-test was performed to analyse statistically significant data (p-value > 0.05). The levels of expression of Let-7 were normalised to endogenous snRNA U6 control.

In summary, we demonstrated that the sponge technology was an efficient means to effectively and specifically sequester endogenous Let-7 levels with our mimics experiment (refer to **Section 3.3.1**). However, our findings showed no sign of improved growth characteristics in CHO-K1 SEAP cells after using Let-7 sponge technology. In the absence of productivity data, we could not establish if Let-7 stable depletion impacted the cellular productivity positively or negatively. The real-time qPCR results to validate the sponge efficacy were consistent with the growth results showing no apparent phenotypic change in the Let-7 sponge clones, in comparison to control clones. In theory, Let-7 sponge should bind all Let-7 family members, however, with RT-qPCR we measured very slightly increased or similar Let-7a expression levels in the Let-7 sponge clones as in the control sponge clones. We believe that the similar Let-7a expression levels in control and Let-7 sponge clones could be due to some unknown feedback mechanism to compensate for the loss of Let-7a in the cells. This observation is discussed in more detail in **Section 4.3**. This

study was a first report for the stable depletion of Let-7 levels in CHO cells, therefore, it could be interesting to observe if we employ a different technique to knockout the Let-7 family members, as opposed to stable depleting Let-7 levels in CHO cells with sponges. In doing so, it would be interesting to see if we could evade the mechanism of compensation of miRNA loss by the cells. Furthermore, it would be interesting to see if Let-7 knockout in CHO cells genome yields any interesting phenotypes.

Section 3.4, 3.5 & 3.6

Applying CRISPR-Cas9 genome editing technology to knockout miR-7 in industrially relevant CHO cell-lines

As the second objective of this thesis, a genome editing technology (CRISPR-Cas9) was employed to knock out the miR-7 locus sequence in the CHO genome. Stable knock down of miR-7 in CHO cells using sponges has been reported as an attractive approach to increase CHO cell growth and protein productivity. We were interested in completely knocking out the miR-7a-5p locus in CHO cells with a view to improve CHO cell bioprocessing phenotypes. The introduction and application of CRISPR-Cas9 technology in CHO cells was fairly new to our laboratory, therefore, firstly we implemented it in a CHO-eGFP cell line (henceforth referred to as eGFP cell line) for an easy read-out of the CRISPR-Cas9 activity. Moreover, this study also allowed us to optimize the CRISPR-Cas9 technology for the later part in this project. Hence, initially as proof-of-concept experiment, CRISPR-Cas9 system was employed in eGFP expressing CHO cell line to induce a functional eGFP gene knock out in the CHO genome.

3.4 Proof of Concept Study: Implementing CRISPR-Cas9 technology to induce functional gene knockout in eGFP cell line.

3.4.1 Designing eGFP sgRNA and cloning into the CRISPR vector (pCas9)

A CRISPR-Cas9 vector (pCas9), sourced from OriGene (9620 Medical Center Drive, Rockville MD, USA) was used as a backbone vector for cloning a 20 nucleotide (nt) guide-RNA sequence that targets the eGFP gene in eGFP cell line. The cloning strategy is outlined in **Figure 3.4.1**. Firstly, the pCas9 vector was doubly digested with *BamHI* and *BsmBI* restriction enzymes to linearize it (refer to **Section 2.8.2.1**). The single guide-RNA (sgRNA) sequences were then designed (refer to **Fig. 3.4.1** for eGFP sgRNA sequence details) (as per OriGene and www.blueheronio.com specifications) to target the eGFP gene at the 5' and 3'-end of the eGFP sequence. Following pCas9 backbone vector linearization, the sgRNA oligo sequence were annealed and then ligated into the linear pCas9 vector (refer to **Section 2.8.2.4**). The sgRNAs were designed to contain an extra 'G' nucleotide (nt) at the 5'-end of the guide sequence since it's a preferred start nucleotide by U6 RNA pol III promoter for efficient transcription (Hwang et al. 2013, Ran et al. 2013). Another extra 'G' nt was appended at the 3'-end of the guide sequence as well as per manufacturer's specifications.

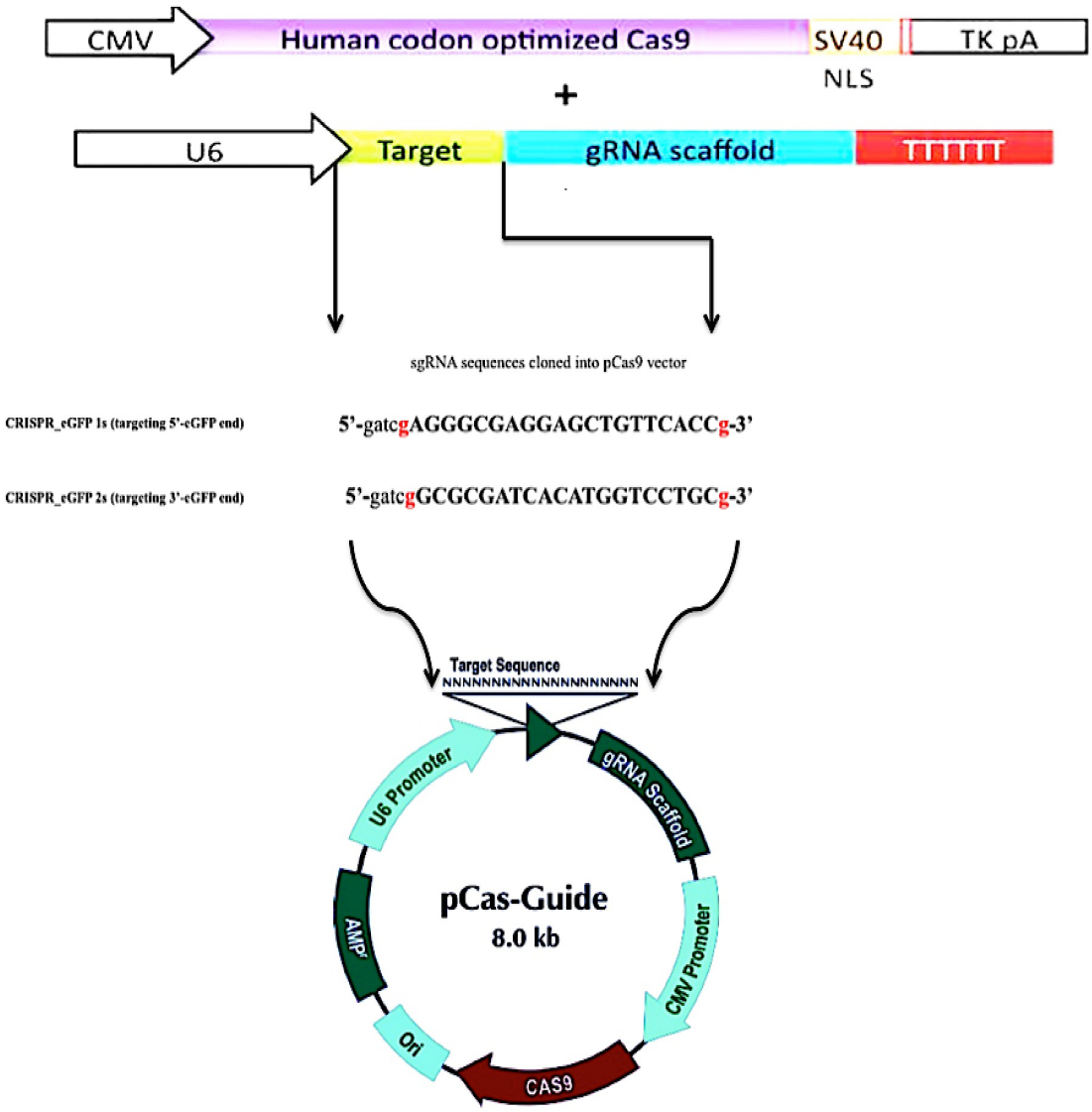


Figure 3.4.1: Schematic of design and cloning strategy for eGFP sgRNAs into pCas9 vector backbone for targeting eGFP gene. The single stranded oligonucleotides (i.e. eGFP 1s and eGFP 2s) were annealed with their respective anti-sense sequences and ligated into two separate linearized pCas9 vectors digested using *BamHI* and *BsmBI*, thereby, resulting in eGFP 1 (5'-eGFP end targeting) and eGFP 2 (3'-eGFP end targeting) CRISPR plasmids. A *BsmBI* (*gatc*) recognition site was also placed at the 5'-end of the guide sequences of both eGFP 1 and eGFP 2 CRISPRs for diagnostic gel purposes. Highlighted letter ***g*** at the 5'-and 3'-end is an extra Guanine, as per manufacturer's gRNA design protocol. The pCas9 vector consists of mammalian U6 RNA pol III promoter to transcribe the ligated target sequence and the gRNA scaffold (crRNA and tracrRNA complex). The pCas9-Guide vector also contains a CMV (cytomegalo viral promoter) driven human-codon optimized Cas9 nuclease (spCas9).

3.4.2 Proof-of-Concept experiment

As a preliminary task i.e. before applying CRISPR-Cas9 technology in the industrially relevant CHO cell line, it was imperative that we optimized and tested the efficiency of CRISPR-Cas9 systems in a model CHO cell line. For this purpose, we used a CHO-K1 cell line with a stable eGFP gene integrated in the genome, hence, referred as CHO-eGFP cell line.

3.4.2.1 CHO-eGFP cell line

CHO-eGFP cell line used in this project contains limited copies of eGFP gene in a transcriptionally active region of the CHO genome and exhibits strong, stable eGFP expression. This cell line was established using adeno-associated virus for the stable integration of the eGFP transgene into the genome without selective pressure. We performed a GFP expression study over 5 days using Guava in order to ascertain the consistency of eGFP. An eGFP negative (i.e. CHO-K1 parental) cell line was used to gate the instrument prior to GFP expression analysis and the same settings were used for each culture time point. For this experiment, CHO-eGFP cells were cultured in a 6-well plate and each culture time point was measured in biological triplicate. During the early stages of the culture period, we observed increasing VCD (**Fig. 3.4.2A**). The mean GFP expression for most of the time of the culture period stayed consistently around 2100 FU. However, on Day 4 and Day 5 the GFP expression dropped ~ 250 FU i.e. from 2146.48 FU to 1900 FU (**Fig. 3.4.2A and Fig. 3.4.2B**) This drop could potentially be attributed to the fact that the culture on Day 5 was already in decline. The VCD and % cell viability on Day 5 were 2.14×10^6 cells/mL (**Fig. 3.4.2A**) and 45 % (data not shown), respectively (**Fig. 3.4.2A**). From this study it can be concluded that the eGFP expression of the CHO-eGFP cell line is quite consistent. It can also be seen that there are two different clusters of eGFP positive cells in **Fig. 3.4.2B**, which suggest that eGFP cell line contains a mix of two eGFP positive populations. These two different populations exhibit varying degrees of GFP fluorescence i.e. one being brighter (higher red cluster) than the other (lower red cluster). Therefore, we believe that eGFP cell line maybe a mixed population of bright

and less bright eGFP cells. Other members of our research group working with eGFP cell line, as well as from our study, we could not completely discern the reason for observing varying degrees of GFP fluorescence.

(A)

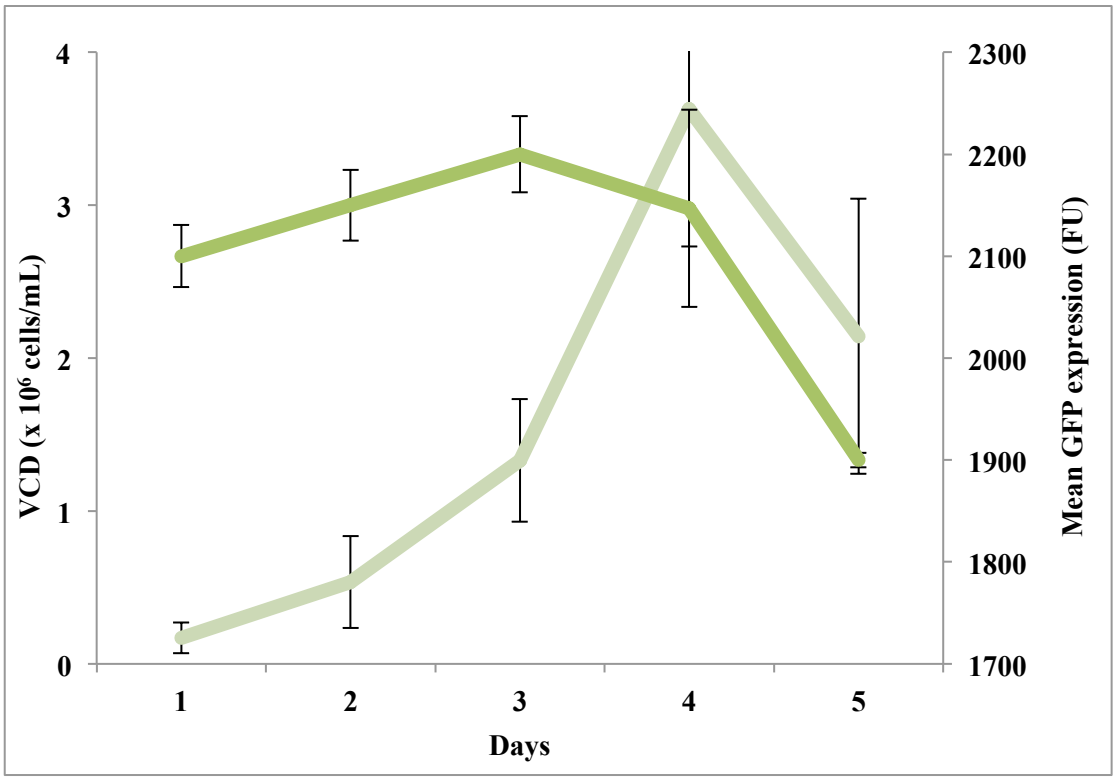


Figure 3.4.2A: VCD and Mean GFP expression analysis. CHO-eGFP VCD (cells/mL) in Light Green Curve and mean GFP expression curve in Dark Green Curve. Cells in the 6-well plate were seeded at 2 × 10^5 cells/mL in biological triplicate. Cell counts and mean GFP expression analysis performed using Guava Viacount and Guava ExpressPlus programme, respectively using the Guava.

(B)

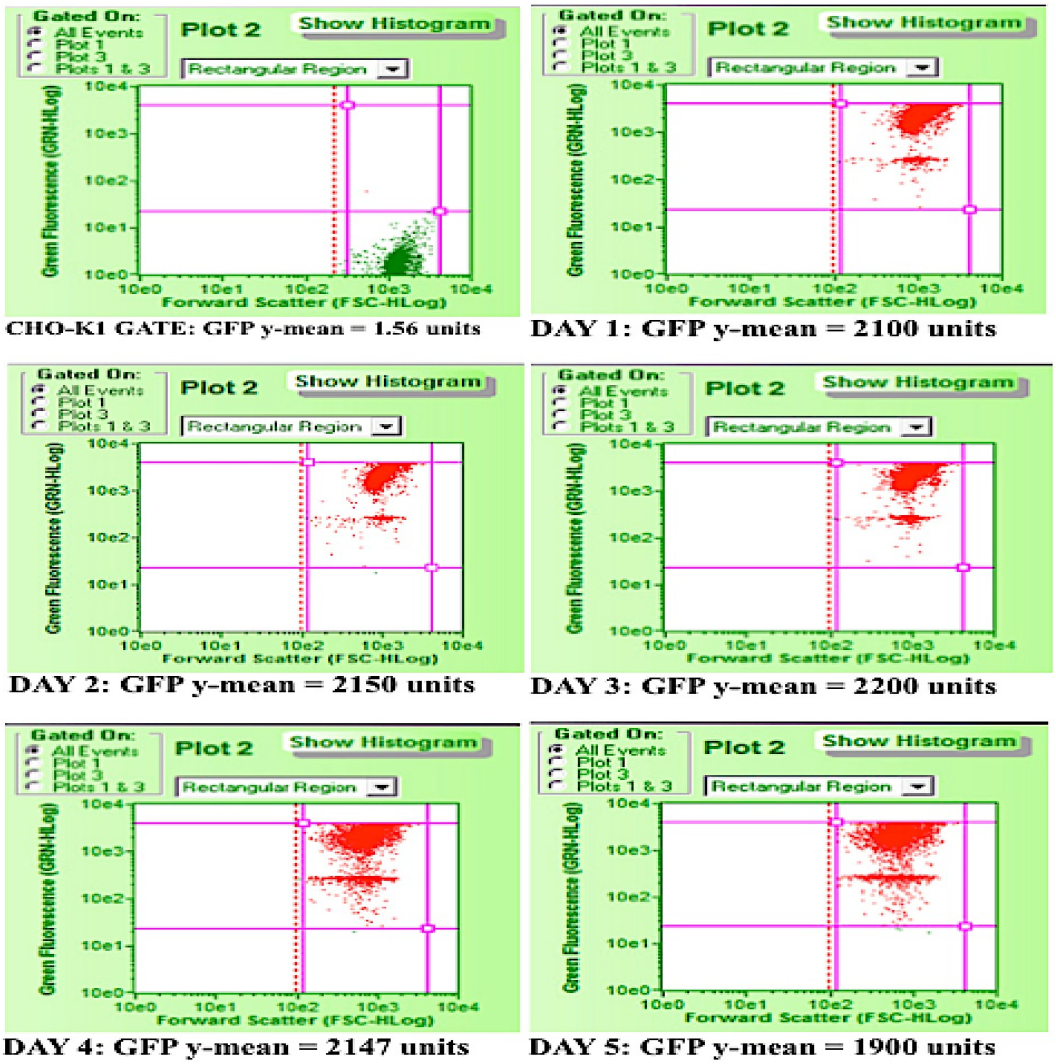


Figure 3.4.2B: Mean GFP expression imaged in the form of dot-plots. Screenshots for ExpressPlus programme representing clusters of GFP positive (Red) and negative cells (Green) from Guava software. The top extreme left screenshot shows the cluster of GFP negative CHO-K1 cells used to Gate the instrument prior GFP expression analysis.

3.4.2.2 Two different eGFP sgRNAs designed to target eGFP gene

The length of each of the two eGFP sgRNA used in this study was 20 nt long. These two eGFP CRISPRs were designed to target two different sites in the eGFP gene i.e. sgGFP 1 targets the 5'-end, while sgGFP 2 targets the 3'-end (**Fig 3.4.3**). Upon successful transfection of eGFP CRISPRs, we expected to observe bright green cells transitioning to dark cells (non-green cells) due to eGFP gene knockout by the Cas9 nuclease component in each of the eGFP CRISPR. We expected Cas9 would induce DSBs in the eGFP gene, which are then repaired by the cells' error prone NHEJ repair mechanism leading to frame shift mutations or indels (causing pre-mature STOP codon formation in a percentage of cells), thereby disrupting the ORF of the eGFP gene. The efficiency of eGFP targeting will depend on a number of factors including the transfection efficiency, sgGFP 1 and sgGFP 2 designs and the degree of error introduced during DSB repair process.

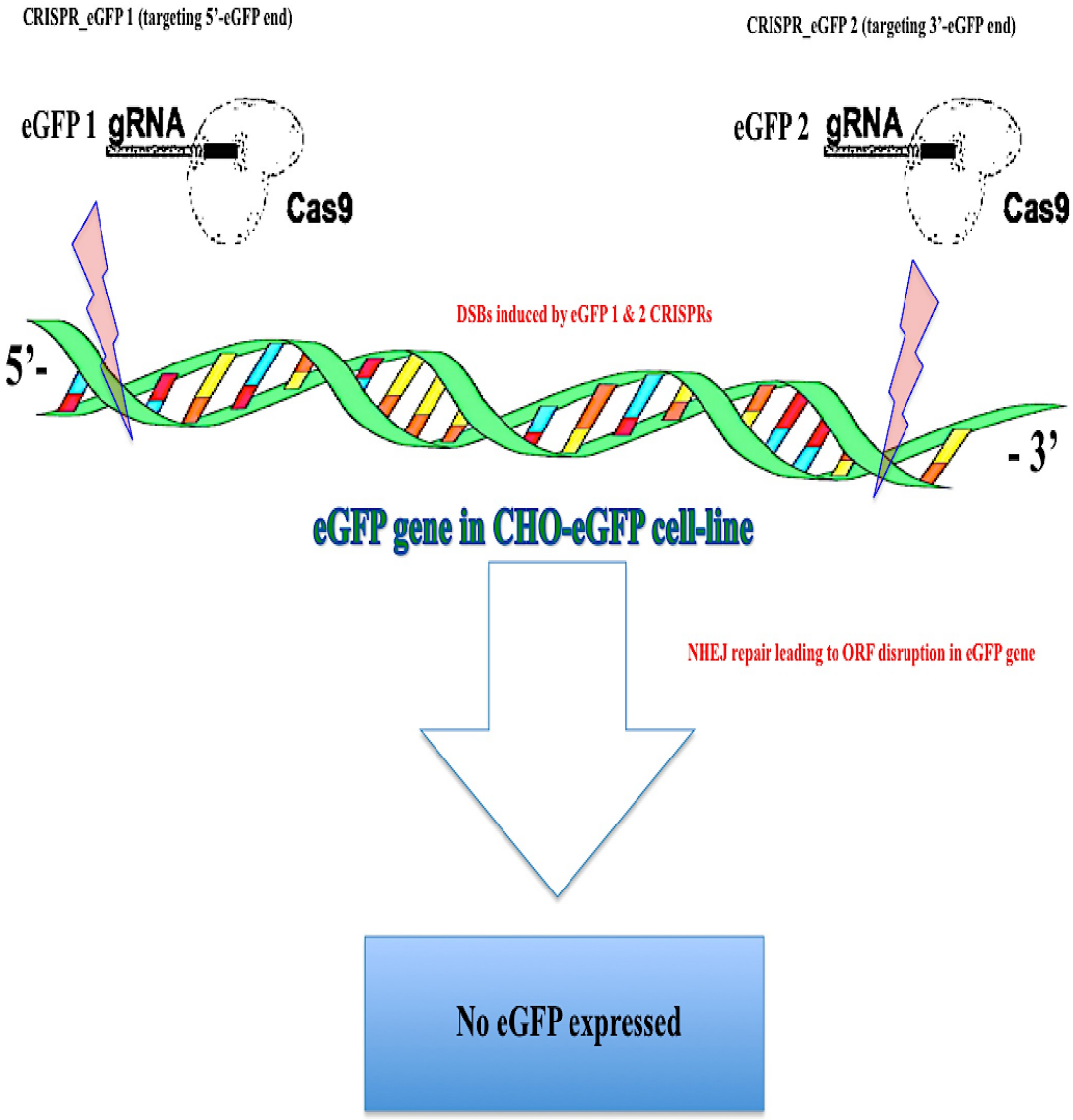


Figure 3.4.3: Outline of the Proof-of-Concept experiment rationale. 20 nt long eGFP 1 and eGFP 2 CRISPRs induce DSBs in the eGFP gene via Cas9 nuclease. The resulting DSBs are then repaired by the cells' error prone NHEJ repair mechanism leading to frame shift mutations or indels, thereby disrupting the ORF of eGFP gene, hence rendering it non-functional. Hence, green cells transitioning to dark cells. DNA cartoon image reproduced from www.neb.com

- Optimization of Transfection conditions

eGFP cells seeded at 5×10^5 cells/mL the day before were transfected with 5 μg of sgGFP 1 and sgGFP 2 in two separate wells of a 6-well plate. 5 μg and 10 μg of negative control vector 1 and 2 (i.e. pCas9 NC 1 and pCas9 NC 2), respectively, were also transfected into eGFP cells in two separate wells of the 6-well plate (refer to **Section 2.6.1.2**). 10 days post transfection, the results were visually analysed using the confocal microscope (**Fig. 3.4.4A**). We observed some CRISPR activity in each of the eGFP CRISPR treated samples. Mean GFP expression was reduced from 3418 FU (99 % GFP positive) in pCas9 NC 1 control to 2700 FU (82 % GFP positive) in sgGFP 1 and 2671 FU (85 % GFP positive) in sgGFP 2. Mean GFP expression of combination sgGFP (1+2) was also low i.e. 2700 FU (80 % GFP positive) in comparison to pCas9 NC 2 control i.e 3386 FU (99 % GFP positive) (**Fig. 3.4.4B**). It can also be seen in **Fig. 3.4.4C** cells with reduced GFP intensities drifting down towards the GFP negative area, however, a few cells with reduced GFP intensities were also observed very close to the GFP negative area in the pCas9 NC 1 and pCas9 NC 2 controls. Therefore, it was hard to infer that these cells in the GFP negative area are due to CRISPR activity or a sub population of less green cells.

We also hypothesised that expanding the cells from 6-well plate wells to T75 flasks could allow for more area for dark cells (targeted cells) to outgrow the eGFP expressing cells. We reasoned that maybe after one or two passages in T75 flasks, dark cells (i.e with eGFP knocked out gene) would outgrow the eGFP positive cells since dark cells won't be expressing the eGFP gene anymore and suffer less translational burden as opposed to eGFP positive cells. However, we did not observe any dramatic impact in the test populations, in terms of mean GFP expression reduction (data not shown). These findings led to speculations about the pCas9 eGFP CRISPR's efficiency in inducing knock out of the eGFP gene. In order to address if there was an issue with pCas9 accessing the eGFP sequence within the genome, we performed co-transfection experiment described in **Section 3.4.2.3**.

(A)

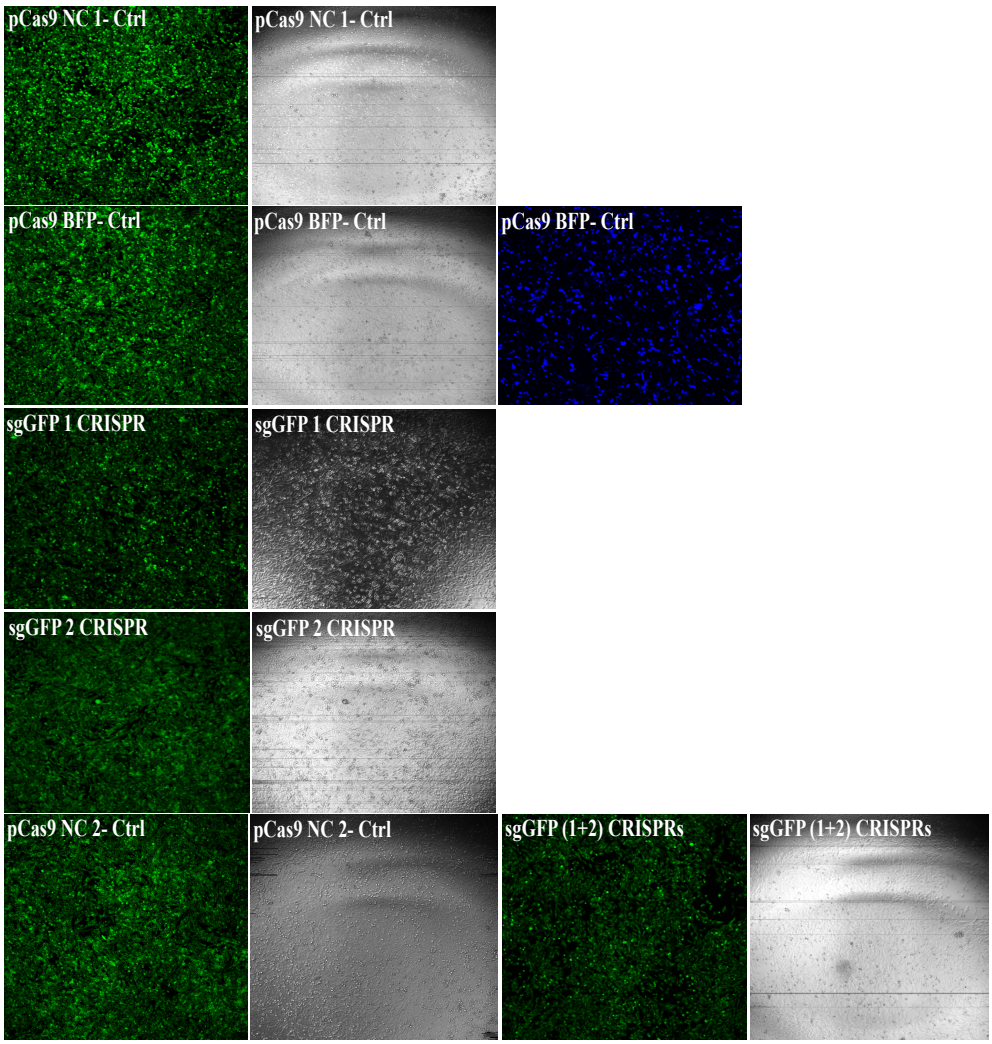


Figure 3.4.4A: Confocal Microscopy Images for the visual inspection of the eGFP 1 and eGFP 2 CRISPR activities. The black and white images with each set are the phase-contrast images. Cells were seeded at 2×10^5 cells/mL per well of a 6-well plate, then expanded to T75 flasks and images were taken using a confocal microscope post transfection in T75 flasks.

(B)

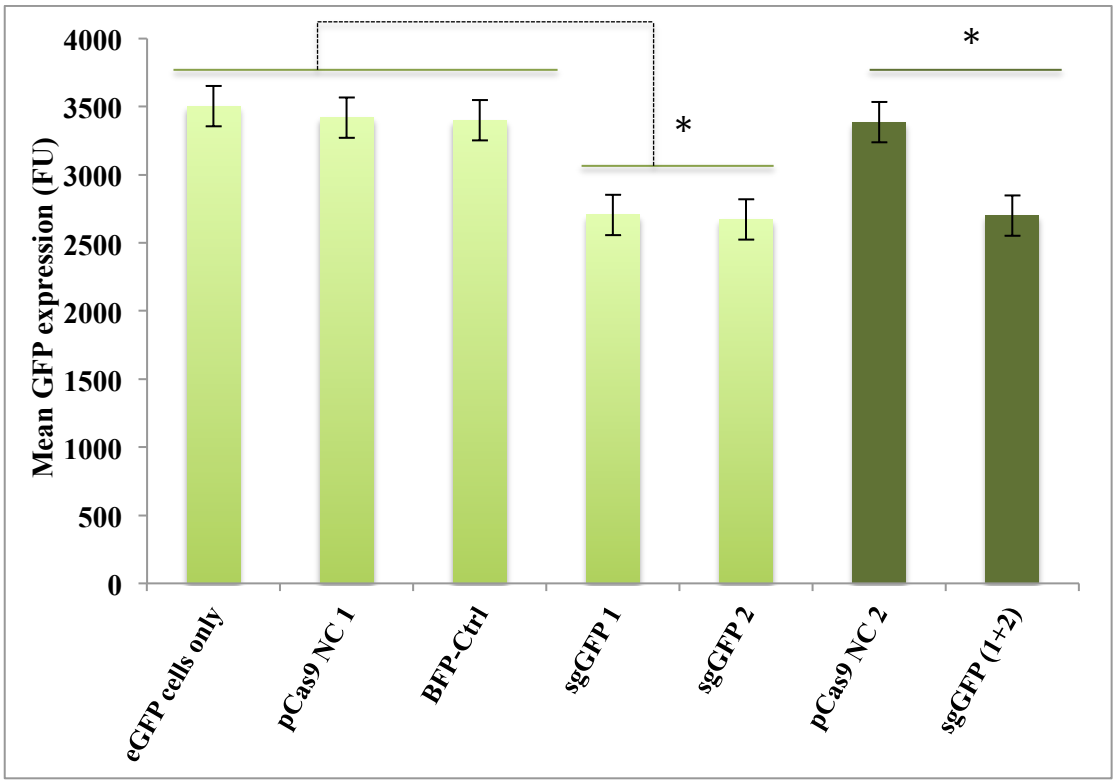


Figure 3.4.4B: Mean GFP expression analysis post eGFP 1 and eGFP 2 CRISPR transfection in eGFP cells. Mean GFP expression data analysed post transfection in control and eGFP CRISPR treated samples. Cells were seeded at 2×10^5 cells/mL per well of a 6-well plate and then expanded to T75 flasks. Mean GFP expression analysed using the Guava. A standard student t-test was performed to calculate statistically significant data (p-value ≤ 0.05).

(C)

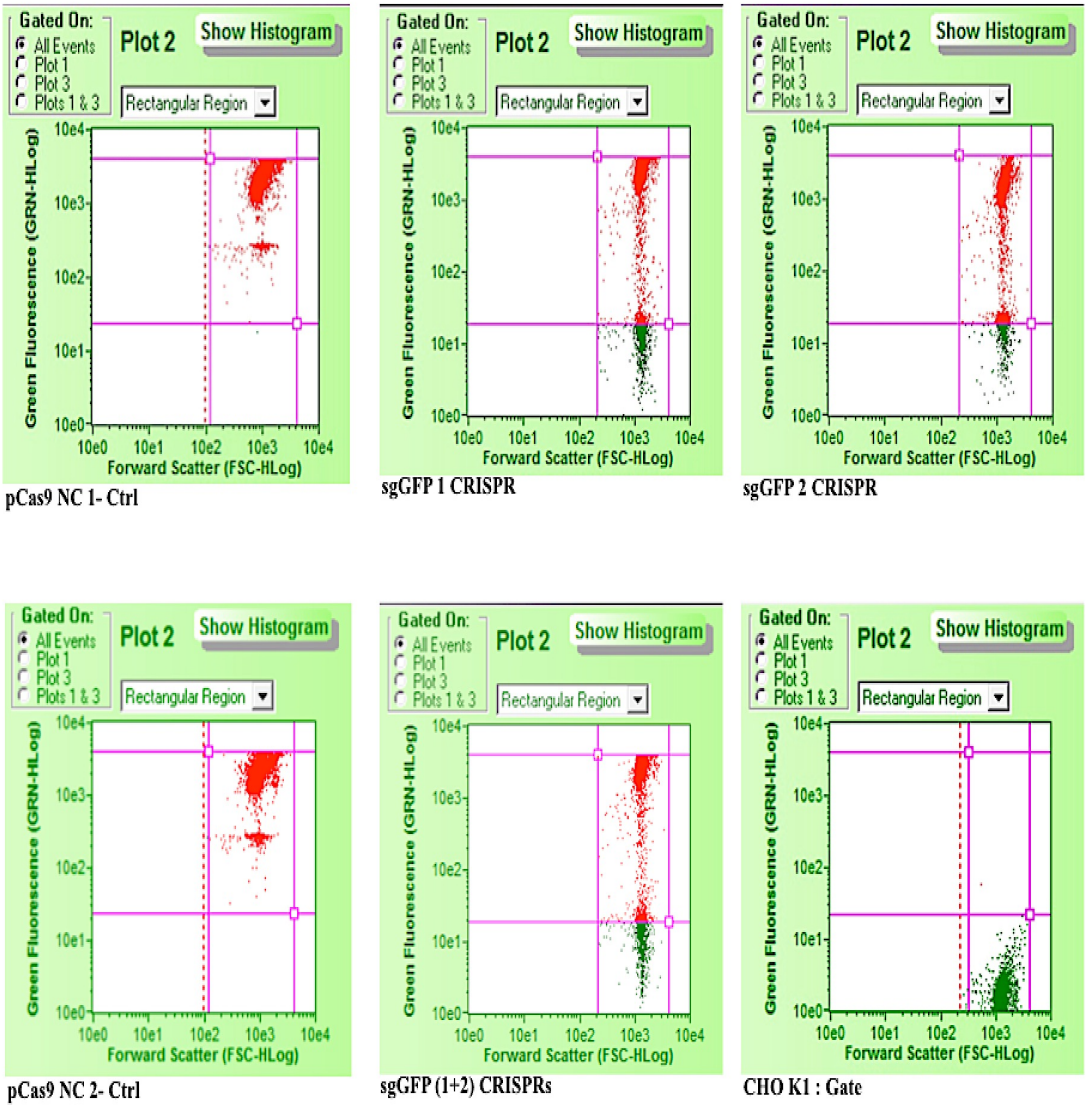


Figure 3.4.4C: Mean GFP expression imaged in the form of dot-plots. Screenshots for ExpressPlus programme representing clusters of GFP positive (Red) and negative (Green) cells in control and test eGFP CRISPR treated samples. The bottom extreme right screenshot shows the cluster of GFP negative CHO-K1 cells used to gate the instrument prior to GFP expression analysis. Cells were seeded at 2×10^5 cells/mL per well of a 6-well plate and then expanded to T75 flasks.

3.4.2.3 pCas9 eGFP CRISPR and eGFP plasmid (peGFP) transient co-transfection

In order to test the efficiency of the sgGFP 1 and sgGFP 2 CRISPRs at targeting a non-genome integrated eGFP sequence, eGFP CRISPRs were transiently co-transfected with peGFP plasmid in CHO-K1 SEAP cells separately i.e. sgGFP 1 + peGFP (5 μg) and sgGFP 2 + peGFP (5 μg) (refer to **Section 2.6.1.2**). This experiment was performed to check if the eGFP CRISPRs were specific in binding to eGFP sequence, and consequently disrupting the eGFP sequence in peGFP plasmid co-transfected in CHO-K1 SEAP cells. Following co-transfection of sgGFP 1 and sgGFP 2, we achieved ~ 50 % transfection efficiency as evidenced from peGFP + pCas9 control vector (no guide sequence) image. The test sgGFP 1 + peGFP sample was observed to have reduced numbers of eGFP positive cells (~ 25 %, data not shown) suggesting that sgGFP 1 CRISPR is specifically binding and disrupting the peGFP. Similar numbers of eGFP positive cells (~ 25 %, data not shown) were observed with sgGFP 2 + peGFP treated sample. The fluorescent microscope imaging demonstrated that the eGFP CRISPRs effectively reduced the eGFP signal in the transiently co-transfected cells (**Fig. 3.4.5**). We also observed that mean GFP expression in test samples was low (i.e. 60-67 FU range) in comparison to the control sample (109 FU) (**Fig. 3.4.5**) i.e. ~ 50 % reduction in GFP fluorescence, which was significant (p-value = 0.025). This study confirmed that eGFP CRISPRs are efficient in targeting the eGFP sequence of peGFP in CHO-K1 SEAP cells.

In summary, although the **Section 3.4.2.2** study showed a reduction in mean GFP expression in the CRISPRs treated cells, which was also significant (p-value ≤ 0.05) (**Fig 3.4.4B**). However, for the amount of CRISPR DNA used, we were unsatisfied with the efficiency eGFP CRISPRs in eGFP cells. Therefore, at this stage we wanted to improve the efficiency and decided to try an alternative expression vector system.

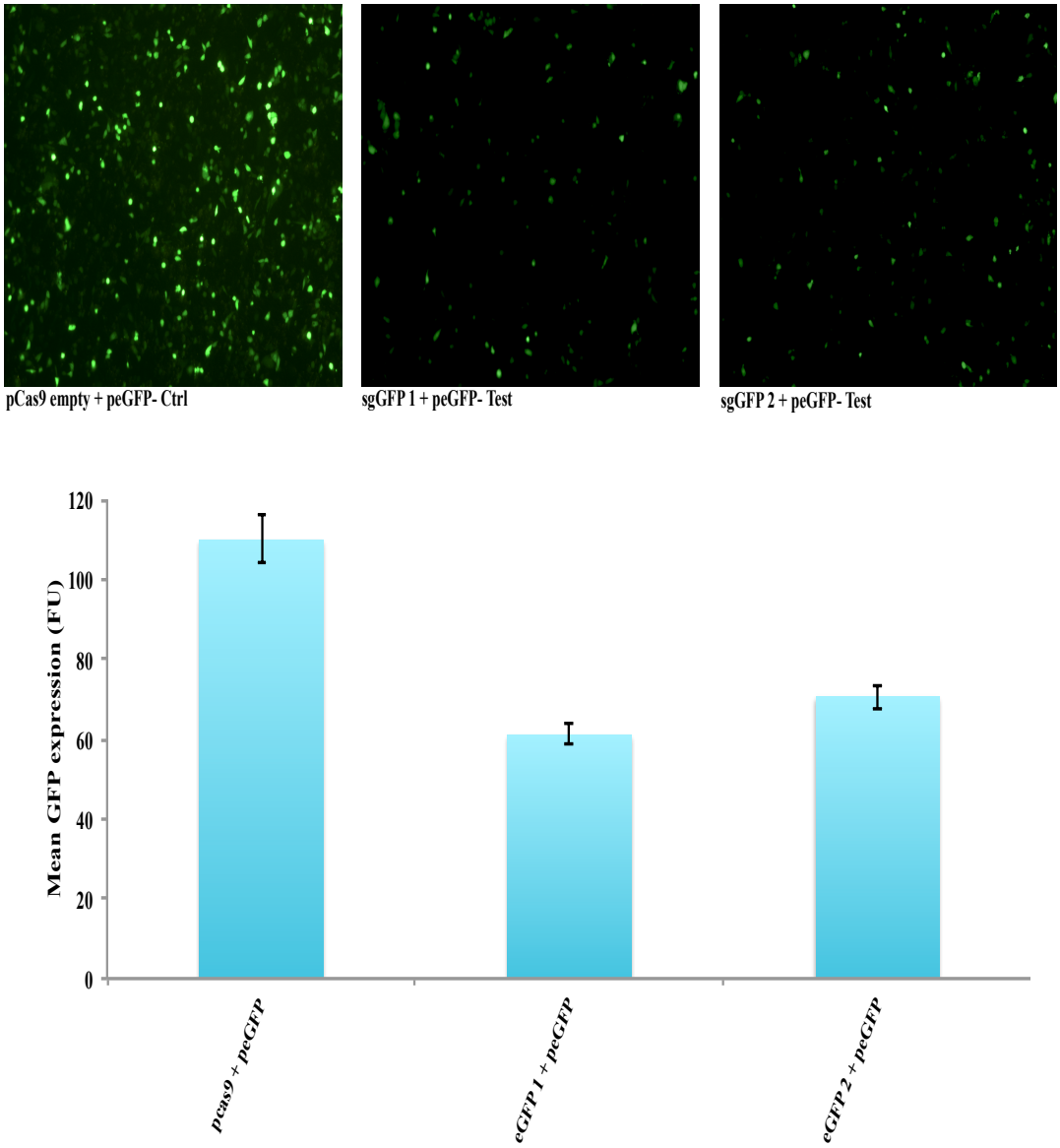


Figure 3.4.5: Fluorescent Microscopy images (Top) and Mean GFP expression (Bottom) analysis in CHO-K1 SEAP cells post co-transfection. Cells were seeded at at 2 × 10^5 cells/mL in biological triplicate. Three days post co-transfection each well of the 6-well plate was imaged using fluorescent microscope. Cells from each well were then analysed for mean GFP expression using the Guava. An eGFP negative (i.e. CHO-K1 parental) cell line was used to gate the instrument prior to GFP expression analysis. A standard student t-test was performed to analyse statistically significant data (p-value = 0.025).

3.4.3 A fresh start: Designing and cloning strategy for new eGFP targeting sgRNA in a different CRISPR vector

As mentioned above the pCas9 eGFP targeting CRISPR-Cas9 system did not induce site-specific eGFP gene disruptions as efficiently as we would have expected. In order to address this issue, we employed a completely new plasmid, px459 (pSpCas9(BB)-2A-Puro, Plasmid #48139) that has proven track record of efficient CRISPR activity as reported in numerous publications. Moreover, we designed two sgRNAs (i.e. 19 nt and 20 nt long) for each of two eGFP sites to be targeted. Therefore, in total four eGFP CRISPR-Cas9 vector constructs were designed (**Table 3.4.3 for sgRNA sequence details**).

Table 3.4.1 List of sgRNA sequences used for targeting eGFP gene. sgRNA sequences (A →19 nt or B →20 nt Highlighted in Yellow). Only sense strands are shown. The *BbsI* site (*cacc*) is also included at the 5'-end of the sequence for ligation into px459 vector. Red highlighted G appended at 5'-end (as preferred base) is the first base of the guide sequence. NGG is PAM sequence in the CHO genome.

CRISPR-Cas9 constructs	sgRNA sense-strand sequences (or guide sequences)	PAM (5'-end upstream of guide sequence)
sgGFP 1A	5' – CACCGGGGCGAGGAGCTGTTCACC – 3'	TGG
sgGFP 1B	5' – CACCGAGGGCGAGGAGCTGTTCACC – 3'	TGG
sgGFP 2A	5' – CACCGCGCGATCACATGGTCCTGC – 3'	TGG
sgGFP 2B	5' – CACCGGCGCGATCACATGGTCCTGC – 3'	TGG

The px459 vector was used as backbone vector for cloning sgGFP 1 and sgGFP 2 as guide sequences (sgRNAs), after linearizing using *BbsI* (refer to **Section 2.8.2.1**). The Cas9 expression cassette construct consists of a CBh (hybrid form of chicken-β actin promoter, a Cas9 ORF human codon-optimized (not shown), SV40 nuclear localization sequence (NLS) and bovine growth hormone (BGH) polyadenylation

signal and transcription termination sequence. The sgRNA expression cassette consisted of a U6 polymerase III promoter, a target sequence, a gRNA scaffold sequence and a poly (T) termination sequence. An additional advantage of using the px459 vector over the pCas9 vector was the presence of a puromycin resistance gene in the selection of transfected cells (**Fig. 3.4.6**).

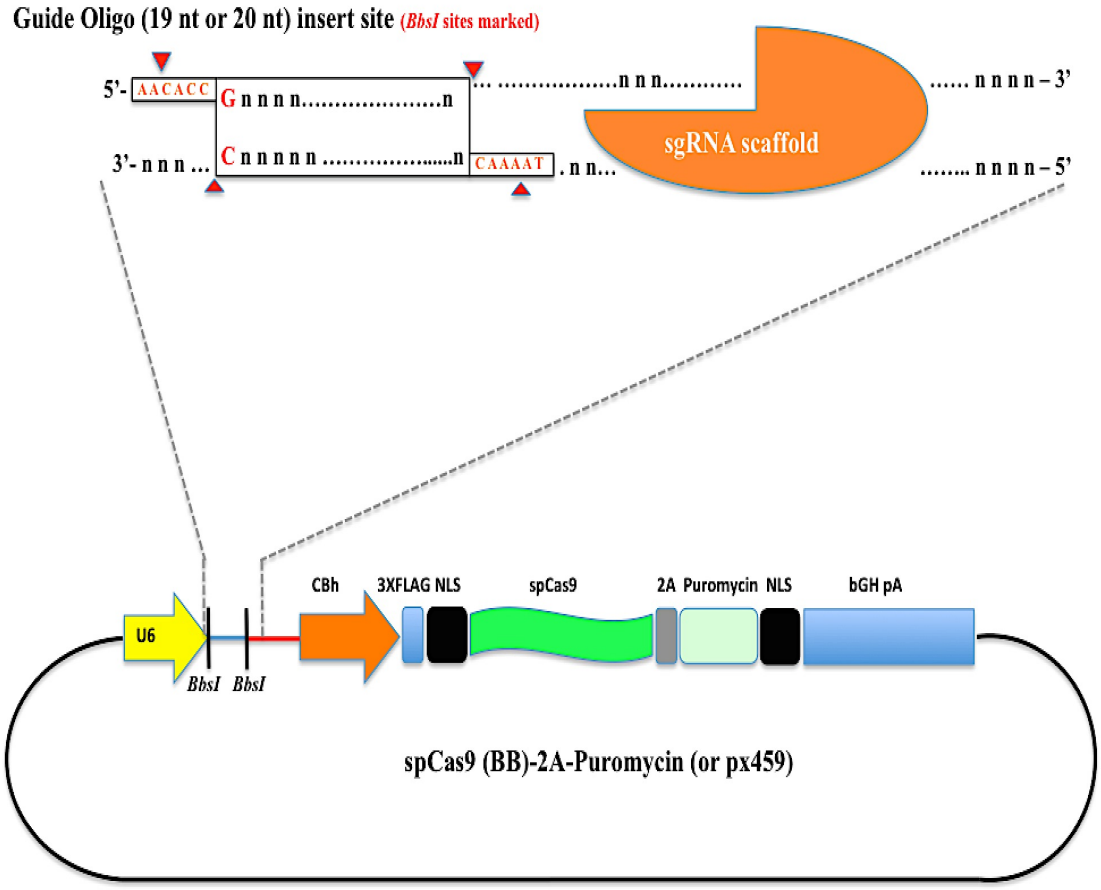


Figure 3.4.6: Schematic of Cas9 and sgRNA expression cassettes in px459 vector backbone. A 19 nt or 20 nt guide sequence containing *BbsI* cut compatible sticky-ends was cloned into a linearized px459 vector backbone (linearized using *BbsI*). The final sgRNA expression cassette and Cas9 construct consisted of a U6 polymerase III promoter, a target sequence, an sgRNA scaffold sequence (Orange structure) and a poly (T) termination sequence. Further downstream are other transgene elements such as CBh promoter, 3X FLAG peptide, SV40, NLS, spCas9 cassette linked with 2A peptide. 2A is also linked to puromycin (Selection Marker Gene), NLS and BGH, polyadenylation signal and transcription termination sequence.

The sgRNA sequences for sgGFP 1A, sgGFP 1B and sgGFP 2A, sgGFP 2B were designed using the CRISPR Design Tool (http://tools.genome-engineering.org) (**Table 3.4.1**). The sgGFP 1 and sgGFP 2 CRISPR systems differ in that sgGFP 1 targets the 5'-end, while sgGFP 2 targets the 3'-end of the eGFP gene. A and B refers to the length of the guide sequence, wherein A is 19 nt long sgRNA and B is 20 nt long sgRNA oligo annealed and ligated into the *BbsI* linearized px459 vector (refer to **Section 2.8.2.4**). The resulting constructs will be referred as sgGFP 1A, sgGFP 1B, sgGFP 2A and sgGFP 2B CRISPRs. We also investigated CRISPR-Cas9 efficiency when different sgRNAs were used for different sites in the eGFP gene.

3.4.3.1 Knocking out eGFP gene in CHO-eGFP cells

As a preliminary task, it was imperative that we test the efficiency of the designed eGFP CRISPR-Cas9 systems in CHO cells. For this purpose, we performed the CRISPR-Cas9 eGFP gene knockouts in the eGFP cell-line with the newly designed eGFP CRISPR constructs. The experimental rationale was the same as outlined in **Figure 3.4.3**, except that on this occasion we tested four sgGFP CRISPRs i.e. sgGFP 1A, sgGFP 1B, sgGFP 2A and sgGFP 2B.

We transfected 5 wells of a 6-well plate containing eGFP cells seeded at 5×10^5 cells/mL one day before transfection with 1 μg of control px459 (empty vector only i.e. no guide sequence) DNA, sgGFP 1A, sgGFP 1B, sgGFP 2A and sgGFP 2B CRISPR DNA (s). We included a control for 'eGFP cells only' to monitor the transfection efficiency when selecting the non-transfected eGFP cells and control px459 only transfected cells using puromycin containing media and comparing the number of viable green cells visually using the fluorescent microscope. Ten days post transfection, we observed ~ 70 % viable eGFP cells in the control px459 vector only treated sample (**Fig. 3.4.7A**), however, the 'cells only' control had all non-green floating dead cells. We also observed that in all eGFP CRISPRs treated cells 65-70 % of eGFP positive cells transitioned to dark, viable cells. The fluorescent microscopy results confirmed significant number of eGFP cells with eGFP gene knocked out (**Fig. 3.4.7A**).

(A)

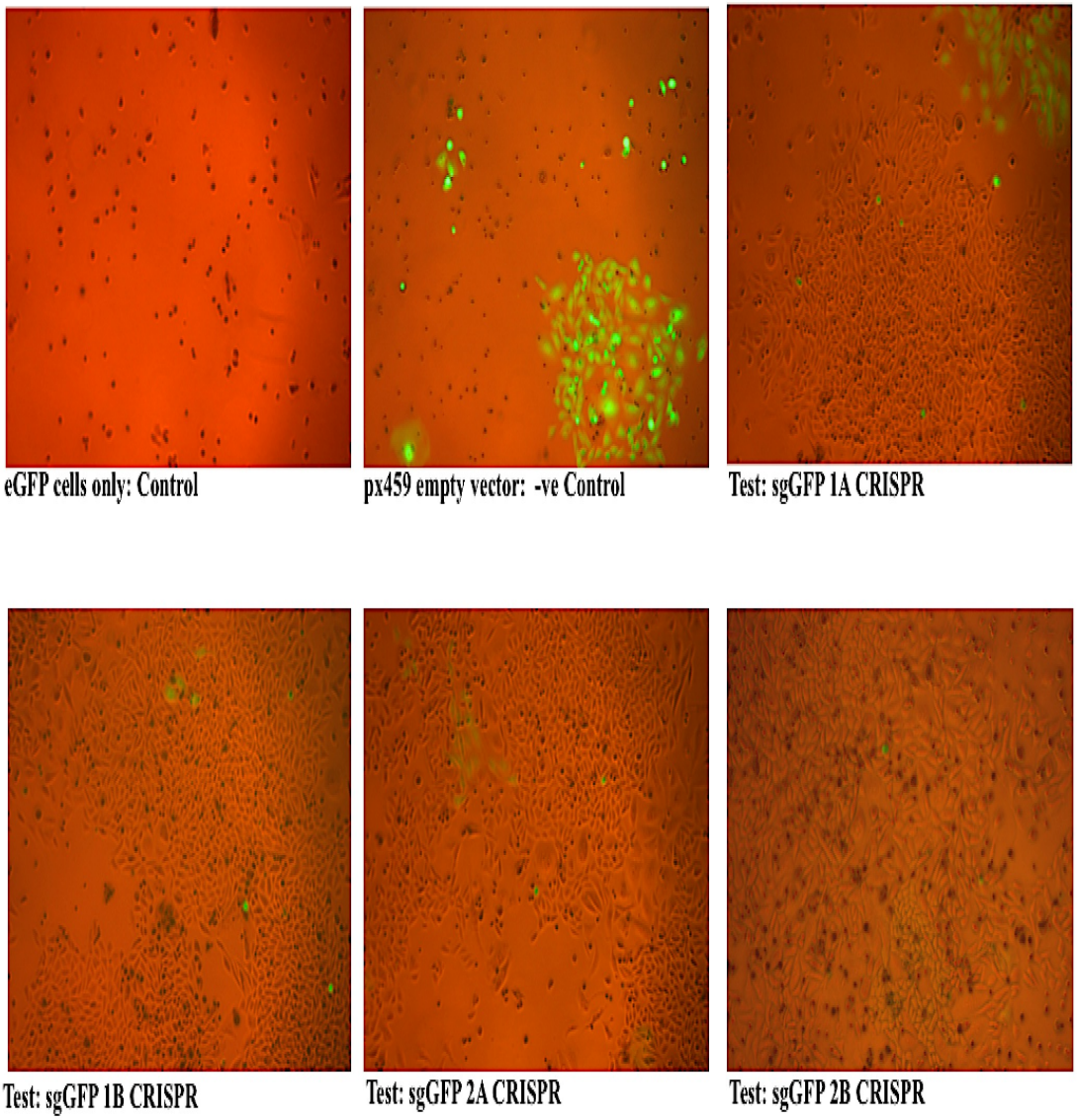


Figure 3.4.7A: Fluorescent Microscopy images post transfection in eGFP cells treated with control px459 vector and eGFP CRISPRs. The top left image is the CHO-eGFP cells only control visualised post transfection showing dead non-transfected cells in the presence of puromycin. The px459 empty vector only control (negative (-ve) control) shows only a few CHO-eGFP green cells due to random area chosen for imaging. The (-ve) control showed 70 % eGFP positive cells. Cells were seeded at 2×10^5 cells/mL per well of a 6-well plate and images were taken post transfection using the fluorescent microscope.

In order to further confirm the visual evidence, we quantified mean GFP expression in each of eGFP CRISPR treated samples. We observed a dramatic drop in mean GFP expression and % GFP positive cells in the eGFP CRISPR treated samples in comparison to the control px459 vector only sample. The mean GFP drop in all the eGFP CRISPR treated samples was 25-fold (average mean GFP across four eGFP test samples was 82 FU) lower than the control px459 vector only treated sample (2282 FU) (**Fig. 3.4.7B**). Moreover, there was a huge drop in the % GFP positive cells in the control px459 vector only sample i.e. 68 % to 4.5 % (an average across all four eGFP CRISPR test samples) in eGFP CRISPR treated samples (**Fig. 3.4.7B**).

(B)

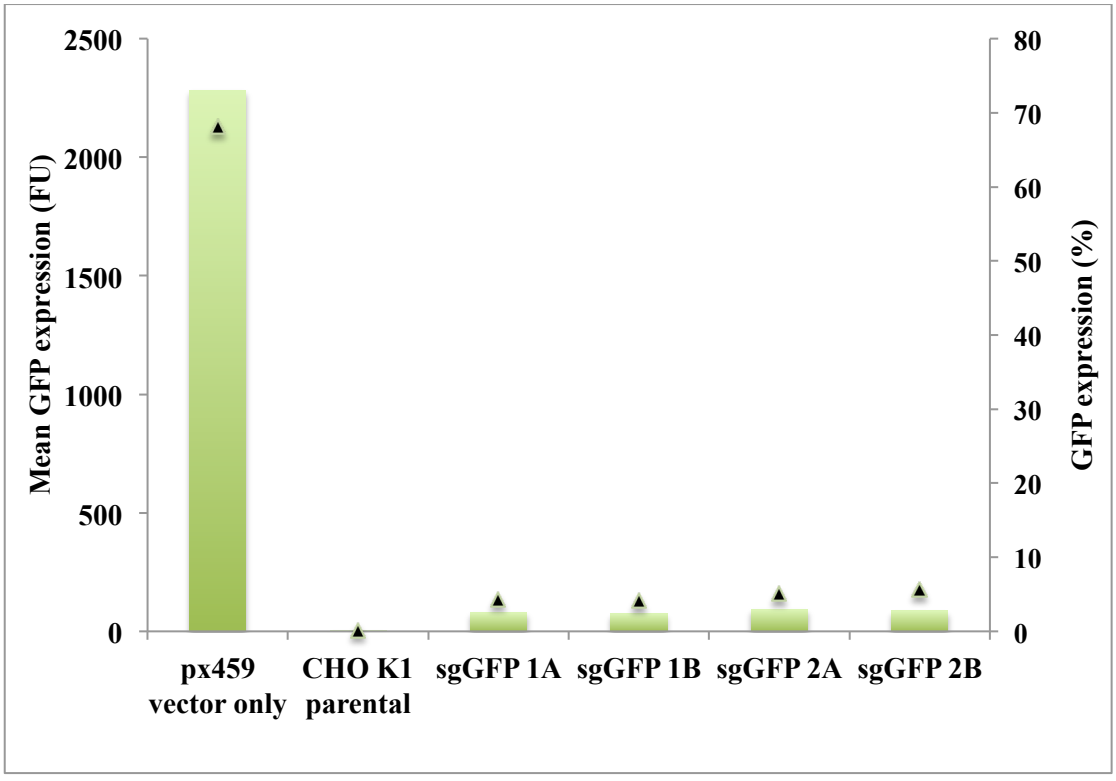
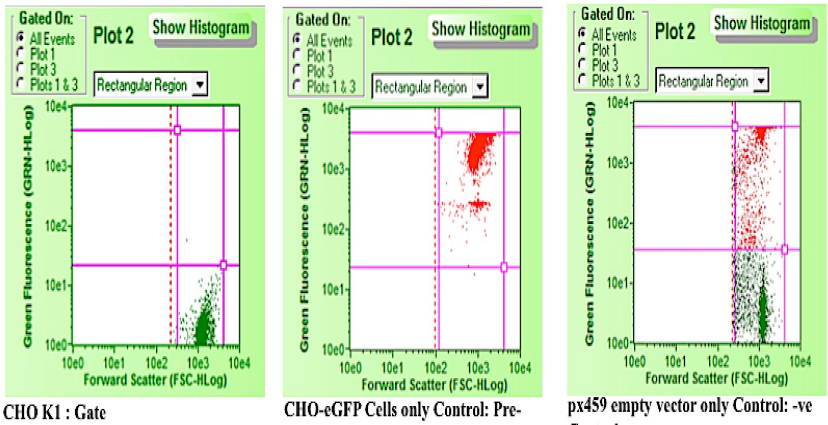


Figure 3.4.7B: Mean GFP expression (Bar Graphs) and GFP expression (%) (Δ Plots) analysed using Guava for control, non-GFP cell-line (CHO-K1) and test CRISPR samples. Cells were seeded at 2 × 10^5 cells/mL per well of a 6-well plate and were analysed for GFP expression using the Guava. An eGFP negative (i.e. CHO-K1 parental) cell line was used to gate the instrument prior to GFP expression analysis.

In **Fig. 3.4.7C**, it can be seen that the px459 vector only control has a cluster of cells in the GFP positive rectangular region, however, there were also cells in the GFP negative quadrant as well. These events could possibly be dead non-transfected eGFP cells. In contrast to control px459 only, all the eGFP CRISPR treated samples show the bulk of the cells clustered in the eGFP negative quadrant. However, eGFP positive events were also acquired representing a puromycin resistant population, which was also evident from the fluorescent microscopy results. The aim of this study was to establish the proof-of-concept that CRISPR-Cas9 system can efficiently and specifically knock out a stably integrated eGFP gene in CHO genome. All the above findings also demonstrated that using either a 19 nt or 20 nt eGFP sgRNA did not impact upon the efficiency of CRISPR induced eGFP knockouts since mean GFP expression and % GFP drop was observed to be similar in all eGFP CRISPR treated samples.

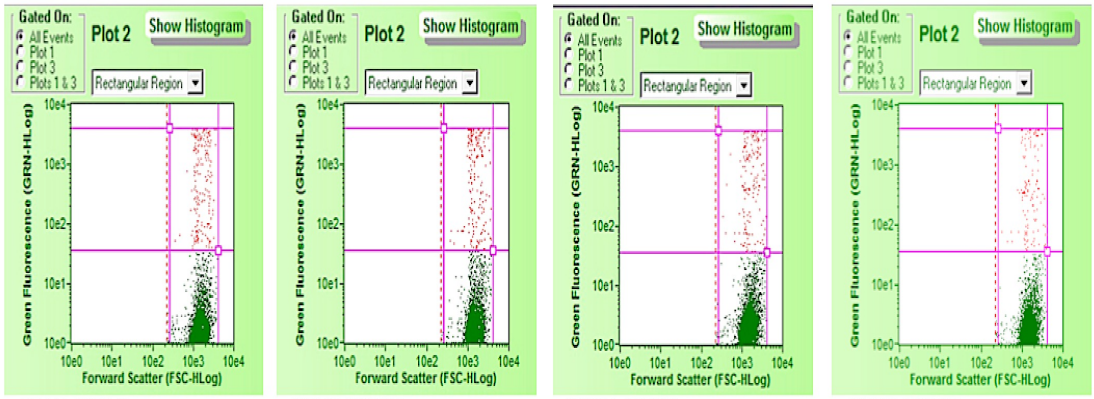
(C)



CHO K1 : Gate

CHO-eGFP Cells only Control: Pre-transfection

px459 empty vector only Control: -ve Control



sgGFP 1A CRISPR

sgGFP 1B CRISPR

sgGFP 2A CRISPR

sgGFP 2B CRISPR

Figure 3.4.7C: Mean GFP expression imaged in the form of dot-plots. Screenshots for ExpressPlus programme representing clusters of GFP positive (Red) and negative (Green) cells in control and test eGFP CRISPR treated samples. Top left screenshot shows the cluster of GFP negative (Green) CHO-K1 cells used to gate the instrument prior to GFP expression analysis. Cells were seeded at 2×10^5 cells/mL per well of a 6-well plate and then expanded to T75 flasks.

3.4.3.2 High indel frequency obtained by all four eGFP targeting sgRNAs

In order to assess the % efficiency of indels induced in the eGFP gene, we performed a Surveyor Assay. This assay is an easy, straightforward and highly sensitive assay to assess for any mutations, including point mutations, in the sequence of interest. This assay uses a highly mismatch specific endonuclease enzyme that scans through the sequence of interest and cleaves the sequence upon mutation detection with respect to a reference DNA sequence (or the wild-type DNA). The cleaved fragments can be visualized on higher resolution agarose or polyacrylamide gels. The methodology of the assay is explained in detail in **Section 2.8.2.8**.

We analysed the genomic DNA from the four eGFP CRISPR transfected populations in order to assess the % indels efficiency attained. For this purpose, we extracted the genomic DNA from the four eGFP CRISPR transfected populations, as well as the genomic DNA from the control px459 vector only cells, according to protocol described in **Section 2.8.3.1**. We then PCR amplified the eGFP (~ 800 bp) DNA fragment-of-interest using flanking primers (refer to **Table 5.1** in **Appendix Section**) from the genomic DNA (as template) extracted from all the four CRISPR and px459 only transfected eGFP cells, according to **Section 2.7.4**. The PCR fragments were then purified (refer to **Section 2.8.3.3**) before performing the Surveyor Assay. Homo/Hetero-duplexes of the reference DNA (i.e. eGFP PCR amplicon from px459 vector only eGFP cells) and test DNA (i.e. eGFP PCR amplicons from CRISPR transfected cells) were formed using equal amounts of DNA, separately (refer to **Section 2.8.2.8**). Finally, the Surveyor assay, whereby the homo/hetero-duplex eGFP fragments from control and test DNA(s) were cleaved by the Nuclease S enzyme and the cleaved fragments were visualized on a 1.2 % high-resolution agarose gel (**Fig. 3.4.8**).

The cleavage intensity of the bands was estimated using Ran et al. (2013b) methodology, wherein the integrated intensity of the PCR amplicon and cleaved bands was measured using *GelAnalyzer* software. For instance, for each lane, the fraction of the PCR product cleaved (f_{cut}) by the nuclease was calculated using the formula: $f_{cut} = (b+c)/(a+b+c)$, where a is the integrated intensity of the undigested PCR product and b and c are the integrated intensities of each cleavage product. Finally, a formula based on the binomial probability distribution of duplex formation

was used to calculate the indel occurrence/ indel (%).

$$Indel\ (\%) = 100 \times (1 - \sqrt{(1 - f_{cut})})$$

We observed that the sgGFP 1A and sgGFP 1B gave rise to the highest (50 %) indel frequency, while sgGFP 2A and sgGFP 2B demonstrated (40 %) indel frequency (**Fig. 3.4.8**).

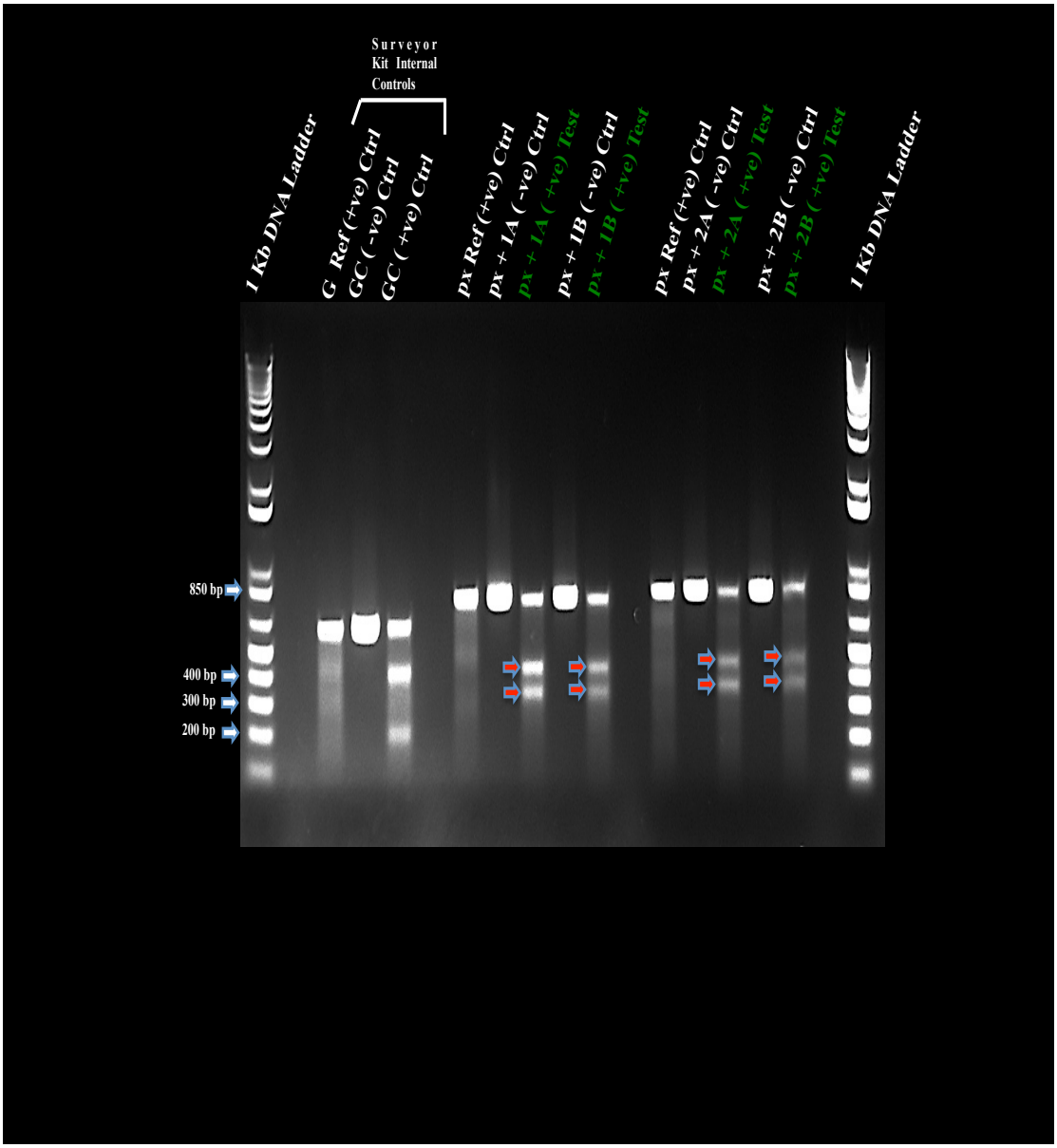


Figure 3.4.8: Agarose gel image post Surveyor Assay. The +ve and -ve stands for samples with Nuclease S and samples without Nuclease S enzyme, respectively. px → Reference DNA Control, (px + 1A) → Homo/Heteroduplex mix formed using PCR amplicons from reference px459 and sgGFP 1A, (px + 1B) → Homo/Heteroduplex mix formed using PCR amplicons from reference px459 and sgGFP 1B, (px + 2A) → Homo/Heteroduplex mix formed using PCR amplicons from reference px459 and sgGFP 2A and (px + 2B) → Homo/Heteroduplex mix formed using PCR amplicons from reference px459 and sgGFP 2B. All Control samples are highlighted with White font, while Green labeled samples are Test eGFP CRISPR samples. Red Arrow Heads represent the bands for the cleaved mutated fragments. The Surveyor kit internal controls were also ran on the gel.

3.4.3.3 Analysis of the nature of indels induced in the eGFP gene

To further assess the nature and frequency of the indels attained with eGFP sgRNAs, TOPO^{TM}-TA cloning (refer to **Section 2.8.2.7**) and sequencing of the gel-purified amplicons from the eGFP genomic site was performed (**Fig. 3.4.9**). The PCR amplicons covering the sgRNA-target site i.e. the eGFP gene fragment (~ 800 bp) were TOPO^{TM} TA-cloned and transformed into DH5α cells (refer to **Section 2.8.2.6**). Ten single colonies for each of the eGFP CRISPRs transformant containing plates were sent for Sanger sequencing to check the frequency and types of genomic (micro) deletion or inversion, indel mutations created by the NHEJ repair mechanism (**Fig. 3.4.9**). We observed almost all the eGFP CRISPRs induced different types of indel mutations, some of them were large deletions in the eGFP sequence and a few were 1 bp or 2 bp insertions at the eGFP target site. We observed that the ten TOPO-cloned samples for the sgGFP 1A CRISPR mostly contained deletions at the target site in the eGFP gene, some were 1-3 bp long, other were 12 bp, 15 bp and 27 bp. In addition to deletions, two samples contained indels, with 73 bp (69 random bp insertion, 4 bp deletion) and large 219 bp (204 random bp insertion, 15 bp deletion) downstream of the PAM sequence (**Fig. 3.4.9**). The majority of indels in sgGFP 1B CRISPR TOPO-cloned samples were > 10 bp. sgGFP 1B CRISPR resulted in eight positives for mutations induced at the eGFP target site, while two sequencing results were reported negative i.e. no indels observed. Similar to sgGFP 1A, 8 sgGFP 1B CRISPR TOPO-cloned samples mostly contained big deletions ranging from 13 bp to 27 bp long, with one TOPO-cloned sgGFP 1B sample resulting a 2 bp insertion at the target eGFP site. Only the sgGFP 1B CRISPR TOPO-cloned samples showed consistent large deletions in all the positive samples (**Fig. 3.4.9**).

All four eGFP CRISPRs resulted > 40 % indel efficiencies, however, as previously observed in the Surveyor assay the eGFP CRISPR (i.e. sgGFP 2A and sgGFP 2B CRISPRs targeting the 3'-end of the eGFP gene sequence resulted in ≤ 40 % indel frequencies (**Fig. 3.4.8**). Therefore, it was interesting to observe the types and frequencies of mutations induced by these eGFP CRISPRs. We observed that both the sgGFP 2A and sgGFP 2B CRISPR TOPO-cloned samples resulted in 1 or 2 bp insertion and/or deletions at the eGFP 3'-end target site. Two TOPO-cloned samples from each sgGFP 2A and sgGFP 2B CRISPR were negative i.e. no indels observed,

while eight samples for each were positive for indels. Of the eight positive indel samples, two samples sgGFP 2A CRISPR resulted 26 bp long sequence deleted at the target site and most of the bases downstream of the PAM sequence were deleted. Moreover, one sample showed 133 bp deleted in the GFP sequence with indels (118 bp deletions and 15 bp random sequence insertion) (**Fig. 3.4.9**). Similar sequencing results for sgGFP 2B CRISPR TOPO-cloned samples were observed, however, with sgGFP 2B CRISPR 1 or 2 bp insertion were reported more often than 1 or 2 bp deletion (**Fig. 3.4.9**). In addition to small indels, one sgGFP 2A CRISPR TOPO-cloned sample resulted in 31 bp indels, while another had 102 bp deletion downstream of the PAM sequence (**Fig. 3.4.9**).

1. sgGFP 1A CRISPR or eGFP 5′-end targeting 19 bp eGFP in CHO-eGFP genome (red underlined 19 bp target sequence & NGG blue-> PAM).

5′- ATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTT…..-3′ GFP1A

5′- ATGGTGAGCAAGGGCGAGGAGCTG - TCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTT….-3′ (1bp del)

5′- ATGGTGAGCAAGGGCGAGGAGCTG- - -ACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTT ….-3′ (3bp del)

5′- ATGGTGAGCAAGGGCGAGG - - - - - - - - - - - - - -TGGTGCCCATCCTGGTCGAGCTGGGAGCAAGGGCGAGGAGCGCGATAGCGG…-3′(15+204bp indel)

5′- ATGGTGAGCAAGGGCGAGGAGCTGTTC - - - - GGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTT….-3′ (4bp del)

5′- ATGGTGAGCAAGGGCGAGGAGCT - - - - ACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTT…..-3′(4bp del)

5′- ACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTT….-3′ (27bp del)

5′- ATGGTGAGCAAGGGCGAGGAGCT - - - - ACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTT….-3′ (4 bp del)

5′- ATGGTGAGCAAGGGCGAGGAGCTGTT - - - - GGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAG..-3′ (4 + 69bp indel)

5′- ATGGTGAGCAAGGGCGAGGAGC - - - TCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTT…-3′ (3 bp del)

5′- ATGGTGAGCAAGGGCGAGGAGC - - - - - - - - - - - - TGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTT…-3′ (12 bp del)

2.sgGFP 1B CRISPR or eGFP 5'-end targeting 20 bp eGFP in CHO-eGFP genome (red underlined 20 bp target sequence & NGG blue-> PAM).

5'- ATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTT…..-3' GFP1B

5'- ATGGTGAGCAAGGGCGA - - - - - - - - - - - - - - - - - - GGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTT…..-3' (18bp del)

5'- ACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTT..-3'(27bp del)

5'- ATGGTGAGCAAGGGCGAGGAG - - - - - - - - - - - - - - - - CCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTT ….-3' (19bp del)

5'- ATGGTGAGCAAGGGCGAGGAGCTGT - - - - - - - - - - - - - - - - - - - TCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTT ….-3' (18bp del)

5'- ATGGTGAGCAAGGGCGA - - - - - - - - - - - - - - GGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTT….-3' (13 bp del)

5'- ATGGTGAGCAAGGGCGAGGAGCTGTTCTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAA..-3'(2 bp insertion)

5'- ACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTT..-3' (27bp del)

5'- ATGGTGAGCAAGGG - - - - - - - - - - - - - - - CGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTT…3' (15 bp del)

5'- ATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTT ….-3' (NO indels)

5'- ATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTT ….-3' (NO indels)

3. sgGFP 2A CRISPR or eGFP 3'-end targeting 19 bp eGFP in CHO-eGFP genome (red underlined 19 bp target sequence & NGG blue-> PAM).

5'-……GAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGTAG -3' GFP 2A

5'- ……GAAGCGCGATCACATGGTC – TGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGTAG-3' (1 bp del)

5'- ……GAAGCGCGATCACATGGTC – TGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGTAG-3' (1 bp del)

5'- ……GAAGCGCGATCACATGGTCC – GCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGTAG-3' (1 bp del)

5'- …----------TTTGGCATGGACGAGCTGTACAAGTAG-3' (118 + 15 bp indels)

5'- ……GAAGCGCGATCA - - - - - - - CTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGTAG-3' (7 bp del)

5'- ……GAAGCGCGATCACA – G - - CC – G - - GGA - T – C- - -ACC -- --- ---- -- TC - -T –T – GGCATGGACGAGCTGTACAAGTAG-3' (26 bp del)

5'- ……GAAGCGCGATCACATGGTCCTGCTGGAGTTCGAAGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGTAG-3' (1 bp insertion)

5'- ……GAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTTTGGCATGGACGAGCTGTACAAGTAG-3' (1 bp insertion)

5'- ……GAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTTGGCATGGACGAGCTGTACAAGTAG-3' (NO indels)

5'- ……GAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTTGGCATGGACGAGCTGTACAAGTAG-3' (NO indels)

Figure 3.4.9: Analysis of indels in eGFP gene of the CHO-eGFP cells. TOPO^{TM} TA – based sequence analysis of eGFP was performed and compared with the reference (or wild-type) eGFP sequence (Top sequence in every panel), with the CRISPR target site Underlined and the PAM site highlighted in Blue. 1 shows the indels generated using sgGFP 1A CRISPR, 2 shows indels generated using sgGFP 1B CRISPR, 3 shows indels generated using sgGFP 2A CRISPR, while 4 shows indels generated using sgGFP 2B CRISPR. Dashed Red line shows the bases deleted from the eGFP sequence (numbers of bases deleted in Red), while Green bases and/or Dashed Green line shows the bases inserted (numbers of bases inserted in Green).

In summary, all four eGFP CRISPR demonstrated significant indel frequencies, with sgGFP 1A (19 nt guide) and sgGFP 1B (20 nt guide) CRISPRs targeting the 5'-end of the eGFP gene, demonstrating highest ~ 50 % indel efficiencies in both. Moreover, majority (more than 60 %) of indels were > 10 bp i.e. 11 out of 18 positive TOPO-cloned sgGFP 1A and sgGFP 1B samples collectively. On the other hand, sgGFP 2A and sgGFP 2B CRISPRs targeting the eGFP 3'-end of the eGFP gene generated ≤ 40 % indels only. Moreover, it was observed that sgGFP 2A and sgGFP 2B CRISPRs induced 1 or 2 bp indels more often (~ 57 %) than big indels (i.e. 9 out of 16 positive TOPO-cloned sgGFP 2A and sgGFP 2B samples collectively, yielded 1 or 2 bp only indels). Only (25 %) 4 out of 16 positive TOPO-cloned eGFP 2A and eGFP 2B

samples collectively, yielded > 20 bp indels (**Fig. 3.4.9**).

3.5 Applying the CRISPR-Cas9 system to target the miR-7 locus in an industrially relevant CHO cell line

There have been several studies published in the literature, wherein the potential of manipulating miR-7 levels is associated with enhancing growth and productivity phenotypes. For instance, lately Sanchez et al. (2014) reported 3-fold improved CHO cell proliferation, ~ 30 % cell viability increase and 2-fold protein productivity increase by using a loss-of-function approach to deplete endogenous miR-7 levels. Therefore, with this in mind, we wished to explore this CRISPR-Cas9 genome modification approach to knockout the miR-7a-5p sequence from the CHO-K1 1.14 cell line. CHO-K1 1.14 cell-line is a monoclonal antibody producing cell line (henceforth referred as 1.14).

3.5.1 Designing and cloning of miR-7 sgRNA into the px459 CRISPR vector

The sgRNA sequences for a miR-7 targeting CRISPR-Cas9 system were designed using the CRISPR Design Tool (http://tools.genome-engineering.org) (**Table 3.5.1**). The miR7 1 and miR7 2 CRISPRs differ in that miR7 1 targets the miR-7a-5p sequence upstream of the mature miR-7 sequence, while miR7 2 targets the miR-7a-5p sequence downstream of the mature miR-7, loop and star-strand sequence. A and B refer to the length of the guide sequence, wherein A is 19 nt long sgRNA and B is 20 nt long sgRNA oligo sequences annealed and ligated into the *BbsI* linearized px459 vector (refer to **Section 2.8.2.1 and Section 2.8.2.4**). We designed four sgRNAs targeting two different sites of the miR-7a-5p sequence in 1.14 cells (**Fig. 3.5**). The resulting constructs will be referred as sgmiR7 1A, sgmiR7 1B, sgmiR7 2A and sgmiR7 2B CRISPRs. This allowed us to investigate the CRISPR-Cas9 efficiency of using different sgRNAs targeting different sites in the gene.

Table 3.5.1 List of miR7 sgRNAs used for targeting cgr-miR-7a-5p sequence. sgRNA sequences (A →19nt or B →20 nt Highlighted in Yellow). Only sense strands are shown. The *BbsI* site (*cacc*) is also included at the 5'-end of the sequence for ligation into px459 vector. Red highlighted G appended at 5'-end (as preferred base) is the first base of the guide sequence. NGG is PAM sequence in the CHO genome.

CRISPR-Cas9 constructs	sgRNA sense-strand sequences (or guide sequences)	PAM (5'-end upstream of guide sequence)
sgmiR7 1A	5' – CACCGTGGACAGGCCAGCCCCGCC – 3'	TGG
sgmiR7 1B	5' – CACCGATGGACAGGCCAGCCCCGCC– 3'	TGG
sgmiR7 2A	5' – CACCGGCGGGTCCTCTGAGCATCA – 3'	GGG
sgmiR7 2B	5' – CACCGGCGGGTCCTCTGAGCATCA – 3'	GGG

The sgRNAs for miR7 1 and miR7 2 were cloned into the px459 vector as described previously in **Section 3.4.3**.

cgr-miR-7a-5p locus sense

GGTGGTGGCCAAGGCCCAGAGGAGGTGGTTAGCAAGGCCATGGACAGGC
CAGCCCCGCCTGGAAGACTAGTGATTTTGTTGTTGTCTCTCTATACTCAAC
AACAAGTCCCAGTCTACCACACGGTGCGGGTCCTCTGAGCATCAGGGGGA
GGTATGGGAGCACCATTTCTGGATGTATTTTCTCTGCCTTCTATGTTCCCTA
GCAAACTACCAAAATATTCCTC

Figure 3.5: Sense cgr-miR-7a-5p locus sequence highlighting in Blue the two miR7 1 and miR7 2 CRISPR target sites upstream and downstream of the mature miR-7a sequence. Sequence highlighted in Cyan represents the miR-7 loop sequence and the Underlined sequence represents the Star strand sequence. TGG and GGG represent the PAM at the 3'-end of miR7 CRISPR sites 1 and 2, respectively. ~~Strike-through three bases~~ upstream and downstream of mature miR-7 sequence represents Cas9 Cleavage sites.

3.5.2 Transfection of px459 miR7 CRISPR constructs in 1.14 cells

The results in **Section 3.4.3.2** and **3.4.3.3** did not assist in the choosing of preferred type of sgRNA to be used for targeting any other gene sequence. For this reason, we employed all four miR7 CRISPRs to disrupt the miR-7a-5p sequence. The sgRNA designed to target the 5'-end of the eGFP gene performed marginally better than sgRNA targeting the 3'-end of the eGFP gene in inducing indels. However, the fact that targeting an ORF of gene encoding protein is easier since only one bp knock out is enough to disrupt the reading frame of eGFP gene, could not be excluded. For non-coding gene like in the case of miR-7 (with two alleles located on two different chromosomal loci), where there are possibly four copies of the gene with functional redundancy due to paralogs, we assumed that it would be a harder task to knock out all the four copies. Moreover, in order to render miRNAs functionless, we required significant deletions in the sequence as opposed to single base indels.

We transfected 5 wells of a 6-well plate containing 1.14 cells seeded at 5×10^5 cells/mL cell density one day before transfection with 1 μg of control px459 vector only (i.e. no guide sequence), sgmiR7 1A, sgmiR7 1B, sgmiR7 2A and sgmiR7 2B CRISPR DNA (s) (refer to **Section 2.6.1.2**). Three days post transfection, we observed ~ 65-70 % viable cells in the px459 only control well of the 6-well plate, on the other hand, 1.14 cells-only control contained only dead cells (data not shown). Similar percentages of viable cells were observed in our test CRISPR wells of the 6-well plate. We then expanded each well of the 6-well plate to T25 flasks and selected cells for another week in order to extract sufficient amount of genomic DNA for further analysis. During this time-period, cells were monitored every day and had their media replenished routinely and were not under any stress.

3.5.3 High indel frequency attained by all miR-7a-5p locus targeting miR7 CRISPRs

A similar workflow as used for analysing eGFP indels was followed to analyse indels induced by miR7 CRISPRs in 1.14 cells. To assess the % efficiency of targeting the miR-7a-5p locus, we performed a Surveyor Assay. The methodology of the assay is explained in detail in **Section 2.8.2.8**.

We analysed the genomic DNA from the four miR7 CRISPR transfected populations in order to assess the % indels efficiency attained. For this purpose, we extracted the genomic DNA from the four miR7 CRISPR transfected populations, as well as the genomic DNA from the control px459 vector only cells. Following this, we PCR amplified the miR7 (~ 855 bp) DNA fragment-of-interest using flanking primers (designed ~ 400 bp upstream and 230 bp downstream of the 224 bp miR-7a-5p sequence) (refer to **Table 5.2** in **Appendix Section**) from genomic DNA (as template). The PCR fragments were then purified before performing the Surveyor assay. Homo/Hetero-duplexes of the reference DNA (i.e. miR7 PCR amplicon from control px459 vector only treated sample) and test DNA (i.e. miR7 PCR amplicons from CRISPR treated samples) were formed using equal amounts of DNA in a thermocycler. Finally, the Surveyor assay, whereby the homo/hetero-duplex miR7 fragments from control and test DNA(s) were cleaved by the Nuclease S enzyme and the cleaved fragments were visualized on a 1.2 % agarose gel (**Fig. 3.5.1**). The cleavage intensity of the bands was estimated as described previously in **Section 3.4.3.2**.

We observed that the sgmiR7 1A and sgmiR7 1B gave rise to (35 %) indel frequency, while sgmiR7 2A and sgmiR7 2B demonstrated (40 %) indel frequency. In summary, we estimated ~ 35-40 % indel frequencies with each of miR7 CRISPR in the miR-7a-5p sequence of CHO-K1 1.14 cells (**Fig. 3.5.1**).

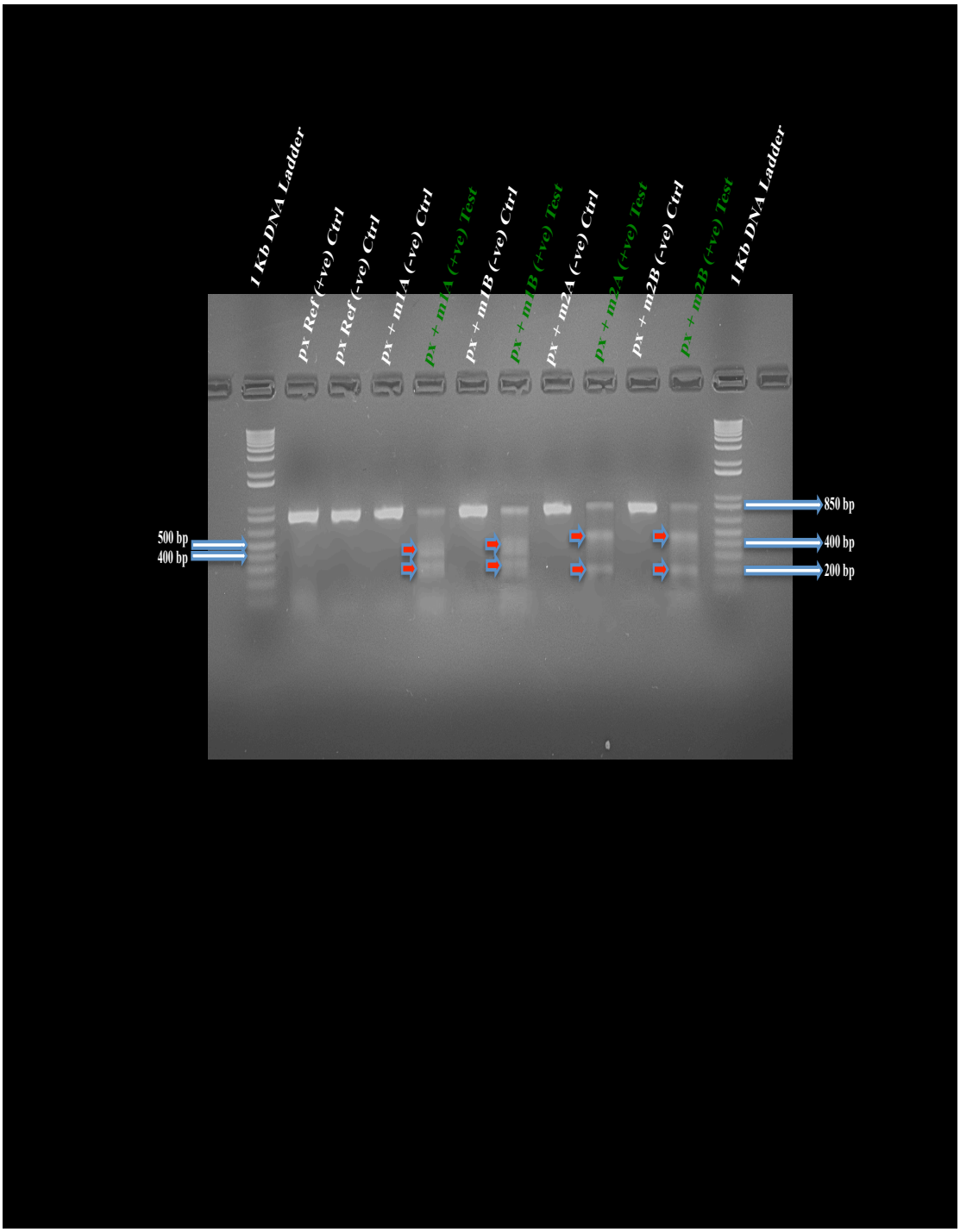


Figure 3.5.1: Agarose gel image post Surveyor Assay. The +ve and -ve stands for samples with Nuclease S and samples without Nuclease S enzyme, respectively. px → Reference DNA Control, (px + m1A) → Homo/Heteroduplex mix formed using PCR amplicons from reference px459 and sgmiR7 1A, (px + m1B) → Homo/Heteroduplex mix formed using PCR amplicons from reference px459 and sgmiR7 1B, (px + m2A) → Homo/Heteroduplex mix formed using PCR amplicons from reference px459 and sgmiR7 2A and (px + m2B) → Homo/Heteroduplex mix formed using PCR amplicons from reference px459 and sgmiR7 2B. All Control samples are highlighted with White font, while Green labeled samples are Test eGFP CRISPR samples. Red Arrow Heads represent the bands for the cleaved mutated fragments.

3.5.4 Analysis of the nature of indels induced at the miR-7 locus

To further assess the indel type and frequency attained with miR7 sgRNAs, TOPO^{TM}-TA cloning and sequencing of the gel-purified amplicons from the miR7 genomic site was performed (**Fig. 3.5.2**). The PCR amplicons covering the sgRNA-target site i.e. the miR-7a-5p flanking gene fragment (~ 855 bp long) were TOPO^{TM} TA-cloned and transformed into DH5α cells. Ten single colonies for each of the miR7 CRISPR transformant containing plates were sent for Sanger sequencing to check the frequency and types of genomic (micro) deletion or inversion, indel mutations created by the NHEJ repair mechanism (**Fig. 3.5.2**). We observed that most of the miR7 CRISPRs generated 1 bp or 2 bp insertion. 80 % of the sgmiR7 2A CRISPR TOPO-cloned samples generated 1 or 2 bp indels (**Fig. 3.5.2**). Similar to sgmiR7 2A, 60 % of the sgmiR7 1A CRISPR TOPO-cloned samples generated 1 or 2 bp indels, however, two sgmiR7 1A CRISPR TOPO-cloned samples showed big deletions of 18 bp and 96 bp (32 bp in miR-7a-5p sequence and 64 bp upstream of miR-7a-5p sequence deleted) each. Therefore, 20 % of the 10 sgmiR7 1A CRISPR TOPO-cloned samples sent for sequencing showed big deletion (**Fig 3.5.2**). The similarity in sgmiR7 1A and sgmiR7 2A guides was the length of the sgRNA, which was 19 nt. Therefore, we could infer that the guide-length here made the difference in miR7 CRISPR activities in generating the magnitude of indels observed. Based on our previous observation in **Section 3.4.3.2**, we hypothesized that the sgRNA with guide length 20 nt long should be more efficient than sgRNA 19 nt long based on results from our 10 TOPO samples for each type of eGFP CRISPR. We observed that sgmiR7 1B CRISPR also generated mostly of 1 or 2 bp deletion, however, less frequently. 80 % of TOPO-cloned sgmiR7 1B samples contained indel sizes in the range of 3 -10 bp. In addition to small indels, two of the TOPO-cloned sgmiR7 1B samples contained big deletions of 25 bp and 175 bp (knocking out the mature miR-7a-5p sequence, loop an star strand sequence downstream of the miR7 1 CRISPR target site) (**Fig. 3.5.2**). The sgmiR7 2B CRISPR demonstrated the highest CRISPR efficiency in terms of generating big deletions in 30 % of the 10 TOPO-cloned 2B samples sent for sequencing. The size of the deletions observed were 127 bp (mature miR-7a-5p sequence along with the loop sequence deleted), 249 bp (89 bp deleted from the miR-7a-5p sequence downstream of the miR7 2 CRISPR target site and another 160 bp deleted downstream of the miR-7a-5p sequence) and 240 bp (169 bp knocked out of the miR-7a-5p sequence and

another 71 bp upstream of the miR-7a-5p sequence). Moreover, the frequency of 1 or 2 bp indels was the lowest i.e. 20 % of 10 TOPO-cloned sgmiR7 2B samples resulting in single base bp indels (**Fig. 3.5.2**). With these observations, we inferred that the location of the sequence against where the guide is designed to bind in relation to miR-7a-5p sequence is also critical, in addition to the sgRNA length, for efficient CRISPR-Cas9 activity.

1. miR-7a sequencing results for sgmiR7 1A CRISPR:

5’GGTAGTATCCAAGGATGATAGACAGGTGGTGGCCAAGGCCCAGAGCCAGGCCTGGTCAGCCTGGAAGACTAGTGATTTTGTTGTTGTCTCTCTATACTCAACAACAAGTCCCAGTCTACCACACGGTGCGGGTCCTCTGAGCATCAGGGGGAGGTATGGGAGCACCATTTCTGGATGTATTTTCTCTGCCTTCTATGTTCCCTAGCAAACTACCAAAATATTCCTC3’ (18 bp mm + 1 bp in + 6 bp del & and 1 bp insertion, 9 bp deletion and 16 bp few bp mismatch mutations upstream of miR-7a-5p sequence).

5’GGTGGTGGCCAAGGCCCAGAGGAGGTGGTTAGCAAGGCCATGGACAGGCCAGCCCCCGCCTGGAAGACTAGTGATTTTGTTGTTGTCTCTCTATACTCAACAACAAGTCCCAGTCTACCACACGGTGCGGGTCCTCTGAGCATCAGGGGGAGGTATGGGAGCACCATTTCTGGATGTATTTTCTCTGCCTTCTATGTTCCCTAGCAAACTACCAAAATATTCCTC3’ (1 bp insertion)

5’GGTGGTGGCCAAGGCCCAGAGGAGGTGGTTAGCAAGGCCATGGACAGGCCAGCCCCGCCTGGAAGACTAGTGATTTTGTTGTTGTCTCTCTATACTCAACAACAAGTCCCAGTCTACCACACGGTGCGGGTCCTCTGAGCATCAGGGGGAGGTATGGGAGCACCATTTCTGGATGTATTTTCTCTGCCTTCTATGTTCCCTAGCAAACTACCAAAATATTCCTC3’ (5 bp deletion)

5’GGTGGTGGCCAAGGCCCAGAGGAGGTGGTTAGCAAGGCCATGGACAGGCCAGCCCCCGCCTGGAAGACTAGTGATTTTGTTGTTGTCTCTCTATACTCAACAACAAGTCCCAGTCTACCACACGGTGCGGGTCCTCTGAGCATCAGGGGGAGGTATGGGAGCACCATTTCTGGATGTATTTTCTCTGCCTTCTATGTTCCCTAGCAAACTACCAAAATATTCCTC3’ (1 bp insertion)

5’GGTGGTGGCCAAGGCCCAGAGGAGGTGGTTAGCAAGGCCATGGACAGGCCAGCCCCCGCCTGGAAGACTAGTGATTTTGTTGTTGTCTCTCTATACTCAACAACAAGTCCCAGTCTACCACACGGTGCGGGTCCTCTGAGCATCAGGGGGAGGTATGGGAGCACCATTTCTGGATGTATTTTCTCTGCCTTCTATGTTCCCTAGCAAACTACCAAAATATTCCTC3’ (1 bp insertion)

5’GGTGGTGGCCAAGGCCCAGAGGAGGTGGTTAGCAAGGCCATGGACAGGCCAGCCCCCGCCTGGAAGACTAGTGATTTTGTTGTTGTCTCTCTATACTCAACAACAAGTCCCAGTCTACCACACGGTGCGGGTCCTCTGAGCATCAGGGGGAGGTATGGGAGCACCATTTCTGGATGTATTTTCTCTGCCTTCTATGTTCCCTAGCAAACTACCAAAATATTCCTC3’ (1 bp insertion)

5’GGTGGTGGCCAAGGCCCAGAGGAGGTGGTTAGCAAGGCCATGGACAGGCCAGCCCCCGCCTGGAAGACTAGTGATTTCGTTGTTGTCTCTCTATACTCAACAACAAGTCCCAGTCTACCACACGGTGCGGGTCCTCTGAGCATCAGGGGGAGGTATGGGAGCACCATTTCTGGATGTATTTTCTCTGCCTTCTATGTTCCCTAGCAAACTACCAAAATATTCCTC3’ (1 bp insertion + 1 bp mm)

5’GGTGGTGGCCAAGGCCCAGAGGAGGTGGTTAGCAAGGCCATGGACAGGCCAGCCCCGCCTGGAAGACTAGTGATTTTGTTGTTGTCTCTCTATACTCAACAACAAGTCCCAGTCTACCACACGGTGCGGGTCCTCTGAGCATCAGGGGGAGGTATGGGAGCACCATTTCTGGATGTATTTTCTCTGCCTTCTATGTTCCCTAGCAAACTACCAAAATATTCCTC3’ (32 bp deletion + 64 bp deleted upstream of miR-7a-5p sequence)

5’GGTGGTGGCCAAGGCCCAGAGGAGGTGGTTAGCAAGGCCATGGACAGGCCAGCCCCCCGCCTGGAAGACTAGTGATTTTGTTGTTGTCTCTCTATACTCAACAACAAGTCCCAGTCTACCACACGGTGCGGGTCCTCTGAGCATCAGGGGGAGGTATGGGAGCACCATTTCTGGATGTATTTTCTCTGCCTTCTATGTTCCCTAGCAAACTACCAAAATATTCCTC3’ (2 bp insertion)

5’GGTGGTGGCCAAGGCCCAGAGGAGGTGGTTAGCAAGGCCATGGACAGGCCAGCCCCGCCTGGAAGACTAGTGATTTTGTTGTTGTCTCTCTATACTCAACAACAAGTCCCAGTCTACCACACGGTGCGGGTCCTCTGAGCATCAGGGGGAGGTATGGGAGCACCATTTCTGGATGTATTTTCTCTGCCTTCTATGTTCCCTAGCAAACTACCAAAATATTCCTC3’ (18 bp deletion)

2. miR-7a sequencing results for sgmiR7 1B CRISPR:

5’GGTGGTGGCCAAGGCCCAGAGGAGGTGGTTAGCAAGGCCATGGACAGGCCAGCCCCGCCTGGAAGACTAGTGATTTTGTTGTTGTCTCTCTATACTCAACAACAAGTCCCAGTCTACCACACGGTGCGGGTCCTCTGAGCATCAGGGGGAGGTATGGGAGCACCATTTCTGGATGTATTTTCTCTGCCTTCTATGTTCCCTAGCAAACTACCAAAATATTCCTC..-3’ (175 bp deletion + 2 bp downstream of miR-7a-5p sequence)

5’GGTGGTGGCCAAGGCCCAGAGGAGGTGGTTAGCAAGGCCATGGACAGGCCAGCCCCGCGCCTGGAAGACTAGTGATTTTGTTGTTGTCTCTCTATACTCAACAACAAGTCCCAGTCTACCACACGGTGCGGGTCCTCTGAGCATCAGGGGGAGGTATGGGAGCACCATTTCTGGATGTATTTTCTCTGCCTTCTATGTTCCCTAGCAAACTACCAAAATATTCCTC3’ (2 bp insertion)

5’GGTGGTGGCCAAGGCCCAGAGGAGGTGGTTAGCAAGGCCATGGACAGGCCAGCCCCGCCTGGAAGACTAGTGATTTTGTTGTTGTCTCTCTATACTCAACAACAAGTCCCAGTCTACCACACGGTGCGGGTCCTCTGAGCATCAGGGGGAGGTATGGGAGCACCATTTCTGGATGTATTTTCTCTGCCTTCTATGTTCCCTAGCAAACTACCAAAATATTCCTC3’ (1 bp deletion)

5’GGTGGTGGCCAAGGCCCAGAGGAGGTGGTTAGCAAGGCCATGGACAGGCCAGCCCCCGCCTGGAAGACTAGTGATTTTGTTGTTGTCTCTCTATACTCAACAACAAGTCCCAGTCTACCACACGGTGCGGGTCCTCTGAGCATCAGGGGGAGGTATGGGAGCACCATTTCTGGATGTATTTTCTCTGCCTTCTATGTTCCCTAGCAAACTACCAAAATATTCCTC3’ (1 bp insertion)

5’GGTGGTGGCCAAGGCCCAGAGGAGGTGGTTAGCAAGGCCATGGACAGGCCAGCCCCCGCCTGGAAGACTAGTGATTTTGTTGTTGTCTCTCTATACTCAACAACAAGTCCCAGTCTACCACACGGTGCGGGTCCTCTGAGCATCAGGGGGAGGTATGGGAGCACCATTTCTGGATGTATTTTCTCTGCCTTCTATGTTCCCTAGCAAACTACCAAAATATTCCTC3’ (1 bp insertion)

5’GGTGGTGGCCAAGGCCCAGAGGAGGTGGTTAGCAAGGCCATGGACAGGCCAGCCCCGCCTGGAAGACTAGTGATTTTGTTGTTGTCTCTCTATACTCAACAACAAGTCCCAGTCTACCACACGGTGCGGGTCCTCTGAGCATCAGGGGGAGGTATGGGAGCACCATTTCTGGATGTATTTTCTCTGCCTTCTATGTTCCCTAGCAAACTACCAAAATATTCCTC3’ (5 bp deletion)

5’GGTGGTGGCCAAGGCCCAGAGGAGGTGGTTAGCAAGGCCATGGACAGGCCAGCCCCCGCCTGGAAGACTAGTGATTTTGTTGTTGTCTCTCTATACTCAACAACAAGTCCCAGTCTACCACACGGTGCGGGTCCTCTGAGCATCAGGGGGAGGTATGGGAGCACCATTTCTGGATGTATTTTCTCTGCCTTCTATGTTCCCTAGCAAACTACCAAAATATTCCTC3’ (1 bp insertion)

5’GGTGGTGGCCAAGGCCCAGAGGAGGTGGTTAGCAAGGCCATGGACAGGCCAGCCCCGCCTGGAAGACTAGTGATGTTTGTTGTTGTCTCTCTATACTCAACAACAAGTCCCAGTCTACCACACGGTGCGGGTCCTCTGAGCATCAGGGGGAGGTATGGGAGCACCATTTCTGGATGTATTTTCTCTGCCTTCTATGTTCCCTAGCAAACTACCAAAATATTCCTC3’ (4 bp del + 1 bp insertion)

5’GGTGGTGGCCAAGGCCCAGAGGAGGTGGTTAGCAAGGCCATGGACAGGCCAGCCCCGCCTGGAAGACTAGTGATTTTGTTGTTGTCTCTCTATACTCAACAACAAGTCCCAGTCTACCACACGGTGCGGGTCCTCTGAGCATCAGGGGGAGGTATGGGAGCACCATTTCTGGATGTATTTTCTCTGCCTTCTATGTTCCCTAGCAAACTACCAAAATATTCCTC3’ (5 bp deletion)

5’GGTGGTGGCCAAGGCCCAGAGGAGGTGGTTAGCAAGGCCATGGACAGGCCAGCCCCGCCTGGAAGACTAGTGATTTTGTTGTTGTCTCTCTATACTCAACAACAAGTCCCAGTCTACCACACGGTGCGGGTCCTCTGAGCATCAGGGGGAGGTATGGGAGCACCATTTCTGGATGTATTTTCTCTGCCTTCTATGTTCCCTAGCAAACTACCAAAATATTCCTC3’ (25 bp deletion)

3. miR-7a sequencing results for sgmiR7 2A CRISPR:

5’GGTGGTGGCCAAGGCCCAGAGGAGGTGGTTAGCAAGGCCATGGACAGGCCAGCCCCGCCTGGAAGACTAGTGATTTTGTTGTTGTCTCTCTATACTCAACAACAAGTCCCAGTCTACCACACGGTGCGGGTCCTCTGAGCATCAGGGGGAGGTATGGGAGCACCATTTCTGGATGTATTTTCTCTGCCTTCTATGTTCCCTAGCAAACTACCAAAATATTCCTC3’ (3 bp deletion)

5’GGTGGTGGCCAAGGCCCAGAGGAGGTGGTTAGCAGAGGCCATGGACAGGCCAGCCCCGCCTGGAAGACTAGTGATTTTGTTGTTGTCTCTCTATACTCAACAACAAGTCCCAGTCTACCACACGGTGCGGGTCCTCTGAGCATCAGGGGGAGGTATGGGAGCACCATTTCTGGATGTATTTTCTCTGCCTTCTATGTTCCCTAGCAAACTACCAAAATATTCCTC3’ (1 bp deletion)

5’GGTGGTGGCCAAGGCCCAGAGGAGGTGGTTAGCAAGGCCATGGACAGGCCAGCCCCGCCTGGAAGACTAGTGATTTTGTTGTTGTCTCTCTATACTCAACAACAAGTCCCAGTCTACCACACGGTGCGGGTCCTCTGAGCAATCAGGGGGAGGTATGGGAGCACCATTTCTGGATGTATTTTCTCTGCCTTCTATGTTCCCTAGCAAACTACCAAAATATTCCTC3’ (1 bp insertion)

5’GGTGGTGGCCAAGGCCCAGAGGAGGTGGTTAGCAAGGCCATGGACAGGCCAGCCCCGCCTGGAAGACTAGTGATTTTGTTGTTGTCTCTCTATACTCAACAACAAGTCCCAGTCTACCACACGGTGCGGGTCCTCTGAGCAATCAGGGGGAGGTATGGGAGCACCATTTCTGGATGTATTTTCTCTGCCTTCTATGTTCCCTAGCAAACTACCAAAATATTCCTC3’ (1 bp insertion)

5’GGTGGTGGCCAAGGCCCAGAGGAGGTGGTTAGCAAGGCCATGGACAGGCCAGCCCCGCCTGGAAGACTAGTGATTTTGTTGTTGTCTCTCTATACTCAACAACAAGTCCCAGTCTACCACACGGTGCGGGTCCTCTGAGCAATCAGGGGGAGGTATGGGAGCACCATTTCTGGATGTATTTTCTCTGCCTTCTATGTTCCCTAGCAAACTACCAAAATATTCCTC3’ (1 bp insertion)

5’GGTGGTGGCCAAGGCCCAGAGGAGGTGGTTAGCAAGGCCATGGACAGGCCAGCCCCGCCTGGAAGACTAGTGATTTTGTTGTTGTCTCTCTATACTCAACAACAAGTCCCAGTCTACCACACGGTGCGGGTCCTCTGAGCATCAGGGGGAGGTATGGGAGCACCATTTCTGGATGTATTTTCTCTGCCTTCTATGTTCCCTAGCAAACTACCAAAATATTCCTC3’ (6 bp deletion)

5’GGTGGTGGCCAAGGCCCAGAGGAGGTGGTTAGAAAGGCTCTGGACAGGTCAGCCCTGCCTGGAAGACTAGTGATTTTGTTGTTGTCTCTCTATACTCAACAACAAGTCCCAGTCTACCACACGGTGCGGGTCCTCTGAGCATCAGGGGGAGGTATGGGAGCACCATTTCTGGATGTATTTTCTCTGCCTTCTATGTTCCCTAGCAAACTACCAAAATATTCCTC3’ (1 bp deletion)

5’GGTGGTGGCCAAGGCCCAGAGGAGGTGGTTAGCAAGGCCATGGACAGGCCAGCCCCGCCTGGAAGACTAGTGATTTTGTTGTTGTCTCTCTATACTCAACAACAAGTCCCAGTCTACCACACGGTGCGGGTCCTCGGAGCATCAGGGGGAGGTATGGGAGCACCATTTCTGGATGTATTTTCTCTGCCTTCTATGTTCCCTAGCAAACTACCAAAATATTCCTC3’ (1 bp mm)

5’GGTGGTGGCCAAGGCCCAGAGGAGGTGGTTAGCAAGGCCATGGACAGGCCAGCCCCGCCTGGAAGACTAGTGATTTTGTTGTTGTCTCTCTATACTCAACAACAAGTCCCAGTCTACCACACGGTGCGGGTCCTCTGAGCAATCAGGGGGAGGTATGGGAGCACCATTTCTGGATGTATTTTCTCTGCCTTCTATGTTCCCTAGCAAACTACCAAAATATTCCTC3’ (1 bp del +1 bp in)

5’GGTGGTGGCCAAGGCCCAGAGGAGGTGGTTAGCAAGGCCATGGACAGGCCAGCCCCGCCTGGAAGACTAGTGATTTTGTTGTTGTCTCTCTATACTCAACAACAAGTCCCAGTCTACCACACGGTGCGGGTCCTCTGAGCAATCAGGGGGAGGTATGGGAGCACCATTTCTGGATGTATTTTCTCTGCCTTCTATGTTCCCTAGCAAACTACCAAAATATTCCTC3’ (1 bp insertion)

4.miR-7a sequencing results for sgmiR7 2B CRISPR:

5'GGTGGTGGCCAAGGCCCAGAGGAGGTGGTTAGCAAGGCCATGGACAGGCCAGCCCCGCCTGGAAGACTAGTGATTTTGTTGTTGTCTCTCTATACTCAACAACAAGTCCCAGTCTACCACAC
GGTGCGGGTCCTCTGAGCATCAGGGGGAGGTATGGGAGCACCATTTCTGGATGTATTTTCTCTGCCTTCTATGTTCCCTAGCAAACTACCAAAATATTCCTC3' (127 bp deletion)

5'GGTGGTGGCCAAGGCCCAGAGGAGGTGGTTAGCAAGGCCATGGACAGGCCAGCCCCGCCTGGAAGACTAGTGATTTTGTTGTTGTCTCTCTATACTCAACAACAAGTCCCAGTCTACCACAC
GGTGCGGGTCCTCTGAGCATCAGGGGGAGGTATGGGAGCACCATTTCTGGATGTATTTTCTCTGCCTTCTATGTTCCCTAGCAAACTACCAAAATATTCCTC3' (8 bp deletion)

5'GGTGGTGGCCAAGGCCCAGAGGAGGTGGTTAGCAAGGCCATGGACAGGCCAGCCCCGCCTGGAAGACTAGTGATTTTGTTGTTGTCTCTCTATACTCAACAACAAGTCCCAGTCTACCACAC
GGTGCGGGTCCTCTGAGCAATCAGGGGGAGGTATGGGAGCACCATTTCTGGATGTATTTTCTCTGCCTTCTATGTTCCCTAGCAAACTACCAAAATATTCCTC3' (1 bp insertion)

5'GGTGGTGGCCAAGGCCCAGAGGAGGTGGTTAGCAAGGCCATGGACAGGCCAGCCCCGCCTGGAAGACTAGTGATTTTGTTGTTGTCTCTCTATACTCAACAACAAGTCCCAGTCTACCACAC
GGTGCGGGTCCTCTGAGCATCAGGGGGAGGTATGGGAGCACCATTTCTGGATGTATTTTCTCTGCCTTCTATGTTCCCTAGCAAACTACCAAAATATTCCTC3' (16 bp deletion)

5'GGTGGTGGCCAAGGCCCAGAGGAGGTGGTTAGCAAGGCCATGGACAGGCCAGCCCCGCCTGGAAGACTAGTGATTTTGTTGTTGTCTCTCTATACTCAACAACAAGTCCCAGTCTACCACAC
GGTGCGGGTCCTCTGAGCACATCAGGGGGAGGTATGGGAGCACCATTTCTGGATGTATTTTCTCTGCCTTCTATGTTCCCTAGCAAACTACCAAAATATTCCTC3' (2 bp insertion)

5'GGTGGTGGCCAAGGCCCAGAGGAGGTGGTTAGCAAGGCCATGGACAGGCCAGCCCCGCCTGGAAGACTAGTGATTTTGTTGTTGTCTCTCTATACTCAACAACAAGTCCCAGTCTACCACAC
GGTGCGGGCTCTCTGAGCATCAGGGGGAGGTATGGGAGCACCATTTCTGGATGTATTTTCTCTGCCTTCTATGTTCCCTAGCAAACTACCAAAATATTCCTC3' (2 bp mm)

5'GGTGGTGGCCAAGGCCCAGAGGAGGTGGTTAGCAAGGCCATGGACAGGCCAGCCCCGCCTGGAAGACTAGTGATTTTGTTGTTGTCTCTCTATACTCAACAACAAGTCCCAGTCTACCACAC
GGTGCGGGTCCTCTGAGCAATCAGGGGGAGGTATGGGAGCACCATTTCTGGATGTATTTTCTCTGCCTTCTATGTTCCCTAGCAAACTACCAAAATATTCCTC3' (1 bp insertion)

5'GGTGGTGGCCAAGGCCCAGAGGAGGTGGTTAGCAAGGCCATGGACAGGCCAGCCCCGCCTGGAAGACTAGTGATTTTGTTGTTGTCTATGTATACTCAACAACAAGTTCCATTTTACCTCAA
CAACAAGTTCCTGTGCCACCAGGGGGAGGTATGGGAGCACCATTTCTGGATGTATTTTCTCTGCCTTCTATGTTCCCTAGCAAACTACCAAAATATTCCTC3' (17 bp mm + 1 bp insertion)

5'GGTGGTGGCCAAGGCCCAGAGGAGGTGGTTAGCAAGGCCATGGACAGGCCAGCCCCGCCTGGAAGACTAGTGATTTTGTTGTTGTCTCTCTATACTCAACAACAAGTCCCAGTCTACCACAC
GGTGCGGGTCCTCTGAGCATCAGGGGGAGGTATGGGAGCACCATTTCTGGATGTATTTTCTCTGCCTTCTATGTTCCCTAGCAAACTACCAAAATATTCCTC3' (89 bp del + 160 bp deleted
downstream of miR-7a-5p sequence)

5'GGTGGTGGCCAAGGCCCAGAGGAGGTGGTTAGCAAGGCCATGGACAGGCCAGCCCCGCCTGGAAGACTAGTGATTTTGTTGTTGTCTCTCTATACTCAACAACAAGTCCCAGTCTACCACAC
GGTGCGGGTCCTCTGAGCATCAGGGGGAGGTATGGGAGCACCATTTCTGGATGTATTTTCTCTGCCTTCTATGTTCCCTAGCAAACTACCAAAATATTCCTC3' (169 bp + 71 bp deleted
downstream of miR-7a-5p sequence)

Fig. 3.5.2: Analysis of indels in the miR-7a-5p sequence. TOPO^{TM} TA – based sequence analysis of miR-7a-5p sequence was performed and compared with the reference (or wild-type) miR-7a-5p sequence. 1 shows the indels generated using sgmiR7 1A CRISPR, 2 shows the indels generated using sgmiR7 1B CRISPR, 3 shows the indels generated using sgmiR7 2A CRISPR, while 4 shows the indels generated using sgmiR7 2B CRISPR. Red letters in the sequence denote the bases deleted from the miR-7a-5p sequence and the numbers of deleted bases mentioned in the closed brackets in Red, Green letters represent the inserted bases in the miR-7a-5p sequence and the number of bases inserted mentioned in the closed brackets in Green, while Orange letters denote the bases substituted (mismatch mutations) in the miR-7a-5p sequence and the number of mismatches introduced mentioned in the closed brackets in Orange. The mutations in the upstream and downstream of the miR-7a-5p sequence are also mentioned in the closed brackets in **Black**. Underlined sequence represents the mature miR-7a sequence.

3.5.5 Isolating single cell clones with big indels in the miR-7a-5p sequence

miR-7 has been reported as one of the critical players in inducing cell cycle arrest in CHO cells, consequently leading to improvement in cellular longevity and productivity in CHO cell (Sanchez et al. 2013, Sanchez et al. 2014). From a bioprocessing context, we hypothesized that picking a CHO clone with mature miR-7a-5p sequence knocked out or altered would yield interesting phenotypes. For this reason, we chose sgmiR7 2B CRISPR 1.14 mixed cells for single cell cloning due to higher probability of finding a potential CHO clone that has a part of or the whole mature miR-7a-5p sequence knocked out. Since with sgmiR7 2B CRISPR we attained ~ 40 % indel frequencies and 30 % of the TOPO-TA cloning sequencing results demonstrated big indels. Once isolated each clone was subjected to real-time qPCR analysis for the quantitation of endogenous mature miR-7a levels.

3.5.6 Endogenous mature miR-7a expression analysis in single cell clones

We hypothesized that deleting one copy of miR-7a might not result in any phenotypic outcome due to compensation by the other copy of the gene or, sometimes, paralogs such as miR-7b in this case. However, it has been reported that miR-7a is more strongly expressed than miR-7b in many CHO cell lines (Hackl et al. 2011b). Therefore, before performing any phenotypic analysis across the 36 clones, we screened the clones on for mature miR-7a expression levels. We performed a real-time qPCR for each of the 36 clones hoping to isolate a clone that had both the miR-7a allele knocked out i.e. both miR-7a copies. We observed that out of 36 clones, 22 clones demonstrated less than 5-fold down regulation in the endogenous mature miR-7a expression levels when compared to the parental 1.14 cell line. However, 11 individual clones reported more than 5-fold down regulation, with Clone 6 and Clone 29 demonstrating ~ 10-fold down-regulation. Clone 14 showed nearly ~ 160-fold down regulation in mature miR-7a expression levels, which seemed to us as the potential clone with both miR-7a copies knocked out in the 1.14 cells genome. In order to confirm the mature mir-7a expression displayed by these clones, we re-analysed a few of these clones again for mature miR-7a expression levels as a validation step. Fresh total RNA samples were harvested again from Clone 6 & 14 (low mature miR-7a expression), Clone 25 & 30 (medium mature miR-7a expression)

and Clone 34 & 36 (high mature miR-7a expression). To our surprise, we observed that the levels of mature miR-7a expression levels were less down regulated for Clone 6 (from 0.107 units to 0.47 units), 14 (from 0.006 units to 0.39 units) and 30 (from 0.328 units to 0.75 units). However, the levels of mature miR-7a expression in Clone 25 did not change significantly (from 0.53 units to 0.62 units), for Clone 34 and Clone 36 stayed the same (i.e. ~ 0.9 units). At this stage, we did not have a single clone that showed more than 10-fold mature miR-7a down regulated since all the clones were still expressing mature miR-7a transcripts (**Fig. 3.5.3a**). We hypothesized that following one miR-7a copy knock out in these clones, the cells are somehow compensating for the lack of miR-7a by up regulating the other mir-7a copy through increased pre-mir-7a transcription. In addition to pre-mir-7a, although from Hackl et al. (2011) data we understood the miR-7a-5p is very strongly expressed as opposed to miR-7b-5p in CHO cell lines, we also hypothesized that up-regulation of miR-7 paralogs could also account for the loss of miR-7a as a compensation mechanism by the cells (explained later in **Section 4.5.5**). Therefore, we performed RT-qPCR for pre-mir-7a and pre-mir-7b expression levels, separately for each Clone 6, 14, 25, 30, 34 and 36. We observed that the pre-mir-7a or pre-mir-7b expression levels were unchanged in all of the clones relative to the control (CHO-K1 1.14 parental cell line) (p-value > 0.05), although insignificant but clone 34 showed slightly increased levels of pre-mir-7b in comparison to the control (**Fig. 3.5.3b**). Therefore, the speculation of possible compensation of miR-7a and/or miR-7b through increased pre-mir-7a and/or pre-mir-7b transcriptional rates through some feedback mechanism was unfounded.

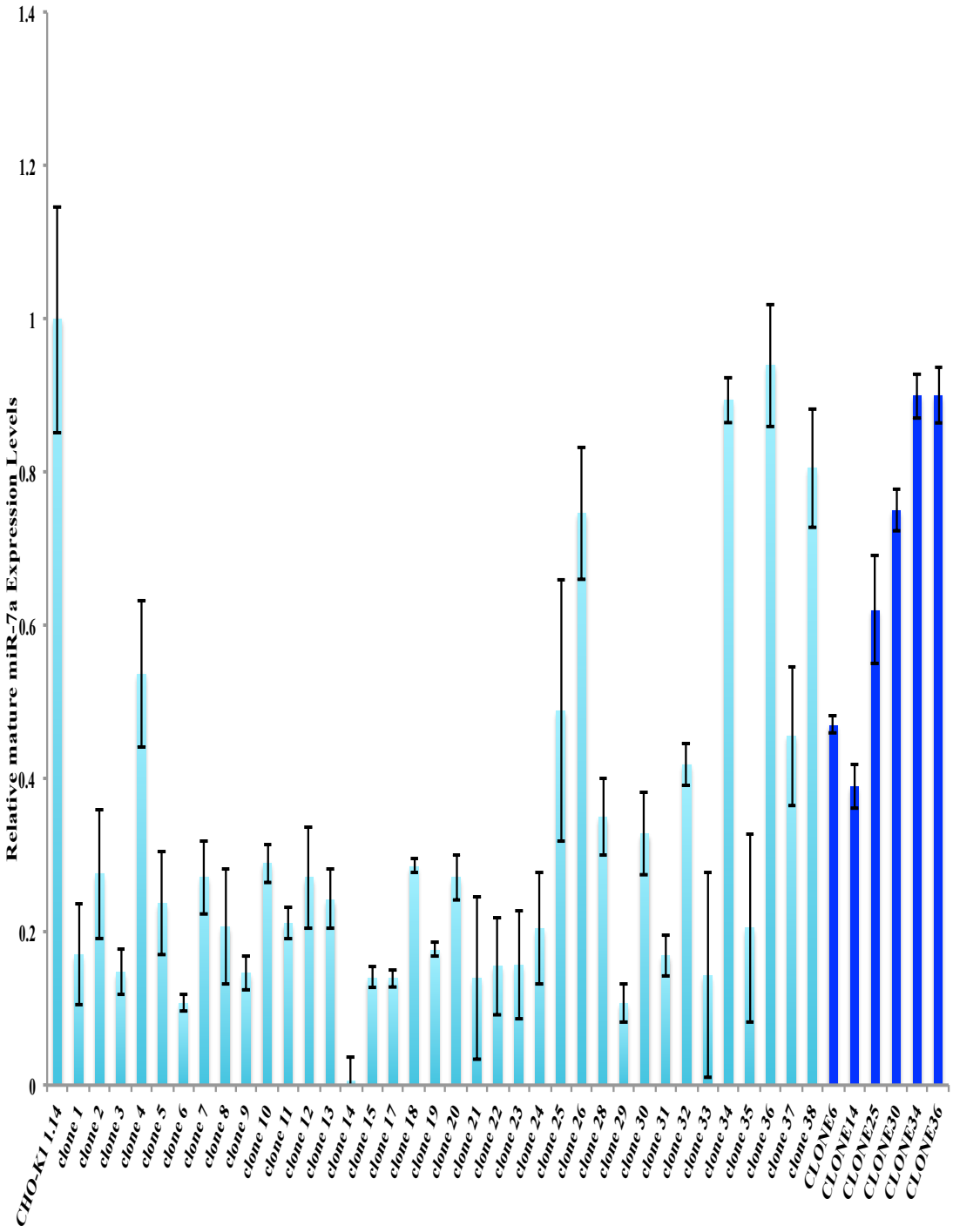


Figure 3.5.3a: Relative Expression levels of the mature miR-7a expression levels across single cell clones assessed using Real-Time qPCR. **Cyan** bars represent the first qPCR run across 38 clones, while **Blue** bars represent second qPCR run for few selected clones from a different total RNA sample for each. A standard student t-test was performed to analyse statistically significant data (p-value > 0.05). The levels of expression of mature miR-7a were normalised to endogenous snRNA U6 control.

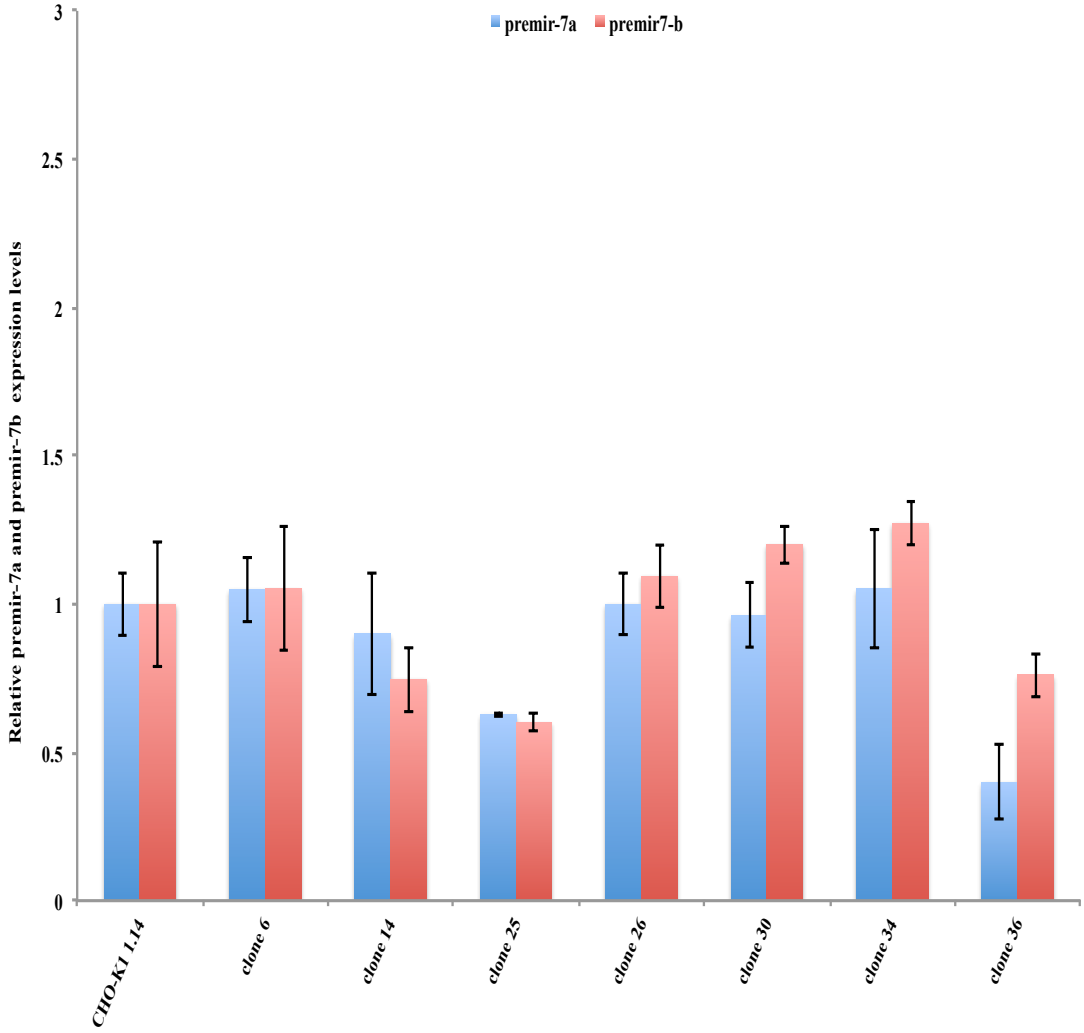


Figure 3.5.3b: Relative Expression levels of the premir-7a and premir-7b expression levels across selected single cell clones assessed using Real-Time qPCR. A standard student t-test was performed to analyse statistically significant data (p-value > 0.05). The levels of premir-7a and premir-7b were normalised to endogenous beta-actin control.

3.5.7 Sanger sequencing to analyse indels in individual isolated clones

The qPCR data above did not clearly report any of the clones contained miR-7a knockout. For this purpose, we chose Clone 6, 14, 25, 30, 34 and 36 to analyse the types of indels induced using Sanger sequencing. From each clone, we sent a 855 bp amplicon, flanking the 224 bp miR-7a-5p sequence, amplified using the primers mentioned in **Table 3.5.2**. We found that Clone 6, 14, 30 and 36 did not show any indels at the target site (i.e. miR-7a-5p sequence) (**Fig. 3.5.4**). Clone 25 showed 8 random bases inserted without disrupting the mature miR-7a sequence. Clone 34 showed 148 bp deletion i.e. 137 bp deleted in the miR-7a-5p sequence (with mature miR-7a, loop and star strand sequence knocked out) and 11 bp upstream of the miR-7a-5p sequence also deleted (**Fig. 3.5.4**). We hypothesized that the up regulation of mature miR-7a expression levels could be due to heterozygous deletion of only one miR-7a copy, whilst the other copy being up regulated as a result of cell's compensation mechanism for the perceived lack of the deleted miR-7a copy. To further check that Clone 34 contained homozygous or heterozygous deletion of mature miR-7a sequence, we PCR amplified the deleted 148 bp in the miR-7a-5p sequence in the wild-type 1.14 cells (wt 1.14), Clone 14 and Clone 34 using combinations of different flanking primers mentioned in the schematic diagram (**Fig. 3.5.5**). The primers flanking deleted 148 bases in the miR-7a-5p sequence should result in strong single 169 bp band with the genomic DNA from 1.14 wt and Clone 14 (since the miR-7a-5p sequence in both of these was intact). For Clone 14 and Clone 34 the expected bands were observed at their respective sizes. We found that Clone 34 had a faint band at the 169 bp mark (**Fig. 3.5.6, Table 3.5.2**). The faint band at 169 bp raised the speculation that the sgmiR7 2B CRISPR induced a heterozygous deletion of one miR-7a copy in Clone 34, while the other copy still either undisrupted or contained insignificant indels. We sequenced the faint 169 bp band gel excision and found that mature miR-7a sequence of the second copy undisrupted, however, there were 2-3 bases inconsequential random mismatch mutations as well 2-3 bases deletion upstream of it (**Fig. 3.5.7**). This observation also explained the unchanged mature miR-7a expression levels in Clone 34 in **Fig. 3.5.3**. For Clone 14, the presence of intact miR-7a-5p sequence was rather surprising since we observed ~ 5-fold down regulation in the mature miR-7a expression in **Section 3.5.6**. In order to address this

anomaly between these two experiments for Clone 14, we performed another validation study as described in the upcoming section.

```
clone34      TAGAGGGCATGGGGAAAGACGCTGGCCTCCCCACAGGAGTATCC----------------
wt1.14       TAGAGGGCATGGGGAAAGACGCTGGCCTCCCCACAGGAGTATCCAGATGATAACAGGTGG
clone25      TAGAGGGCATGGGGAAAGACGCTGGCCTCCCCACAGGAGTATCCAGATGATAACAGGTGG
clone6       TAGAGGGCATGGGGAAAGACGCTGGCCTCCCCACAGGAGTATCCAGATGATAACAGGTGG
clone30      TAGAGGGCATGGGGAAAGACGCTGGCCTCCCCACAGGAGTATCCAGATGATAACAGGTGG
clone14      TAGAGGGCATGGGGAAAGACGCTGGCCTCCCCACAGGAGTATCCAGATGATAACAGGTGG
clone36      TAGAGGGCATGGGGAAAGACGCTGGCCTCCCCACAGGAGTATCCAGATGATAACAGGTGG
             ****************************************

clone34      ------------------------------------------------------------
wt1.14       TGGCCAAGGCCCAGAGGAGGTGGTTAGCAAGGCCATGGACAGGCCAGCCCCGCCTGGAAG
clone25      TGGCCAAGGCCCAGAGGAGGTGGTTAGCAAGGCCATGGACAGGCCAGCCCCGCCTGGAAG
clone6       TGGCCAAGGCCCAGAGGAGGTGGTTAGCAAGGCCATGGACAGGCCAGCCCCGCCTGGAAG
clone30      TGGCCAAGGCCCAGAGGAGGTGGTTAGCAAGGCCATGGACAGGCCAGCCCCGCCTGGAAG
clone14      TGGCCAAGGCCCAGAGGAGGTGGTTAGCAAGGCCATGGACAGGCCAGCCCCGCCTGGAAG
clone36      TGGCCAAGGCCCAGAGGAGGTGGTTAGCAAGGCCATGGACAGGCCAGCCCCGCCTGGAAG

clone34      ------------------------------------------------------------
wt1.14       ACTAGTGATTTTGTTGTTGTCTCTCTATACTCAACAACAAGTCCCAGTCTACCACACGGT
clone25      ACTAGTGATTTTGTTGTTGTCTCTCTATACTCAACAACAAGTCCCAGTCTACCACACGGT
clone6       ACTAGTGATTTTGTTGTTGTCTCTCTATACTCAACAACAAGTCCCAGTCTACCACACGGT
clone30      ACTAGTGATTTTGTTGTTGTCTCTCTATACTCAACAACAAGTCCCAGTCTACCACACGGT
clone14      ACTAGTGATTTTGTTGTTGTCTCTCTATACTCAACAACAAGTCCCAGTCTACCACACGGT
clone36      ACTAGTGATTTTGTTGTTGTCTCTCTATACTCAACAACAAGTCCCAGTCTACCACACGGT

clone34      --------------------AGTATCAGGGGGAGGTATGGGAGCACCATTTCTGGATGTA
wt1.14       GCGGGTCCTCTGA--------GCATCAGGGGGAGGTATGGGAGCACCATTTCTGGATGTA
clone25      GCGGGTCCTCTGAGCATCTGAGCATCAGGGGGAGGTATGGGAGCACCATTTCTGGATGTA
clone6       GCGGGTCCTCTGA--------GCATCAGGGGGAGGTATGGGAGCACCATTTCTGGATGTA
clone30      GCGGGTCCTCTGA--------GCATCAGGGGGAGGTATGGGAGCACCATTTCTGGATGTA
clone14      GCGGGTCCTCTGA--------GCATCAGGGGGAGGTATGGGAGCACCATTTCTGGATGTA
clone36      GCGGGTCCTCTGA--------GCATCAGGGGGAGGTATGGGAGCACCATTTCTGGATGTA
                                 * ***********************************

clone34      TTTTCTCTGCCTTCTATGTTCCCTAGCAAACTACCAAAATATTCCTCTAAGGCCTGGGCT
wt1.14       TTTTCTCTGCCTTCTATGTTCCCTAGCAAACTACCAAAATATTCCTCTAAGGCCTGGGCT
clone25      TTTTCTCTGCCTTCTATGTTCCCTAGCAAACTACCAAAATATTCCTCTAAGGCCTGGGCT
clone6       TTTTCTCTGCCTTCTATGTTCCCTAGCAAACTACCAAAATATTCCTCTAAGGCCTGGGCT
clone30      TTTTCTCTGCCTTCTATGTTCCCTAGCAAACTACCAAAATATTCCTCTAAGGCCTGGGCT
clone14      TTTTCTCTGCCTTCTATGTTCCCTAGCAAACTACCAAAATATTCCTCTAAGGCCTGGGCT
clone36      TTTTCTCTGCCTTCTATGTTCCCTAGCAAACTACCAAAATATTCCTCTAAGGCCTGGGCT
             ************************************************************
```

Figure 3.5.4: Multiple sequence alignment results for individual clones 855 bp amplicon sequences versus wt-1.14 854 bp amplicon sequence. Only the sequence wherein indels were observed is shown. Highlighted in **Green** represent the insertion in Clone 25, Dashed **Black Lines** represent deleted bases and **Asterisks (*)** represent strong similarity in bases with the wt1.14.

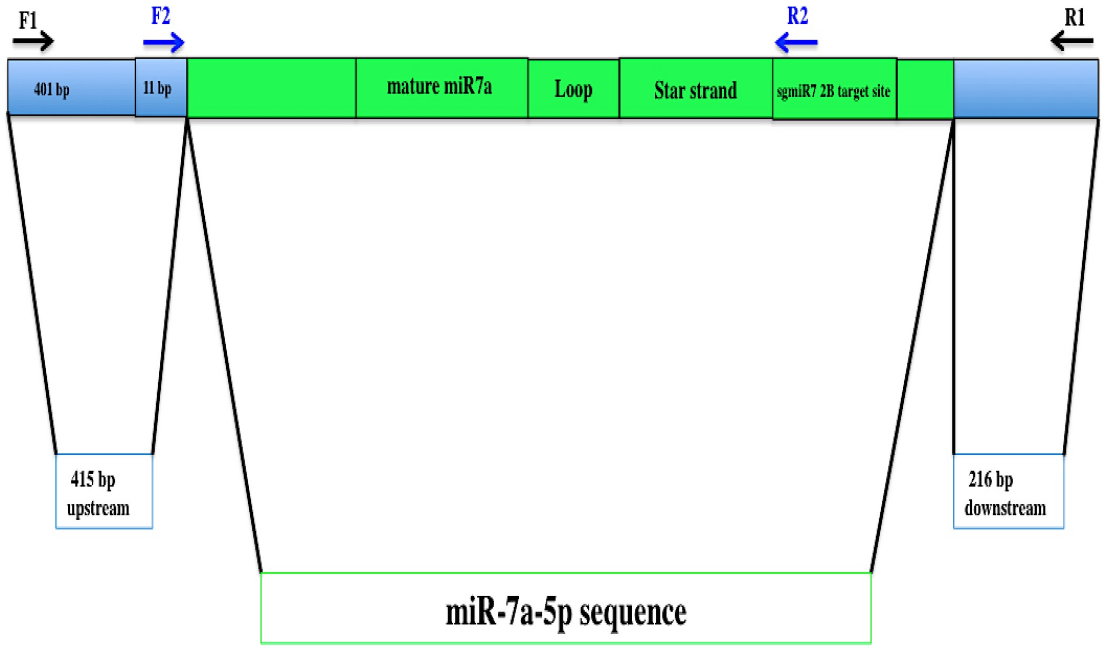


Figure 3.5.5: Schematic diagram of miR-7a-5p sequence amplified using different flanking primers. Green solid boxes represent the 224 bp miR-7a-5p sequence showing the mature, loop, star and sgmiR7 2B CRISPR target site. **F1** and **R1** are Forward 1 and Reverse 1 flanking primers, respectively. **F2 and R2** are Forward 2 and Reverse 2 flanking primers, respectively.

Table 3.5.2: List of PCR products obtained from wt 1.14, Clone 14 and Clone 34 templates using different combinations of flanking primers

Primers	**Expected PCR product size for wt 1.14 template**	**Observed PCR product size for wt 1.14 template**
F1 + R1	854 bp	854 bp
F2 + R2	169 bp	169 bp
F1 + R2	575 bp	575 bp
F2 + R1	449 bp	449 bp
Primers	**Expected PCR product size for Clone 14 template**	**Observed PCR product size for Clone 14 template**
F1 + R1	854 bp	854 bp
F2 + R2	169 bp	169 bp
F1 + R2	575 bp	575 bp
F2 + R1	449 bp	449 bp
Primers	**Expected PCR product size for Clone 34 template**	**Observed PCR product size for Clone 34 template**
F1 + R1	706 bp	706 bp & very faint band at 850 bp
F2 + R2	No band	169 bp faint band
F1 + R2	404 bp	404 bp & 575 bp faint band
F2 + R1	301 bp	449 bp band

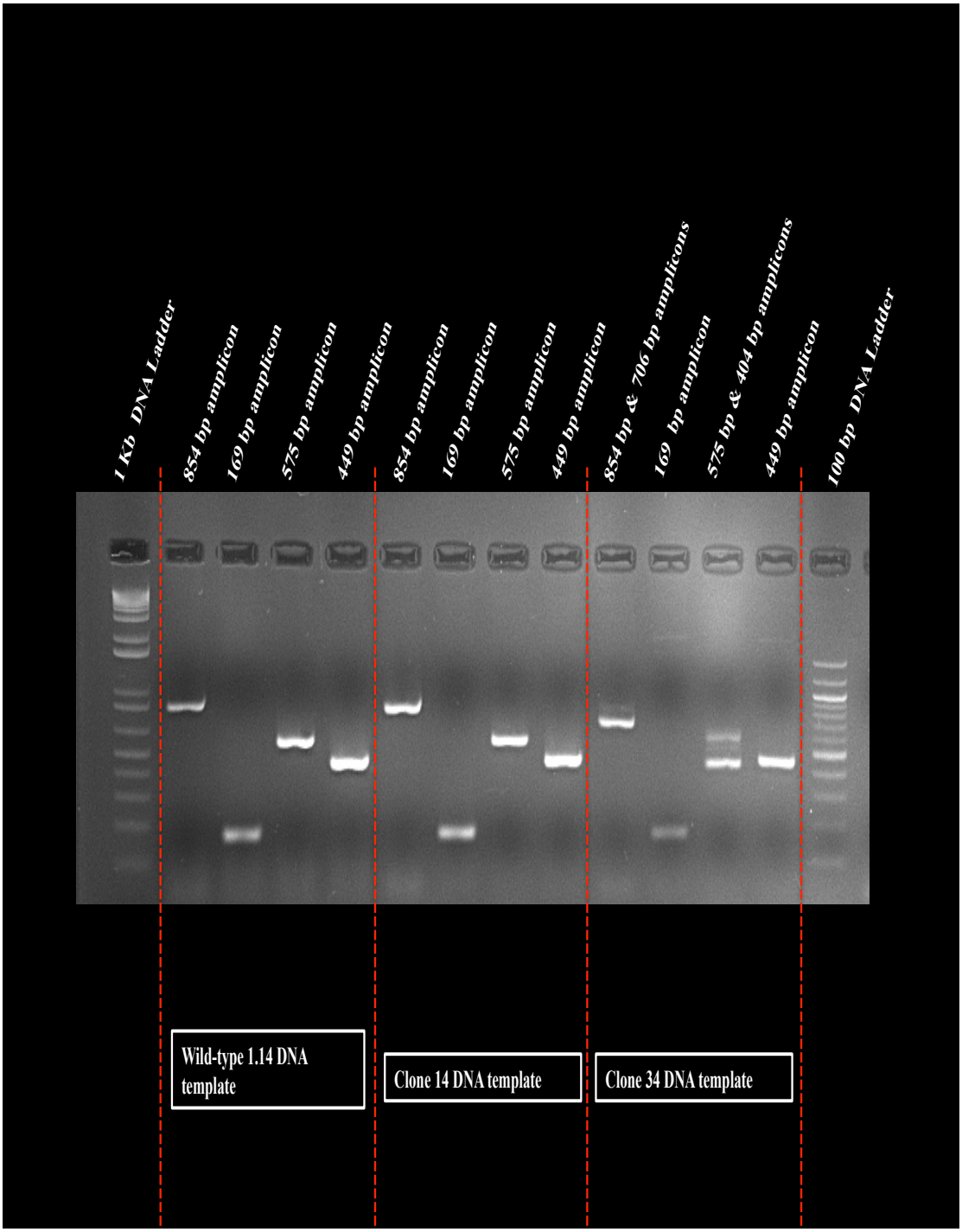


Figure 3.5.6: Agarose gel image showing PCR amplicons from wt 1.14, Clone 14 and Clone 34 as DNA template with different combinations of flanking primers (refer to Fig. 3.5.5, Table 3.5.1 for primers details).

Sequence ID: lcl|Query_231349 **Length:** 169 **Number of Matches:** 2

Range 1: 4 to 119 Graphics ▼ Next Match ▲ Previous Match

Score	Expect	Identities	Gaps	Strand
150 bits(166)	9e-42	106/116(91%)	5/116(4%)	Plus/Plus

```
Query  11   ATAACAGGTGG-GCCCAAGGCCCAGAGGAGGTGGTAAGCAAG-CCA--GAACGCCCAGCC  66
            |||||||||||| | ||||||||||||||||||||||| ||||||| |||  ||  | ||||||
Sbjct  4    ATAACAGGTGGTGGCCAAGGCCCAGAGGAGGTGGTTAGCAAGGCCATGGACAGGCCAGCC  63

Query  67   CCGCCTGGAAGACTAGTGATTTTGTTGTTGTCTCTCTATACTC-ACAACAAGTCCC  121
            |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Sbjct  64   CCGCCTGGAAGACTAGTGATTTTGTTGTTGTCTCTCTATACTCAACAACAAGTCCC  119
```

Range 2: 64 to 92 Graphics ▼ Next Match ▲ Previous Match ▲ First Match

Score	Expect	Identities	Gaps	Strand
53.6 bits(58)	2e-12	29/29(100%)	0/29(0%)	Plus/Plus

```
Query  140  CCGCCTGGAAGACTAGTGATTTTGTTGTT  168
            |||||||||||||||||||||||||||||
Sbjct  64   CCGCCTGGAAGACTAGTGATTTTGTTGTT  92
```

Figure 3.5.7: Sequence alignment of the 169 bp gel excision of Clone 34. Highlighted sequence in Cyan represents the mature miR-7a sequence.

3.5.8 Impact of miR-7a-5p knockout on miR-7 downstream targets

We performed real time qPCR on two previously validated miR-7 downstream targets from our laboratory in order to observe if deleting one copy of miR-7 would result in changes in the expression levels of DNA-repair and recombination protein 54-like (Rad54L) and Proteosome (prosome, macropain) activator subunit 3 (Psme3) in wt 1.14 cells, Clone 14 and Clone 34. Moreover, this study was also important to address the anomaly observed in **Section 3.5.7** for Clone 14. miR-7 has been reported to play a key role in the regulation of Rad 54L and Psme3 (Sanchez et al. 2013). Upon miR-7 mimics transfection Rad 54L and Psme3 were ~ 21-fold and 19-fold down regulated (data not shown) as previously validated in our laboratory. Therefore, from the observations in **Section 3.5.6 and 3.5.7**, we expected that Clone 14 would result in increased levels of expression of Rad54L and Psme3 than Clone 34. Since Clone 14 had 5-fold less mature miR-7a expression than Clone 34, in which the mature miR-7a expression level was insignificantly down regulated. To our surprise, we observed that Clone 14 showed lowered expression levels (i.e. 10-fold and 2.5-fold down regulated) of Rad54L and Psme3, respectively. Clone 34 showed similar expression levels of Psme3 as observed in the wt 1.14 cells. Moreover, Clone 34 showed Rad54L gene slightly overexpressed, although insignificant (p-value > 0.05) (**Fig. 3.5.8** and **Fig. 3.5.9**). From this study, we inferred that mature miR-7a expression in Clone 14 was a false positive result when compared with the sequencing and validation data. For Clone 34, we observed that we were successful in achieving a heterozygous miR-7a deletion as demonstrated by Sanger sequencing and PCR studies.

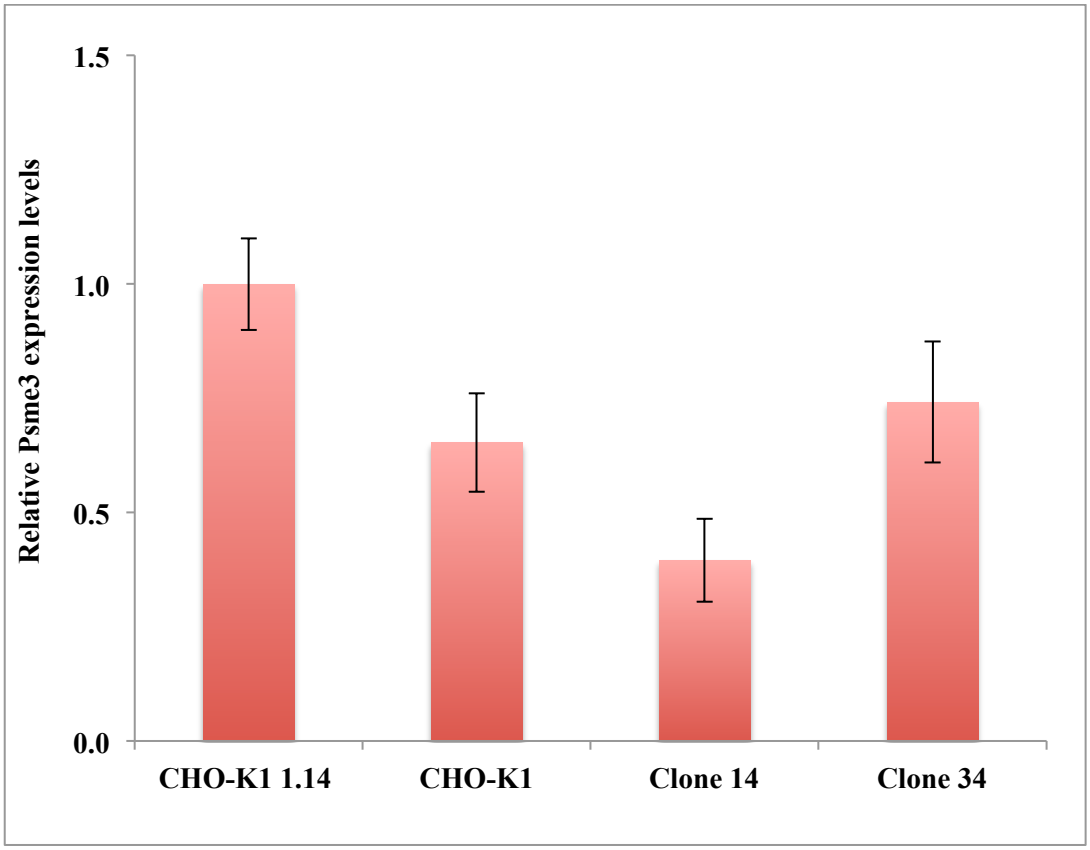


Figure 3.5.8: Relative expression levels of Psme3 measured in wt 1.14 cells, CHO-K1 parental cells, Clone 14 and Clone 34 using real time qPCR. A standard student t-test was performed to analyse statistically significant data (p-value > 0.05). The levels of Psme3 were normalised to endogenous beta-actin control.

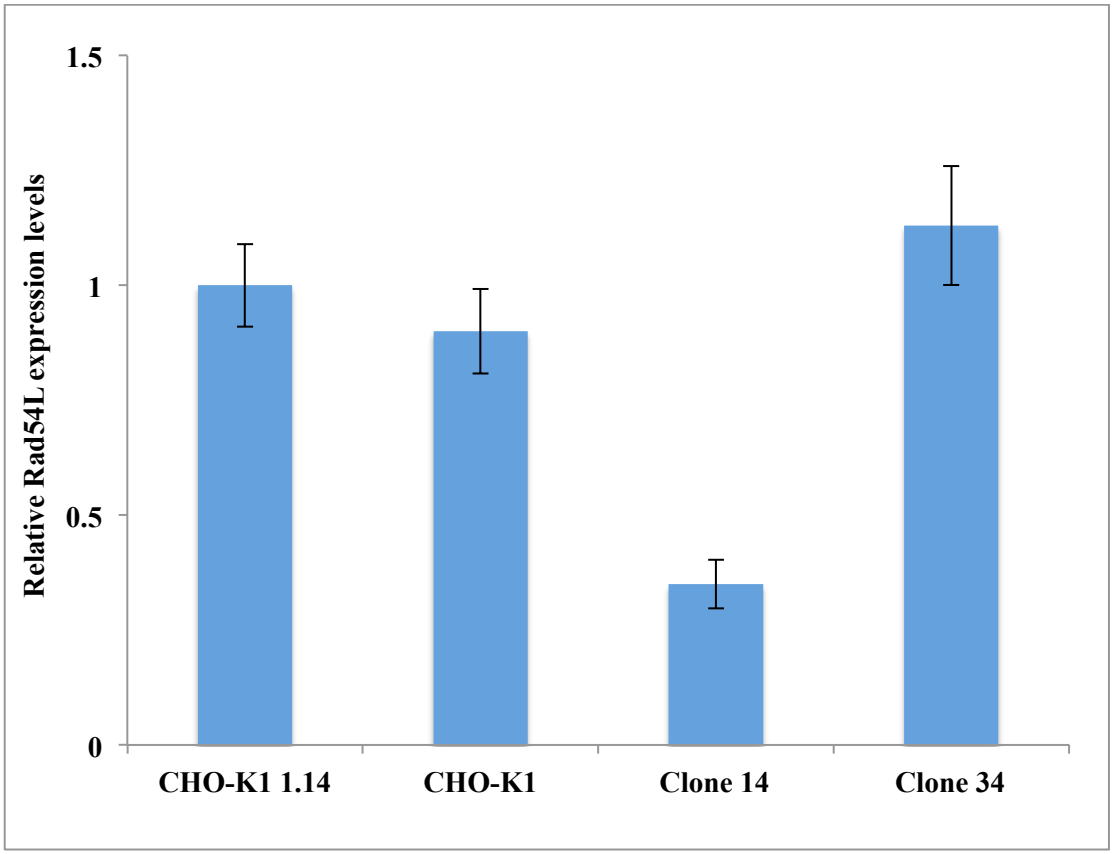


Figure 3.5.9: Relative expression levels of Rad54L measured in wt 1.14 cells, CHO-K1 parental cells, Clone 14 and Clone 34 using real time qPCR. A standard student t-test was performed to analyse statistically significant data (p-value > 0.05). The levels of Rad54L were normalised to endogenous beta-actin control.

In summary, we were able to induce ~ 50 % indels in the stable eGFP gene integrated in CHO-K1 cells' genome using CRISPR-Cas9 system. Following this, we also attained 35-40 % indels frequencies in the miR-7a-5p locus in industrially relevant CHO cell line using four different CRISPR-Cas9 systems. However, due to less number of single clones screened we were not able to find clone that contained homozygous miR-7a allele knockout. We were only able to find a clone containing heterozygous miR-7a knockout, which was evident from both the sequencing as well as the mature miR-7a expression levels RT-qPCR data. We inferred from this study that despite similar targeting efficiencies of the CRISPR as was the case for protein-coding gene, for non-protein coding gene i.e. miRNA, an effective knockout requires more than single base indel to cause a frameshift mutation. This illustrates the additional challenge in effectively knocking out a miRNA compared to a protein-coding gene. Therefore, in such a case we believe a multiplex CRISPR strategy could prove instrumental in enabling an effective miRNA sequence knockout.

We also strongly believe that we could have found interesting candidates from the rest of the 32 clones, which were not analysed further using Sanger sequencing due to time constraint for the project.

3.6 Multiplex CRISPR-Cas9 systems

3.6.1 Duplex px459 CRISPR-Cas9 constructs designing and cloning strategy

Following successful gene knockouts of the eGFP gene in CHO-eGFP cells, as well as of the miR-7a-5p sequence in the 1.14 cells via single sgRNA CRISPR, we hypothesized that a duplex CRISPR system (i.e. containing two sgRNAs) would be an attractive approach to effect a complete gene sequence knockout. Initially as prototype to check the % efficiency induced by a Duplex system, we designed a duplex CRISPR system using sgGFP 1B px459 as the backbone vector (described above in **Section 3.4.3**) and linearized it using *XbaI* and *KpnI* (refer to **Section 2.8.2.1**). Following sgGFP 1B px459 linearization, we cloned and ligated the sgmiR7 2B sgRNA sequence (along with U6 promoter sequence upstream of it) sequence into the linearized sgGFP 1B px459 vector. This resulted in a Duplex sgGFP 1B + miR7 2B CRISPR-Cas9 system (henceforth referred to as GFP Duplex) (**Fig. 3.6.1**) with two sgRNAs in one plasmid DNA.

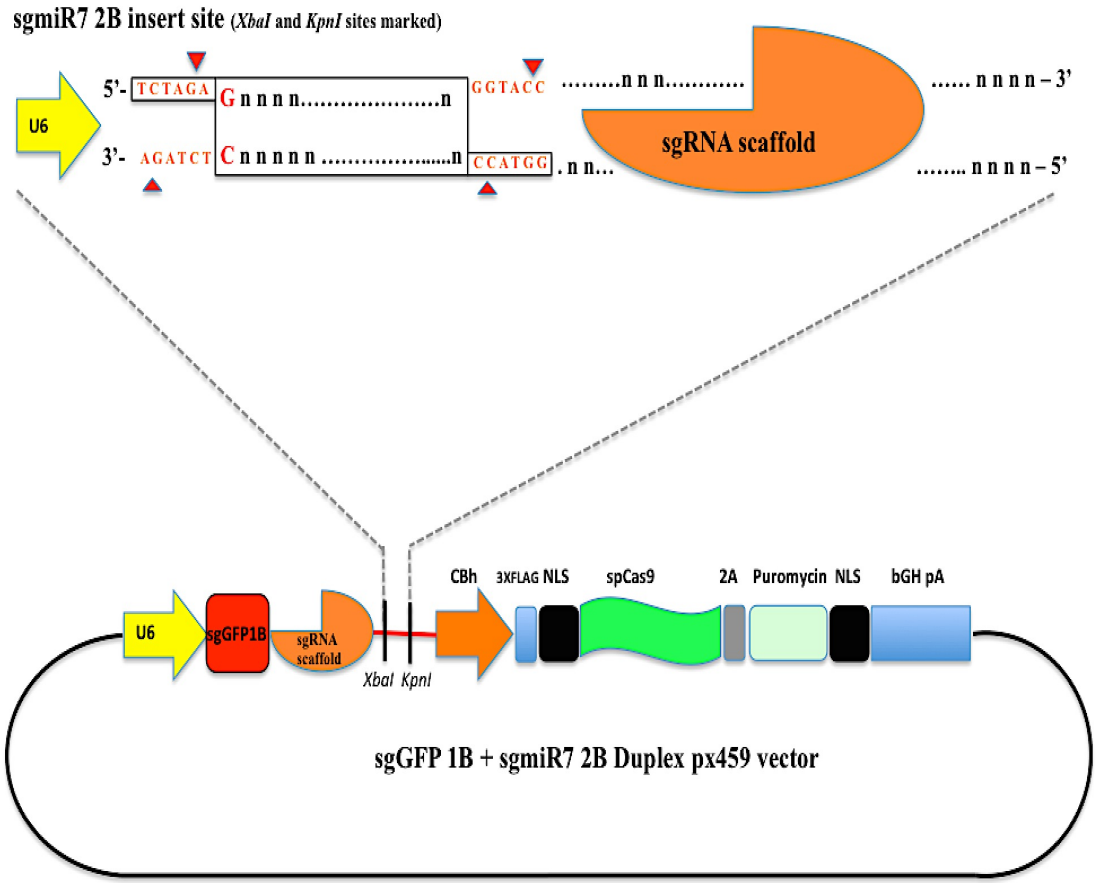


Figure 3.6.1: Schematic of Cas9 and sgRNA expression cassettes in eGFP 1B px459 vector backbone. A 19 nt or 20 nt guide sequence containing *BbsI* cut compatible sticky-ends was cloned into a linearized sgGFP 1B px459 vector backbone (linearized using the same *BbsI*). The final sgRNA expression cassette and Cas9 construct consisted of a U6 polymerase III promoter, two guide RNAs (sgGFP 1B backbone as well as sgmiR7 2B), a sgRNA scaffold sequence (Orange shaped structure) and a poly (T) termination sequence. Further downstream are other transgene elements such as hybrid form of chicken beta-actin (CBh) promoter, 3X FLAG peptide, SV40, NLS, spCas9 cassette linked with 2A peptide, which is also linked to Puromycin (Selection Marker Gene), NLS and BGH, polyadenylation signal and transcription termination sequence.

3.6.2 Testing the efficiency of the GFP Duplex CRISPR system

We tested the efficiency of the GFP Duplex CRISPR-Cas9 system in the eGFP cell line (used earlier in **Section 3.4.3**). We transfected eGFP cells with 1 μg of px459 GFP Duplex CRISPR, as well as the px459 empty vector separately to monitor the transfection efficiency. Post transfection (~ 10 days), we observed ~ 60 % of viable green cells in the control px459 vector only treated sample (data not shown). Similarly in the duplex CRISPR treateded cells, we observed ~ 60 % of non-green viable cells, suggesting that the sgGFP 1B guide in the duplex CRISPR worked successfully in knocking out eGFP gene in the cells (**Fig. 3.6.2**).

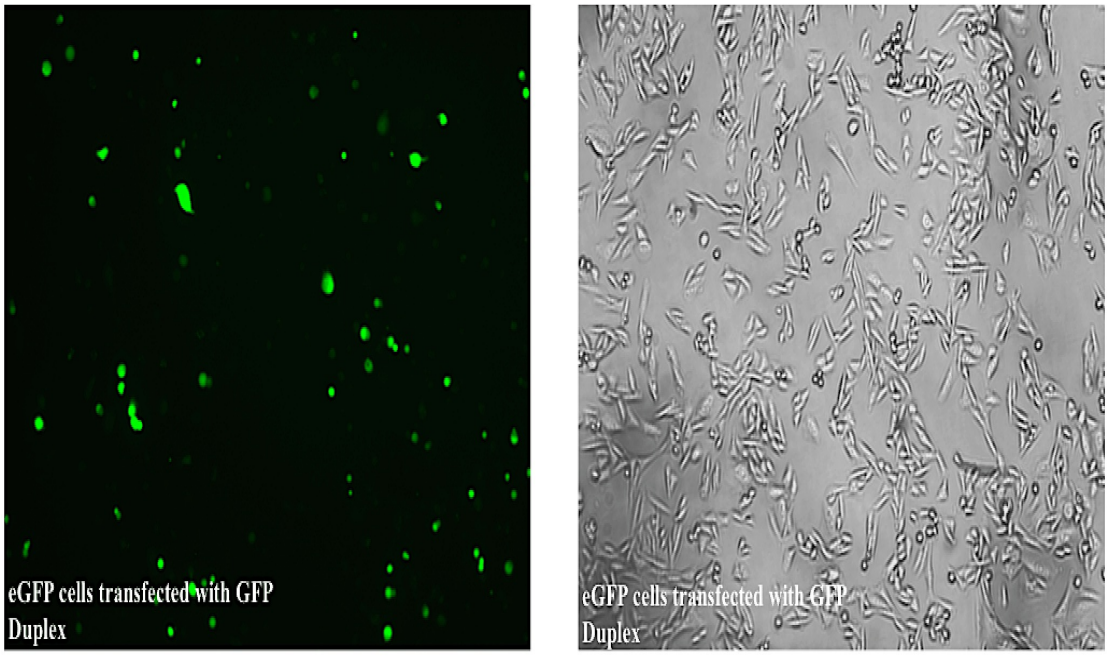


Figure 3.6.2: Fluorescent Microcopy images of the eGFP cells transfected with GFP Duplex CRISPR system. Image on the left is imaged on fluorescent lamp, while image on the right is the phase contrast image of the image in left. Cells were seeded at 2×10^5 cells/mL per well of a 6-well plate and images were taken post transfection using the fluorescent microscope.

Following the visual confirmation of sgGFP 1B guide working efficiently, we tested the % indels frequencies induced by both the eGFP 1B and miR7 2B encoding sgRNAs in the GFP Duplex CRISPR system using the Surveyor Assay. Using the same methodology of calculating % indels as described in **Section 3.4.3.2 and Section 3.5.3**. We observed that the sgeGFP 1B resulted in (~ 35-40 %) indels efficiency, however, the second guide i.e. sgmiR7 2B generated low % indels in the range of (15-20 %) (**Fig. 3.6.3**).

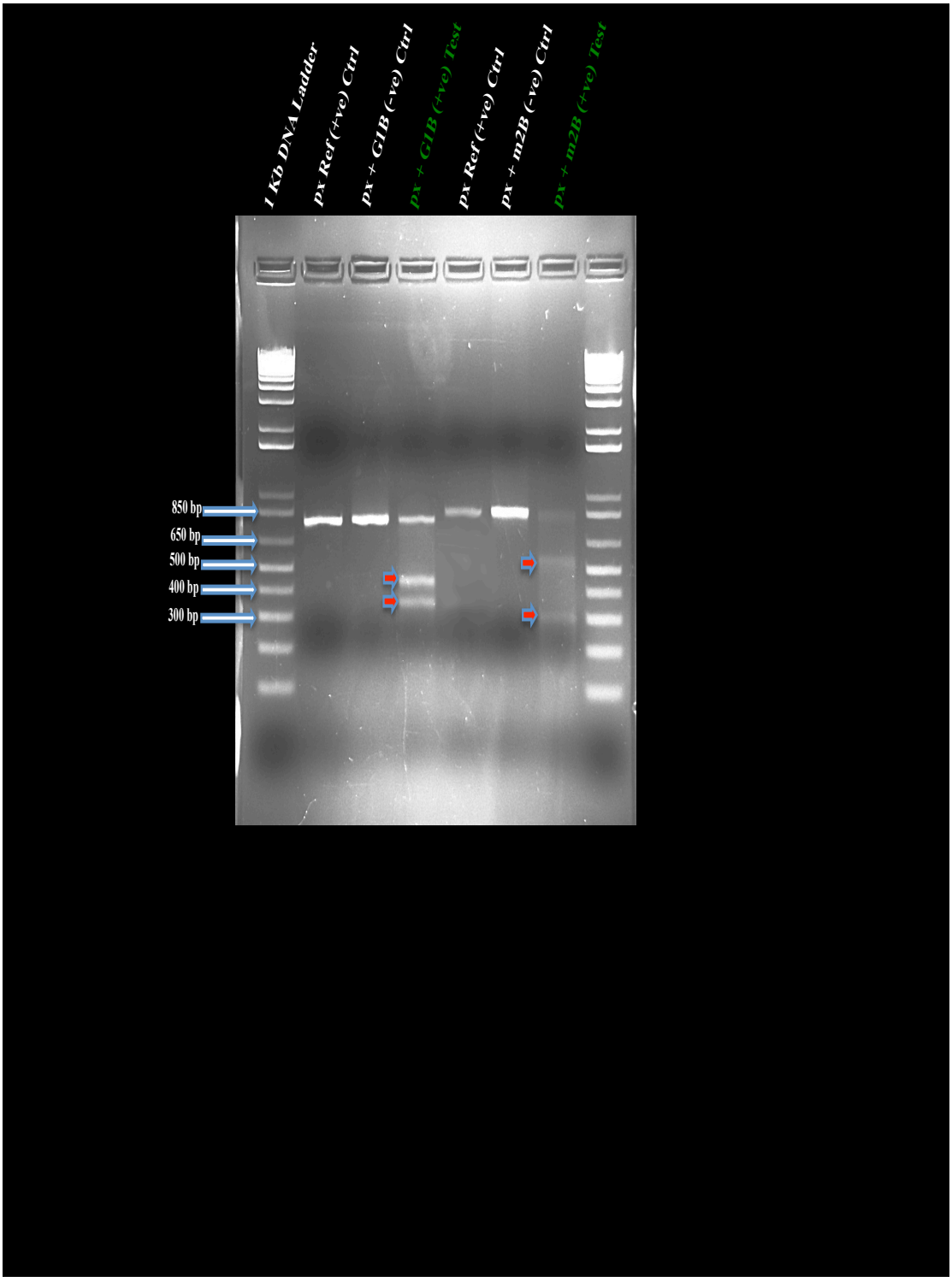


Figure 3.6.3: Agarose gel image post Surveyor Assay. The +ve and -ve stands for samples with Nuclease S and samples without Nuclease S enzyme, respectively. px → Reference DNA Control, (px + G1A) → Homo/Heteroduplex mix formed using PCR amplicons from reference px459 and sgGFP 1B and (px + m2B) → Homo/Heteroduplex mix formed using PCR amplicons from reference px459 and sgmiR7 2B. All Control samples are highlighted with White font, while Green labeled samples are Test eGFP CRISPR samples. Red Arrow Heads represent the bands for the cleaved mutated fragments.

As mentioned above in **Section 3.5.2**, we hypothesized that it is necessary to target miR-7a-5p sequence to introduce significant indels with greater frequency since 1 or 2 bp indel might not be enough to impact miRNA expression. Based on the same protocol, we also designed a miR-7 Duplex CRISPR system that contained two guides i.e. sgmiR7 1B (as the backbone vector) with sgmiR7 2B alongwith U6 promoter sequence upstream of the sgmiR7 2B guide. However, due to time constraints we did not reach to a stage to implement and check the efficiency of this miR7 Duplex CRISPR-Cas9 system. We anticipate that multiplex CRISPR-Cas9 system would generate greater magnitude of indels in miRNA sequence that could provide us with interesting CHO phenotypes, when the CHO cells with big knockouts are evaluated for bioprocess-related phenotypes.

Section 4.0

Discussion

4. Discussion

In the past two decades, significant progress has been witnessed in the field of production of biopharmaceuticals in order to meet the increasing market needs. Biotechnologists have been able to keep-up with demand by improving the product titres by a commendable 100-fold. This significant improvement in product titres is mainly attributed due to the development of serum-free media compositions and feeding strategies. The remaining 20-fold biologics production improvement is due to improved vector expression and genetic engineering approaches commonly employed during the development of recombinant CHO cell line (Hacker, De Jesus and Wurm 2009). However, certain recombinant proteins that have to be administered at a higher dose e.g. mAbs and with the current product yield of 5 g/L, would mean that there still at large an issue that is required to be addressed in the field of biopharmaceuticals production process. Therefore, for this purpose single gene engineering approaches are employed in CHO cells with a view to enhance bioprocess-related CHO cell phenotypes and to fast track the production process of biologics (Hacker, De Jesus and Wurm 2009). However, single and/or multi-gene engineering approaches are limited in their ability to modify entire cellular pathways or processes and could potentially result in undesirable side effects (Borth et al. 2005). An alternative and more attractive solution to single and/or multi-gene engineering approaches is through the exploration and manipulation of endogenous microRNA(s). miRNAs offer two main advantages over conventional gene engineering approaches; (i) miRNA are capable of targeting hundreds of genes as they are involved in regulation of many cellular pathways, (ii) their manipulation does not impose any burden on the cell's translational machinery as they are not processed by the translational machinery of the cell (Bartel 2004b). All these positive points have encouraged many research groups across the globe in the CHO area to explore these beneficial tools to positively influence CHO cell phenotypes such as cellular growth and productivity by searching for potential engineering targets.

With a similar objective, this project was mainly focused on improving CHO cell phenotypes (in bioprocess context) by manipulating two different miRNAs using two different gene-engineering approaches.

4.1 Sponge technology: GFP an effective mean to gauge sponge activity in CHO-K1 SEAP cells

The 'miRNA sponge' technology was previously demonstrated to result in efficient inhibition of endogenous miRNA by competing with the endogenous mRNA targets for miRNA binding (Ebert, Neilson and Sharp 2007). The application of both transient and stable knockdown of miRNA using miRNA sponges has been extensively described (Ebert and Sharp 2010). The stable knockdown of endogenous miRNAs using miRNA sponge vectors in CHO cells has been reported to enhance CHO VCD and/or protein productivity in various reports from our research group (Sanchez et al. 2013, Kelly et al. 2015b, Kelly et al. 2015a). These reports also led our interest to use the Let-7 sponge approach with a view to increase CHO cell growth and productivity characteristics in this project.

In our study, the Let-7 sponge was designed in order to soak-up endogenous Let-7 by competing with mRNA targets for Let-7 binding. To ensure efficient endogenous Let-7 sequestration by the Let-7 sponge, four Let-7 binding sites were placed in tandem in the 3'-UTR of the destabilized GFP (deGFP). Ebert et al. (2007) reported that there is a threshold for the number of binding sites that can be incorporated in a single miRNA sponge system and suggested a range of 4-10 binding sites separated by a few nucleotides each. Ebert et al. (2007) also reported that increasing the number of binding sites increases the susceptibility of the miRNA sponge system to RNAi-mediated degradation. Moreover, it was also reported that a bulge or "wobble" in at nt position 9-11 could increase the stability of miRNA sponge by evading RNAi-mediated cleavage since perfect complementarity at nt position 9-11 is recognised by the Ago 2 protein (Ebert, Neilson and Sharp 2007). Let-7 is a big family of miRNAs, hence we anticipated the combined impact of using Let-7 sponge to deplete all endogenous Let-7 members in CHO cells and would generate interesting phenotypes. The Let-7 sponge also contained a deGFP upstream of the four Let-7 binding sites as a reporter gene. This unstable deGFP has a short half-life of 2 hours and provides a perfect read-out of the miRNA sponge effective binding since there is a change in fluorescence upon binding of the endogenous Let-7. When the Let-7 and NC sponge systems were stably transfected into CHO-K1 SEAP cells separately, we observed a range of different mean GFP expression in the both NC and Let-7 mixed SEAP populations. The mean GFP expression in the Let-7 SEAP mix population was lower

than the NC SEAP mix population potentially due to the endogenous Let-7 binding to the four binding sites in the Let-7 sponge vector, leading to deGFP translational repression. For our phenotypic analysis, only medium to high GFP expressing clones were picked in order to ensure that they were all expressing the excess sponge transcript above the level of saturation by endogenous Let-7.

To ensure that the Let-7 sponge was efficient and specific in binding the endogenous Let-7 family members, we conducted a validation experiment by transiently transfecting Let-7 mimics in the Let-7 stable sponge clones. Post Let-7 mimic transfection, a further reduction in GFP expression in these clones verified that the Let-7 sponge was efficiently binding the endogenous Let-7, hence reducing the Let-7 levels available to bind to its endogenous mRNA targets.

4.2 Stable Let-7 depletion using sponge technology did not enhance growth in CHO-K1 SEAP cells

As mentioned above in **Section 1.6.9.1** that Let-7 miRNAs have been implicated in regulating an array of genes involved in various cellular pathways. Let-7 plays an important role in the development of *C.elegans* and various other organisms. Let-7 miRNAs have also been reported to act as tumour suppressors in many cancer types (Roush and Slack 2008). Let-7 is reported to be involved in regulating the expression of a multitude of genes involved in the process of cell division and cell proliferation (Johnson et al. 2007) (**Fig. 4.4**). Recently, a research group reported a more than 60 % increase in CHO cell specific productivity in two mAb producing cell lines by using anti-let-7a miRNA inhibitor molecules (Greenlees et al. 2014). Moreover, previously our laboratory research group reported an anti-proliferative role of Let-7e in CHO cells through gain-of-function (use of Let-7 mimics) studies i.e. transient up-regulation of Let-7e decreased the VCD by 3 to 4-fold. Additionally, the cells were observed to be growth arrested for 5 days of the culture period and this resulted in increased > 4-fold improvement in the normalised productivity in two CHO cell lines (unpublished data).

Initially, we observed that the Let-7 sponge expressing cells showed enhanced VCD and improved cell viability compared to the NC sponge mixed pool of cells during the later stages of the culture. Following this we FACs sorted lones, but did not observe any significant improvement in VCD in the Let-7 sponge stable clones with all clones

displaying similar growth profiles to the NC sponge stable clones. However, we also observed that the stable knock down of endogenous Let-7 improved the % cell viability during the later stages of the culture in 24-well plate format for some clones, in contrast to the NC sponge clones. Greenlees et al. (2014) observed a similar improvement in cell viability during the later stages of culture. They employed anti-let-7a inhibitor molecules that led to the improvement in cell viability as well as the specific protein productivity in two CHO cell-lines, in comparison to the controls (Greenlees et al. 2014). Moreover, increased cell survival in two carcinoma cell lines, MT-1 and HeLa, was also reported with the use of Let-7 sponge vector (Yang et al. 2012a). In addition to the growth data, we also observed increased volumetric SEAP productivity in the Let-7 sponge in comparison to NC sponge mixed pool of cells, which was due to enhanced specific productivity.

Furthermore, we were interested in confirming this enhanced viability data in scale-up studies as well. We could not choose our best performing clones on the basis of both the growth and productivity data due to loss of the SEAP transcript in both sets of NC and Let-7 sponge clones. From our speculation, we believe this could have occurred due to changes in the methylation state of the SEAP transgene resulting in silencing, which was further confirmed by qPCR of SEAP mRNA expression levels. We observed the levels of SEAP transgene were 25-fold lower in both the stable NC and Let-7 sponge clones compared to the reference control population. The loss of SEAP transgene in both the NC and Let-7 sponge clones is inexplicable since all the clones from both NC and Let-7 panels were routinely pulsed separately with hygromycin and geneticin containing media. However, production instability due to loss of transgene expression is not a novel phenomenon since CHO cells are genetically instable. The loss of cell-specific productivity during sub-culture is completely unpredictable in that it may occur, it may not occur or it may occur after several passages. For instance, Heller-Harrison et al. (2009) reported that 8-63 % of all recombinant CHO cell lines suffered from production instability. Moreover, the cell lines established using GS and DHFR based gene expression systems are more susceptible, whether in the presence or absence of selection agent, to loss of transgene expression (Fann C.H. et al. 2000, Jun S.C. et al. 2006). The molecular mechanisms underlying mosaicism or variegation as well as the complete cessation of transgene expression have not been completely deciphered. However, methylation of CpG islands, thereby inhibiting transcription factors recruitment or promoting transcription repressor binding could

result in epigenetic gene silencing. Promoter methylation and other histone modifications such as deacetylation, hypoacetylation of DNA proteins are among the reported mechanisms responsible for transgene silencing in CHO and other mammalian cells as well (Yang et al. 2010). For instance, recently, CHO cell line expressing a recombinant protein, established independently of any of the traditional gene amplification processes, was reported to have its recombinant protein-encoding gene silenced due to hypoacetylation of H3 histone protein (Paredes et al. 2013). To further investigate the reason for SEAP transgene loss, was beyond the scope of this project, therefore, no such studies were conducted.

In the absence of productivity data, we chose a set of four clones from each NC and Let-7 sponge stable clonal panels. This choice was based on improved VCD on Day 3 of the culture period and cell viability on Day 6. We observed all the clones failed to demonstrate improved cell viability and VCD. In summary, none of the Let-7 clones exhibited statistically significantly improved growth characteristics in comparison to the NC sponge stable clones. The improved cell viability for two clones in 24-well plate format could have been a result of either clonal variation or 24-well plate versus 5 mL tube growth format difference. It has been reported that commercially available adhesive plate seals could result in two outcomes, in terms of volume and oxygen transfer rate i.e. (i) plate seals wherein volume can reduce due to evaporation (edge effects in shallow plates, mostly), but oxygen transfer is comparable to that of unsealed plates, and (ii) plate seals wherein volume preservation is higher (no loss due to evaporation), but oxygen transfer is slower (Zimmermann et al. 2003, Patel, Tuckerman and Dong 2005). We surmised that in 24-well plate format due to edge effect (since both the clones that showed improved cellular viability were around the edge of the plate) resulted in evaporation of media and accounted for the improved cellular viability, as a calculation error. Chaturvedi et al. (2014) recently reported that in 24-well shallow plates the small volume i.e. 0.7 mL to 1 mL capacities could account for VCD calculation inconsistencies due to multiple sampling reducing the volume further, in addition to evaporation (Chaturvedi et al. 2014).

4.3 Validation of the sponge technology: Understanding the reason for 'No phenotypic outcome'

With the aforementioned observations, we were interested in verifying whether the Let-7 sponge vector in the Let-7 sponge clones was effectively and specifically binding endogenous Let-7 or not. For this purpose, we transiently transfected two of our clones with Let-7 mimics and non-specific control miRNA mimics. We observed that the transient up-regulation of Let-7 induced further significant reduction in the GFP signal fluorescence in the Let-7 sponge expressing cells but not in the control cells. Therefore, these observations confirmed that Let-7 sponge was indeed specifically binding endogenous Let-7, thereby reducing the endogenous Let-7 levels available to bind to their mRNA targets. Both the tested Let-7 stable clones, upon exogenous Let-7 mimics introduction, displayed significantly reduced GFP signal fluorescence. Therefore, it was safe to say that theoretically all Let-7 clones had enough sponge transcript copies to fully sequester the endogenous Let-7 available. Similar to our study, this method of transfecting exogenous molecules for sponge validation have been performed in our laboratory previously by Sanchez et al. (2014), Kelly et al. (2015a) and Kelly et al. (2015b) for miR-7, miR-23b and miR-34, respectively.

In order to further confirm that Let-7 is deflected from its usual function upon Let-7 depletion by the Let-7 sponge in CHO cells, we conducted two separate real time qPCR experiments for the quantitation of Bcl2 gene and CDK6 gene expression levels. Both of these two genes are marked as predicted genes in www.chogenome.org database. We examined these two predicted and RT-qPCR validated targets of Let-7a due to their potential roles in a myriad of cell types and disease settings in regulation of cell-cycle pathways, in humans (Garzon et al. 2007, Johnson et al. 2005). The expression levels of both of these genes, in Let-7 sponge clones, were more or less similar to the controls.

After the previous observations in the Let-7 mimic transfection experiment versus Bcl2 and CDK6 expression level quantitation, we were interested in characterising the manner of Let-7 sponge in sequestering the endogenous Let-7 in CHO cells. For this purpose, we quantified the amount of mature endogenous Let-7a expression in the stable Let-7 clones. To our surprise, we observed that the stable Let-7 sponge clones had virtually similar endogenous Let-7a levels as the NC stable sponge clones as

measured by RT-qPCR. Moreover, one of the clones showed only slightly increased, although statistically insignificant, endogenous Let-7 expression levels, in comparison to all NC stable sponge clones. Similar observations, wherein the endogenous mature miRNA expression levels in stable miRNA sponge expressing CHO cells were higher in contrast to control sponge expressing CHO cells was reported by (Yang et al. 2012a, Sanchez et al. 2014, Kelly et al. 2015a, Kelly et al. 2015b) for Let-7, miR-7, miR-23b and miR-34, respectively. However, these findings of slightly higher or similar endogenous Let-7a levels in stable Let-7 clones seem inconsistent with the results of gain-of-function experiment using Let-7 mimics, which showed that the sponge was effective and specific to Let-7 binding in the stably expressing Let-7 CHO clones. The somewhat increased or similar Let-7 levels in the Let-7 sponge expressing cells could also be explained by the fact that endogenous Let-7 are single-stranded molecules bound to the Let-7 sponge, which could dissociate from the sponge during total RNA extraction process i.e. the sponge loses its sequestration capability outside the cellular context. Subsequently, during the RT step the specific Let-7a primer could bind to the now-released Let-7a miRNA resulting in an apparent slight up-regulation of Let-7a levels or similar Let-7a levels in Let-7 sponge clones, compared to control sponge clones (Sanchez et al. 2014).

Recently, a phenomenon known as target-mediated miRNA protection (TMMP) has come to light. TMMP may enable miRNA targets to protect their cognate miRNAs against cytoplasmic nucleases such as exoribonuclease XRN-2 (in *C.elegans*) (Chatterjee et al. 2011) and XRN-1 (in humans) (Bail et al. 2010), hence leading to accumulation of mature target-bound miRNAs (**Fig. 4.1**). Briefly, miRNA turnover is regulated by the XRN-2 via TMMP. Due to mRNA target availability, the active turnover of miRNAs and consequently, the degradation of miRNA duplex components are prevented (Chatterjee and Großhans 2009, Chatterjee et al. 2011). This could be very well be the case with our study wherein the sponge was efficiently soaking-up and protecting endogenous Let-7, however, the cells were compensating for the lack of the endogenous Let-7 levels in order to maintain normal cellular homeostasis through the experimentally validated mechanism of TMMP. It has also been reported that target availability could enforce accumulation of certain miR passenger (miR*) strands through TMMP. This could mean that independent of the thermodynamic asymmetry of miR: miR* duplex, for certain miRNAs the bias between miR and miR* (Chatterjee and Großhans 2009, Chatterjee et al. 2011) can

occur after incorporation of Ago into RISC. This adds more complexity to the regulatory role of miRNAs whereby, depending on mRNA availability, could result in further expression of a miRNAs and/or the miR: miR* ratio is maybe altered (Chatterjee et al. 2011). From these observations, it would be interesting to quantitate Let-7 levels immediately (or after fixed time intervals) after endogenous Let-7 sequestration by the Let-7 sponge to assess and evaluate the role of TMMP in protecting endogenous Let-7 levels and the generation of the mature form of Let-7. Such a study could allow for a better understanding of TMMP in influencing miRNA levels and at what magnitude and after what time post-sponge transfection.

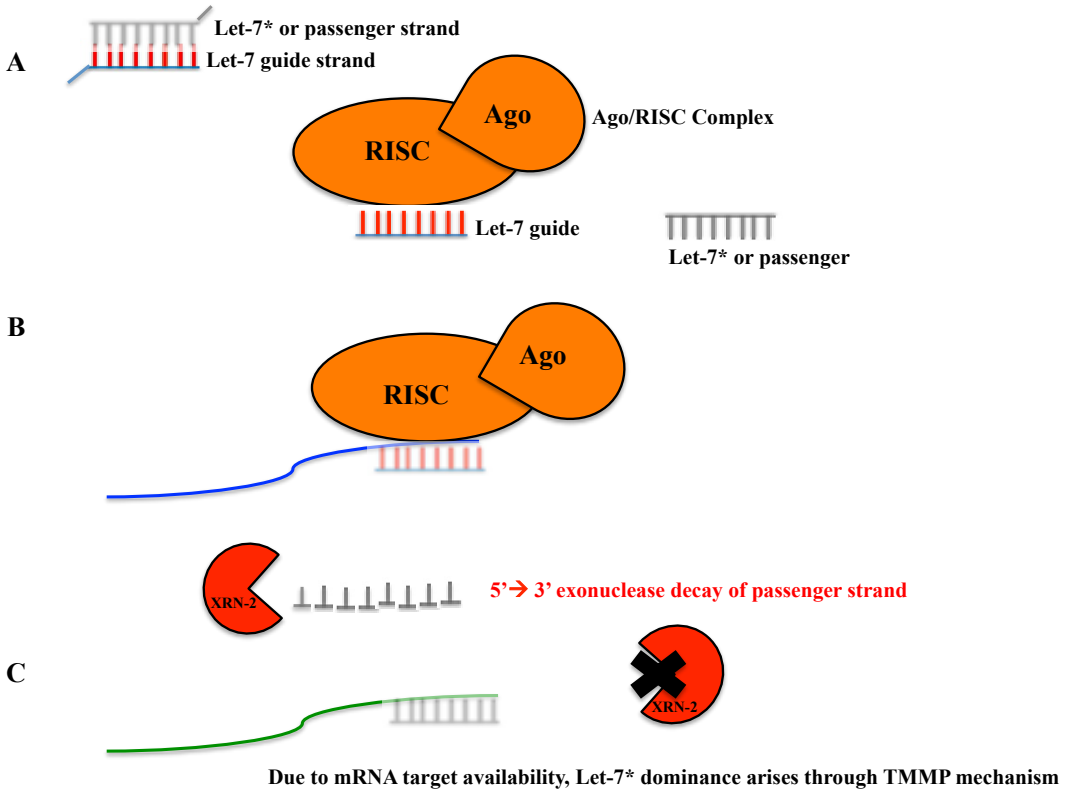


Figure 4.1: Outlines the process of normal miRNA regulation of their targets and the process of TMMP. (A) Shows the thermodynamically stable pre-miRNA (Let-7 in our case) duplex with 3' nt overhangs. (B) Normally, the guide strand (Let-7 guide in our case) is loaded onto the Ago/RISC complex for the translational repression of their mRNA targets and the passgenger (miR*, or Let-7* strand in our case) is degraded through cytoplasmic exoribonuclease enzyme, XRN-2 (*C.elegans*) and/or XRN-1 (in humans) through its 5'→ 3' nuclease activity. However, (C) TMMP mechanism when the mRNA targets are available, then passenger strand accumulation prevents their decay by XRN-2 enzyme by abolishing XRN-2 activity. This image is reproduced from Thesis of Dr. Paul S Kelly 2013.

In addition to TMMP, there have been reports suggesting that for highly abundant miRNAs like the Let-7 family (12 members distributed across 8 loci in mouse), the genetic inactivation of miRNA results in very modest de-repression of its direct targets i.e. less than 2-fold. On many occasions, such modest de-repression of miRNA targets are well tolerated by cells, hence resulting in no obvious phenotypes. These observations sparks the idea that miRNAs act as rheostats to fine-tune the expression of hundreds of gene in order to reinforce cellular fates, rather than acting as genetic switches. As rheostats, miRNAs are able to buffer the cell against stochastic fluctuations that could perturb normal cellular homeostasis and provide robustness in spite of internal or external perturbations (Kitano 2004, Klein et al. 2007). We speculate that the miRNA buffering mechanism of cells (via feedback loops) against any environmental stress and/or miRNA manipulation (e.g. sponge-based miRNA depletion) includes the known and experimentally validated TMMP mechanism (Bartel and Chen 2004, Hornstein and Shomron 2006, Baek et al. 2008, Mu et al. 2009, Ebert and Sharp 2012).

In contrast to our results of no significant change in growth profile with CHO cells expressing the stable Let-7 sponge transcript, Greenlees et al. (2014) have reported improved CHO cell phenotypes, as mentioned above, using anti-Let-7a synthetic molecules (Greenlees et al. 2014). However, reports have suggested that ectopic expression studies through the use of inhibitors molecules, often lead to supra-physiological levels of miRNA and therefore, stronger repression of its targets and sometimes repression of mRNAs that might not even be their biological targets (Baek et al. 2008, Selbach et al. 2008, Guo et al. 2010, Ebert and Sharp 2012).

4.4 Let-7 as a potential candidate for CHO cell engineering

Although our study to stably knockdown Let-7 in CHO cells did not induce enhanced growth and/or productivity characteristics, it did improve cell viability. However, enhanced viability could not be replicated in the scale-up studies. In the absence of SEAP expression data, we could not test whether Let-7 stable knockdown impacted productivity. There are studies reporting that knocking down Let-7 family members resulted in > 60 % improved antibody specific productivity in two different CHO cell-lines, and improved cell viability during the later stages of culture (Greenlees et al. 2014). Moreover, the manipulation of Let-7 using sponge decoy vector has been

reported to increase cell survival in carcinoma cell lines (Yang et al. 2012b). We believe that the productivity data would have allowed us to pick clones that maybe performed better in contrast to control clones. However, due to epigenetic gene silencing of the SEAP transgene, we were unable to examine this. As mentioned above, Let-7 is the most highly studied miRNA and plays a vital role in various cellular pathways such as cell division, cell proliferation and induction of apoptosis (Johnson et al. 2007). Moreover, we feel that Let-7 manipulation by means of other genome modifying techniques such as CRISPR-Cas9 to knock out Let-7 locus could be interesting to explore. In contrast to stable Let-7 depletion, the knock out approach differs in that the gene is knocked out in the subsequent generations in the particular cell line and could provide an effective alternative to stable miRNA-based depletion, which has the drawback of compensation for the depletion of miRNA through TMMP.

Keeping this in mind, in the second part of this project we employed a novel genome editing system to manipulate a widely published miRNA with a proven track record in CHO cells to improve CHO cell growth and productivity characteristics.

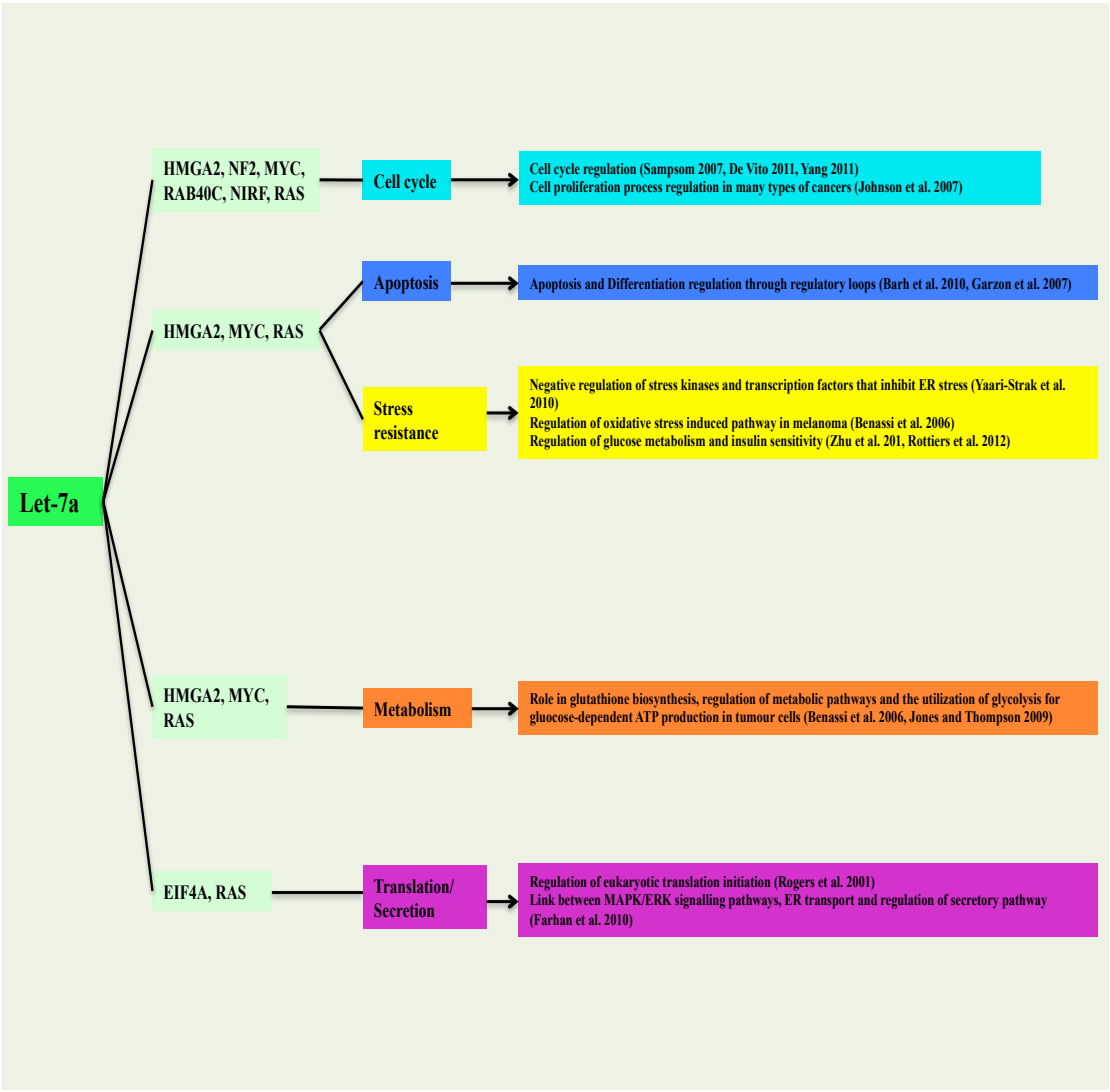


Figure 4.2: Outlines the impact of Let-7a alterations in effecting multiple potentially important cellular pathways by regulating an array of genes involved in these pathways in two antibody producing CHO cell lines. Image reproduced from an article by Greenlees et al. 2014.

4.5 Exploring the CRISPR-Cas9 system for CHO cell engineering

In contrast to other genome editing techniques such as ZFNs, TALENs and various other nuclease-based techniques, CRISPR-Cas9 stands out as the latest 'go-to' gene-editing tool. This is essentially due to its greater specificity, efficiency, time and cost economy and less cytotoxic effects when applied to mammalian cell lines. CRISPR-Cas9 also allows for multiplex genome engineering to induce functional gene knockouts in the genome at multiple sites by using an all-in-one CRISPR-Cas9 system, which is a tedious, expensive and time-consuming process with ZFNs and TALENs. All these points make CRISPR-Cas9 the method of choice for the next generation of genome editing in mammalian cells (Cong et al. 2013, Mali et al. 2013b, Sakuma et al. 2014), and in regard to this project, for CHO cell engineering specifically. CRISPR-Cas9 technology is beginning to be explored in CHO cells in order to improve bioprocess-related and therapeutic protein quality related characteristics. Working towards this goal, we aimed to implement the CRISPR-Cas9 technology in CHO cells in this project, firstly, in a test CHO cell line (CHO-eGFP cell line), then finally to induce miR-7a-5p gene locus knockout in industrially relevant CHO cel lines with a view to improve CHO cell growth and productivity characteristics.

4.5.1 Implementation of CRISPR-Cas9 system in CHO-eGFP cell line

For the proof-of-concept, we employed CRISPR-Cas9 to attempt eGFP gene knockout in the CHO-eGFP cell line (or eGFP cell line). This cell line has very high levels of GFP expression and contains one or more copies of eGFP gene, possibly integrated into rare transcriptional "hot spot" in the genome. It expresses stable eGFP invariably as confirmed by our 5-day mean GFP expression analysis.

Initially, we used two 20 nt long sgRNAs designed to target eGFP gene at 5'- and 3'-end, respectively, cloned into a pCas9 CRISPR vector separately. We transfected the eGFP cells with these two eGFP targeting CRISPR systems, however, we did not observe satisfactory reduction in mean GFP expression in either of the wells transfected with eGFP 1 and eGFP 2 CRISPR system. We conducted a transient co-transfection experiment, in which we found that the eGFP CRISPR was efficient in knocking out the eGFP gene sequence of peGFP plasmid. However, similar

satisfactory reduction in mean GFP expression could not be achieved in eGFP cells. We believe that whilst designing the sgRNA sequences 20 nt long according to the manufacturer's guidelines we appended an extra 'G' nt at 3-end of the guide sequence (in addition to appending an extra 'G' at the 5'-end of the guide sequence), making the sgRNA guide length 21 nt long. The sgRNAs with 21 nt long sequence has been reported to be far less effective than 20 nt sgRNA CRISPR systems (Ren et al. 2014). Moreover, even after appending extra 'G' nt at the 5'-end of the guide sequence as preferential and efficient U6 transcription start site, certain sgRNAs are reported not functioning on rare occasions for unknown reasons.

4.5.2 Opening up the prospects for the generation of isogenic CHO cell lines via CRISPR-Cas9 technology

In relation to this project, it was imperative for us to demonstrate CRISPR-Cas9 genome technology can be implemented for the task of knocking out the eGFP gene in CHO cells. We used a different CRIPSR-Cas9 expression vector system to induce eGFP gene knockouts in the cells. We designed four eGFP targeting sgRNAs i.e. sgGFP 1A, sgGFP 1B, sgGFP 2A and sgGFP 2B and cloned into the px459 vector.
With these eGFP CRISPRs we demonstrated high eGFP knock out efficiency as mean GFP expression was reduced by more than 25-fold in eGFP CRISPR-treated samples. We also reported indels frequency in the range of 40-50 %. Moreover, from sequencing results we observed that the majority of eGFP CRISPRs induced deletions, as opposed to insertions. Ronda et al. (2014) reported similar percentage of indels frequencies (i.e. 47 %) in the COSMC gene in CHO cells. The sgGFP 19 nt long demonstrated highest ~ 50 % indel frequencies and with more than 60 % of the TOPO samples with indels in double-digit numbers, as opposed to sgGFP 20 nt long, which mostly generated indels of sizes 1 or 2 bp. The sgRNAs that Ronda et al. (2014) used also generated single base pair insertions, in the two targeted genes, more frequently than big indels (Ronda et al. 2014). From these findings we infer that the length of sgRNA used did not impact the performance of sgRNAs to a great deal. However, the location of the genomic site to be targeted was critical since we observed marginally higher eGFP CRISPR targeting efficiency with the eGFP 1 CRISPRs targeting the 5'-end of the eGFP gene. This could be due to the fact that sgGFP 1A and sgGFP 1B were designed to target the 5'end of the eGFP gene, where

a frameshift mutation is more likely to result downstream effect. On the other hand, eGFP 2 CRISPRs targets the 3'-end of the eGFP gene, wherein, for instance, in case of a single bp indels a part of the eGFP sequence would still be transcribed and could result in a functional GFP (Lawhorn, Ferreira and Wang 2014). It has been reported that mutations such as indels in the ORF of GFP gene differentially reduce the GFP mRNA transcription rate, translation efficiency due to protein instability but still allows expression of very faint GFP (Sacchetti et al. 2001). Additionally, sequence surrounding the translation start site (i.e. near the 5'-end) such as the Kozak motif if disrupted could directly affect its transcription and translation abilities (Kozak 1991, Afshar-Kharghan et al. 1999). With the successful implementation of CRISPR-Cas9 to specifically target the stably integrated eGFP gene in CHO-eGFP cell line, we demonstrated that the CRISPR-Cas9 system could efficiently target genes in CHO cells. Moreover, another spin-off project from this proof-of-concept study could be the use of eGFP CRISPR to induce eGFP functional knockout in CHO-eGFP cell line, followed by high fidelity HDR-based transgene insertion with donor DNA (encoding another GOI) (Smith et al. 2008, Chu et al. 2015, Lee et al. 2015, Maruyama et al. 2015). Therefore, making CRISPR-Cas9 as an attractive option for the generation of a 'truly hyperproducing' isogenic CHO cell line that has homogenous and sustained expression of the introduced GOI into a transcriptionally 'active zone' of the CHO genome. This could also save time and labor by facilitating generation of isogenic recombinant CHO cell lines without using the traditional DHFR or GS-based selection and amplification system.

4.5.3 CRISPR-Cas9 genome editing technology to manipulate miR-7 for improved CHO cell phenotypes

In the first half of the project, we employed sponge technology to stably deplete endogenous Let-7 levels in CHO-K1 SEAP expressing cells. However, due to possible compensation via TMMP for the depletion of endogenous Let-7 levels, we were not able to observe any significantly improved growth and/or productivity phenotypes. For this reason, we chose to employ CRISPR-Cas9 system to completely knockout a previously well-characterised miRNA i.e. miR-7, in industrially relevant CHO cell line. As opposed to miR-7 sponge stable depletion (which already has been reported in our laboratory), we were interested in completely knocking out the miR-

7a-5p locus in CHO cells using CRISPR-Cas9. miR-7 is reported to be one of the key regulators of cell processes such as cell proliferation/division and apoptosis. Numerous reports have suggested miR-7 role as tumour suppressor in cancer of different types (Webster et al. 2009, Giles et al. 2013, Liu et al. 2013). However, very recently, miR-7 has also been reported as an oncogene since its up-regulated in renal carcinoma. These reports demonstrate cell- and organ-specificity of miR-7 action (Gu, Huang and Tian 2015b).

From the biotechnological point of view, the importance of miR-7 as a potential candidate for CHO cell engineering has been widely published due to the fact that it is an important cell cycle regulator miRNA. As reported by Sanchez et al. (2013) miR-7 triggers CHO cell cycle arrest by targeting a multitude of cell cycle regulator components such as S-phase kinase-associated protein-2 (Skp2), myc and $p27^{KIP1}$ for temporary growth arrest by down-regulating pro-apoptotic factors, DNA repair factors, as well by up-regulating anti-apoptotic factors (Sanchez et al. 2013). These findings demonstrated what a critically important role miR-7 plays in the regulation of cell cycle. Barron et al. (2011) transiently overexpressed miR-7 to induce cell cycle growth arrest and observed the temperature shift phenotype, resulting in increased normalised productivity (Barron et al. 2011b). These findings are further supported from miR-7 sponge work by Sanchez et al. (2014), who reported 3-fold improved cell proliferation and 2-fold increase in protein productivity by stably depleting endogenous miR-7 levels in CHO cells (Sanchez et al. 2014). Therefore, for this half of the project we were interested in applying a different gene engineering strategy to knockout the miR-7a alleles altogether. The reason for choosing miR-7a to be knocked out, and not miR-7b, also stems from the data published by Hackl et al. (2011) in CHO cell lines. Their next generation sequencing data reported stronger expression of miR-7a in comparison to miR-7b in CHO cell lines (Hackl et al. 2011a). In order to cause miR-7a-5p locus knock out in CHO-K1 1.14 cells, we designed four sgRNAs CRISPR-Cas9 systems targeting the miR-7a-5p sequence. Keeping in mind that targeting an ORF of a protein gene is easier since only one bp knock out is enough to disrupt the reading frame e.g. in the case of eGFP. For miR-7 (with two alleles located on two different chromosomal loci), where there are possibly four copies of the gene with functional redundancy due to paralogs, we assumed that it would be a harder task to knock out all the four copies of a non-coding gene. Moreover, in order to render miRNAs functionless, we required significant deletions

in the mature miR-7a-5p sequence as opposed to single base indels. Therefore, we were interested in knocking out the mature or loop sequence in the miR-7a-5p sequence so that subsequently it would not be recognised by the Dicer complex for processing. No miR-7 would be available to bind to its target mRNAs once the mature sequence or loop structure is disrupted, and such a disruption would make it functionless. miR-7 is a mirtronic non-coding RNA molecule, whose pre-mir-like transcripts by-pass Drosha processing (Berezikov et al. 2007). The four miR7 CRISPR-Cas9 systems, when transfected into 1.14 cells, resulted in 35-40 % targeting efficiency. Inexplicably, the location of the target site showed preferentially higher CRISPR activity since sgmiR7(s) 20 nt in length targeting downstream of the mature miR-7a-5p sequence demonstrated highest indel frequencies, and were marginally better than the sgmiR7(s) targeting upstream of the mature miR-7a-5p sequence. Reports have suggested that the 20-nt sgRNA CRISPR systems work better in most cases. However, many studies have also shown the importance of certain major parameters such as the number, position and distribution of nucleotide mismatches between the sgRNAs and genomic DNA in mammalian cells, can impact the targeting efficiency significantly (Fu et al. 2013, Hsu et al. 2013, Mali et al. 2013a, Pattanayak et al. 2013, Wu et al. 2014). For instance, Fu et al. (2014) reported that truncated 18 nt long sgRNAs showed similar targeting efficiency as 19 or 20 nt truncated sgRNAs (Fu Y. et al. 2014).

Here as well, sequencing results reported that majority of the indels generated by all the four CRISPRs were single base pair insertions, as observed for protein coding COSMC gene in CHO cells by Ronda et al. (2014). From this study, the knowledge and further detailed analyses of the types of indels induced by these four sgmiR7 based CRISPRs could allow for the design of better (maybe multiplex) CRISPR systems in future, and such systems could be used to induce bigger deletions of the mature miR-7a-5p sequence.

4.5.4 Picking the population for the isolation of potential single cell CHO clone/s

We were interested in picking clones that contained large indels in the mature miR-7 and/or loop sequence from the mix of these four 1.14 cell populations transfected separately with the four aforementioned CRISPRs. We believed that in order to have a functional knockout of miR-7, we needed clones that showed a large part of the mature miR-7a-5p sequence mutagenized and/or deleted due to indels induced by the CRISPR-Cas9 system. In these clones, the altered sequence (with the seed region deleted or random bp inserted) would not act as Dicer substrates and would not be processed, and hence would not lead to translational repression of their biological targets. The sgmiR7 2B CRISPR with 40% indel frequencies and 30 % of the TOPO sequencing samples demonstrating large miR-7a-5p sequence deletions, was the 1.14 mixed pool of cells chosen to be analyzed further due to higher probability of finding a clone that had a part or the whole of mature miR-7a-5p sequence knocked out.

4.5.5 Partially down regulated miR-7 expression in single cell clones

Following limited dilution cloning, we screened 38 single clones for endogenous mature miR-7a expression levels. Encouragingly, we observed a range of single cell clones demonstrating 5-fold and more down-regulation in endogenous mature miR-7a expression levels. A few of the clones showed more than 10-fold down regulation of mature miR-7a expression levels. However, for a few selected clones the mature miR-7a expression levels were re-analysed as a rigorous validation step using fresh total RNA samples. Upon re-analysis, the mature miR-7a levels were less down regulated for three of the clones (i.e. Clone 6, Clone 14 and Clone 30), however, the levels in Clone 25, Clone 34 and Clone 36 remained unchanged, in comparison to levels in the control cells. This change in mature expression level could be potentially due to timing of the total RNA extraction. Nevertheless, theoretically we were expecting to observe significant down regulation of mature miR-7a in these clones, as opposed to partial or no change of mature miR-7a expression levels. From the studies by Sanchez et al. (2014), Kelly et al. (2015a and 2015b) on the stable depletion of miR-7, miR-23b and miR-34, respectively, they observed compensation (via TMMP) in response to the depletion of these miRNAs, whilst using specific miRNA sponge for their studies. (Sanchez et al. 2014, Kelly et al. 2015a, Kelly et al. 2015b). We speculated

that due to miR-7a knockout or disruption by the CRISPR-Cas9, the cells were perhaps compensating for the perceived lack of miR-7a by increasing the transcriptional rate of other alleles. This speculation is true for many miRNAs as they are vital for the normal functioning in the cell and can sense any internal or external perturbations that can have a large downstream effect. miRNAs (e.g. here miR-7) regulate many critical cell cycle related processes and confer cells with phenotypic robustness by buffering the cell against any cell-to-cell variability arising from any stochastic fluctuations in gene transcription through unknown feedback or feed forward loop mechanisms (Bartel 2004a, Hornstein and Shomron 2006, Klein et al. 2007, Ebert and Sharp 2012). Moreover, we hypothesized that the compensation might be contributed by miR-7 paralogs i.e. miR-7b. The presence of these paralogs with high degree of sequence homology could compensate for the knockout of a miR-7a allele since miR-7b is a functional paralog for miR-7a and has the same seed region with similar affinities for miR-7 target mRNAs. The presence of miRNAs with redundant functions, in certain scenarios, means that several members of the family need to be deleted in order to observe any phenotypic consequence. For instance, in vertebrate multiciliated cells, six miRNAs comprising the conserved miR-34/449 family encoded by miR-34a, miR-34b/c, and miR-449a/b/c- can coordinately repress cp110 expression during ciliogenesis. Animals with deleted single genes of miR-34/449 family were viable and phenotypically normal, however, deletion of all three loci leads to animals' survival with defective ciliogenesis (Bao et al. 2012, Song et al. 2014). We also quantitated the nascent transcript of miR-7 paralog i.e. pre-mir-7b expression levels in a few selected clones. Based on next generation sequencing data from Hackl et al. (2011), we understood that miR-7a-5p is more strongly expressed in comparison to miR-7b-5p across seven CHO cell lines (Hackl et al. 2011b) and expected that this latter speculation about miR-7b to not be responsible for the compensation of loss of miR-7a.

We did not observe any significant differences in the expression of pre-mir-7a and pre-mir-7b in Clone 14 and Clone 34, in comparison to levels in the control cells. Clone 34 showed slightly increased levels of pre-mir-7b, although insignificant. Therefore, the speculation of possible compensation of miR-7a and/or miR-7b through increased pre-mir-7a and/or pre-mir-7b transcriptional rates through some feedback mechanism was unfounded.

4.5.6 Heterozygous deletion of miR-7a

The miR-7a-5p sequence region for a few selected clones (i.e, Clone 6, 14, 25, 30, 34 and 36) was analysed by Sanger sequencing. Of the 6 clones, we found that Clone 14 contained the wild-type miR-7a-5p sequence, while Clone 34 contained a heterozygous deletion of miR-7a. The increased mature miR-7a expression levels in Clone 34 was rather surprising since even after heterozygous deletion of miR-7a, the mature miR-7a was expressed in abundance. One possible explanation for this could be that the deleted miR-7a copy was silent due to gene imprinting phenomenon (McGrath and Solter 1984, Bartolomei and Ferguson-Smith 2011). We also speculate that even if we were able to induce a homozygous deletion of miR-7a alleles, we believe such a deletion was lethal to cells and these cells with homozygous miR-7a deletions never survived, hence never analysed. This could be true in the case of miR-7, as it plays a critical role in cellular and many other pathways (Sanchez et al. 2013). Hence, maybe an absolute requirement for cells to survive. The another possible, however, less likely explanation could be the fact that the levels of pre-mir-7b were slightly high in Clone 34, although insignificant, as a compensation mechanism by cells to cater for the loss of one miR-7a copy in order to buffer the cell against any internal/external cellular perturbations and providing ‘robustness’ to cell (Bao et al. 2012, Ebert and Sharp 2012, Song et al. 2014).

For Clone 14, the presence of wild-type miR-7a-5p sequence was rather surprising since we observed ~ 5-fold down regulation in the mature miR-7a expression levels. In order to address the anomaly between the sequencing and qPCR data for Clone 14, we performed a validation study to quantitate the expression levels of Psme3 and Rad54L as two potential miR-7 downstream targets, previously validated in our laboratory (Sanchez et al. 2013). In theory, Clone 14 would result in increased levels of expression of Rad54L and Psme3 compared to Clone 34 (if the qPCR data was not a false positive). Since Clone 14 had 5-fold less mature miR-7a expression than Clone 34, in which the mature miR-7a expression levels were insignificantly down regulated. The down-regulated expression of both the genes in Clone 14, in comparison to control confirmed that it contains mature miR-7a-5p sequence intact and the RT-qPCR result for partial down regulation of mature miR-7a was a false positive. For Clone 34, the slightly increased levels of Rad54L and the sequencing

results further suggest the fact that it contains a heterozygous deletion in one of the miR-7a copy while the other was still expressed due to which the overexpression of both of these genes was not to a greater magnitude. However, we also believe that some genes require more than one miRNA to bind to regulate them. Perhaps the reduction in miR-7 via knockout strategy was not enough to observe upregulation of these two target genes. Moreover, these genes were validated using mimic molecules, which could often lead to supra-physiological levels of miRNA and therefore, stronger repression of its targets, in contrast to stable depletion via miR-7 sponge or in our case of knocking miR-7a out (Baek et al. 2008, Selbach et al. 2008, Guo et al. 2010, Ebert and Sharp 2012).

We strongly believe that more questions regarding the nature and number of miR-7a copies knocked out, in the Clone 34 and rest of the 32, clones will be answered through next generation sequencing, which was beyond the scope of this project. As for future work, we believe it would be interesting to evaluate Clone 34 phenotype in a batch culture study. PCR and/or Sanger sequencing analysis of the rest of the 32 clones would allow us perhaps to find more interesting clones. We also believe that increasing the number of single cell clones, from the mixed CHO 1.14 population chosen here, to be analyzed using RT-qPCR, PCR, and Sanger as well as maybe next generation sequencing would increase the probability of isolating clone/s containing a homozygous miR-7a knockout. If this is proved unsuccessful it may suggest that homozygous miR-7a deletion is lethal, and cannot be tolerated by cells.

4.6 Multiplex CRISPR-Cas9 tools for the generation of isogenic CHO cell lines

In addition to single guide CRISPR-Cas9 systems, we designed duplex CRISPR-Cas9 system- GFP Duplex CRISPR containing sgGFP 1B and sgmiR7 2B guides. We achieved a range of indel efficiencies with the two-guide CRISPR-Cas9 system in-one, i.e. one guide resulting in 35-40 %, while the other yielded ~ 20 % targeting efficiency. As mentioned earlier targeting an ORF of a protein gene is easier since only one bp knock out is enough to disrupt the reading frame of e.g. in the case of eGFP gene. For non-coding genes such as in the case of miR-7 (with two alleles located on two different chromosomal loci), where there are possibly four copies of the gene, we infer that it was a harder task to knock out all the four copies, hence, resulted in low indel efficiency of the second guide. We believe that multiplexing

CRISPR-Cas9 systems with further improvement in designing strategy could be an attractive tool to target other relevant miRNAs or any gene in one-go (Cong et al. 2013). From our single guide miR7 CRISPR systems, we believe it is necessary to target miR-7a-5p sequence to introduce significant indels with greater frequency since single indels might not be enough to impact miRNA expression, hence, to render them functionless. For miRNA knockout we believe multiplex CRISPR-Cas9 technology could, particularly be useful in deleting whole mature miRNA sequence in CHO cells. When these CHO cells with large knockouts are evaluated for bioprocess-related phenotypes could yield interesting phenotypes in terms of growth and protein productivity.

Irrespective of the technology used, miRNA manipulation to improve CHO cell phenotypes is a very promising tool to be further examined. Their striking ability to regulate hundreds of genes involved in complex cellular pathways such as cell cycle, metabolism, secretion and apoptosis, without imposing any burden on the cell's translational machinery, make them attractive tools for the CHO cell engineering. Rightly so, investigators are continuing the process of evaluating miRNA roles in CHO cell cultures through the analysis of up/down regulation of endogenous miRNAs in CHO production cultures. Several groups have also successfully implemented the use of ectopic expression of miRs (mimics) or anti-miRs (inhibitors) synthetic molecules in CHO cells to improved bioprocess-related phenotypes (Greenlees et al. 2014). However, until recently the lack of publically available CHO genomic data, as well as miRNA seriously stymied the progress of using miRNA for CHO cell engineering. We hope that with the availability of CHO genomic data and with novel genome-editing technology to manipulate miRNA in CHO cells we could achieve this objective.

Conclusions

Study I: Stable depletion of endogenous Let-7 miRNA in CHO-K1 SEAP cells using sponge "decoy" vectors

- Stable depletion of endogenous Let-7 levels was achieved and validated in CHO-K1 SEAP cells using a Let-7 sponge vector.
- Let-7 depletion did not enhance growth characteristics in CHO-K1 SEAP clones though there was some evidence of improved viability and SEAP productivity in mixed population before cloning.
- Impact of productivity could not be evaluated due to loss of SEAP transgene expression in CHO-K1 SEAP cells.

Study II: CRISPR-Cas9 genome editing technology to knockout miR-7a-5p sequence in industrially relevant CHO cell line

- 50 % targeting efficiency could be achieved using CRISPR-Cas9 to target a stably integrated eGFP gene in CHO-K1 cells.
- 40 % targeting efficiency could be achieved using CRISPR-Cas9 to target miR-7a-5p locus in mAb-producing CHO cell line.
- The length of sgRNA (19 or 20 nt) did not have a big impact on the nature and extent indels efficiency.
- The location of target site (i.e. 5' or 3'end of gene) did have an impact, with 5'-end targeting resulting in marginally better targeting efficiency.
- Isolated a clone containing heterozygous deletion of miR-7a.
- ~ 30 % targeting efficiency could be achieved with a Duplex CRISPR-Cas9 system in CHO-eGFP cell line.

Future Work

- Analyse the impact of stable Let-7 depletion in other industrially relevant protein producing CHO cell lines.
- Explore the potential of TMMP: By quantitating Let-7 levels immediately (or after fixed time intervals) after endogenous Let-7 sequestration by the Let-7 sponge to assess and evaluate the role of TMMP in protecting endogenous Let-7 levels and the generation of the mature form of Let-7. This study could allow for a better understanding of TMMP in influencing miRNA levels and at what magnitude and after what time post-sponge transfection.
- Use CRISPR-Cas9 system to knockout mature Let-7a sequence and/or other Let-7 family members in other relevant CHO cell lines using an efficient multiplex CRISPR-Cas9 system. Following this, studying the impact such knockout on CHO cell phenotypes in Batch and Fed-Batch mode of culture.
- Use the CRISPR-Cas9 system to target the eGFP, and subsequently, in conjunction with HDR introduce a GOI (another protein encoding gene) at that target site in CHO-eGFP cell line to establish a targetable hyperproducing CHO cell line.
- Evaluate the Clone 34 for growth and productivity characteristics.
- Analyse the rest of the 32 single cell clones using PCR and Sanger as well as next generation sequencing to study the number and type of miR-7a indels observed.
- Analyse more single cell clones from CHO 1.14 sgmiR7 2B mixed population using RT-qPCR, PCR, and next generation sequencing for the isolation of clone/s containing homozygous miR-7a knockout.

Bibliography

Aboobaker, A.A., Tomancak, P., Patel, N., Rubin, G.M. and Lai, E.C. 2005. Drosophila microRNAs exhibit diverse spatial expression patterns during embryonic development. *Proceedings of the National Academy of Sciences of the United States of America,* 102(50), pp.18017-18022.

Afshar-Kharghan, V., Li, C.Q., Khoshnevis-Asl, M. and López, J.A. 1999. Kozak sequence polymorphism of the glycoprotein (GP) iba gene is a major determinant of the plasma membrane levels of the platelet GP ib-IX-V complex. *Blood,* 94(1), pp.186-191.

Ambros, V., Bartel, B., Bartel, D.P., Burge, C.B., Carrington, J.C., Chen, X., Dreyfuss, G., Eddy, S.R., Griffiths-Jones, S., Marshall, M., Matzke, M., Ruvkun, G. and Tuschl, T. 2003. A uniform system for microRNA annotation. *RNA,* 9(3), pp.277-279.

Animal Cell Technology Industrial Platform (2012) Biopharmaceuticals produced with animal cell technology.

Ansai, S. and Kinoshita, M. 2014. Targeted mutagenesis using CRISPR/Cas system in medaka. *Biol.* 3, pp. 362-371.

Astley, K. and Al-Rubeai, M. 2008. The role of bcl-2 and its combined effect with p21CIP1 in adaptation of CHO cells to suspension and protein-free culture. *Applied Microbiology and Biotechnology,* 78(3), pp.391-399.

Auer, T.O., Duroure, K., De Cian, A., Concordet, J.-. and Del Bene, F. 2014. Highly efficient CRISPR/Cas9-mediated knock-in in zebrafish by homology-independent DNA repair. *Genome Research,* 24(1), pp.142-153.

Bachu, R., Bergareche, I. and Chasin, L.A. 2015. CRISPR-cas targeted plasmid integration into mammalian cells via non-homologous end joining. *Biotechnology and Bioengineering,*

Baek, D., Villén, J., Shin, C., Camargo, F.D., Gygi, S.P. and Bartel, D.P. 2008. The impact of microRNAs on protein output. *Nature,* 455(7209), pp.64-71.

Bagga, S., Bracht, J., Hunter, S., Massirer, K., Holtz, J., Eachus, R. and Pasquinelli, A.E. 2005. Regulation by let-7 and lin-4 miRNAs results in target mRNA degradation. *Cell,* 122(4), pp.553-563.

Baik, J.Y. and Lee, G.M. 2010. A DIGE approach for the assessment of differential expression of the CHO proteome under sodium butyrate addition: Effect of bcl-xL overexpression. *Biotechnology and Bioengineering,* 105(2), pp.358-367.

Bail, S., Swerdel, M., Liu, H., Jiao, X., Goff, L.A., Hart, R.P. and Kiledjian, M. 2010. Differential regulation of microRNA stability. *RNA,* 16(5), pp.1032-1039.

Bao, J., Li, D., Wang, L., Wu, J., Hu, Y., Wang, Z., Chen, Y., Cao, X., Jiang, C., Yan, W. and Xu, C. 2012. MicroRNA-449 and MicroRNA-34b/c function redundantly in murine testes by targeting E2F transcription factor-retinoblastoma protein (E2F-pRb) pathway. *Journal of Biological Chemistry,* 287(26), pp.21686-21698.

Barnes, D.E. 2001. Non-homologous end joining as a mechanism of DNA repair. *Current Biology,* 11(12), pp.R455-R457.

Barron, N., Kumar, N., Sanchez, N., Doolan, P., Clarke, C., Meleady, P., O'Sullivan, F. and Clynes, M. 2011a. Engineering CHO cell growth and recombinant protein productivity by overexpression of miR-7. *Journal of Biotechnology,* 151(2), pp.204-211.

Barron, N., Sanchez, N., Kelly, P. and Clynes, M. 2011c. MicroRNAs: Tiny targets for engineering CHO cell phenotypes? *Biotechnology Letters,* 33(1), pp.11-21.

Bartel, D.P. 2004a. MicroRNAs: Genomics, biogenesis, mechanism, and function. *Cell,* 116(2), pp.281-297.

Bartel, D.P. and Chen, C.-. 2004. Micromanagers of gene expression: The potentially widespread influence of metazoan microRNAs. *Nature Reviews Genetics,* 5(5), pp.396-400.

Bartolomei, M.S. and Ferguson-Smith, A.C. 2011. Mammalian genomic imprinting. *Cold Spring Harbor Perspectives in Biology,* 3(7), pp.1-17.

Bashirullah, A., Pasquinelli, A.E., Kiger, A.A., Perrimon, N., Ruvkun, G. and Thummel, C.S. 2003. Coordinate regulation of small temporal RNAs at the onset of drosophila metamorphosis. *Developmental Biology,* 259(1), pp.1-8.

Baskerville, S. and Bartel, D.P. 2005a. Microarray profiling of microRNAs reveals frequent coexpression with neighboring miRNAs and host genes. *RNA,* 11(3), pp.241-247.

BCC Research. 2012. Anitbody drugs: Technologies and global markets.

Becker, E., Florin, L., Pfizenmaier, K. and Kaufmann, H. 2010. Evaluation of a combinatorial cell engineering approach to overcome apoptotic effects in XBP-1(s) expressing cells. *Journal of Biotechnology,* 146(4), pp.198-206.

Becker, J., Hackl, M., Rupp, O., Jakobi, T., Schneider, J., Szczepanowski, R., Bekel, T., Borth, N., Goesmann, A., Grillari, J., Kaltschmidt, C., Noll, T., Pühler, A., Tauch, A. and Brinkrolf, K. 2011. Unraveling the chinese hamster ovary cell line transcriptome by next-generation sequencing. *Journal of Biotechnology,* 156(3), pp.227-235.

Bedell, V.M., Wang, Y., Campbell, J.M., Poshusta, T.L., Starker, C.G., Krug Ii, R.G., Tan, W., Penheiter, S.G., Ma, A.C., Leung, A.Y.H., Fahrenkrug, S.C., Carlson, D.F., Voytas, D.F., Clark, K.J., Essner, J.J. and Ekker, S.C. 2012. In vivo genome editing using a high-efficiency TALEN system. *Nature,* 491(7422), pp.114-118.

Beloglazova, N., Petit, P., Flick, R., Brown, G., Savchenko, A. and Yakunin, A.F. 2011. Structure and activity of the Cas3 HD nuclease MJ0384, an effector enzyme of the CRISPR interference. *EMBO Journal,* 30(22), pp.4616-4627.

Berezikov, E., Chung, W.-., Willis, J., Cuppen, E. and Lai, E.C. 2007. Mammalian mirtron genes. *Molecular Cell,* 28(2), pp.328-336.

Bernardi, C., Soffientini, U., Piacente, F. and Tonetti, M.G. 2013. Effects of MicroRNAs on fucosyltransferase 8 (FUT8) expression in hepatocarcinoma cells. *PLoS ONE,* 8(10),

Bibikova, M., Beumer, K., Trautman, J.K. and Carroll, D. 2003. Enhancing gene targeting with designed zinc finger nucleases. *Science,* 300(5620), pp.764.

Bibikova, M., Golic, M., Golic, K.G. and Carroll, D. 2002. Targeted chromosomal cleavage and mutagenesis in drosophila using zinc-finger nucleases. *Genetics,* 161(3), pp.1169-1175.

Birzele, F., Schaub, J., Rust, W., Clemens, C., Baum, P., Kaufmann, H., Weith, A., Schulz, T.W. and Hildebrandt, T. 2010. Into the unknown: Expression profiling without genome sequence information in CHO by next generation sequencing. *Nucleic Acids Research,* 38(12), pp.3999-4010.

Bitinaite, J., Wah, D.A., Aggarwal, A.K. and Schildkraut, I. 1998. Foki dimerization is required for dna cleavage. *Proceedings of the National Academy of Sciences of the United States of America,* 95(18), pp.10570-10575.

Blenkiron, C., Goldstein, L.D., Thorne, N.P., Spiteri, I., Chin, S.-., Dunning, M.J., Barbosa-Morais, N.L., Teschendorff, A.E., Green, A.R., Ellis, I.O., Tavaré, S., Caldas, C. and Miska, E.A. 2007. MicroRNA expression profiling of human breast cancer identifies new markers of tumor subtype. *Genome Biology,* 8(10),

Boch, J. and Bonas, U. 2010. *Xanthomonas AvrBs3 family-type III effectors: Discovery and function.*

Bohnsack, M.T., Czaplinski, K. and Görlich, D. 2004. Exportin 5 is a RanGTP-dependent dsRNA-binding protein that mediates nuclear export of pre-miRNAs. *RNA,* 10(2), pp.185-191.

Bonas, U., Stall, R.E. and Staskawicz, B. 1989. Genetic and structural characterization of the avirulence gene avrBs3 from xanthomonas campestris pv. vesicatoria. *MGG Molecular & General Genetics,* 218(1), pp.127-136.

Bort, J.A.H., Hackl, M., Höflmayer, H., Jadhav, V., Harreither, E., Kumar, N., Ernst, W., Grillari, J. and Borth, N. 2012. Dynamic mRNA and miRNA profiling of CHO-K1 suspension cell cultures. *Biotechnology Journal,* 7(4), pp.500-515.

Borth, N., Mattanovich, D., Kunert, R. and Katinger, H. 2005. Effect of increased expression of protein disulfide isomerase and heavy chain binding protein on antibody secretion in a recombinant CHO cell line. *Biotechnology Progress,* 21(1), pp.106-111.

Boyerinas, B., Park, S.-., Hau, A., Murmann, A.E. and Peter, M.E. 2010. The role of let-7 in cell differentiation and cancer. *Endocrine-Related Cancer,* 17(1), pp.F19-F36.

Bradley, S.A., Ouyang, A., Purdie, J., Smitka, T.A., Wang, T. and Kaerner, A. 2010. Fermentanomics: Monitoring mammalian cell cultures with NMR spectroscopy. *Journal of the American Chemical Society,* 132(28), pp.9531-9533.

Bratkovic, T., Glavan, G., Štrukelj, B., Živin, M. and Rogelj, B. 2012. Exploiting microRNAs for cell engineering and therapy. *Biotechnology Advances,* 30(3), pp.753-765.

Browne, S.M. and Al-Rubeai, M. 2007a. Selection methods for high-producing mammalian cell lines. *Trends in Biotechnology,* 25(9), pp.425-432.

Brueckner, B., Stresemann, C., Kuner, R., Mund, C., Musch, T., Meister, M., Sültmann, H. and Lyko, F. 2007. The human let-7a-3 locus contains an epigenetically regulated microRNA gene with oncogenic function. *Cancer Research,* 67(4), pp.1419-1423.

Butler, M. 2005a. Animal cell cultures: Recent achievements and perspectives in the production of biopharmaceuticals. *Applied Microbiology and Biotechnology,* 68(3), pp.283-291.

Butler, M. and Meneses-Acosta, A. 2012. Recent advances in technology supporting biopharmaceutical production from mammalian cells. *Applied Microbiology and Biotechnology,* 96(4), pp.885-894.

Byrd, A.E., Aragon, I.V. and Brewer, J.W. 2012. MicroRNA-30c-2* limits expression of proadaptive factor XBP1 in the unfolded protein response. *Journal of Cell Biology,* 196(6), pp.689-698.

Calin, G.A., Cimmino, A., Fabbri, M., Ferracin, M., Wojcik, S.E., Shimizu, M., Taccioli, C., Zanesi, N., Garzon, R., Aqeilan, R.I., Alder, H., Volinia, S., Rassenti, L., Liu, X., Liu, C.-., Kipps, T.J., Negrini, M. and Croce, C.M. 2008. MiR-15a and miR-16-1 cluster functions in human leukemia. *Proceedings of the National Academy of Sciences of the United States of America,* 105(13), pp.5166-5171.

Carlage, T., Hincapie, M., Zang, L., Lyubarskaya, Y., Madden, H., Mhatre, R. and Hancock, W.S. 2009. Proteomic profiling of a high-producing chinese hamster ovary cell culture. *Analytical Chemistry,* 81(17), pp.7357-7362.

Carrio, M., Arderiu, G., Myers, C. and Boudreau, N.J. 2005. Homeobox D10 induces phenotypic reversion of breast tumor cells in a three-dimensional culture model. *Cancer Research,* 65(16), pp.7177-7185.

Carte, J., Wang, R., Li, H., Terns, R.M. and Terns, M.P. 2008. Cas6 is an endoribonuclease that generates guide RNAs for invader defense in prokaryotes. *Genes and Development,* 22(24), pp.3489-3496.

Caygill, E.E. and Johnston, L.A. 2008. Temporal regulation of metamorphic processes in drosophila by the let-7 and miR-125 heterochronic MicroRNAs. *Current Biology,* 18(13), pp.943-950.

Cermak, T., Doyle, E.L., Christian, M., Wang, L., Zhang, Y., Schmidt, C., Baller, J.A., Somia, N.V., Bogdanove, A.J. and Voytas, D.F. 2011. Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. *Nucleic Acids Research,* 39(12),

Chakrabarti, A., Chen, A.W. and Varner, J.D. 2011. A review of the mammalian unfolded protein response. *Biotechnology and Bioengineering,* 108(12), pp.2777-2793.

Chan, J.A., Krichevsky, A.M. and Kosik, K.S. 2005. MicroRNA-21 is an antiapoptotic factor in human glioblastoma cells. *Cancer Research,* 65(14), pp.6029-6033.

Chang, N., Sun, C., Gao, L., Zhu, D., Xu, X., Zhu, X., Xiong, J.-. and Xi, J.J. 2013. Genome editing with RNA-guided Cas9 nuclease in zebrafish embryos. *Cell Research,* 23(4), pp.465-472.

Chatterjee, S., Fasler, M., Büssing, I. and Großhans, H. 2011. Target-mediated protection of endogenous MicroRNAs in C. elegans. *Developmental Cell,* 20(3), pp.388-396.

Chatterjee, S. and Großhans, H. 2009. Active turnover modulates mature microRNA activity in caenorhabditis elegans. *Nature,* 461(7263), pp.546-549.

Chaturvedi, K., Sun, S.Y., O'Brien, T., Liu, Y.J. and Brooks, J.W. 2014. Comparison of the behavior of CHO cells during cultivation in 24-square deep well microplates and conventional shake flask systems. *Biotechnology Reports,* 1-2pp.22-26.

Chen, B., Gilbert, L.A., Cimini, B.A., Schnitzbauer, J., Zhang, W., Li, G.-., Park, J., Blackburn, E.H., Weissman, J.S., Qi, L.S. and Huang, B. 2013. Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/cas system. *Cell,* 155(7), pp.1479-1491.

Chen, K., Liu, Q., Xie, L., Sharp, P.A. and Wang, D.I.C. 2001. Engineering of a mammalian cell line for reduction of lactate formation and high monoclonal antibody production. *Biotechnology and Bioengineering,* 72(1), pp.55-61.

Chendrimada, T.P., Gregory, R.I., Kumaraswamy, E., Norman, J., Cooch, N., Nishikura, K. and Shiekhattar, R. 2005. TRBP recruits the dicer complex to Ago2 for microRNA processing and gene silencing. *Nature,* 436(7051), pp.740-744.

Cheng, A.M., Byrom, M.W., Shelton, J. and Ford, L.P. 2005. Antisense inhibition of human miRNAs and indications for an involvement of miRNA in cell growth and apoptosis. *Nucleic Acids Research,* 33(4), pp.1290-1297.

Chiang, G.G. and Sisk, W.P. 2005. Bcl-xL mediates increased production of humanized monoclonal antibodies in chinese hamster ovary cells. *Biotechnology and Bioengineering,* 91(7), pp.779-792.

Chivukula, R.R. and Mendell, J.T. 2008. Circular reasoning: microRNAs and cell-cycle control. *Trends in Biochemical Sciences,* 33(10), pp.474-481.

Cho, S.W., Kim, S., Kim, J.M. and Kim, J.-. 2013. Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. *Nature Biotechnology,* 31(3), pp.230-232.

Chong, W.P.K., Reddy, S.G., Yusufi, F.N.K., Lee, D.-., Wong, N.S.C., Heng, C.K., Yap, M.G.S. and Ho, Y.S. 2010. Metabolomics-driven approach for the improvement of chinese hamster ovary cell growth: Overexpression of malate dehydrogenase II. *Journal of Biotechnology,* 147(2), pp.116-121.

Chou, Y.-., Lin, H.-., Lien, Y.-., Wang, Y.-., Hong, C.-., Kao, Y.-., Lin, S.-., Chang, Y.-., Lin, S.-., Chen, S.-., Chen, H.-., Yeh, S.-. and Wu, C.-. 2010. EGFR promotes lung tumorigenesis by activating miR-7 through a ras/ERK/myc pathway that targets the Ets2 transcriptional repressor ERF. *Cancer Research,* 70(21), pp.8822-8831.

Chu, V.T., Weber, T., Wefers, B., Wurst, W., Sander, S., Rajewsky, K. and Kühn, R. 2015. Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells. *Nature Biotechnology,* 33(5), pp.543-548.

Chung, J.Y., Lim, S.W., Hong, Y.J., Hwang, S.O. and Lee, G.M. 2004a. Effect of doxycycline-regulated calnexin and calreticulin expression on specific thrombopoietin productivity of recombinant chinese hamster ovary cells. *Biotechnology and Bioengineering,* 85(5), pp.539-546.

Chusainow, J., Yang, Y.S., Yeo, J.H.M., Ton, P.C., Asvadi, P., Wong, N.S.C. and Yap, M.G.S. 2009. A study of monoclonal antibody-producing CHO cell lines: What makes a stable high producer? *Biotechnology and Bioengineering,* 102(4), pp.1182-1196.

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. and Barron, N. 2012. Integrated miRNA, mRNA and protein expression analysis reveals the role of post-transcriptional regulation in controlling CHO cell growth rate. *BMC Genomics,* 13(1),

Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A. and Zhang, F. 2013. Multiplex genome engineering using CRISPR/cas systems. *Science,* 339(6121), pp.819-823.

Coppo, R. and Amore, A. 2004. Aberrant glycosylation in IgA nephropathy (IgAN). *Kidney International,* 65(5), pp.1544-1547.

Cost, G.J., Freyvert, Y., Vafiadis, A., Santiago, Y., Miller, J.C., Rebar, E., Collingwood, T.N., Snowden, A. and Gregory, P.D. 2010a. BAK and BAX deletion using zinc-finger nucleases yields apoptosis-resistant CHO cells. *Biotechnology and Bioengineering,* 105(2), pp.330-340.

Courtes, F.C., Lin, J., Lim, H.L., Ng, S.W., Wong, N.S.C., Koh, G., Vardy, L., Yap, M.G.S, Loo, B. and Lee D.Y. 2013. Translatome analysis of CHO cells to identify key growth genes. *Journal of Biotechnology*, 167, pp. 215-224.

Cradick, T.J., Ambrosini, G., Iseli, C., Bucher, P. and McCaffrey, A.P. 2011. ZFN-site searches genomes for zinc finger nuclease target sites and off-target sites. *BMC Bioinformatics,* 12

Creighton, C.J., Fountain, M.D., Yu, Z., Nagaraja, A.K., Zhu, H., Khan, M., Olokpa, E., Zariff, A., Gunaratne, P.H., Matzuk, M.M. and Anderson, M.L. 2010. Molecular profiling uncovers a p53-associated role for microRNA-31 in inhibiting the proliferation of serous ovarian carcinomas and other cancers. *Cancer Research,* 70(5), pp.1906-1915.

Dang, C.V. 2010. Rethinking the warburg effect with myc micromanaging glutamine metabolism. *Cancer Research,* 70(3), pp.859-862.

Davies, S.L., Lovelady, C.S., Grainger, R.K., Racher, A.J., Young, R.J. and James, D.C. 2013. Functional heterogeneity and heritability in CHO cell populations. *Biotechnology and Bioengineering,* 110(1), pp.260-274.

Davis, R., Schooley, K., Rasmussen, B., Thomas, J. and Reddy, P. 2000. Effect of PDI overexpression on recombinant protein secretion in CHO cells. *Biotechnology Progress,* 16(5), pp.736-743.

De Jesus, M. and Wurm, F.M. 2011. Manufacturing recombinant proteins in kg-ton quantities using animal cells in bioreactors. *European Journal of Pharmaceutics and Biopharmaceutics,* 78(2), pp.184-188.

Deer, J.R. and Allison, D.S. 2004. High-level expression of proteins in mammalian cells using transcription regulatory sequences from the chinese hamster EF-1a gene. *Biotechnology Progress,* 20(3), pp.880-889.

Dicarlo, J.E., Norville, J.E., Mali, P., Rios, X., Aach, J. and Church, G.M. 2013. Genome engineering in saccharomyces cerevisiae using CRISPR-cas systems. *Nucleic Acids Research,* 41(7), pp.4336-4343.

Dietmair, S., Hodson, M.P., Quek, L.-., Timmins, N.E., Gray, P. and Nielsen, L.K. 2012. A multi-omics analysis of recombinant protein production in Hek293 cells. *PLoS ONE,* 7(8),

Dong, Q., Meng, P., Wang, T., Qin, W., Qin, W., Wang, F., Yuan, J., Chen, Z., Yang, A. and Wang, H. 2010. MicroRNA let-7a inhibits proliferation of human prostate cancer cells in vitro and in vivo by targeting E2F2 and CCND2. *PLoS ONE,* 5(4),

Doolan, P., Clarke, C., Kinsella, P., Breen, L., Meleady, P., Leonard, M., Zhang, L., Clynes, M., Aherne, S.T. and Barron, N. 2013. Transcriptomic analysis of clonal growth rate variation during CHO cell line development. *Journal of Biotechnology,* 166(3), pp.105-113.

Doolan, P., Meleady, P., Barron, N., Henry, M., Gallagher, R., Gammell, P., Melville, M., Sinacore, M., McCarthy, K., Leonard, M., Charlebois, T. and Clynes, M. 2010. Microarray and proteomics expression profiling identifies several candidates, including the valosin-containing protein (VCP), involved in regulating high cellular growth rate in production CHO cell lines. *Biotechnology and Bioengineering,* 106(1), pp.42-56.

Doolan, P., Melville, M., Gammell, P., Sinacore, M., Meleady, P., McCarthy, K., Francullo, L., Leonard, M., Charlebois, T. and Clynes, M. 2008. Transcriptional profiling of gene expression changes in a PACE-transfected CHO DUKX cell line secreting high levels of rhBMP-2. *Molecular Biotechnology,* 39(3), pp.187-199.

Druz, A., Chu, C., Majors, B., Santuary, R., Betenbaugh, M. and Shiloach, J. 2011. A novel microRNA mmu-miR-466h affects apoptosis regulation in mammalian cells. *Biotechnology and Bioengineering,* 108(7), pp.1651-1661.

Druz, A., Son, Y., Betenbaugh, M. and Shiloach, J. 2013. Stable inhibition of mmu-miR-466h-5p improves apoptosis resistance and protein production in CHO cells. *Metabolic Engineering,* 16(1), pp.87-94.

Du, Z., Treiber, D., Mccarter, J.D., Fomina-Yadlin, D., Saleem, R.A., Mccoy, R.E., Zhang, Y., Tharmalingam, T., Leith, M., Follstad, B.D., Dell, B., Grisim, B., Zupke, C., Heath, C., Morris, A.E. and Reddy, P. 2014. Use of a small molecule cell cycle inhibitor to control cell growth and improve specific productivity and product quality of recombinant proteins in CHO cell cultures. *Biotechnology and Bioengineering,*

Duda, K., Lonowski, L.A., Kofoed-Nielsen, M., Ibarra, A., Delay, C.M., Kang, Q., Yang, Z., Pruett-Miller, S.M., Bennett, E.P., Wandall, H.H., Davis, G.D., Hansen, S.H. and Frödin, M. 2014. High-efficiency genome editing via 2A-coupled co-expression of fluorescent proteins and zinc finger nucleases or CRISPR/Cas9 nickase pairs. *Nucleic Acids Research,* 42(10),

Ebert, M.S., Neilson, J.R. and Sharp, P.A. 2007. MicroRNA sponges: Competitive inhibitors of small RNAs in mammalian cells. *Nature Methods,* 4(9), pp.721-726.

Ebert, M.S. and Sharp, P.A. 2012. Roles for microRNAs in conferring robustness to biological processes. *Cell,* 149(3), pp.515-524.

Ebert, M.S. and Sharp, P.A. 2010. MicroRNA sponges: Progress and possibilities. *RNA,* 16(11), pp.2043-2050.

Eeken, J.C.J. and Sobels, F.H. 1983. The effect of two chemical mutagens ENU and MMS on MR-mediated reversion of an insertion-sequence mutation in drosophila melanogaster. *Mutation Research - Fundamental and Molecular Mechanisms of Mutagenesis,* 110(2), pp.297-310.

Elmén, J., Lindow, M., Schütz, S., Lawrence, M., Petri, A., Obad, S., Lindholm, M., Hedtjärn, M., Hansen, H.F., Berger, U., Gullans, S., Kearney, P., Sarnow, P., Straarup, E.M. and Kauppinen, S. 2008. LNA-mediated microRNA silencing in non-human primates. *Nature,* 452(7189), pp.896-899.

Esquela-Kerscher, A. and Slack, F.J. 2006. Oncomirs - MicroRNAs with a role in cancer. *Nature Reviews Cancer,* 6(4), pp.259-269.

Eulalio, A., Huntzinger, E., Nishihara, T., Rehwinkel, J., Fauser, M. and Izaurralde, E. 2009. Deadenylation is a widespread effect of miRNA regulation. *RNA,* 15(1), pp.21-32.

Fan, L., Kadura, I., Krebs, L.E., Hatfield, C.C., Shaw, M.M. and Frye, C.C. 2012. Improving the efficiency of CHO cell line generation using glutamine synthetase gene knockout cells. *Biotechnology and Bioengineering,* 109(4), pp.1007-1015.

Fang, Y., Xue, J.-., Shen, Q., Chen, J. and Tian, L. 2012. MicroRNA-7 inhibits tumor growth and metastasis by targeting the phosphoinositide 3-kinase/akt pathway in hepatocellular carcinoma. *Hepatology,* 55(6), pp.1852-1862.

Fann C.H., Guirgis F., Chen G., Lao M.S. and Piret J.M. 2000. Limitations to the amplification and stability of human tissue-type plasminogen activator expression by chinese hamster ovary cells. *Biotechnology and Bioengineering,* 69(2), pp.204-212.

Ferrara, C., Brünker, P., Suter, T., Moser, S., Püntener, U. and Umaña, P. 2006. Modulation of therapeutic antibody effector functions by glycosylation engineering: Influence of golgi enzyme localization domain and co-expression of heterologous ß1, 4-N-acetylglucosaminyltransferase III and golgi a-mannosidase II. *Biotechnology and Bioengineering,* 93(5), pp.851-861.

Fire, A., Xu, S., Montgomery, M.K., Kostas, S.A., Driver, S.E. and Mello, C.C. 1998. Potent and specific genetic interference by double-stranded RNA in caenorhabditis elegans. *Nature,* 391(6669), pp.806-811.

Florin, L., Pegel, A., Becker, E., Hausser, A., Olayioye, M.A. and Kaufmann, H. 2009. Heterologous expression of the lipid transfer protein CERT increases therapeutic protein productivity of mammalian cells. *Journal of Biotechnology,* 141(1-2), pp.84-90.

Fox, S.R., Patel, U.A., Yap, M.G.S. and Wang, D.I.C. 2004. Maximizing interferon-γ production by chinese hamster ovary cells through temperature shift optimization: Experimental and modeling. *Biotechnology and Bioengineering,* 85(2), pp.177-184.

Fu Y., Sander J.D., Reyon D., Cascio V.M. and Joung J.K. 2014. Improving CRISPR-cas nuclease specificity using truncated guide RNAs. *Nature Biotechnology,* 32(3), pp.279-284.

Fu, Y., Foden, J.A., Khayter, C., Maeder, M.L., Reyon, D., Joung, J.K. and Sander, J.D. 2013. High-frequency off-target mutagenesis induced by CRISPR-cas nucleases in human cells. *Nature Biotechnology,* 31(9), pp.822-826.

Fussenegger, M., Mazur, X. and Bailey, J.E. 1997. A novel cytostatic process enhances the productivity of chinese hamster ovary cells. *Biotechnology and Bioengineering,* 55(6), pp.927-939.

Gaj, T., Gersbach, C.A. and Barbas, C.F. 2013. ZFN, TALEN, and CRISPR/cas-based methods for genome engineering. *Trends in Biotechnology,* 31(7), pp.397-405.

Galetto, R., Duchateau, P. and Pâques, F. 2009. Targeted approaches for gene therapy and the emergence of engineered meganucleases. *Expert Opinion on Biological Therapy,* 9(10), pp.1289-1303.

Gammell, P., Barron, N., Kumar, N. and Clynes, M. 2007. Initial identification of low temperature and culture stage induction of miRNA expression in suspension CHO-K1 cells. *Journal of Biotechnology,* 130(3), pp.213-218.

Gao, P., Tchernyshyov, I., Chang, T.-., Lee, Y.-., Kita, K., Ochi, T., Zeller, K.I., De Marzo, A.M., Van Eyk, J.E., Mendell, J.T. and Dang, C.V. 2009. C-myc suppression of miR-23a/b enhances mitochondrial glutaminase expression and glutamine metabolism. *Nature,* 458(7239), pp.762-765.

Garg, A., Lohmueller, J.J., Silver, P.A. and Armel, T.Z. 2012. Engineering synthetic TAL effectors with orthogonal target sites. *Nucleic Acids Research,* 40(15), pp.7584-7595.

Garneau, J.E., Dupuis, M.-., Villion, M., Romero, D.A., Barrangou, R., Boyaval, P., Fremaux, C., Horvath, P., Magadán, A.H. and Moineau, S. 2010. The CRISPR/cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature,* 468(7320), pp.67-71.

Garzon, R., Pichiorri, F., Palumbo, T., Visentini, M., Aqeilan, R., Cimmino, A., Wang, H., Sun, H., Volinia, S., Alder, H., Calin, G.A., Liu, C.-., Andreeff, M. and Croce, C.M. 2007. MicroRNA gene expression during retinoic acid-induced differentiation of human acute promyelocytic leukemia. *Oncogene,* 26(28), pp.4148-4157.

Gasiunas, G., Barrangou, R., Horvath, P. and Siksnys, V. 2012. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proceedings of the National Academy of Sciences of the United States of America,* 109(39), pp.E2579-E2586.

Gaziel-Sovran, A. and Hernando, E. 2012. Mirna-mediated GALNT modulation of invasion and immune suppression: A sweet deal for metastatic cells. *OncoImmunology,* 1(5), pp.746-748.

Gilbert, L.A., Larson, M.H., Morsut, L., Liu, Z., Brar, G.A., Torres, S.E., Stern-Ginossar, N., Brandman, O., Whitehead, E.H., Doudna, J.A., Lim, W.A., Weissman, J.S. and Qi, L.S. 2013. XCRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell,* 154(2), pp.X442-451.

Giles, K.M., Brown, R.A.M., Epis, M.R., Kalinowski, F.C. and Leedman, P.J. 2013. MiRNA-7-5p inhibits melanoma cell migration and invasion. *Biochemical and Biophysical Research Communications,* 430(2), pp.706-710.

Goeddel, D.V. (ed.) 1990. *Methods in Enzymology*, 185, pp. 543-551.

Goudar, C., Biener, R., Boisart, C., Heidemann, R., Piret, J., de Graaf, A. and Konstantinov, K. 2010. Metabolic flux analysis of CHO cells in perfusion culture by metabolite balancing and 2D [13C, 1H] COSY NMR spectroscopy. *Metabolic Engineering,* 12(2), pp.138-149.

Grav, L.M., Lee, J.S., Gerling, S., Beuchert Kallehauge, T., Hansen, A.H., Kol, S., Lee, G.M., Pedersen, L.E. and Kildegaard, H.F. 2015. One-step generation of triple knockout CHO cell lines using CRISPR/Cas9 and fluorescent enrichment. *Biotechnology Journal,*

Gregory, R.I., Yan, K.-., Amuthan, G., Chendrimada, T., Doratotaj, B., Cooch, N. and Shiekhattar, R. 2004. The microprocessor complex mediates the genesis of microRNAs. *Nature,* 432(7014), pp.235-240.

Greenlees, L., Georgantas III, R.W., Zhu, J., Dong, H., Roy, G., Jacobs, J., Clarke, L., Stracener, C., Feng, H., Yao, Y., Bowen, M.A., Ranade, K., Striecher, K. 2014. Inhibition of microRNA Let-7a increases the specific productivity of antibody-producing CHO cell lines. *Genomics and Applied Biology.* 5(1), pp. 1-15.

Griffiths-Jones, S., Grocock, R.J., van Dongen, S., Bateman, A. and Enright, A.J. 2006. miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Research.,* 34(Database issue), pp.D140-144.

Grimson, A., Farh, K.K.-., Johnston, W.K., Garrett-Engele, P., Lim, L.P. and Bartel, D.P. 2007. MicroRNA targeting specificity in mammals: Determinants beyond seed pairing. *Molecular Cell,* 27(1), pp.91-105.

Großhans, H., Johnson, T., Reinert, K.L., Gerstein, M. and Slack, F.J. 2005. The temporal patterning microRNA let-7 regulates several transcription factors at the larval to adult transition in C. elegans. *Developmental Cell,* 8(3), pp.321-330.

Groth, A.C., Fish, M., Nusse, R. and Calos, M.P. 2004. Construction of transgenic drosophila by using the site-specific integrase from phage fC31. *Genetics,* 166(4), pp.1775-1782.

Gu, D.-., Huang, Q. and Tian, L. 2015a. The molecular mechanisms and therapeutic potential of microRNA-7 in cancer. *Expert Opinion on Therapeutic Targets,* 19(3), pp.415-426.

Guo, H., Ingolia, N.T., Weissman, J.S. and Bartel, D.P. 2010. Mammalian microRNAs predominantly act to decrease target mRNA levels. *Nature,* 466(7308), pp.835-840.

Gupta, P. and Lee, K.H. 2007. Genomics and proteomics in process development: Opportunities and challenges. *Trends in Biotechnology,* 25(7), pp.324-330.

Hacker, D.L., De Jesus, M. and Wurm, F.M. 2009. 25 years of recombinant proteins from reactor-grown cells - where do we go from here? *Biotechnology Advances,* 27(6), pp.1023-1027.

Hackl, M., Jakobi, T., Blom, J., Doppmeier, D., Brinkrolf, K., Szczepanowski, R., Bernhart, S.H., Siederdissen, C.H.Z., Bort, J.A.H., Wieser, M., Kunert, R., Jeffs, S., Hofacker, I.L., Goesmann, A., Pühler, A., Borth, N. and Grillari, J. 2011a. 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), pp.62-75.

Hale, C.R., Majumdar, S., Elmore, J., Pfister, N., Compton, M., Olson, S., Resch, A.M., Glover, V.C., Graveley, B.R., Terns, R.M. and Terns, M.P. 2012. Essential features and rational design of CRISPR RNAs that function with the cas RAMP module complex to cleave RNAs. *Molecular Cell,* 45(3), pp.292-302.

Hale, C.R., Zhao, P., Olson, S., Duff, M.O., Graveley, B.R., Wells, L., Terns, R.M. and Terns, M.P. 2009. RNA-guided RNA cleavage by a CRISPR RNA-cas protein complex. *Cell,* 139(5), pp.945-956.

Han, J., Lee, Y., Yeom, K.-., Kim, Y.-., Jin, H. and Kim, V.N. 2004. The drosha-DGCR8 complex in primary microRNA processing. *Genes and Development,* 18(24), pp.3016-3027.

Hartig, J.S., Grüne, I., Najafi-Shoushtari, S.H. and Famulok, M. 2004. Sequence-specific detection of MicroRNAs by signal-amplifying ribozymes. *Journal of the American Chemical Society,* 126(3), pp.722-723.

Hatfield, S. and Ruohola-Baker, H. 2008. microRNA and stem cell function. *Cell and Tissue Research,* 331(1), pp.57-66.

Hayduk, E.J. and Lee, K.H. 2005. Cytochalasin D can improve heterologous protein productivity in adherent chinese hamster ovary cells. *Biotechnology and Bioengineering,* 90(3), pp.354-364.

Heller-Harrison, R.A., Crowe, K., Cooley, C., Hone, M., McCarthy, K., Leonard, M. 2009. Managing cell line instability and its impact during cell line development. *BioPharm Int. Suppl.* 22, pp. 16-27.

Henke, J.I., Goergen, D., Zheng, J., Song, Y., Schüttler, C.G., Fehr, C., Jünemann, C. and Niepmann, M. 2008. microRNA-122 stimulates translation of hepatitis C virus RNA. *EMBO Journal,* 27(24), pp.3300-3310.

Hennessy, E., Clynes, M., Jeppesen, P.B. and O'Driscoll, L. 2010. Identification of microRNAs with a role in glucose stimulated insulin secretion by expression profiling of MIN6 cells. *Biochemical and Biophysical Research Communications,* 396(2), pp.457-462.

Hermeking, H. 2010. The miR-34 family in cancer and apoptosis. *Cell Death and Differentiation,* 17(2), pp.193-199.

Hilton, I.B., D'Ippolito, A.M., Vockley, C.M., Thakore, P.I., Crawford, G.E., Reddy, T.E. and Gersbach, C.A. 2015. Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. *Nature Biotechnology,* 33(5), pp.510-517.

Horii, T., Tamura, D., Morita, S., Kimura, M. and Hatada, I. 2013. Generation of an ICF syndrome model by efficient genome editing of human induced pluripotent stem cells using the CRISPR system. *International Journal of Molecular Sciences,* 14(10), pp.19774-19781.

Hornstein, E. and Shomron, N. 2006. Canalization of development by microRNAs. *Nature Genetics,* 38(SUPPL. 1), pp.S20-S24.

Hsu, P.D., Scott, D.A., Weinstein, J.A., Ran, F.A., Konermann, S., Agarwala, V., Li, Y., Fine, E.J., Wu, X., Shalem, O., Cradick, T.J., Marraffini, L.A., Bao, G. and Zhang, F. 2013. DNA targeting specificity of RNA-guided Cas9 nucleases. *Nature Biotechnology,* 31(9), pp.827-832.

Hsu, P.D. and Zhang, F. 2012. Dissecting neural function using targeted genome engineering technologies. *ACS Chemical Neuroscience,* 3(8), pp.603-610.

Huang, Y., Li, Y., Wang, Y.G., Gu, X., Wang, Y. and Shen, B.F. 2007a. An efficient and targeted gene integration system for high-level antibody expression. *Journal of Immunological Methods,* 322(1-2), pp.28-39.

Hwang, S.O., Chung, J.Y. and Lee, G.M. 2003. Effect of doxycycline-regulated ERp57 expression on specific thrombopoietin productivity of recombinant CHO cells. *Biotechnology Progress,* 19(1), pp.179-184.

Hwang, S.O. and Lee, G.M. 2009. Effect of akt overexpression on programmed cell death in antibody-producing chinese hamster ovary cells. *Journal of Biotechnology,* 139(1), pp.89-94.

Hwang, W.Y., Fu, Y., Reyon, D., Maeder, M.L., Kaini, P., Sander, J.D., Joung, J.K., Peterson, R.T. and Yeh, J.-.J. 2013. Heritable and precise zebrafish genome editing using a CRISPR-cas system. *PLoS ONE,* 8(7),

Ivanovska, I., Ball, A.S., Diaz, R.L., Magnus, J.F., Kibukawa, M., Schelter, J.M., Kobayashi, S.V., Lim, L., Burchard, J., Jackson, A.L., Linsley, P.S. and Cleary, M.A. 2008. MicroRNAs in the miR-106b family regulate p21/CDKN1A and promote cell cycle progression. *Molecular and Cellular Biology,* 28(7), pp.2167-2174.

Jadhav, V., Hackl, M., Druz, A., Shridhar, S., Chung, C.-., Heffner, K.M., Kreil, D.P., Betenbaugh, M., Shiloach, J., Barron, N., Grillari, J. and Borth, N. 2013. CHO microRNA engineering is growing up: Recent successes and future challenges. *Biotechnology Advances,* 31(8), pp.1501-1513.

Jahn, R. and Scheller, R.H. 2006. SNAREs - engines for membrane fusion. *Nature Reviews Molecular Cell Biology,* 7(9), pp.631-643.

Jansen, R., Van Embden, J.D.A., Gaastra, W. and Schouls, L.M. 2002. Identification of genes that are associated with DNA repeats in prokaryotes. *Molecular Microbiology,* 43(6), pp.1565-1575.

Jayapal, K.P., Wlaschin, K.F., Hu, W.-. and Yap, M.G.S. 2007a. Recombinant protein therapeutics from CHO cells - 20 years and counting. *Chemical Engineering Progress,* 103(10), pp.40-47.

Jeong, D.-., Kim, T.S., Lee, J.W., Kim, K.T., Kim, H.J., Kim, I.-. and Kim, I.Y. 2001. Blocking of acidosis-mediated apoptosis by a reduction of lactate dehydrogenase activity through antisense mRNA expression. *Biochemical and Biophysical Research Communications,* 289(5), pp.1141-1149.

Jiang, L., Liu, X., Chen, Z., Jin, Y., Heidbreder, C.E., Kolokythas, A., Wang, A., Dai, Y. and Zhou, X. 2010. MicroRNA-7 targets IGF1R (insulin-like growth factor 1 receptor) in tongue squamous cell carcinoma cells. *Biochemical Journal,* 432(1), pp.199-205.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A. and Charpentier, E. 2012. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science,* 337(6096), pp.816-821.

Jinek, M., East, A., Cheng, A., Lin, S., Ma, E. and Doudna, J. 2013. RNA-programmed genome editing in human cells. *eLife,* 2013(2),

Johnson, C.D., Esquela-Kerscher, A., Stefani, G., Byrom, M., Kelnar, K., Ovcharenko, D., Wilson, M., Wang, X., Shelton, J., Shingara, J., Chin, L., Brown, D. and Slack, F.J. 2007. The let-7 microRNA represses cell proliferation pathways in human cells. *Cancer Research,* 67(16), pp.7713-7722.

Johnson, K.C., Jacob, N.M., Nissom, P.M., Hackl, M., Lee, L.H., Yap, M. and Hu, W.-. 2011. Conserved MicroRNAs in chinese hamster ovary cell lines. *Biotechnology and Bioengineering,* 108(2), pp.475-480.

Johnson, S.M., Grosshans, H., Shingara, J., Byrom, M., Jarvis, R., Cheng, A., Labourier, E., Reinert, K.L., Brown, D. and Slack, F.J. 2005. RAS is regulated by the let-7 microRNA family. *Cell,* 120(5), pp.635-647.

Joon, C.Y., Gatti, M.D.L., Philp, R.J., Yap, M. and Hu, W.-. 2008. Genomic and proteomic exploration of CHO and hybridoma cells under sodium butyrate treatment. *Biotechnology and Bioengineering,* 99(5), pp.1186-1204.

Jore, M.M., Lundgren, M., Van Duijn, E., Bultema, J.B., Westra, E.R., Waghmare, S.P., Wiedenheft, B., Pul, U., Wurm, R., Wagner, R., Beijer, M.R., Barendregt, A., Zhou, K., Snijders, A.P.L., Dickman, M.J., Doudna, J.A., Boekema, E.J., Heck, A.J.R., Van Der Oost, J. and Brouns, S.J.J. 2011. Structural basis for CRISPR RNA-

guided DNA recognition by cascade. *Nature Structural and Molecular Biology,* 18(5), pp.529-536.

Jun S.C., Kim M.S., Hong H.J. and Lee G.M. 2006. Limitations to the development of humanized antibody producing chinese hamster ovary cells using glutamine synthetase-mediated gene amplification. *Biotechnology Progress,* 22(3), pp.770-780.

Kameyama, Y., Kawabe, Y., Ito, A. and Kamihira, M. 2010. An accumulative site-specific gene integration system using cre recombinase-mediated cassette exchange. *Biotechnology and Bioengineering,* 105(6), pp.1106-1114.

Kang, S.M, Choi, J.W, Hong, S.H. and Lee, H.J. 2013. Up-regulation of microRNA* strands by their target transcripts. *International Journal of Molecular Science* 14(7), pp. 31-40.

Kantardjieff, A., Jacob, N.M., Yee, J.C., Epstein, E., Kok, Y.-., Philp, R., Betenbaugh, M. and Hu, W.-. 2010a. Transcriptome and proteome analysis of chinese hamster ovary cells under low temperature and butyrate treatment. *Journal of Biotechnology,* 145(2), pp.143-159.

Kantardjieff, A., Nissom, P.M., Chuah, S.H., Yusufi, F., Jacob, N.M., Mulukutla, B.C., Yap, M. and Hu, W.-. 2009. Developing genomic platforms for chinese hamster ovary cells. *Biotechnology Advances,* 27(6), pp.1028-1035.

Karsy, M., Arslan, E. and Moy, F. 2012. Current progress on understanding MicroRNAs in glioblastoma multiforme. *Genes and Cancer,* 3(1), pp.3-15.

Kaufmann, H., Mazur, X., Fussenegger, M. and Bailey, J.E. 1999. Influence of low temperature on productivity, proteome and protein phosphorylation of CHO cells. *Biotechnology and Bioengineering,* 63(5), pp.573-582.

Kaur, H., Babu, B.R. and Maiti, S. 2007. Perspective on chemistry and therapeutic applications of locked nucleic acid (LNA). *Chemical Reviews,* 107(11), pp.4672-4697.

Kefas, B., Godlewski, J., Comeau, L., Li, Y., Abounader, R., Hawkinson, M., Lee, J., Fine, H., Chiocca, E.A., Lawler, S. and Purow, B. 2008. microRNA-7 inhibits the epidermal growth factor receptor and the akt pathway and is down-regulated in glioblastoma. *Cancer Research,* 68(10), pp.3566-3572.

Kelly, P.S., Breen, L., Gallagher, C., Kelly, S., Henry, M., Lao, N.T., Meleady, P., O'Gorman, D., Clynes, M. and Barron, N. 2015a. Re-programming CHO cell metabolism using miR-23 tips the balance towards a highly productive phenotype. *Biotechnology Journal,* 10(7), pp.1029-1040.

Kelly, P.S., Gallagher, C., Clynes, M. and Barron, N. 2015b. Conserved microRNA function as a basis for chinese hamster ovary cell engineering. *Biotechnology Letters,* 37(4), pp.787-798.

Kennard, M.L., Goosney, D.L., Monteith, D., Zhang, L., Moffat, M., Fischer, D. and Mott, J. 2009. The generation of stable, high MAb expressing CHO cell lines based on the artificial chromosome expression (ACE) technology. *Biotechnology and Bioengineering,* 104(3), pp.540-553.

Kildegaard, H.F., Baycin-Hizal, D., Lewis, N.E. and Betenbaugh, M.J. 2013. The emerging CHO systems biology era: Harnessing the 'omics revolution for biotechnology. *Current Opinion in Biotechnology,* 24(6), pp.1102-1107.

Kim, H. and Kim, J.-. 2014a. A guide to genome engineering with programmable nucleases. *Nature Reviews Genetics,* 15(5), pp.321-334.

Kim, H. and Kim, J.-. 2014b. A guide to genome engineering with programmable nucleases. *Nature Reviews Genetics,* 15(5), pp.321-334.

Kim, J.Y., Kim, Y.-., Han, Y.K., Choi, H.S., Kim, Y.H. and Lee, G.M. 2011a. Proteomic understanding of intracellular responses of recombinant chinese hamster ovary cells cultivated in serum-free medium supplemented with hydrolysates. *Applied Microbiology and Biotechnology,* 89(6), pp.1917-1928.

Kim, J.Y., Kim, Y.-. and Lee, G.M. 2012. CHO cells in biotechnology for production of recombinant proteins: Current state and further potential. *Applied Microbiology and Biotechnology,* 93(3), pp.917-930.

Kim, M., O'Callaghan, P.M., Droms, K.A. and James, D.C. 2011b. A mechanistic understanding of production instability in CHO cell lines expressing recombinant monoclonal antibodies. *Biotechnology and Bioengineering,* 108(10), pp.2434-2446.

Kim, N.S. and Lee, G.M. 2000a. Overexpression of bcl-2 inhibits sodium butyrate-induced apoptosis in chinese hamster ovary cells resulting in enhanced humanized antibody production. *Biotechnology and Bioengineering,* 71(3), pp.184-193.

Kim, S.H. and Lee, G.M. 2007. Differences in optimal pH and temperature for cell growth and antibody production between two chinese hamster ovary clones derived from the same parental clone. *Journal of Microbiology and Biotechnology,* 17(5), pp.712-720.

Kim, Y.-., Cha, J. and Chandrasegaran, S. 1996. Hybrid restriction enzymes: Zinc finger fusions to fok I cleavage domain. *Proceedings of the National Academy of Sciences of the United States of America,* 93(3), pp.1156-1160.

Kinoshita, T., Nohata, N., Yoshino, H., Hanazawa, T., Kikawa, N., Fujimura, L., Chiyomaru, T., Kawakami, K., Enokida, H., Nakagawa, M., Okamoto, Y. and Seki, N. 2012. Tumor suppressive microRNA-375 regulates lactate dehydrogenase B in maxillary sinus squamous cell carcinoma. *International Journal of Oncology,* 40(1), pp.185-193.

Kiss, Z., Elliott, S., Jedynasty, K., Tesar, V. and Szegedi, J. 2010. Discovery and basic pharmacology of erythropoiesis-stimulating agents (ESAs)'including the

hyperglycosylated ESA, darbepoetin alfa: An update of the rationale and clinical impact. *European Journal of Clinical Pharmacology,* 66(4), pp.331-340.

Kitano, H. 2004. Biological robustness. *Nature Reviews Genetics,* 5(11), pp.826-837.

Kito, M., Itami, S., Fukano, Y., Yamana, K. and Shibui, T. 2003. Construction of engineered cho strains for high-level production of recombinant proteins. *Applied Microbiology and Biotechnology,* 60(4), pp.442-448.

Klein, M.E., Lioy, D.T., Ma, L., Impey, S., Mandel, G. and Goodman, R.H. 2007. Homeostatic regulation of MeCP2 expression by a CREB-induced microRNA. *Nature Neuroscience,* 10(12), pp.1513-1514.

Kluiver, J., Gibcus, J.H., Hettinga, C., Adema, A., Richter, M.K.S., Halsema, N., Slezak-Prochazka, I., Ding, Y., Kroesen, B.-. and van den Berg, A. 2012. Rapid generation of microRNA sponges for microRNA inhibition. *PLoS ONE,* 7(1),

Kozak, M. 1991. Structural features in eukaryotic mRNAs that modulate the initiation of translation. *Journal of Biological Chemistry,* 266(30), pp.19867-19870.

Krämer, O., Klausing, S. and Noll, T. 2010. Methods in mammalian cell line engineering: From random mutagenesis to sequence-specific approaches. *Applied Microbiology and Biotechnology,* 88(2), pp.425-436.

Krützfeldt, J., Rajewsky, N., Braich, R., Rajeev, K.G., Tuschl, T., Manoharan, M. and Stoffel, M. 2005. Silencing of microRNAs in vivo with 'antagomirs'. *Nature,* 438(7068), pp.685-689.

Ku, S.C.Y., Toh, P.C., Lee, Y.Y., Chusainow, J., Yap, M.G.S. and Chao, S.-. 2010. Regulation of XBP-1 signaling during transient and stable recombinant protein production in CHO cells. *Biotechnology Progress,* 26(2), pp.517-526.

Kumar, N., Gammell, P., Meleady, P., Henry, M. and Clynes, M. 2008. Differential protein expression following low temperature culture of suspension CHO-K1 cells. *BMC Biotechnology,* 8

Lagos-Quintana, M., Rauhut, R., Lendeckel, W. and Tuschl, T. 2001. Identification of novel genes coding for small expressed RNAs. *Science,* 294(5543), pp.853-858.

Lai, T., Yang, Y. and Ng, S.K. 2013. Advances in mammalian cell line development technologies for recombinant protein production. *Pharmaceuticals,* 6(5), pp.579-603.

Lal, A., Navarro, F., Maher, C.A., Maliszewski, L.E., Yan, N., O'Day, E., Chowdhury, D., Dykxhoorn, D.M., Tsai, P., Hofmann, O., Becker, K.G., Gorospe, M., Hide, W. and Lieberman, J. 2009. miR-24 inhibits cell proliferation by targeting E2F2, MYC, and other cell-cycle genes via binding to "seedless" 3'UTR MicroRNA recognition elements. *Molecular Cell,* 35(5), pp.610-625.

Landthaler, M., Yalcin, A. and Tuschl, T. 2004. The human DiGeorge syndrome critical region gene 8 and its D. melanogaster homolog are required for miRNA biogenesis. *Current Biology,* 14(23), pp.2162-2167.

Lao, M.-. and Toth, D. 1997. Effects of ammonium and lactate on growth and metabolism of a recombinant chinese hamster ovary cell culture. *Biotechnology Progress,* 13(5), pp.688-691.

Lau, N.C., Lim, L.P., Weinstein, E.G. and Bartel, D.P. 2001. An abundant class of tiny RNAs with probable regulatory roles in caenorhabditis elegans. *Science,* 294(5543), pp.858-862.

Lawhorn, I.E.B., Ferreira, J.P. and Wang, C.L. 2014. Evaluation of sgRNA target sites for CRISPR-mediated repression of TP53. *PLoS ONE,* 9(11),

Lee, J.S., Kallehauge, T.B., Pedersen, L.E. and Kildegaard, H.F. 2015. Site-specific integration in CHO cells mediated by CRISPR/Cas9 and homology-directed DNA repair pathway. *Scientific Reports,* 5

Lee, K.H., Harrington, M.G. and Bailey, J.E. 1996. Two-dimensional electrophoresis of proteins as a tool in the metabolic engineering of cell cycle regulation. *Biotechnology and Bioengineering,* 50(3), pp.336-340.

Lee, K.H., Sburlati, A., Renner, W.A. and Bailey, J.E. 1996. Deregulated expression of cloned transcription factor E2F-1 in chinese hamster ovary cells shifts protein patterns and activates growth in protein- free medium. *Biotechnology and Bioengineering,* 50(3), pp.273-279.

Lee, M.S., Kim, K.W., Kim, Y.H. and Lee, G.M. 2003. Proteome analysis of antibody-expressing CHO cells in response to hyperosmotic pressure. *Biotechnology Progress,* 19(6), pp.1734-1741.

Lee, R.C., Feinbaum, R.L. and Ambros, V. 1993. The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell,* 75(5), pp.843-854.

Lee, Y., Jeon, K., Lee, J.-., Kim, S. and Kim, V.N. 2002. MicroRNA maturation: Stepwise processing and subcellular localization. *EMBO Journal,* 21(17), pp.4663-4670.

Lee, Y.S. and Dutta, A. 2006. MicroRNAs: Small but potent oncogenes or tumor suppressors. *Current Opinion in Investigational Drugs,* 7(6), pp.560-564.

Lewis, N.E., 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. and Palsson, B.O. 2013. Genomic landscapes of chinese hamster ovary cell lines as revealed by the cricetulus griseus draft genome. *Nature Biotechnology,* 31(8), pp.759-765.

Ley, D., Harraghy, N., Le Fourn, V., Bire, S., Girod, P.-., Regamey, A., Rouleux-Bonnin, F., Bigot, Y. and Mermod, N. 2013. MAR elements and transposons for improved transgene integration and expression. *PLoS ONE,* 8(4),

Li, F., Vijayasankaran, N., Shen, A.Y., Kiss, R. and Amanullah, A. 2010. Cell culture processes for monoclonal antibody production. *mAbs*. 2(5), pp. 466-479.

Li, J., Huang, Z., Sun, X., Yang, P. and Zhang, Y. 2006. Understanding the enhanced effect of dimethyl sulfoxide on hepatitis B surface antigen expression in the culture of chinese hamster ovary cells on the basis of proteome analysis. *Enzyme and Microbial Technology,* 38(3-4), pp.372-380.

Li, L., Xu, J., Yang, D., Tan, X. and Wang, H. 2010. Computational approaches for microRNA studies: A review. *Mammalian Genome,* 21(1-2), pp.1-12.

Li, X., Cassidy, J.J., Reinke, C.A., Fischboeck, S. and Carthew, R.W. 2009. A MicroRNA imparts robustness against environmental fluctuation during development. *Cell,* 137(2), pp.273-282.

Lieber, M.R. 2010. *The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway.*

Lieu, P.T., MacHleidt, T., Thyagarajan, B., Fontes, A., Frey, E., Fuerstenau-Sharp, M., Thompson, D.V., Swamilingiah, G.M., Derebail, S.S., Piper, D. and Chesnut, J.D. 2009. Generation of site-specific retargeting platform cell lines for drug discovery using phiC31 and R4 integrases. *Journal of Biomolecular Screening,* 14(10), pp.1207-1215.

Lim, S.F., Chuan, K.H., Liu, S., Loh, S.O.H., Chung, B.Y.F., Ong, C.C. and Song, Z. 2006. RNAi suppression of bax and bak enhances viability in fed-batch cultures of CHO cells. *Metabolic Engineering,* 8(6), pp.509-522.

Lin, Q., Gao, Z., Alarcon, R.M., Ye, J. and Yun, Z. 2009. A role of miR-27 in the regulation of adipogenesis. *FEBS Journal,* 276(8), pp.2348-2358.

Little, P. 1993. Small and perfectly formed. *Nature,* 366(6452), pp.204-205.

Liu, C.-., Tsai, M.-., Hung, P.-., Kao, S.-., Liu, T.-., Wu, K.-., Chiou, S.-., Lin, S.-. and Chang, K.-. 2010a. miR-31 ablates expression of the HIF regulatory factor FIH to activate the HIF pathway in head and neck carcinoma. *Cancer Research,* 70(4), pp.1635-1644.

Liu, P.-., Chan, E.M., Cost, G.J., Zhang, L., Wang, J., Miller, J.C., Guschin, D.Y., Reik, A., Holmes, M.C., Mott, J.E., Collingwood, T.N. and Gregory, P.D. 2010b. Generation of a triple-gene knockout mammalian cell line using engineered zinc-finger nucleases. *Biotechnology and Bioengineering,* 106(1), pp.97-105.

Liu, Q., Segal, D.J., Ghiara, J.B. and Barbas III, C.F. 1997. Design of polydactyl zinc-finger proteins for unique addressing within complex genomes. *Proceedings of the National Academy of Sciences of the United States of America,* 94(11), pp.5525-5530.

Liu, S., Zhang, P., Chen, Z., Liu, M., Li, X. and Tang, H. 2013. MicroRNA-7 downregulates XIAP expression to suppress cell growth and promote apoptosis in cervical cancer cells. *FEBS Letters,* 587(14), pp.2247-2253.

Lu, Y., Xiao, J., Lin, H., Bai, Y., Luo, X., Wang, Z. and Yang, B. 2009. A single anti-microRNA antisense oligodeoxyribonucleotide (AMO) targeting multiple microRNAs offers an improved approach for microRNA interference. *Nucleic Acids Research,* 37(3),

Lund, E. and Dahlberg, J.E. 2006. *Substrate selectivity of exportin 5 and Dicer in the biogenesis of microRNAs.*

Lund, E., Güttinger, S., Calado, A., Dahlberg, J.E. and Kutay, U. 2004. Nuclear export of MicroRNA precursors. *Science,* 303(5654), pp.95-98.

Magadán, A.H., Dupuis, M.-., Villion, M. and Moineau, S. 2012. Cleavage of phage DNA by the streptococcus thermophilus CRISPR3-cas system. *PLoS ONE,* 7(7),

Majors, B.S., Betenbaugh, M.J., Pederson, N.E. and Chiang, G.G. 2009. Mcl-1 overexpression leads to higher viabilities and increased production of humanized monoclonal antibody in chinese hamster ovary cells. *Biotechnology Progress,* 25(4), pp.1161-1168.

Makarova, K.S., Grishin, N.V., Shabalina, S.A., Wolf, Y.I. and Koonin, E.V. 2006. A putative RNA-interference-based immune system in prokaryotes: Computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. *Biology Direct,* 1

Makarova, K.S., Haft, D.H., Barrangou, R., Brouns, S.J.J., Charpentier, E., Horvath, P., Moineau, S., Mojica, F.J.M., Wolf, Y.I., Yakunin, A.F., Van Der Oost, J. and Koonin, E.V. 2011. Evolution and classification of the CRISPR-cas systems. *Nature Reviews Microbiology,* 9(6), pp.467-477.

Mali, P., Aach, J., Stranges, P.B., Esvelt, K.M., Moosburner, M., Kosuri, S., Yang, L. and Church, G.M. 2013a. CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nature Biotechnology,* 31(9), pp.833-838.

Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E., Norville, J.E. and Church, G.M. 2013b. RNA-guided human genome engineering via Cas9. *Science,* 339(6121), pp.823-826.

Mansfield, J.H., Harfe, B.D., Nissen, R., Obenauer, J., Srineel, J., Chaudhuri, A., Farzan-Kashani, R., Zuker, M., Pasquinelli, A.E., Ruvkun, G., Sharp, P.A., Tabin, C.J. and McManus, M.T. 2004. MicroRNA-responsive 'sensor' transgenes uncover hox-like and other developmentally regulated patterns of vertebrate microRNA expression. *Nature Genetics,* 36(10), pp.1079-1083.

Maresca, M., Lin, V.G., Guo, N. and Yang, Y. 2013. Obligate ligation-gated recombination (ObLiGaRe): Custom-designed nuclease-mediated targeted integration through nonhomologous end joining. *Genome Research,* 23(3), pp.539-546.

Marraffini, L.A. and Sontheimer, E.J. 2008. CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. *Science,* 322(5909), pp.1843-1845.

Maruyama, T., Dougan, S.K., Truttmann, M.C., Bilate, A.M., Ingram, J.R. and Ploegh, H.L. 2015. Increasing the efficiency of precise genome editing with CRISPR-Cas9 by inhibition of nonhomologous end joining. *Nature Biotechnology,* 33(5), pp.538-542.

Marx, J.L. 1982. Gene transfer into the drosophila germ line. *Science,* 218(4570), pp.364-365.

Matasci, M., Hacker, D.L., Baldi, L. and Wurm, F.M. 2009. Recombinant therapeutic protein production in cultivated mammalian cells: Current status and future prospects. *Drug Discovery Today: Technologies,* 5(2-3), pp.e37-e42.

Mazur, X., Fussenegger, M., Renner, W.A. and Bailey, J.E. 1998a. Higher productivity of growth-arrested chinese hamster ovary cells expressing the cyclin-dependent kinase inhibitor p27. *Biotechnology Progress,* 14(5), pp.705-713.

McClellan, A.J., Tam, S., Kaganovich, D. and Frydman, J. 2005. Protein quality control: Chaperones culling corrupt conformations. *Nature Cell Biology,* 7(8), pp.736-741.

McGrath, J. and Solter, D. 1984. Completion of mouse embryogenesis requires both the maternal and paternal genomes. *Cell,* 37(1), pp.179-183.

Meents, H., Enenkel, B., Eppenberger, H.M., Werner, R.G. and Fussenegger, M. 2002a. Impact of coexpression and coamplification of sICAM and antiapoptosis determinants bcl-2/bcl-xL on productivity, cell survival, and mitochondria number in CHO-DG44 grown in suspension and serum-free media. *Biotechnology and Bioengineering,* 80(6), pp.706-716.

Meents, H., Enenkel, B., Werner, R.G. and Fussenegger, M. 2002b. P27Kip1-mediated controlled proliferation technology increases constitutive sICAM production in CHO-DUKX adapted for growth in suspension and serum-free media. *Biotechnology and Bioengineering,* 79(6), pp.619-627.

Meleady, P., Hoffrogge, R., Henry, M., Rupp, O., Bort, J.H., Clarke, C., Brinkrolf, K., Kelly, S., Müller, B., Doolan, P., Hackl, M., Beckmann, T.F., Noll, T., Grillari, J., Barron, N., Pühler, A., Clynes, M. and Borth, N. 2012. Utilization and evaluation of CHO-specific sequence databases for mass spectrometry based proteomics. *Biotechnology and Bioengineering,* 109(6), pp.1386-1394.

Miller, J.C., Tan, S., Qiao, G., Barlow, K.A., Wang, J., Xia, D.F., Meng, X., Paschon, D.E., Leung, E., Hinkley, S.J., Dulay, G.P., Hua, K.L., Ankoudinova, I., Cost, G.J., Urnov, F.D., Zhang, H.S., Holmes, M.C., Zhang, L., Gregory, P.D. and Rebar, E.J.

2011a. A TALE nuclease architecture for efficient genome editing. *Nature Biotechnology,* 29(2), pp.143-150.

Mimura, Y., Ashton, P.R., Takahashi, N., Harvey, D.J. and Jefferis, R. 2007. Contrasting glycosylation profiles between fab and fc of a human IgG protein studied by electrospray ionization mass spectrometry. *Journal of Immunological Methods,* 326(1-2), pp.116-126.

Mohan, C., Soon, H.P., Joo, Y.C. and Lee, G.M. 2007. Effect of doxycycline-regulated protein disulfide isomerase expression on the specific productivity of recombinant CHO cells: Thrombopoietin and antibody. *Biotechnology and Bioengineering,* 98(3), pp.611-615.

Mu, P., Han, Y.-., Betel, D., Yao, E., Squatrito, M., Ogrodowski, P., De Stanchina, E., D'Andrea, A., Sander, C. and Ventura, A. 2009. Genetic dissection of the miR-17-92 cluster of microRNAs in myc-induced B-cell lymphomas. *Genes and Development,* 23(24), pp.2806-2811.

Müller, D., Katinger, H. and Grillari, J. 2008. MicroRNAs as targets for engineering of CHO cell factories. *Trends in Biotechnology,* 26(7), pp.359-365.

Muniyappa, M.K., Dowling, P., Henry, M., Meleady, P., Doolan, P., Gammell, P., Clynes, M. and Barron, N. 2009. MiRNA-29a regulates the expression of numerous proteins and reduces the invasiveness and proliferation of human carcinoma cell lines. *European Journal of Cancer,* 45(17), pp.3104-3118.

NCBI Database, Gene. Available from: http://www.ncbi.nlm.nih.gov/gene/?term=hsa-mir-7

Nemudryi, A.A., Valetdinova, K.R., Medvedev, S.P. and Zakian, S.M. 2014. TALEN and CRISPR/cas genome editing systems: Tools of discovery. *Acta Naturae,* 6(22), pp.19-40.

Ng, S.K., Tan, T.R.M., Wang, Y., Ng, D., Goh, L.-., Bardor, M., Wong, V.V.T. and Lam, K.P. 2012. Production of functional soluble dectin-1 glycoprotein using an IRES-linked destabilized-dihydrofolate reductase expression vector. *PLoS ONE,* 7(12),

Ng, S.K., Wang, D.I.C. and Yap, M.G.S. 2007. Application of destabilizing sequences on selection marker for improved recombinant protein productivity in CHO-DG44. *Metabolic Engineering,* 9(3), pp.304-316.

Nissom, P.M., Sanny, A., Kok, Y.J., Hiang, Y.T., Chuah, S.H., Shing, T.K., Lee, Y.Y., Wong, K.T.K., Hu, W.-., Sim, M.Y.G. and Philp, R. 2006. Transcriptome and proteome profiling to understanding the biology of high productivity CHO cells. *Molecular Biotechnology,* 34(2), pp.125-140.

O'Donnell, K.A., Wentzel, E.A., Zeller, K.I., Dang, C.V. and Mendell, J.T. 2005. C-myc-regulated microRNAs modulate E2F1 expression. *Nature,* 435(7043), pp.839-843.

Ohya, T., Hayashi, T., Kiyama, E., Nishii, H., Miki, H., Kobayashi, K., Honda, K., Omasa, T. and Ohtake, H. 2008. Improved production of recombinant human antithrombin III in chinese hamster ovary cells by ATF4 overexpression. *Biotechnology and Bioengineering,* 100(2), pp.317-324.

Okazaki, A., Shoji-Hosaka, E., Nakamura, K., Wakitani, M., Uchida, K., Kakita, S., Tsumoto, K., Kumagai, I. and Shitara, K. 2004. Fucose depletion from human IgG1 oligosaccharide enhances binding enthalpy and association rate between IgG1 and FcγRIIIa. *Journal of Molecular Biology,* 336(5), pp.1239-1249.

Omasa, T., Takami, T., Ohya, T., Kiyama, E., Hayashi, T., Nishii, H., Miki, H., Kobayashi, K., Honda, K. and Ohtake, H. 2008. Overexpression of GADD34 enhances production of recombinant human antithrombin III in chinese hamster ovary cells. *Journal of Bioscience and Bioengineering,* 106(6), pp.568-573.

Ozturk, S.S., Riley, M.R. and Palsson, B.O. 1992. Effects of ammonia and lactate on hybridoma growth, metabolism, and antibody production. *Biotechnology and Bioengineering,* 39(4), pp.418-431.

Pan, Y., Xiao, L., Li, A.S.S., Zhang, X., Sirois, P., Zhang, J. and Li, K. 2013. Biological and biomedical applications of engineered nucleases. *Molecular Biotechnology,* 55(1), pp.54-62.

Paredes, V., Park, J.S., Jeong, Y., Yoon, J. and Baek, K. 2013. Unstable expression of recombinant antibody during long-term culture of CHO cells is accompanied by histone H3 hypoacetylation. *Biotechnology Letters,* 35(7), pp.987-993.

Pascoe, D.E., Arnott, D., Papoutsakis, E.T., Miller, W.M. and Andersen, D.C. 2007. Proteome analysis of antibody-producing CHO cell lines with different metabolic profiles. *Biotechnology and Bioengineering,* 98(2), pp.391-410.

Pasquinelli, A.E., Reinhart, B.J., Slack, F., Martindale, M.Q., Kuroda, M.I., Maller, B., Hayward, D.C., Ball, E.E., Degnan, B., Müller, P., Spring, J., Srinivasan, A., Fishman, M., Finnerty, J., Corbo, J., Levine, M., Leahy, P., Davidson, E. and Ruvkun, G. 2000. Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA. *Nature,* 408(6808), pp.86-89.

Pasquinelli, A.E. and Ruvkun, G. 2002. *Control of developmental timing by microRNAs and their targets.*

Patel, M.I., Tuckerman, R. and Dong, Q. 2005. A pitfall of the 3-(4,5-dimethylthiazol-2-yl)-5(3-carboxymethonyphenol)-2- (4-sulfophenyl)-2H-tetrazolium (MTS) assay due to evaporation in wells on the edge of a 96 well plate. *Biotechnology Letters,* 27(11), pp.805-808.

Pattanayak, V., Lin, S., Guilinger, J.P., Ma, E., Doudna, J.A. and Liu, D.R. 2013. High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nature Biotechnology,* 31(9), pp.839-843.

Peng, R.-., Abellan, E. and Fussenegger, M. 2011. Differential effect of exocytic SNAREs on the production of recombinant proteins in mammalian cells. *Biotechnology and Bioengineering,* 108(3), pp.611-620.

Peng, R.-. and Fussenegger, M. 2009. Molecular engineering of exocytic vesicle traffic enhances the productivity of chinese hamster ovary cells. *Biotechnology and Bioengineering,* 102(4), pp.1170-1181.

Perez, E.E., Wang, J., Miller, J.C., Jouvenot, Y., Kim, K.A., Liu, O., Wang, N., Lee, G., Bartsevich, V.V., Lee, Y.-., Guschin, D.Y., Rupniewski, I., Waite, A.J., Carpenito, C., Carroll, R.G., Orange, J.S., Urnov, F.D., Rebar, E.J., Ando, D., Gregory, P.D., Riley, J.L., Holmes, M.C. and June, C.H. 2008. Establishment of HIV-1 resistance in CD4+ T cells by genome editing using zinc-finger nucleases. *Nature Biotechnology,* 26(7), pp.808-816.

Peters, L. and Meister, G. 2007. Argonaute proteins: Mediators of RNA silencing. *Molecular Cell,* 26(5), pp.611-623.

Porkka, K.P., Pfeiffer, M.J., Waltering, K.K., Vessella, R.L., Tammela, T.L.J. and Visakorpi, T. 2007. MicroRNA expression profiling in prostate cancer. *Cancer Research,* 67(13), pp.6130-6135.

Qiu, P., Shandilya, H., D'Alessio, J.M., O'Connor, K., Durocher, J. and Gerard, G.F. 2004. Mutation detection using surveyor™ nuclease. *BioTechniques,* 36(4), pp.702-707.

Rahimpour, A., Vaziri, B., Moazzami, R., Nematollahi, L., Barkhordari, F., Kokabee, L., Adeli, A. and Mahboudi, F. 2013. Engineering the cellular protein secretory pathway for enhancement of recombinant tissue plasminogen activator expression in chinese hamster ovary cells: Effects of CERT and XBP1s genes. *Journal of Microbiology and Biotechnology,* 23(8), pp.1116-1122.

Rai, K., Takigawa, N., Ito, S., Kashihara, H., Ichihara, E., Yasuda, T., Shimizu, K., Tanimoto, M. and Kiura, K. 2011. Liposomal delivery of MicroRNA-7-expressing plasmid overcomes epidermal growth factor receptor tyrosine kinase inhibitor-resistance in lung cancer cells. *Molecular Cancer Therapeutics,* 10(9), pp.1720-1727.

Ran, F.A., Hsu, P.D., Lin, C.-., Gootenberg, J.S., Konermann, S., Trevino, A.E., Scott, D.A., Inoue, A., Matoba, S., Zhang, Y. and Zhang, F. 2013. Double nicking by RNA-guided CRISPR cas9 for enhanced genome editing specificity. *Cell,* 154(6), pp.1380-1389.

Rathore, M.G., Saumet, A., Rossi, J.-., De Bettignies, C., Tempé, D., Lecellier, C.-. and Villalba, M. 2012. The NF-κB member p65 controls glutamine metabolism through miR-23a. *International Journal of Biochemistry and Cell Biology,* 44(9), pp.1448-1456.

Raver-Shapira, N., Marciano, E., Meiri, E., Spector, Y., Rosenfeld, N., Moskovits, N., Bentwich, Z. and Oren, M. 2007. Transcriptional activation of miR-34a contributes to p53-mediated apoptosis. *Molecular Cell,* 26(5), pp.731-743.

Reddy, S.D.N., Ohshiro, K., Rayala, S.K. and Kumar, R. 2008. MicroRNA-7, a homeobox D10 target, inhibits p21-activated kinase 1 and regulates its functions. *Cancer Research,* 68(20), pp.8195-8200.

Reinhart, B.J., Slack, F.J., Basson, M., Pasquienelll, A.E., Bettlnger, J.C., Rougvle, A.E., Horvitz, H.R. and Ruvkun, G. 2000. The 21-nucleotide let-7 RNA regulates developmental timing in caenorhabditis elegans. *Nature,* 403(6772), pp.901-906.

Ren, X., Yang, Z., Xu, J., Sun, J., Mao, D., Hu, Y., Yang, S.-., Qiao, H.-., Wang, X., Hu, Q., Deng, P., Liu, L.-., Ji, J.-., Li, J.B. and Ni, J.-. 2014. Enhanced specificity and efficiency of the CRISPR/Cas9 system with optimized sgRNA parameters in drosophila. *Cell Reports,* 9(3), pp.1151-1162.

Repligen (2011) Annual Report. www.repligen.com. Accessed August 2012.

Rita Costa, A., Elisa Rodrigues, M., Henriques, M., Azeredo, J. and Oliveira, R. 2010. Guidelines to cell engineering for monoclonal antibody production. *European Journal of Pharmaceutics and Biopharmaceutics,* 74(2), pp.127-138.

Rodriguez, A., Griffiths-Jones, S., Ashurst, J.L. and Bradley, A. 2004. Identification of mammalian microRNA host genes and transcription units. *Genome Research,* 14(10 A), pp.1902-1910.

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

Roush, S. and Slack, F.J. 2008. The let-7 family of microRNAs. *Trends in Cell Biology,* 18(10), pp.505-516.

Sacchetti, A., El Sewedy, T., Nasr, A.F. and Alberti, S. 2001. Efficient GFP mutations profoundly affect mRNA transcription and translation rates. *FEBS Letters,* 492(1-2), pp.151-155.

Sakuma, T., Nishikawa, A., Kume, S., Chayama, K. and Yamamoto, T. 2014. Multiplex genome engineering in human cells using all-in-one CRISPR/Cas9 vector system. *Scientific Reports,* 4

Sanchez, N., Gallagher, M., Lao, N., Gallagher, C., Clarke, C., Doolan, P., Aherne, S., Blanco, A., Meleady, P., Clynes, M. and Barron, N. 2013. MiR-7 triggers cell cycle arrest at the G1/S transition by targeting multiple genes including Skp2 and Psme3. *PLoS ONE,* 8(6),

Sanchez, N., Kelly, P., Gallagher, C., Lao, N.T., Clarke, C., Clynes, M. and 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(3), pp.396-404.

Santiago, Y., Chan, E., Liu, P.-., Orlando, S., Zhang, L., Urnov, F.D., Holmes, M.C., Guschin, D., Waite, A., Miller, J.C., Rebar, E.J., Gregory, P.D., Klug, A. and Collingwood, T.N. 2008. Targeted gene knockout in mammalian cells by using engineered zinc-finger nucleases. *Proceedings of the National Academy of Sciences of the United States of America,* 105(15), pp.5809-5814.

Sapranauskas, R., Gasiunas, G., Fremaux, C., Barrangou, R., Horvath, P. and Siksnys, V. 2011. The streptococcus thermophilus CRISPR/cas system provides immunity in escherichia coli. *Nucleic Acids Research,* 39(21), pp.9275-9282.

Sauerwald, T.M., Oyler, G.A. and Betenbaugh, M.J. 2003. Study of caspase inhibitors for limiting death in mammalian cell culture. *Biotechnology and Bioengineering,* 81(3), pp.329-340.

Scherr, M., Venturini, L., Battmer, K., Schaller-schoenitz, M., Schaefer, D., Dallmann, I., Ganser, A. and Eder, M. 2007. Lentivirus-mediated antagomir expression for specific inhibition of miRNA function. *Nucleic Acids Research,* 35(22),

Schiffer, J.T., Aubert, M., Weber, N.D., Mintzer, E., Stone, D. and Jerome, K.R. 2012. Targeted DNA mutagenesis for the cure of chronic viral infections. *Journal of Virology,* 86(17), pp.8920-8936.

Schröder, M. and Kaufman, R.J. 2005. ER stress and the unfolded protein response. *Mutation Research - Fundamental and Molecular Mechanisms of Mutagenesis,* 569(1-2), pp.29-63.

Schröder, M., Schäfer, R. and Friedl, P. 2002. Induction of protein aggregation in an early secretory compartment by elevation of expression level. *Biotechnology and Bioengineering,* 78(2), pp.131-140.

Schubert, U., Antón, L.C., Gibbs, J., Norbury, C.C., Yewdell, J.W. and Bennink, J.R. 2000. Rapid degradation of a large fraction of newly synthesized proteins by proteasomes. *Nature,* 404(6779), pp.770-774.

Schulman, B.R.M., Esquela-Kerscher, A. and Slack, F.J. 2005. Reciprocal expression of lin-41 and the microRNAs let-7 and mir-125 during mouse embryogenesis. *Developmental Dynamics,* 234(4), pp.1046-1054.

Schwank, G., Koo, B.-., Sasselli, V., Dekkers, J.F., Heo, I., Demircan, T., Sasaki, N., Boymans, S., Cuppen, E., Van Der Ent, C.K., Nieuwenhuis, E.E.S., Beekman, J.M. and Clevers, H. 2013. Functional repair of CFTR by CRISPR/Cas9 in intestinal stem cell organoids of cystic fibrosis patients. *Cell Stem Cell,* 13(6), pp.653-658.

Schwarz, D.S., Hutvágner, G., Du, T., Xu, Z., Aronin, N. and Zamore, P.D. 2003. Asymmetry in the assembly of the RNAi enzyme complex. *Cell,* 115(2), pp.199-208.

Sedivy, J.M. and Sharp, P.A. 1989. Positive genetic selection for gene disruption in mammalian cells by homologous recombination. *Proceedings of the National Academy of Sciences of the United States of America,* 86(1), pp.227-231.

Selbach, M., Schwanhäusser, B., Thierfelder, N., Fang, Z., Khanin, R. and Rajewsky, N. 2008. Widespread changes in protein synthesis induced by microRNAs. *Nature,* 455(7209), pp.58-63.

Sempere, L.F., Dubrovsky, E.B., Dubrovskaya, V.A., Berger, E.M. and Ambros, V. 2002. The expression of the let-7 small regulatory RNA is controlled by ecdysone during metamorphosis in drosophila melanogaster. *Developmental Biology,* 244(1), pp.170-179.

Sempere, L.F., Sokol, N.S., Dubrovsky, E.B., Berger, E.M. and Ambros, V. 2003. Temporal regulation of microRNA expression in drosophila melanogaster mediated by hormonal signals and broad-complex gene activity. *Developmental Biology,* 259(1), pp.9-18.

Shaffer, A.L., Shapiro-Shelef, M., Iwakoshi, N.N., Lee, A.-., Qian, S.-., Zhao, H., Yu, X., Yang, L., Tan, B.K., Rosenwald, A., Hurt, E.M., Petroulakis, E., Sonenberg, N., Yewdell, J.W., Calame, K., Glimcher, L.H. and Staudt, L.M. 2004. XBP1, downstream of blimp-1, expands the secretory apparatus and other organelles, and increases protein synthesis in plasma cell differentiation. *Immunity,* 21(1), pp.81-93.

Shalem, O., Sanjana, N.E., Hartenian, E., Shi, X., Scott, D.A., Mikkelsen, T.S., Heckl, D., Ebert, B.L., Root, D.E., Doench, J.G. and Zhang, F. 2014. Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science,* 343(6166), pp.84-87.

Shen, D., Kiehl, T.R., Khattak, S.F., Li, Z.J., He, A., Kayne, P.S., Patel, V., Neuhaus, I.M. and Sharfstein, S.T. 2010. Transcriptomic responses to sodium chloride-induced osmotic stress: A study of industrial fed-batch CHO cell cultures. *Biotechnology Progress,* 26(4), pp.1104-1115.

Shmakov, S., Abudayyeh, O.O., Makarova, K.S., Wolf, Y.I., Gootenberg, J.S., Semenova, E., Minakhin, L., Joung, J., Konermann, S., Severinov, K., Zhang, F. and Koonin, E.V. 2015. Discovery and functional characterization of diverse class 2 CRISPR-cas systems. *Molecular Cell,* 60(3), pp.385-397.

Shuman, S. 1994. Novel approach to molecular cloning and polynucleotide synthesis using vaccinia DNA topoisomerase. *Journal of Biological Chemistry,* 269(51), pp.32678-32684.

Shuman, S. 1991. Recombination mediated by vaccinia virus DNA topoisomerase I in escherichia coli is sequence specific. *Proceedings of the National Academy of Sciences of the United States of America,* 88(22), pp.10104-10108.

Sinkunas, T., Gasiunas, G., Fremaux, C., Barrangou, R., Horvath, P. and Siksnys, V. 2011. Cas3 is a single-stranded DNA nuclease and ATP-dependent helicase in the CRISPR/cas immune system. *EMBO Journal,* 30(7), pp.1335-1342.

Slack, F.J. and Weidhass, J.B. 2006. MicroRNAs as a potential magic bullet in cancer. *Future Oncology,* 2(1), pp.73-82.

Smales, C.M., Dinnis, D.M., Stansfield, S.H., Alete, D., Sage, E.A., Birch, J.R., Racher, A.J., Marshall, C.T. and James, D.C. 2004. Comparative proteomic analysis of GS-NSO murine myeloma cell lines with varying recombinant monoclonal antibody production rate. *Biotechnology and Bioengineering,* 88(4), pp.474-488.

Smith, J.R., Maguire, S., Davis, L.A., Alexander, M., Yang, F., Chandran, S., Ffrench-Constant, C. and Pedersen, R.A. 2008. Robust, persistent transgene expression in human embryonic stem cells is achieved with AAVS1-targeted integration. *Stem Cells,* 26(2), pp.496-504.

Sokol, N.S., Xu, P., Jan, Y.-. and Ambros, V. 2008. Drosophila let-7 microRNA is required for remodeling of the neuromusculature during metamorphosis. *Genes and Development,* 22(12), pp.1591-1596.

Solnica-Krezel, L., Schier, A.F. and Driever, W. 1994. Efficient recovery of ENU-induced mutations from the zebrafish germline. *Genetics,* 136(4), pp.1401-1420.

Solomides, C.C., Evans, B.J., Navenot, J.-., Vadigepalli, R., Peiper, S.C. and Wang, Z.-. 2012. MicroRNA profiling in lung cancer reveals new molecular markers for diagnosis. *Acta Cytologica,* 56(6), pp.645-654.

Song, R., Walentek, P., Sponer, N., Klimke, A., Lee, J.S., Dixon, G., Harland, R., Wan, Y., Lishko, P., Lize, M., Kessel, M. and He, L. 2014. MiR-34/449 miRNAs are required for motile ciliogenesis by repressing cp110. *Nature,* 510(7503), pp.115-120.

Spenger, A., Ernst, W., Condreay, J.P., Kost, T.A. and Grabherr, R. 2004. Influence of promoter choice and trichostatin a treatment on expression of baculovirus delivered genes in mammalian cells. *Protein Expression and Purification,* 38(1), pp.17-23.

Spradling, A.C. and Rubin, G.M. 1982. Transposition of cloned P elements into drosophila germ line chromosomes. *Science,* 218(4570), pp.341-347.

Sun, N., Liang, J., Abil, Z. and Zhao, H. 2012. Optimized TAL effector nucleases (TALENs) for use in treatment of sickle cell disease. *Molecular BioSystems,* 8(4), pp.1255-1263.

Sung, Y.H., Baek, I.-., Kim, D.H., Jeon, J., Lee, J., Lee, K., Jeong, D., Kim, J.-. and Lee, H.-. 2013. Knockout mice created by TALEN-mediated gene targeting. *Nature Biotechnology,* 31(1), pp.23-24.

Sung, Y.H., Lee, J.S., Park, S.H., Koo, J. and Lee, G.M. 2007. Influence of co-down-regulation of caspase-3 and caspase-7 by siRNAs on sodium butyrate-induced apoptotic cell death of chinese hamster ovary cells producing thrombopoietin. *Metabolic Engineering,* 9(5-6), pp.452-464.

Sunley, K. and Butler, M. 2010. Strategies for the enhancement of recombinant protein production from mammalian cells by growth arrest. *Biotechnology Advances,* 28(3), pp.385-394.

Takamizawa, J., Konishi, H., Yanagisawa, K., Tomida, S., Osada, H., Endoh, H., Harano, T., Yatabe, Y., Nagino, M., Nimura, Y., Mitsudomi, T. and Takahashi, T. 2004. Reduced expression of the let-7 microRNAs in human lung cancers in association with shortened postoperative survival. *Cancer Research,* 64(11), pp.3753-3756.

Tang, F., Hajkova, P., Barton, S.C., Lao, K. and Surani, M.A. 2006. MicroRNA expression profiling of single whole embryonic stem cells. *Nucleic Acids Research.,* 34(2),

Tartaglia, G.G., Pechmann, S., Dobson, C.M. and Vendruscolo, M. 2007. Life on the edge: A link between gene expression levels and aggregation rates of human proteins. *Trends in Biochemical Sciences,* 32(5), pp.204-206.

Thaisuchat, H., Baumann, M., Pontiller, J., Hesse, F. and Ernst, W. 2011. Identification of a novel temperature sensitive promoter in cho cells. *BMC Biotechnology,* 11

Thomson, J.M., Newman, M., Parker, J.S., Morin-Kensicki, E.M., Wright, T. and Hammond, S.M. 2006. Extensive post-transcriptional regulation of microRNAs and its implications for cancer. *Genes and Development,* 20(16), pp.2202-2207.

Tigges, M. and Fussenegger, M. 2006. Xbp1-based engineering of secretory capacity enhances the productivity of chinese hamster ovary cells. *Metabolic Engineering,* 8(3), pp.264-272.

Toonen, R.F.G. and Verhage, M. 2003. Vesicle trafficking: Pleasure and pain from SM genes. *Trends in Cell Biology,* 13(4), pp.177-186.

Tsuji, T. and Niida, Y. 2008. Development of a simple and highly sensitive mutation screening system by enzyme mismatch cleavage with optimized conditions for standard laboratories. *Electrophoresis,* 29(7), pp.1473-1483.

Tupler, R., Perini, G. and Green, M.R. 2001. Expressing the human genome. *Nature,* 409(6822), pp.832-833.

Urnov, F.D., Rebar, E.J., Holmes, M.C., Zhang, H.S. and Gregory, P.D. 2010. Genome editing with engineered zinc finger nucleases. *Nature Reviews Genetics,* 11(9), pp.636-646.

van den Bosch, M., Lohman, P.H.M. and Pastink, A. 2002. DNA double-strand break repair by homologous recombination. *Biological Chemistry,* 383(6), pp.873-892.

Van Dyk, D.D., Misztal, D.R., Wilkins, M.R., Mackintosh, J.A., Poljak, A., Varnai, J.C., Teber, E., Walsh, B.J. and Gray, P.P. 2003. Identification of cellular changes associated with increased production of human growth hormone in a recombinant chinese hamster ovary cell line. *Proteomics,* 3(2), pp.147-156.

Vidigal, J.A. and Ventura, A. 2015. The biological functions of miRNAs: Lessons from in vivo studies. *Trends in Cell Biology,* 25(3), pp.137-147.

Vishwanathan, N., Yongky, A., Johnson, K.C., Fu, H.Y., Jacob, N.M., Le, H., Yusufi, F.N.K., Lee, D.Y. and Hu, W.S. 2015. Global Insights into the Chinese Hamster and CHO cell transcriptomes. *Biotechnology and Bioengineering*, 112(5). pp. 965-976.

Walsh (2011) New Biopharmaceuticals: A review of new biologic drug approvals over the years, featuring highlights from 2010 and 2011. *BioPharm International*.

Walsh (2012) New Biopharmaceuticals: A review of new biologic drug approvals over the years, featuring highlights from 2011 and 2012. *BioPharm International*.

Wang, J. and Quake, S.R. 2014. RNA-guided endonuclease provides a therapeutic strategy to cure latent herpesviridae infection. *Proceedings of the National Academy of Sciences of the United States of America,* 111(36), pp.13157-13162.

Wang, S., Aurora, A.B., Johnson, B.A., Qi, X., McAnally, J., Hill, J.A., Richardson, J.A., Bassel-Duby, R. and Olson, E.N. 2008. The endothelial-specific MicroRNA miR-126 governs vascular integrity and angiogenesis. *Developmental Cell,* 15(2), pp.261-271.

Wang, Y., Ju, T., Ding, X., Xia, B., Wang, W., Xia, L., He, M. and Cummings, R.D. 2010. Cosmc is an essential chaperone for correct protein O-glycosylation. *Proceedings of the National Academy of Sciences of the United States of America,* 107(20), pp.9228-9233.

Wang, Z. 2011. The guideline of the design and validation of MiRNA mimics. *Methods in Molecular Biology (Clifton, N.J.),* 676pp.211-223.

Weber, M.J. 2005. New human and mouse microRNA genes found by homology search. *FEBS Journal,* 272(1), pp.59-73.

Weber, W. and Fussenegger, M. 2007. Inducible product gene expression technology tailored to bioprocess engineering. *Current Opinion in Biotechnology,* 18(5), pp.399-410.

Webster, R.J., Giles, K.M., Price, K.J., Zhang, P.M., Mattick, J.S. and Leedman, P.J. 2009. Regulation of epidermal growth factor receptor signaling in human cancer cells by MicroRNA-7. *Journal of Biological Chemistry,* 284(9), pp.5731-5741.

Wei, C., Liu, J., Yu, Z., Zhang, B., Gao, G. and Jiao, R. 2013. TALEN or Cas9 - rapid, efficient and specific choices for genome modifications. *Journal of Genetics and Genomics,* 40(6), pp.281-289.

West, A.G. and Fraser, P. 2005. Remote control of gene transcription. *Human Molecular Genetics,* 14(SPEC. ISS. 1), pp.R101-R111.

Wiedenheft, B., Sternberg, S.H. and Doudna, J.A. 2012. RNA-guided genetic silencing systems in bacteria and archaea. *Nature,* 482(7385), pp.331-338.

Wlaschin, K.F. and Hu, W.-. 2007. Engineering cell metabolism for high-density cell culture via manipulation of sugar transport. *Journal of Biotechnology,* 131(2), pp.168-176.

Wolfe, S.A., Nekludova, L. and Pabo, C.O. 2000. *DNA recognition by Cys2His2 zinc finger proteins.*

Wong, D.C.F., Wong, K.T.K., Nissom, P.M., Heng, C.K. and Yap, M.G.S. 2006. Targeting early apoptotic genes in batch and fed-batch CHO cell cultures. *Biotechnology and Bioengineering,* 95(3), pp.350-361.

Wu, L., Zhou, H., Zhang, Q., Zhang, J., Ni, F., Liu, C. and Qi, Y. 2010. DNA methylation mediated by a MicroRNA pathway. *Molecular Cell,* 38(3), pp.465-475.

Wu, S.-. 2009. RNA interference technology to improve recombinant protein production in chinese hamster ovary cells. *Biotechnology Advances,* 27(4), pp.417-422.

Wu, X., Scott, D.A., Kriz, A.J., Chiu, A.C., Hsu, P.D., Dadon, D.B., Cheng, A.W., Trevino, A.E., Konermann, S., Chen, S., Jaenisch, R., Zhang, F. and Sharp, P.A. 2014. Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. *Nature Biotechnology,* 32(7), pp.670-676.

Wurm, F.M. 2004. Production of recombinant protein therapeutics in cultivated mammalian cells. *Nature Biotechnology,* 22(11), pp.1393-1398.

Xi, J.J. 2013. *MicroRNAs in cancer.*

Xiong, S., Zheng, Y., Jiang, P., Liu, R., Liu, X., Qian, J., Gu, J., Chang, L., Ge, D. and Chu, Y. 2014. PA28gamma emerges as a novel functional target of tumour suppressor microRNA-7 in non-small-cell lung cancer. *British Journal of Cancer,* 110(2), pp.353-362.

Xu, C., Lu, Y., Pan, Z., Chu, W., Luo, X., Lin, H., Xiao, J., Shan, H., Wang, Z. and Yang, B. 2007. The muscle-specific microRNAs miR-1 and miR-133 produce opposing effects on apoptosis by targeting HSP60, HSP70 and caspase-9 in cardiomyocytes. *Journal of Cell Science,* 120(17), pp.3045-3052.

Xu, P., Vernooy, S.Y., Guo, M. and Hay, B.A. 2003. The drosophila microRNA mir-14 suppresses cell death and is required for normal fat metabolism. *Current Biology,* 13(9), pp.790-795.

Xu, X., Nagarajan, H., Lewis, N.E., Pan, S., Cai, Z., Liu, X., Chen, W., Xie, M., Wang, W., Hammond, S., Andersen, M.R., Neff, N., Passarelli, B., Koh, W., Fan, H.C., Wang, J., Gui, Y., Lee, K.H., Betenbaugh, M.J., Quake, S.R., Famili, I., Palsson, B.O. and Wang, J. 2011. The genomic sequence of the chinese hamster ovary (CHO)-K1 cell line. *Nature Biotechnology,* 29(8), pp.735-741.

Yang, L., Guell, M., Byrne, S., Yang, J.L., De Los Angeles, A., Mali, P., Aach, J., Kim-Kiselak, C., Briggs, A.W., Rios, X., Huang, P.-., Daley, G. and Church, G. 2013.

Optimization of scarless human stem cell genome editing. *Nucleic Acids Research,* 41(19), pp.9049-9061.

Yang, M. and Butler, M. 2000. Effects of ammonia on CHO cell growth, erythropoietin production, and glycosylation. *Biotechnology and Bioengineering,* 68(4), pp.370-380.

Yang, W. and Paschen, W. 2008. Conditional gene silencing in mammalian cells mediated by a stress-inducible promoter. *Biochemical and Biophysical Research Communications,* 365(3), pp.521-527.

Yang, X., Rutnam, Z.J., Jiao, C., Wei, D., Xie, Y., Du, J., Zhong, L. and Yang, B.B. 2012a. An anti-let-7 sponge decoys and decays endogenous let-7 functions. *Cell Cycle,* 11(16), pp.3097-3108.

Yang, X., Rutnam, Z.J., Jiao, C., Wei, D., Xie, Y., Du, J., Zhong, L. and Yang, B.B. 2012b. An anti-let-7 sponge decoys and decays endogenous let-7 functions. *Cell Cycle,* 11(16), pp.3097-3108.

Yang, Y., Mariati, Chusainow, J. and Yap, M.G.S. 2010. DNA methylation contributes to loss in productivity of monoclonal antibody-producing CHO cell lines. *Journal of Biotechnology,* 147(3-4), pp.180-185.

Yewdell, J.W. and Nicchitta, C.V. 2006. The DRiP hypothesis decennial: Support, controversy, refinement and extension. *Trends in Immunology,* 27(8), pp.368-373.

Yi, R., Qin, Y., Macara, I.G. and Cullen, B.R. 2003. Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. *Genes and Development,* 17(24), pp.3011-3016.

Yoo, A.S., Staahl, B.T., Chen, L. and Crabtree, G.R. 2009. MicroRNA-mediated switching of chromatin-remodelling complexes in neural development. *Nature,* 460(7255), pp.642-646.

Yoon, S.K., Song, J.Y. and Lee, G.M. 2003. Effect of low culture temperature on specific productivity, transcription level, and heterogeneity of erythropoietin in chinese hamster ovary cells. *Biotechnology and Bioengineering,* 82(3), pp.289-298.

Zanghi, J.A., Mendoza, T.P., Knop, R.H. and Miller, W.M. 1998. Ammonia inhibits neural cell adhesion molecule polysialylation in chinese hamster ovary and small cell lung cancer cells. *Journal of Cellular Physiology,* 177(2), pp.248-263.

Zeng, A.-., Deckwer, W.-. and Hu, W.-. 1998. Determinants and rate laws of growth and death of hybridoma cells in continuous culture. *Biotechnology and Bioengineering,* 57(6), pp.642-654.

Zeng, Y. and Cullen, B.R. 2004. Structural requirements for pre-microRNA binding and nuclear export by exportin 5. *Nucleic Acids Research,* 32(16), pp.4776-4785.

Zhang, F., Sun, X., Yi, X. and Zhang, Y. 2006. Metabolic characteristics of recombinant chinese hamster ovary cells expressing glutamine synthetase in presence and absence of glutamine. *Cytotechnology,* 51(1), pp.21-28.

Zhang, M., Wang, F., Li, S., Wang, Y., Bai, Y. and Xu, X. 2014. TALE: A tale of genome editing. *Progress in Biophysics and Molecular Biology,* 114(1), pp.25-32.

Zhang, N., Li, X., Wu, C.W., Dong, Y., Cai, M., Mok, M.T.S., Wang, H., Chen, J., Ng, S.S.M., Chen, M., Sung, J.J.Y. and Yu, J. 2013. MicroRNA-7 is a novel inhibitor of YY1 contributing to colorectal tumorigenesis. *Oncogene,* 32(42), pp.5078-5088.

Zhao, X., Dou, W., He, L., Liang, S., Tie, J., Liu, C., Li, T., Lu, Y., Mo, P., Shi, Y., Wu, K., Nie, Y. and Fan, D. 2013. MicroRNA-7 functions as an anti-metastatic microRNA in gastric cancer by targeting insulin-like growth factor-1 receptor. *Oncogene,* 32(11), pp.1363-1372.

Zhou, J., Li, Y.-., Nguyen, P., Wang, K.-., Weiss, A., Kuo, Y.-., Chiu, J.-., Shyy, J.-. and Chien, S. 2013. Regulation of vascular smooth muscle cell turnover by endothelial cell-secreted microRNA-126 role of shear stress. *Circulation Research,* 113(1), pp.40-51.

Zhu, H., Ng, S.-., Segr, A.V., Shinoda, G., Shah, S.P., Einhorn, W.S., Takeuchi, A., Engreitz, J.M., Hagan, J.P., Kharas, M.G., Urbach, A., Thornton, J.E., Triboulet, R., Gregory, R.I., Altshuler, D. and Daley, G.Q. 2011. The Lin28/let-7 axis regulates glucose metabolism. *Cell,* 147(1), pp.81-94.

Zimmermann, H.F., John, G.T., Trauthwein, H., Dingerdissen, U. and Huthmacher, K. 2003. Rapid evaluation of oxygen and water permeation through microplate sealing tapes. *Biotechnology Progress,* 19(3), pp.1061-1063.

Zu, Y., Tong, X., Wang, Z., Liu, D., Pan, R., Li, Z., Hu, Y., Luo, Z., Huang, P., Wu, Q., Zhu, Z., Zhang, B. and Lin, S. 2013. TALEN-mediated precise genome modification by homologous recombination in zebrafish. *Nature Methods,* 10(4), pp.329-331.

Section 5.0

Appendix

(A) **(B)**

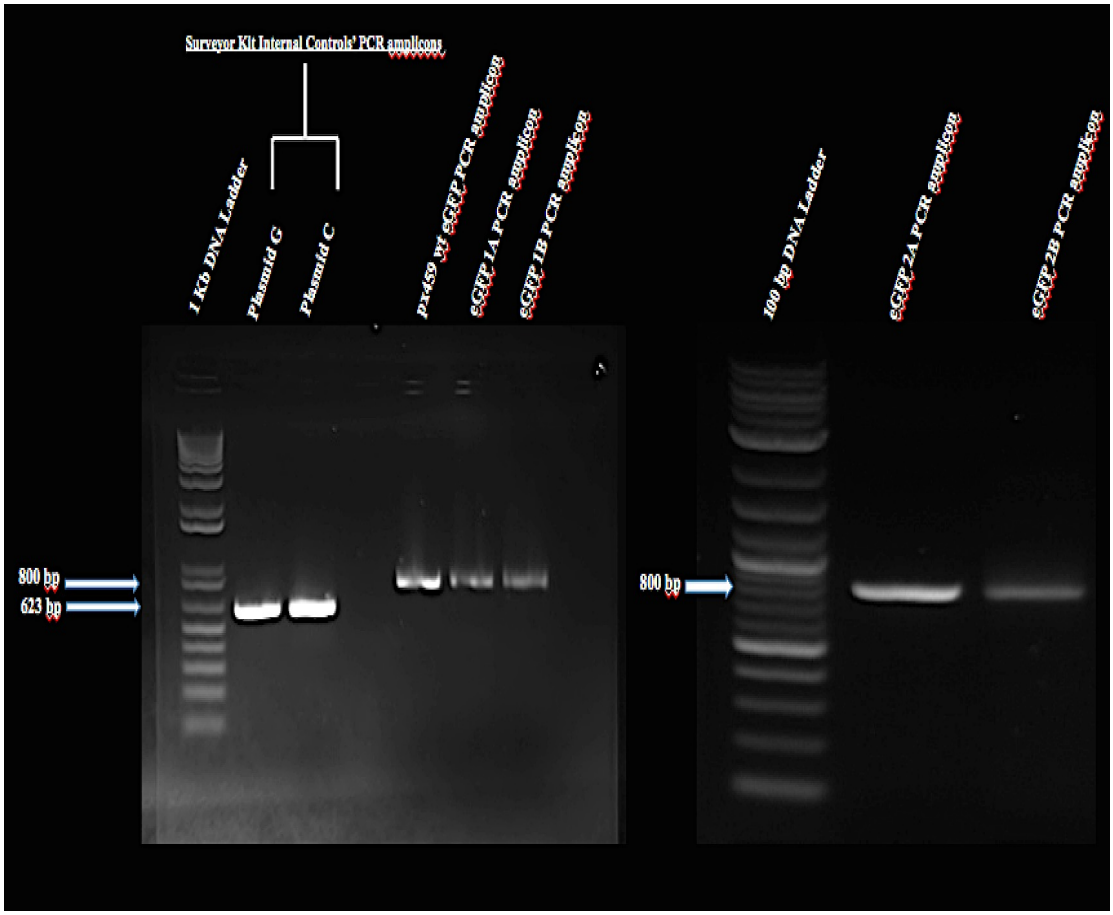


Figure A1: PCR optimization for Surveyor Assay. (A) Gel image showing plasmid G and plasmid C from the Surveyor kit's Internal control plasmids used to form amplicon of size 623 bp and px459 wt or reference eGFP, eGFP 1A and eGFP 1B PCR amplicons of size 800 bp from the genomic DNA extracted from px459 empty vector only plasmid DNA, eGFP 1A and eGFP 1B CRISPR transfected eGFP cells, respectively (B) Gel image showing eGFP 2A and eGFP 2B PCR amplicons of size 800 bp from the genomic DNA extracted from eGFP 2A and eGFP 2B CRISPR transfected eGFP cells, respectively.

(C)

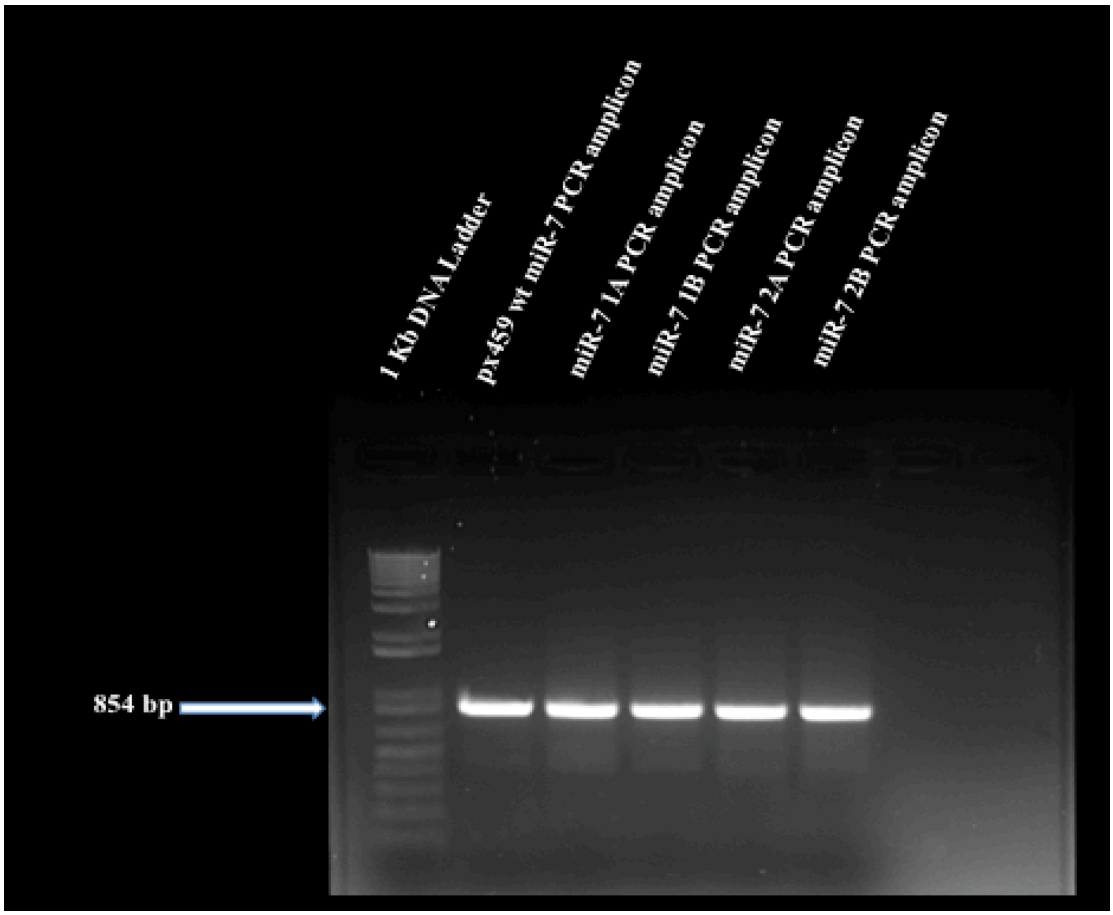


Figure A2: PCR optimization for Surveyor Assay. (C) Gel image showing miR-7 1A, miR-7 1B, miR-7 2A and miR-7 2B PCR amplicons of size 854 bp from the genomic DNA extracted from px459 empty vector only plasmid DNA, miR-7 1A, miR-7 1B, miR-7 2A and miR-7 2B CRISPR transfected 1.14 cells, respectively.

(A) **(B)**

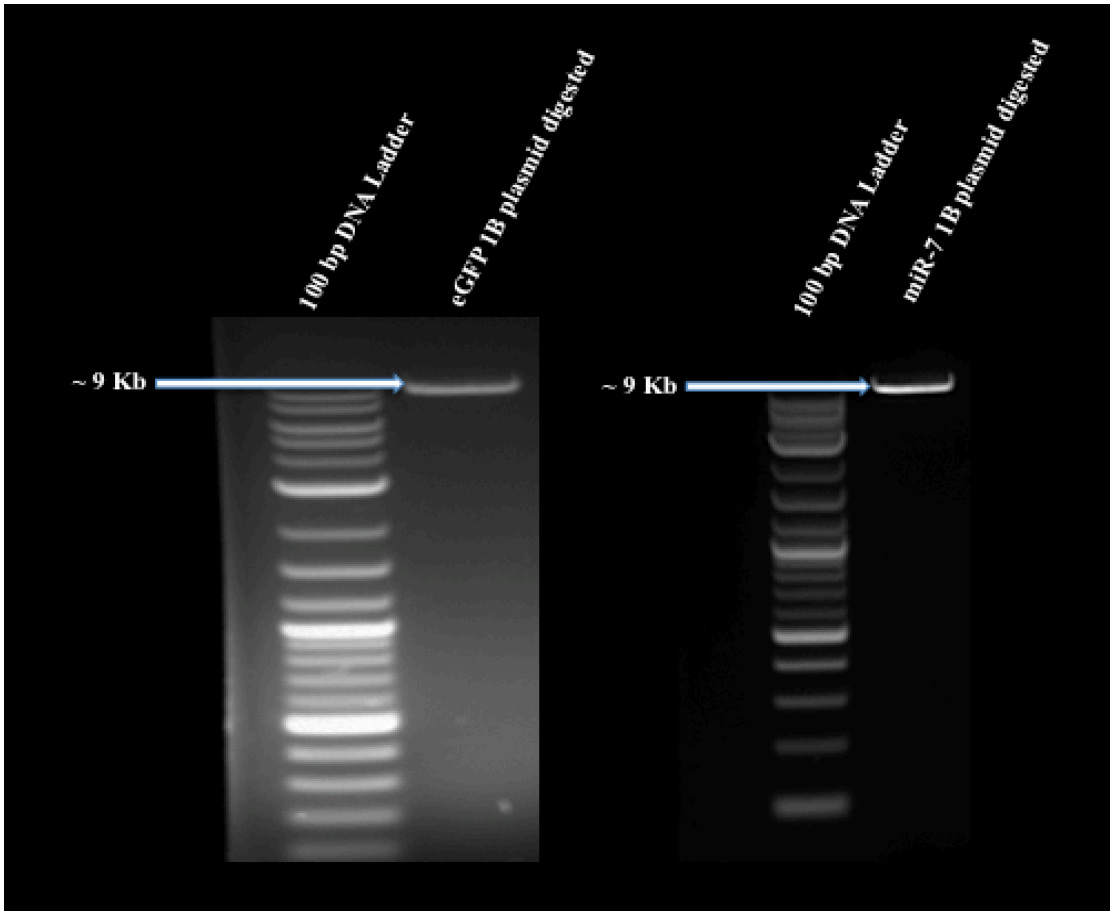


Figure A3: Diagnostic gels for eGFP 1B and miR-7 1B plasmid CRISPR DNA linearization. (A) eGFP 1B plasmid doubly digested with *XbaI* + *Kpn I* to linearize and generating ~ 9 Kb eGFP 1B backbone vector (B) miR-7 1B plasmid doubly digested with *XbaI* + *Kpn I* to linearize and generating ~ 9 Kb miR-7 1B backbone vector.

(C)

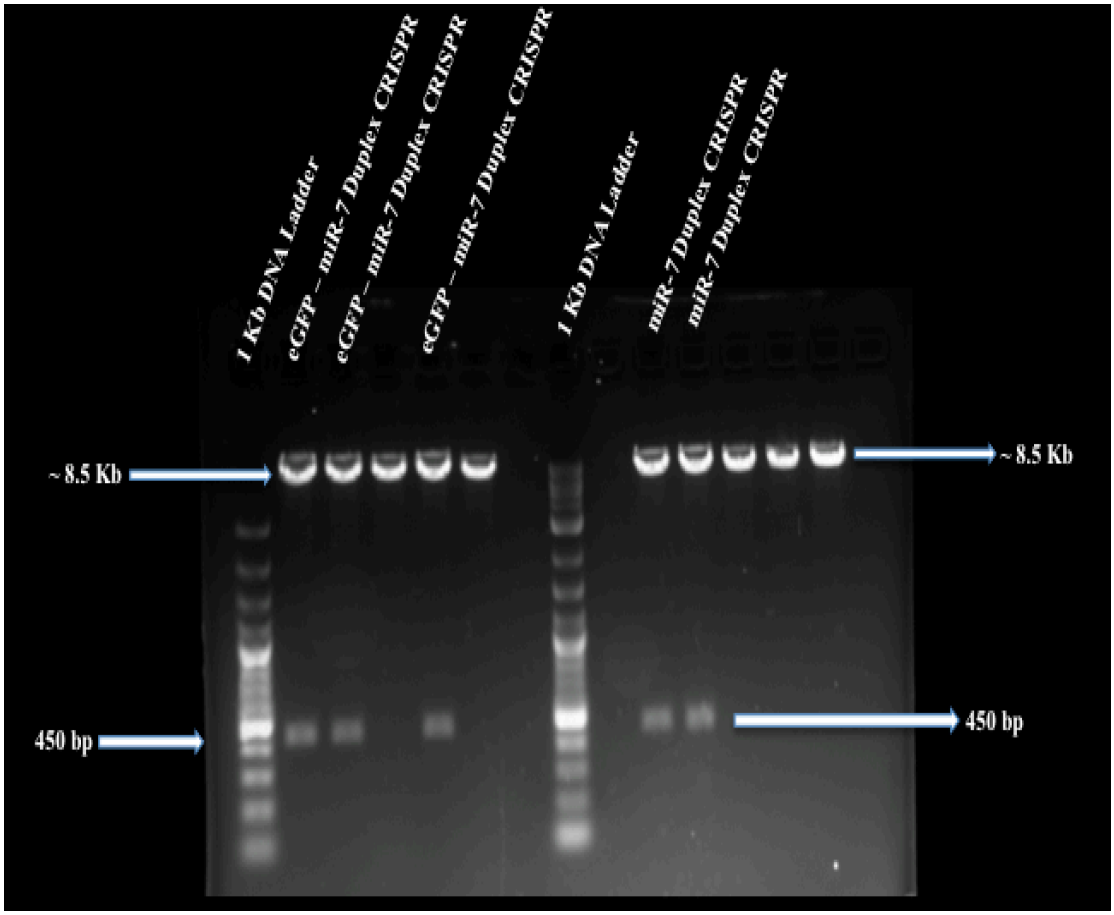


Figure A4: Diagnostic gels for eGFP and miR-7 duplex CRISPR plasmid DNA preparation. (C) Left hand side of the gel image showing three positive eGFP duplex CRISPR samples with 450 bp miR-7 2B guide RNA and U6 promoter sequence ligated along with *XbaI* + *KpnI* recognition sites, while right hand side of the image showing two positive miR-7 duplex CRISPR samples with 450 bp miR-7 2B guide RNA and U6 promoter sequence ligated along with *XbaI* + *KpnI* recognition sites.

Table A1: List of all the primers and oligo sequences used in this project

Primer & Oligos Name	Type	Sequence (5'→3' direction)
GFP_Cas1s	Sense sequence for cloning	GATCGAGGGCGAGGAGCTGTTCACCG
GFP_Cas1as	Sequence for cloning	AAAACGGTGAACAGCTCCTCGCCCTC
GFP_Cas2s	Sequence for cloning	GATCGGCGCGATCACATGGTCCTGCG
GFP_Cas2as	Sequence for cloning	AAAACGCAGGACCATGTGATCGCGCC
GFP1F	GFP sequence primers	GACTCCACGGAGTACCGG
GFP1R	GFP sequence primers	GTGGGCCATGATATAGACGTT
GFP2F	GFP sequence primers	GCGCACCATCTTCTTCAAG
GFP2R	GFP sequence primers	GCTGAGCCTGGTCATGCAT
px459_R640	Duplex CRISPR for sequencing	GCCATTTACCGTCATTGACGTC
px459_R423_K	Duplex CRISPR for sequencing	GGGGTACCTTGTCTGCAGAATTGGCGCACG
px459_F1_Xba	Duplex CRISPR for sequencing	GCTCTAGAGAGGGCCTATTTCCCATGATTC
miR7_F_a459	miR7 sequence primers	CCTGTCCATGCAATTAGAGGC
miR7_R_a1520	miR7 sequence primers	CAGACATTGCAGACCATCCAG
miR7_F_b586	miR7 sequence primers	AGCTATGGGTTATGCGCTCT
miR_7_R_b1440	miR7 sequence primers	TGTTACCTGCCCTCCATGAG
miR7_F_131 upstr	miR7 sequence primers	CAAGGAGTTTTGGGGTGTGG
miR7_R_seed seq	miR7 sequence	CAACAAAATCACTAGTCTTCC

	primers	
miR7_CRIS_R	miR7 sequence primers	CTCCCATACCTCCCCCTGATG
miR7_Loop_R	miR7 sequence primers	TGAGTATAGAGAGACAACAAC
Cgr-premir7a_F	Premir7a sequence primer	CCACAGGAGGTGGTTAGCAA
Cgr-premir7a_R	Premir7a sequence primer	ATCCAGAAATGGTGCTCCCA
Cgr-premir7b_F	Premir7b sequence primer	TGAGCGTGTGAGCCAGTGCTAT
Cgr-premir7b_R	Premir7b sequence primer	GGAGAAAGGTGAAAGGAGTCCAC
Cgr-miR7b_F	miR7b sequence primer	TCCTCCTGCCTCTACTTCCT
Cgr-miR7b_R	miR7b sequence primer	CGTTGACTTTGCCAGACTCC
CDC7_F	CDC7 gene primer	CATCTGACCTTCCCTCCAAA
CDC7_R	CDC7 gene primer	GATACGGCATGGCAATAACC
Rad54_F	Rad54L gene primer	CCATTAAGAAGCGAGCCAAG
Rad54_R	Rad54L gene primer	GGGTTCCAGTCAGGATCAAA

wt eGFP sequence

ATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCT
GGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCG
GCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTC
ATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGAC
CACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACA
TGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTC
CAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCG
CGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGC
TGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAG
CTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCGACAA
GCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCG
AGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCC
ATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCAC
CCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGG
TCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGAC
GAGCTGTACAAGTAG

wt eGFP sequence reverse complement

CTACTTGTACAGCTCGTCCATGCCGAGAGTGATCCCGGCGGCGGTCA
CGAACTCCAGCAGGACCATGTGATCGCGCTTCTCGTTGGGGTCTTTG
CTCAGGGCGGACTGGGTGCTCAGGTAGTGGTTGTCGGGCAGCAGCAC
GGGGCCGTCGCCGATGGGGGTGTTCTGCTGGTAGTGGTCGGCGAGC
TGCACGCTGCCGTCCTCGATGTTGTGGCGGATCTTGAAGTTCACCTT
GATGCCGTTCTTCTGCTTGTCGGCCATGATATAGACGTTGTGGCTGTT
GTAGTTGTACTCCAGCTTGTGCCCCAGGATGTTGCCGTCCTCCTTGAA
GTCGATGCCCTTCAGCTCGATGCGGTTCACCAGGGTGTCGCCCTCGA
ACTTCACCTCGGCGCGGGTCTTGTAGTTGCCGTCGTCCTTGAAGAAG
ATGGTGCGCTCCTGGACGTAGCCTTCGGGCATGGCGGACTTGAAGAA
GTCGTGCTGCTTCATGTGGTCGGGGTAGCGGCTGAAGCACTGCACGC
CGTAGGTCAGGGTGGTCACGAGGGTGGGCCAGGGCACGGGCAGCTT
GCCGGTGGTGCAGATGAACTTCAGGGTCAGCTTGCCGTAGGTGGCAT
CGCCCTCGCCCTCGCCGGACACGCTGAACTTGTGGCCGTTTACGTCG
CCGTCCAGCTCGACCAGGATGGGCACCACCCCGGTGAACAGCTCCTC
GCCCTTGCTCACCAT

GFP 1A2- 1 bp deletion

ATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCT
GGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCG
GCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTC
ATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGAC
CACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACA
TGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTC
CAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCG
CGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGC
TGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAG
CTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCGACAA
GCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCG
AGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCC
ATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCAC
CCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGG
TCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGAC
GAGCTGTACAAGTAG

GFP 1A3- 3 bp deletion

ATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCT
GGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCG
GCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTC
ATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGAC
CACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACA
TGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTC
CAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCG
CGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGC
TGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAG
CTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCGACAA
GCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCG
AGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCC
ATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCAC
CCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGG
TCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGAC
GAGCTGTACAAGTAG

GFP 1A4- 15 bp deletion + 204 bp eGFP sequence insertion

ATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCT
GGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCG
GCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTC
ATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGAC
CACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACA
TGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTC
CAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCG

CGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGC
TGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAG
CTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCGACAA
GCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCG
AGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCC
ATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCAC
CCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGG
TCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGAC
GAGCTGTACAAGTAG

GFP 1A6- 4 bp deletion

ATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCT
GGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCG
GCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTC
ATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGAC
CACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACA
TGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTC
CAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCG
CGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGC
TGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAG
CTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCGACAA
GCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCG
AGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCC
ATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCAC
CCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGG
TCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGAC
GAGCTGTACAAGTAG

GFP 1A7- 4 bp deletion

ATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCT
GGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCG
GCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTC
ATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGAC
CACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACA
TGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTC
CAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCG
CGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGC
TGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAG
CTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCGACAA
GCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCG
AGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCC
ATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCAC
CCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGG
TCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGAC
GAGCTGTACAAGTAG

GFP 1A9- 27 bp deletion

ATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCT
GGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCG
GCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTC
ATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGAC
CACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACA
TGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTC
CAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCG
CGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGC
TGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAG
CTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCGACAA
GCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCG
AGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCC
ATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCAC
CCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGG
TCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGAC
GAGCTGTACAAGTAG

GFP 1A12- 4 bp deletion + 69 bp insertion

ATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCT
GGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCG
GCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTC
ATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGAC
CACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACA
TGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTC
CAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCG
CGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGC
TGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAG
CTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCGACAA
GCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCG
AGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCC
ATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCAC
CCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGG
TCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGAC
GAGCTGTACAAGTAG

GFP 1A13- 1 bp deletion Exactly the same as GFP 1A2.

GFP 1A14- 3 bp deletion

ATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCT
GGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCG
GCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTC
ATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGAC
CACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACA
TGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTC
CAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCG
CGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGC
TGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAG

CTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCGACAA
GCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCG
AGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCC
ATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCAC
CCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGG
TCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGAC
GAGCTGTACAAGTAG

GFP 1A15- 12 bp deletion

ATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCT
GGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCG
GCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTC
ATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGAC
CACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACA
TGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTC
CAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCG
CGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGC
TGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAG
CTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCGACAA
GCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCG
AGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCC
ATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCAC
CCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGG
TCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGAC
GAGCTGTACAAGTAG

GFP 1B1- 18 bp deletion

ATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCT
GGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCG
GCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTC
ATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGAC
CACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACA
TGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTC
CAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCG
CGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGC
TGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAG
CTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCGACAA
GCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCG
AGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCC
ATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCAC
CCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGG
TCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGAC
GAGCTGTACAAGTAG

GFP 1B4- 27 bp deletion

ATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCT
GGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCG
GCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTC
ATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGAC
CACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACA
TGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTC
CAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCG
CGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGC
TGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAG
CTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCGACAA
GCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCG
AGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCC
ATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCAC
CCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGG
TCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGAC
GAGCTGTACAAGTAG

GFP 1B6- 19 bp deletion

ATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCT
GGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCG
GCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTC
ATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGAC
CACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACA
TGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTC
CAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCG
CGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGC
TGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAG
CTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCGACAA
GCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCG
AGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCC
ATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCAC
CCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGG
TCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGAC
GAGCTGTACAAGTAG

GFP 1B9- 18 bp deletion

ATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCT
GGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCG
GCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTC
ATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGAC
CACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACA
TGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTC
CAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCG
CGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGC
TGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAG
CTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCGACAA
GCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCG
AGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCC

ATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCAC
CCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGG
TCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGAC
GAGCTGTACAAGTAG

GFP 1B12- 13 bp deletion

ATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCT
GGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCG
GCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTC
ATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGAC
CACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACA
TGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTC
CAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCG
CGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGC
TGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAG
CTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCGACAA
GCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCG
AGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCC
ATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCAC
CCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGG
TCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGAC
GAGCTGTACAAGTAG

GFP 1B11- 2 bp insertion

ATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCT
GGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCG
GCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTC
ATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGAC
CACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACA
TGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTC
CAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCG
CGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGC
TGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAG
CTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCGACAA
GCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCG
AGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCC
ATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCAC
CCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGG
TCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGAC
GAGCTGTACAAGTAG

GFP 1B7- 27 bp deletion

ATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCT
GGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCG
GCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTC
ATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGAC

CACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACA
TGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTC
CAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCG
CGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGC
TGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAG
CTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCGACAA
GCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCG
AGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCC
ATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCAC
CCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGG
TCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGAC
GAGCTGTACAAGTAG

GFP 1B8- 15bp deletion

ATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCT
GGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCG
GCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTC
ATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGAC
CACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACA
TGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTC
CAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCG
CGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGC
TGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAG
CTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCGACAA
GCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCG
AGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCC
ATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCAC
CCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGG
TCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGAC
GAGCTGTACAAGTAG

GFP 2A1- 1 bp deletion

ATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCT
GGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCG
GCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTC
ATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGAC
CACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACA
TGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTC
CAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCG
CGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGC
TGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAG
CTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCGACAA
GCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCG
AGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCC
ATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCAC

CCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGG
TCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGAC
GAGCTGTACAAGTAG

GFP 2A2- 1 bp deletion

ATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCT
GGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCG
GCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTC
ATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGAC
CACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACA
TGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTC
CAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCG
CGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGC
TGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAG
CTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCGACAA
GCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCG
AGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCC
ATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCAC
CCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGG
TCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGAC
GAGCTGTACAAGTAG

GFP 2A3- 1 bp deletion

ATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCT
GGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCG
GCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTC
ATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGAC
CACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACA
TGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTC
CAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCG
CGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGC
TGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAG
CTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCGACAA
GCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCG
AGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCC
ATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCAC
CCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGG
TCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGAC
GAGCTGTACAAGTAG

GFP2A4- 15 bp insertion & 118 bp deletion = 132 indels

ATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCT
GGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCG
GCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTC
ATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGAC
CACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACA

TGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTC
CAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCG
CGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGC
TGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAG
CTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCGACAA
GCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCG
AGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCC
ATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCAC
CCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGG
TCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTC.....TT
GGCATGGACGAGCTGTACAAGTAG

GFP2A5- 7 bp deletion

ATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCT
GGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCG
GCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTC
ATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGAC
CACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACA
TGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTC
CAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCG
CGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGC
TGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAG
CTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCGACAA
GCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCG
AGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCC
ATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCAC
CCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGG
TCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGAC
GAGCTGTACAAGTAG

GFP2A6- 26 bp deletion- randomly

ATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCT
GGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCG
GCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTC
ATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGAC
CACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACA
TGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTC
CAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCG
CGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGC
TGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAG
CTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCGACAA
GCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCG
AGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCC
ATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCAC
CCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGG
TCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGAC
GAGCTGTACAAGTAG

GFP2A7 & 2A8- No indels detected.

GFP2A9- 1 bp insertion + sequence not good enough

ATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCT
GGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCG
GCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTC
ATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGAC
CACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACA
TGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTC
CAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCG
CGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGC
TGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAG
CTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCGACAA
GCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCG
AGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCC
ATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCAC
CCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGG
TCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTTGGCATGGAC
GAGCTGTACAAGTAG

GFP2A10- No indels detected.

GFP2A11- 1 bp insertion

ATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCT
GGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCG
GCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTC
ATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGAC
CACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACA
TGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTC
CAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCG
CGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGC
TGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAG
CTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCGACAA
GCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCG
AGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCC
ATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCAC
CCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGG
TCCTGCTGGAGTTCGAGACCGCCGCCGGGATCACTCTCGGCATGGAC
GAGCTGTACAAGTAG

GFP2B3- 1 bp insertion

ATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCT
GGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCG
GCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTC
ATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGAC
CACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACA
TGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTC

CAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCG
CGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGC
TGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAG
CTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCGACAA
GCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCG
AGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCC
ATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCAC
CCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGG
CCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGAC
GAGCTGTACAAGTAG

GFP2B4- BAD Sequence

GFP2B5- 1 bp insertion

ATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCT
GGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCG
GCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTC
ATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGAC
CACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACA
TGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTC
CAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCG
CGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGC
TGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAG
CTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCGACAA
GCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCG
AGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCC
ATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCAC
CCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGG
CCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGAC
GAGCTGTACAAGTAG

GFP2B6- 1 bp insertion off-target & 102 bp deletion on target, however, deletion downstream of eGFP target sequence as well.

ATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCT
GGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCG
GCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTC
ATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGAC
CACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACA
TGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTC
CAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCG
CGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGC
TGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAG
CTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCGACAA
GCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCG
AGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACTCCCCC
ATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCAC

CCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGG
TCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGAC
GAGCTGTACAAGTAG

GFP2B7- no indels detected.

GFP2B8 – 2 bp insertion

ATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCT
GGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCG
GCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTC
ATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGAC
CACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACA
TGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTC
CAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCG
CGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGC
TGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAG
CTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCGACAA
GCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCG
AGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCC
ATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCAC
CCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGG
TCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGAC
GAGCTGTACAAGTAG

GFP2B9- 3 bp deletion

ATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCT
GGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCG
GCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTC
ATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGAC
CACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACA
TGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTC
CAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCG
CGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGC
TGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAG
CTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCGACAA
GCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCG
AGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCC
ATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCAC
CCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGG
TCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGAC
GAGCTGTACAAGTAG

GFP2B10- no indels detected.

GFP2B11- 12 bp insertion & 19 bp deletion = 31 bp indels

ATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCT
GGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCG
GCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTC
ATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGAC
CACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACA
TGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTC
CAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCG
CGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGC
TGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAG
CTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCGACAA
GCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCG
AGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCC
ATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCAC
CCAGTCCGCCCTGAGCAAAGACCCCATCTAGAAGCGCGATCACATGG
CCCACATGGAGTTCAAGTTCGCCGCCGGGCTCACTCTTGGCATGGAC
GAGCTGTACAAGTAG

GFP2B12- 1 bp deletion

ATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCT
GGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCG
GCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTC
ATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGAC
CACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACA
TGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTC
CAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCG
CGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGC
TGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAG
CTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCGACAA
GCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCG
AGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCC
ATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCAC
CCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGG
TCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGAC
GAGCTGTACAAGTAG

GFP2B13- no indels detected

GFP2B14- Sequence not good enough.

ATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCT
GGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCG
GCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTC
ATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGAC

CACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACA
TGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTC
CAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCG
CGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGC
TGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAG
CTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCGACAA
GCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCG
AGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCC
ATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCAC
CCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGG
TCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGAC
GAGCTGTACAAGTAG

ATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCT
GGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCG
GCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTC
ATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGAC
CACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACA
TGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTC
CAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCG
CGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGC
TGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAG
CTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCGACAA
GCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCG
AGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCC
ATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCAC
CCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGG
TCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGAC
GAGCTGTACAAGTAG

Cgr-miR-7 CRISPR sequencing results

cgr-miR-7a-5p locus sense

GGTGGTGGCCAAGGCCCAGAGGAGGTGGTTAGCAAGGCCATGGACAGGC
CAGCCCC~~GCC~~TGGAAGACTAGTGATTTTGTTGTTGTCTCTCTATACTCAAC
AACAAGTCCCAGTCTACCACACGGTGCGGGTCCTCTGAGCA~~TCA~~GGGGGA
GGTATGGGAGCACCATTTCTGGATGTATTTTCTCTGCCTTCTATGTTCCCTA
GCAAACTACCAAAATATTCCTC

Protospacer
~~Cleavage site~~
Mature miR-7
Loop

Star Strand

miR-7a flanking sequence ~2.224 kb

>gi|351515069:117788-120011 Cricetulus griseus unplaced genomic scaffold, CriGri_1.0 scaffold4923, whole genome shotgun sequence

AAAAGTATCAACTGTTCCTACTACCCACTTTTGGCTCAGCAGCCCAGT
TAGACTTGTTCTTTAGCATACA
CTGAAGTCGATCCTCACAGTGGTCCTGGGTTAGAACTGGGGACGTAC
CTTGCATCCTCCACAAAGCGCTT
TTTAGAGTGTAGAACTGTGCCCCATCCCAGGCTCTGGGCTATGAAAG
GGTTGGACTGAGAGGCCTTCCCC
ACCCCATTTGTCGATGACTTTCTCCATGTCCAGGCAAGCCCTTCCCCA
GTCTAGATGAACAGAACAAATC
AAAGGATGGAGGAGGGGTGGGGCAAAGAGAATGAGCTAGCAATTAA
AAGGCAAAAGAATTAAAAGCACCA
ATCAGGAAGGGTAATAATCCTTCAGCTGCTGCAGAAATACAATTAGA
GCCATGCGCTCTTGGAGTTGCCA
GGATATTGTGGCTTTTCCTAATTAAGGAGCGGATGTGCCCTGTCCATG
CAATTAGAGGCTGGCTCAGGCA
GTGGGGGCCCTGCCTACCTTTTATCAACGTACTACAAGGGGAATTAA
CCAACAATACAGCCTATGTCCCC
TGCCTAGTAGTCCCACTTGACTCTGAGCTATGGGTTATGCGCTCTGG
GAAGATGTTCACCTTGGGTTCTG
GGGACCCTTGGGAAGGGGAGAAGAGGCCTGGGGTAGGCACTACTTG
GAGCTTTTTCTTTTATCTATTCTC
TGTGGTAGTGCAGGGTTTTATGATACCTGAAGAAAATATGGAGACAT
GAGCTGCCCACCTGCTTTCCATG
AGAAATAAGATTCCCCTAAGATTAAAGGAAGCGTCTTCACTGTGGAA
AATAGGAGGAGAAAAAAAAGGGG
ATTTCTCCTCCATTTGGAGCCAGAGACTCCAAGGAGTTTTGGGGTGT
GGGGGGTACATGGGGCGGGGAGG
TGTATGGGGTGGTGCAGTGGAGAACATAAGGAAAGTAGAGGGCATG
GGGAAAGACGCTGGCCTCCCCACA
GGAGTATCCAG~~ATG~~ATAAC~~AGGTGG~~~~TGGCCAAGGCCCAGAGGAGGT
GGTTAGCAAGGCCATGGACAGGCC
AGCCCCGCCTGGAAGACTAGTGATTTTGTTGTTGTCTCTCTATACTCA
ACAACAAGTCCCAGTCTACCAC
ACGGTGC~~GGGTCCTCTGAGCATCAGGGGGAGGTATGGGAG~~CACCATT
TCTGGATGTATTTTCTCTGCCTT
CTATGTTCCCTAGCAAACTACCAAAATATTCCTCTAAGGCCTGGGCTG
AAGACACCTCAGGGGGACTCCT
CCAGAAACGAGTACACTGAATCTAGAGTGGGGCAGGGGACAGGATGT
CTGTCCTCAGGGAATGGTGAGCG
ACTCTGCATCCACACACCCTCCCATGGGCTTATTCATTCACCCTGCAA
GCTCCTCCTTGGTGAGAGACAC
CACAGGGCCAACACACAGGGCTCATGGAGGGCAGGTAACAGTGTGTA
CTAGCACAGGCACATGTCCCTAT

CACCTTTTGGGCAAGTGGGCCACCCGATCCTGGATGGTCTGCAATGT
CTGTGCCTGCCTGTGGGGCATAA
TGATATAGAAAAAAACTCCTCACTGCCAATAAATGGAATTTTGGAAGC
CCTGTCTGCTGCCTGTGCTTGG
AGGCCCCCTCAATGGCAGAAGTAGTCATTTCCCCACAAATGTGGCAG
CATATGCCTTTAATTCCATCACT
CAGTAAGGCTTCTGTGTGTAAACCTGTCTTAAAAGGGGAAGGAAGGA
AGGAAGGAAGGAAGGAAGGAAGG
AAGGAAGGAAGGAAGGAAGGAAAGAAGGAAAGAAGGAAAGAAGGAA
GGAAGGAAGGAAACCCTTCACAGC
CCTCCATCCTTTCTGGGTGTCATGTAGCCCAGGCTGGTCTTCAGTTAC
CTACACAGCTCTCCATCCTTTC
TGGGTGTCATGTAGCCCAGGCTGGTCTTCAGTTACCTATGTAGCTGA
GGATGATCCTGAACTCTGTTCTT
CCTGCTTCTGCCCCAAATGCTGGAATTACAGGCATGTGCAATCATACT
CGGTTTATGCAGTGCTGGGGCA
CAAACCCAAGGCTTCACACAAGAGAAGCAAGCACTCTACCACTTGAG
TCCTTCAGCCCTTCTCTTTGCTT
CTCTTCTTTTCCTCCTCTCCCACTCCCTCCATCCTGACATTCATTCATT
CATTTGCACTACTGGAAATAG
AACCTAGGGTTCTGTGCTCTCTTCTTTTCCTCCTCTCCCACTCCCTCC
TTCCTG

miR-7 1B1: 177 bp deletion

GGTGGTGGCCAAGGCCCAGAGGAGGTGGTTAGCAAGGCCATGGACAGGC
CAGCCCCGCCTGGAAGACTAGTGATTTTGTTGTTGTCTCTCTATACTCAAC
AACAAGTCCCAGTCTACCACACGGTGCGGGTCCTCTGAGCATCAGGGGGA
GGTATGGGAGCACCATTTCTGGATGTATTTTCTCTGCCTTCTATGTTCCCTA
GCAAACTACCAAAATATTCCTC---------

miR-7 1B2: Non-specific PCR product into TOPO-TA vector

miR-7 1B3: Incomplete sequence.

miR-7 1B4: 2 bp insertion

GGTGGTGGCCAAGGCCCAGAGGAGGTGGTTAGCAAGGCCATGGACAGGC
CAGCCCCGCGCCTGGAAGACTAGTGATTTTGTTGTTGTCTCTCTATACTCA
ACAACAAGTCCCAGTCTACCACACGGTGCGGGTCCTCTGAGCATCAGGGG
GAGGTATGGGAGCACCATTTCTGGATGTATTTTCTCTGCCTTCTATGTTCC
CTAGCAAACTACCAAAATATTCCTC

miR-7 1B5: 1 bp deletion

GGTGGTGGCCAAGGCCCAGAGGAGGTGGTTAGCAAGGCCATGGACAGGC
CAGCCCCGCCTGGAAGACTAGTGATTTTGTTGTTGTCTCTCTATACTCAAC
AACAAGTCCCAGTCTACCACACGGTGCGGGTCCTCTGAGCATCAGGGGGA
GGTATGGGAGCACCATTTCTGGATGTATTTTCTCTGCCTTCTATGTTCCCTA
GCAAACTACCAAAATATTCCTC

miR-7 1B6: Incomplete sequence.

miR-7 1B7: 1 bp insertion

GGTGGTGGCCAAGGCCCAGAGGAGGTGGTTAGCAAGGCCATGGACAGGC
CAGCCCCCGCCTGGAAGACTAGTGATTTTGTTGTTGTCTCTCTATACTCAA
CAACAAGTCCCAGTCTACCACACGGTGCGGGTCCTCTGAGCATCAGGGGG
AGGTATGGGAGCACCATTTCTGGATGTATTTTCTCTGCCTTCTATGTTCCCT
AGCAAACTACCAAAATATTCCTC

miR-7 1B8: 1 bp insertion

GGTGGTGGCCAAGGCCCAGAGGAGGTGGTTAGCAAGGCCATGGACAGGC
CAGCCCCCGCCTGGAAGACTAGTGATTTTGTTGTTGTCTCTCTATACTCAA
CAACAAGTCCCAGTCTACCACACGGTGCGGGTCCTCTGAGCATCAGGGGG
AGGTATGGGAGCACCATTTCTGGATGTATTTTCTCTGCCTTCTATGTTCCCT
AGCAAACTACCAAAATATTCCTC

miR-7 1B9: 1 bp insertion

GGTGGTGGCCAAGGCCCAGAGGAGGTGGTTAGCAAGGCCATGGACAGGC
CAGCCCCCGCCTGGAAGACTAGTGATTTTGTTGTTGTCTCTCTATACTCAA
CAACAAGTCCCAGTCTACCACACGGTGCGGGTCCTCTGAGCATCAGGGGG
AGGTATGGGAGCACCATTTCTGGATGTATTTTCTCTGCCTTCTATGTTCCCT
AGCAAACTACCAAAATATTCCTC

miR-7 1B10: Non-specific PCR product into TOPO-TA vector

3652440 order

miR-7 1B7: negative

miR-7 1B8: 1 bp mismatch mutation + 1 bp insertion + 4 bp deletion

GGTGGTGGCCAAGGCCCAGGGGAGGTGGTTAGCAAGGCCATGGACAGGC
CAGCCCCGCCTGGAAGACTAGTGATGTTGTTGTTGTCTCTCTATACTCAAC
AACAAGTCCCAGTCTACCACACGGTGCGGGTCCTCTGAGCATCAGGGGGA
GGTATGGGAGCACCATTTCTGGATGTATTTTCTCTGCCTTCTATGTTCCCTA
GCAAACTACCAAAATATTCCTC

miR-7 1B9: 1 bp insertion

GGTGGTGGCCAAGGCCCAGAGGAGGTGGTTAGCAAGGCCATGGACAGGC
CAGCCCCCGCCTGGAAGACTAGTGATTTTGTTGTTGTCTCTCTATACTCAA
CAACAAGTCCCAGTCTACCACACGGTGCGGGTCCTCTGAGCATCAGGGGG
AGGTATGGGAGCACCATTTCTGGATGTATTTTCTCTGCCTTCTATGTTCCCT
AGCAAACTACCAAAATATTCCTC

miR-7 1B10 & 11: 5 bp deletion same for both the samples.

GGTGGTGGCCAAGGCCCAGAGGAGGTGGTTAGCAAGGCCATGGACAGGC
CAGCCCCGCCTGGAAGACTAGTGATTTTGTTGTTGTCTCTCTATACTCAAC
AACAAGTCCCAGTCTACCACACGGTGCGGGTCCTCTGAGCATCAGGGGGA
GGTATGGGAGCACCATTTCTGGATGTATTTTCTCTGCCTTCTATGTTCCCTA
GCAAACTACCAAAATATTCCTC

miR-7 1B12: 25 bp deletion

GGTGGTGGCCAAGGCCCAGAGGAGGTGGTTAGCAAGGCCATGGACAGGC
CAGCCCCGCCTGGAAGACTAGTGATTTTGTTGTTGTCTCTCTATACTCAAC
AACAAGTCCCAGTCTACCACACGGTGCGGGTCCTCTGAGCATCAGGGGGA
GGTATGGGAGCACCATTTCTGGATGTATTTTCTCTGCCTTCTATGTTCCCTA
GCAAACTACCAAAATATTCCTC

miR-7 1B13: 6 bp mismatch mutations

GGTGGTGGCCAAGGCCCAGAGGAGGTGGTTAGCAAGGCCATGGACAGGC
CAGCCCCCGCCTGGAAGACTAGTGATTTTGTTGTTGTCCCTCTACACTCAA
CAACAAGTCCCAGCCGACCACACGGAGCGGGTCCACTGAGCATCAGGGG
GAGGTATGGGAGCACCATTTCTGGATGTATTTTCTCTGCCTTCTATGTTCC
CTAGCAAACTACCAAAATATTCCTC

miR-7 2B1: 127 bp deletion

GGTGGTGGCCAAGGCCCAGAGGAGGTGGTTAGCAAGGCCATGGACAGGC
CAGCCCCGCCTGGAAGACTAGTGATTTTGTTGTTGTCTCTCTATACTCAAC
AACAAGTCCCAGTCTACCACACGGTGCGGGTCCTCTGAGCATCAGGGGGA
GGTATGGGAGCACCATTTCTGGATGTATTTTCTCTGCCTTCTATGTTCCCTA
GCAAACTACCAAAATATTCCTC

miR-7 2B2: 8 bp deletion

GGTGGTGGCCAAGGCCCAGAGGAGGTGGTTAGCAAGGCCATGGACAGGC
CAGCCCCGCCTGGAAGACTAGTGATTTTGTTGTTGTCTCTCTATACTCAAC

AACAAGTCCCAGTCTACCACACGGTGCGGGTCCTCTGAGCATCAGGGGGA
GGTATGGGAGCACCATTTCTGGATGTATTTTCTCTGCCTTCTATGTTCCCTA
GCAAACTACCAAAATATTCCTC

miR-7 2B3: 16 bp deletion

GGTGGTGGCCAAGGCCCAGAGGAGGTGGTTAGCAAGGCCATGGACAGGC
CAGCCCCGCCTGGAAGACTAGTGATTTTGTTGTTGTCTCTCTATACTCAAC
AACAAGTCCCAGTCTACCACACGGTGCGGGTCCTCTGAGCATCAGGGGGA
GGTATGGGAGCACCATTTCTGGATGTATTTTCTCTGCCTTCTATGTTCCCTA
GCAAACTACCAAAATATTCCTC

miR-7 2B4: Incomplete sequence

miR-7 2B5: 2 bp insertion

GGTGGTGGCCAAGGCCCAGAGGAGGTGGTTAGCAAGGCCATGGACAGGC
CAGCCCCGCCTGGAAGACTAGTGATTTTGTTGTTGTCTCTCTATACTCAAC
AACAAGTCCCAGTCTACCACACGGTGCGGGTCCTCTGAGCACATCAGGGG
GAGGTATGGGAGCACCATTTCTGGATGTATTTTCTCTGCCTTCTATGTTCC
CTAGCAAACTACCAAAATATTCCTC

miR-7 2B6: 7 bp mismatch mutations + possibility for more, however could not observe due incomplete amplification of sequencing results.

GGTGGTGGCCAAGGCCTGGAGGAGGTGGTCCGTAGGGCTATGGACAGGC
CAGCCCCGCCTGGAAGACTAGTGATTTTGTTGTTGTCTCTCTATACTCAAC
AACAAGTCCCAGTCTACCACACGGTGCGGGTCCTCTGAGCATCAGGGGGA
GGTATGGGAGCACCATTTCTGGATGTATTTTCTCTGCCTTCTATGTTCCCTA
GCAAACTACCAAAATATTCCTC

miR-7 2B7: 1 bp insertion

GGTGGTGGCCAAGGCCCAGAGGAGGTGGTTAGCAAGGCCATGGACAGGC
CAGCCCCGCCTGGAAGACTAGTGATTTTGTTGTTGTCTCTCTATACTCAAC
AACAAGTCCCAGTCTACCACACGGTGCGGGTCCTCTGAGCAATCAGGGGG
AGGTATGGGAGCACCATTTCTGGATGTATTTTCTCTGCCTTCTATGTTCCCT
AGCAAACTACCAAAATATTCCTC

miR-7 2B8: 17 bp mismatch mutations and 1bp insertion

GGTGGTGGCCAAGGCCCAGAGGAGGTGGTTAGCAAGGCCATGGACAGGC
CAGCCCCGCCTGGAAGACTAGTGATTTTGTTGTTGTCTATGTATACTCAAC
AACAAGTTCCATTTTACCTCAACAACAAGTTCCTGTGTGCCACCAGGGGG

AGGTATGGGAGCACCATTTCTGGATGTATTTTCTCTGCCTTCTATGTTCCCT
AGCAAACTACCAAAATATTCCTC

miR-7 2B9: DID NOT ALIGN→ BAD

miR-7 2B10: 89 bp and 160 bp deletion downstream of miR-7a sequence deletion

GGTGGTGGCCAAGGCCCAGAGGAGGTGGTTAGCAAGGCCATGGACAGGC
CAGCCCCGCCTGGAAGACTAGTGATTTTGTTGTTGTCTCTCTATACTCAAC
AACAAGTCCCAGTCTACCACACGGTGCGGGTCCTCTGAGCATCAGGGGGA
GGTATGGGAGCACCATTTCTGGATGTATTTTCTCTGCCTTCTATGTTCCCTA
GCAAACTACCAAAATATTCCTC……..160 bp downstream miR-7a sequence….

miR-7 2B11: 1 bp insertion

GGTGGTGGCCAAGGCCCAGAGGAGGTGGTTAGCAAGGCCATGGACAGGC
CAGCCCCGCCTGGAAGACTAGTGATTTTGTTGTTGTCTCTCTATACTCAAC
AACAAGTCCCAGTCTACCACACGGTGCGGGTCCTCTGAGCAATCAGGGGG
AGGTATGGGAGCACCATTTCTGGATGTATTTTCTCTGCCTTCTATGTTCCCT
AGCAAACTACCAAAATATTCCTC

miR-7 2B12: 240 bp deletion : 169 bp in the miR-7a sequence + 71 bp upstream of the miR-7 sequence.

GGTGGTGGCCAAGGCCCAGAGGAGGTGGTTAGCAAGGCCATGGACAGGC
CAGCCCCGCCTGGAAGACTAGTGATTTTGTTGTTGTCTCTCTATACTCAAC
AACAAGTCCCAGTCTACCACACGGTGCGGGTCCTCTGAGCATCAGGGGGA
GGTATGGGAGCACCATTTCTGGATGTATTTTCTCTGCCTTCTATGTTCCCTA
GCAAACTACCAAAATATTCCTC

miR-7 1A1: 1 bp insertion + 6 bp deletion + 18 bp mismatch and 1 bp insertion, 9 bp deletion and 16 bp mismatch mutations + just few bp upstream of the sequence

upstrmidm…….GGTAGTATCCAAGGATGATAACAGGTGGTGGCCAAGGCCC
AGAGGCAGGCCTGGTCAGCCTGGAAGACTAGTGATTTTGTTGTTGTCTCTC
TATACTCAACAACAAGTCCCAGTCTACCACACGGTGCGGGTCCTCTGAGC
ATCAGGGGGAGGTATGGGAGCACCATTTCTGGATGTATTTTCTCTGCCTTC
TATGTTCCCTAGCAAACTACCAAAATATTCCTC

miR-7 1A2: 1 bp insertion

GGTGGTGGCCAAGGCCCAGAGGAGGTGGTTAGCAAGGCCATGGACAGGC
CAGCCCCCGCCTGGAAGACTAGTGATTTTGTTGTTGTCTCTCTATACTCAA
CAACAAGTCCCAGTCTACCACACGGTGCGGGTCCTCTGAGCATCAGGGGG
AGGTATGGGAGCACCATTTCTGGATGTATTTTCTCTGCCTTCTATGTTCCCT
AGCAAACTACCAAAATATTCCTC

miR-7 1A3: Incomplete sequence.

GGTGGTGGCCAAGGCCCAGAGGAGGTGGTTAGCAAGGCCATGGACAGGC
CAGCCCCGCCTGGAAGACTAGTGATTTTGTTGTTGTCTCTCTATACTCAAC
AACAAGTCCCAGTCTACCACACGGTGCGGGTCCTCTGAGCATCAGGGGGA
GGTATGGGAGCACCATTTCTGGATGTATTTTCTCTGCCTTCTATGTTCCCTA
GCAAACTACCAAAATATTCCTC

Order 3652440

miR-7 1A3: 5 bp deletion + 1 bp mismatch mutation
GGTGGTGGCCAAGGCCCAGAGGAGGTGGTTAGCAAGGCCATGGACAGGC
CAGCCCCGCCTGGAAGACTAGTGATTTTGTTGTTGTCTCTCTATACTCAAC
AACAAGTCCCAGTCTACCACACGGTGCGGGTCCTCTGAGCATCAGGGGGA
GGTATGGGAGCGCCATTTCTGGATGTATTTTCTCTGCCTTCTATGTTCCCTA
GCAAACTACCAAAATATTCCTC

miR-7 1A4: 1bp insertion

GGTGGTGGCCAAGGCCCAGAGGAGGTGGTTAGCAAGGCCATGGACAGGC
CAGCCCCCGCCTGGAAGACTAGTGATTTTGTTGTTGTCTCTCTATACTCAA
CAACAAGTCCCAGTCTACCACACGGTGCGGGTCCTCTGAGCATCAGGGGG
AGGTATGGGAGCACCATTTCTGGATGTATTTTCTCTGCCTTCTATGTTCCCT
AGCAAACTACCAAAATATTCCTC

miR-7 1A5: 1bp insertion

GGTGGTGGCCAAGGCCCAGAGGAGGTGGTTAGCAAGGCCATGGACAGGC
CAGCCCCCGCCTGGAAGACTAGTGATTTTGTTGTTGTCTCTCTATACTCAA
CAACAAGTCCCAGTCTACCACACGGTGCGGGTCCTCTGAGCATCAGGGGG
AGGTATGGGAGCACCATTTCTGGATGTATTTTCTCTGCCTTCTATGTTCCCT
AGCAAACTACCAAAATATTCCTC

miR-7 1A6: 1bp insertion

GGTGGTGGCCAAGGCCCAGAGGAGGTGGTTAGCAAGGCCATGGACAGGC
CAGCCCCCGCCTGGAAGACTAGTGATTTTGTTGTTGTCTCTCTATACTCAA
CAACAAGTCCCAGTCTACCACACGGTGCGGGTCCTCTGAGCATCAGGGGG
AGGTATGGGAGCACCATTTCTGGATGTATTTTCTCTGCCTTCTATGTTCCCT
AGCAAACTACCAAAATATTCCTC

miR-7 1A7: 1bp insertion

GGTGGTGGCCAAGGCCCAGAGGAGGTGGTTAGCAAGGCCATGGACAGGC
CAGCCCCCGCCTGGAAGACTAGTGATTTTGTTGTTGTCTCTCTATACTCAA
CAACAAGTCCCAGTCTACCACACGGTGCGGGTCCTCTGAGCATCAGGGGG
AGGTATGGGAGCACCATTTCTGGATGTATTTTCTCTGCCTTCTATGTTCCCT
AGCAAACTACCAAAATATTCCTC

miR-7 1A8: 1bp insertion + 1 bp mismatch mutation

GGTGGTGGCCAAGGCCCAGAGGAGGTGGTTAGCAAGGCCATGGACAGGC
CAGCCCCCGCCTGGAAGACTAGTGATTTCGTTGTTGTCTCTCTATACTCAA
CAACAAGTCCCAGTCTACCACACGGTGCGGGTCCTCTGAGCATCAGGGGG
AGGTATGGGAGCACCATTTCTGGATGTATTTTCTCTGCCTTCTATGTTCCCT
AGCAAACTACCAAAATATTCCTC

miR-7 1A9: 64 bp deletion + 32 bp upstream of miR-7a sequence

GGTGGTGGCCAAGGCCCAGAGGAGGTGGTTAGCAAGGCCATGGACAGGC
CAGCCCCGCCTGGAAGACTAGTGATTTTGTTGTTGTCTCTCTATACTCAAC
AACAAGTCCCAGTCTACCACACGGTGCGGGTCCTCTGAGCATCAGGGGGA
GGTATGGGAGCACCATTTCTGGATGTATTTTCTCTGCCTTCTATGTTCCCTA
GCAAACTACCAAAATATTCCTC

miR-7 1A10: 2bp insertion

GGTGGTGGCCAAGGCCCAGAGGAGGTGGTTAGCAAGGCCATGGACAGGC
CAGCCCCCGCCTGGAAGACTAGTGATTTTGTTGTTGTCTCTCTATACTCA
ACAACAAGTCCCAGTCTACCACACGGTGCGGGTCCTCTGAGCATCAGGGG
GAGGTATGGGAGCACCATTTCTGGATGTATTTTCTCTGCCTTCTATGTTCC
CTAGCAAACTACCAAAATATTCCTC

miR-7 1A11: 18 bp deletion

GGTGGTGGCCAAGGCCCAGAGGAGGTGGTTAGCAAGGCCATGGACAGGC
CAGCCCCGCCTGGAAGACTAGTGATTTTGTTGTTGTCTCTCTATACTCAAC
AACAAGTCCCAGTCTACCACACGGTGCGGGTCCTCTGAGCATCAGGGGGA
GGTATGGGAGCACCATTTCTGGATGTATTTTCTCTGCCTTCTATGTTCCCTA
GCAAACTACCAAAATATTCCTC

miR-7 2A1: 3 bp deletion

GGTGGTGGCCAAGGCCCAGAGGAGGTGGTTAGCAAGGCCATGGACAGGC
CAGCCCCGCCTGGAAGACTAGTGATTTTGTTGTTGTCTCTCTATACTCAAC
AACAAGTCCCAGTCTACCACACGGTGCGGGTCCTCTGAGCATCAGGGGGA
GGTATGGGAGCACCATTTCTGGATGTATTTTCTCTGCCTTCTATGTTCCCTA
GCAAACTACCAAAATATTCCTC

miR-7 2A2: 1 bp deletion + 1 bp insertion and possibility for more, however, incomplete sequence further downstream…..

GGTGGTGGCCAAGGCCCAGAGGAGGTGGTTAGCAGAGGCCATGGACAGG
CCAGCCCCGCCTGGAAGACTAGTGATTTTGTTGTTGTCTCTCTATACTCAA
CAACAAGTCCCAGTCTACCACACGGTGCGGGTCCTCTGAGCATCAGGGGG
AGGTATGGGAGCACCATTTCTGGATGTATTTTCTCTGCCTTCTATGTTCCCT
AGCAAACTACCAAAATATTCCTC

miR-7 2A3: 1 bp insertion

GGTGGTGGCCAAGGCCCAGAGGAGGTGGTTAGCAAGGCCATGGACAGGC
CAGCCCCGCCTGGAAGACTAGTGATTTTGTTGTTGTCTCTCTATACTCAAC
AACAAGTCCCAGTCTACCACACGGTGCGGGTCCTCTGAGCAATCAGGGGG
AGGTATGGGAGCACCATTTCTGGATGTATTTTCTCTGCCTTCTATGTTCCCT
AGCAAACTACCAAAATATTCCTC

miR-7 2A4: 1 bp insertion same as above

GGTGGTGGCCAAGGCCCAGAGGAGGTGGTTAGCAAGGCCATGGACAGGC
CAGCCCCGCCTGGAAGACTAGTGATTTTGTTGTTGTCTCTCTATACTCAAC
AACAAGTCCCAGTCTACCACACGGTGCGGGTCCTCTGAGCATCAGGGGGA
GGTATGGGAGCACCATTTCTGGATGTATTTTCTCTGCCTTCTATGTTCCCTA
GCAAACTACCAAAATATTCCTC

miR-7 2A5: 1 bp insertion same as above

GGTGGTGGCCAAGGCCCAGAGGAGGTGGTTAGCAAGGCCATGGACAGGC
CAGCCCCGCCTGGAAGACTAGTGATTTTGTTGTTGTCTCTCTATACTCAAC
AACAAGTCCCAGTCTACCACACGGTGCGGGTCCTCTGAGCATCAGGGGGA
GGTATGGGAGCACCATTTCTGGATGTATTTTCTCTGCCTTCTATGTTCCCTA
GCAAACTACCAAAATATTCCTC

miR-7 2A6: 6 bp deletion

GGTGGTGGCCAAGGCCCAGAGGAGGTGGTTAGCAAGGCCATGGACAGGC
CAGCCCCGCCTGGAAGACTAGTGATTTTGTTGTTGTCTCTCTATACTCAAC
AACAAGTCCCAGTCTACCACACGGTGCGGGTCCTCTGAGCATCAGGGGGA
GGTATGGGAGCACCATTTCTGGATGTATTTTCTCTGCCTTCTATGTTCCCTA
GCAAACTACCAAAATATTCCTC

miR-7 2A7: 5bp mismatch + 1 bp deletion and possibility for more, however, downstream seq info incomplete.

GGTGGTGGCCAAGGCCCAGAGGAGGTGGTTAGAAAGGCTCTGGACAGGT
CAGCCCTGCCTGGAAGACTAGTGATTTTGTTGTTGTCTCTCTATACTCAAC
AACAAGTCCCAGTCTACCACACGGTGCGGGTCCTCTGAGCATCAGGGGGA
GGTATGGGAGCACCATTTCTGGATGTATTTTCTCTGCCTTCTATGTTCCCTA
GCAAACTACCAAAATATTCCTC

miR-7 2A8: 1 bp mismatch, however, possibility for more incomplete sequence downstream

GGTGGTGGCCAAGGCGCAGAGGAGGTGGTTAGCAAGGCCATGGACAGGC
CAGCCCCGCCTGGAAGACTAGTGATTTTGTTGTTGTCTCTCTATACTCAAC
AACAAGTCCCAGTCTACCACACGGTGCGGGTCCTCTGAGCATCAGGGGGA
GGTATGGGAGCACCATTTCTGGATGTATTTTCTCTGCCTTCTATGTTCCCTA
GCAAACTACCAAAATATTCCTC

miR-7 2A9: 1 bp deletion + 1 bp insertion

GGTGGTGGCCAAGGCCCAGAGGAGGTGGTTAGCAAGGCCATGGACAGGC
CAGCCCCGCCTGGAAGACTAGTGATTTTGTTGTTGTCTCTCTATACTCAAC
AACAAGTCCCAGTCTACCACACGGTGCGGGTCCTCTGAGCAATCAGGGGG
AGGTATGGGAGCACCATTTCTGGATGTATTTTCTCTGCCTTCTATGTTCCCT
AGCAAACTACCAAAATATTCCTC

miR-7 2A10: 1 bp insertion

GGTGGTGGCCAAGGCCCAGAGGAGGTGGTTAGCAAGGCCATGGACAGGC
CAGCCCCGCCTGGAAGACTAGTGATTTTGTTGTTGTCTCTCTATACTCAAC
AACAAGTCCCAGTCTACCACACGGTGCGGGTCCTCTGAGCAATCAGGGGG
AGGTATGGGAGCACCATTTCTGGATGTATTTTCTCTGCCTTCTATGTTCCCT
AGCAAACTACCAAAATATTCCTC